Adnectin Solubility and Dynamics

by

Kyle Trainor

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemistry

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	R. Scott Prosser Professor, Department of Chemistry University of Toronto
Supervisor	Elizabeth Meiering Professor, Department of Chemistry University of Waterloo
Internal Member	Thorsten Dieckmann Associate Professor, Department of Chemistry University of Waterloo
Internal Member	John Honek Professor, Department of Chemistry University of Waterloo
Internal-External Member	Andrew Doxey Associate Professor, Department of Biology University of Waterloo

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Rapid growth of the global market for monoclonal antibodies (mAbs) has generated considerable interest in the development of alternative molecules that facilitate rapid discovery and manufacturing, while replicating the low toxicity/immunogenicity and tight, specific binding of mAbs. One such molecule is the tenth human fibronectin type III domain (^{10}Fn3), which has solvent accessible loops resembling the V_H complementarity-determining regions H1, H2, and H3 of immunoglobulin. ^{10}Fn3-based binding proteins called Adnectins have been engineered to bind with high affinity to diverse targets using *in vitro* evolution methods such as mRNA display, yeast display, and phage display.

Adnectins are known to vary in aggregation propensity, sometimes despite exceptionally high amino acid sequence identity, and have been used as a basis for protein aggregation/solubility research. Aggregation of therapeutic proteins can provoke a protein-specific immune response, and the solubility of Adnectins is therefore of immediate practical interest. The aggregation of proteins in general is a complicated and incompletely understood phenomenon, the study of which we advance using Adnectins as a model system. We also investigate protein dynamics (which can be related to protein aggregation, but additionally has enormous impact on how we think about protein structure and function) through nuclear magnetic resonance (NMR) spectroscopic and computational study of Adnectins.

Here, we first present solubility data for a reference set of 41 Adnectins and use them to screen computational solubility/aggregation prediction methods. On the basis of these results, we select the CamSol prediction method for use in a protein engineering project that applies the principles of consensus design to enhance the solubility of the Adnectin scaffold. Furthermore, we demonstrate that hydrogen/deuterium exchange by hydrogen bonded amides in the C-terminal β-strand of the original scaffold is induced by transient inter-Adnectin association, and that equivalent exchange is not observed in our solubility-enhanced scaffold.

Next, we describe the results of variable-temperature solution NMR experiments that probe Adnectin dynamics. Each resonance observable by NMR spectroscopy is composed of contributions from structurally equivalent nuclei in a vast number of proteins, the conformations of which may both differ from each other at any particular instant and evolve over the timescale of the experiment. The temperature dependences of amide proton and nitrogen chemical shifts are due to differences in the conformations sampled and their probabilities of occupation. Empirically, these temperature dependences are well-approximated by fits to a linear model, the slopes of which (known as temperature coefficients) may report on protein dynamics in the vicinity of each backbone amide. We explore possible determinants of amide proton and nitrogen temperature coefficients using a combination of molecular dynamics simulations and quantum chemical (density functional theory) calculations. In the directly detected (high resolution) proton dimension, we also analyze deviations from linearity, which may be attributable to fast exchange between protein conformations with distinct chemical shifts and a temperature-dependent difference in free energy.

Acknowledgements

I would like to thank my supervisor, Dr. Elizabeth Meiering, and all members of the Meiering lab, past and present - a strong team makes all the difference - particularly Colleen Doyle, Aron Broom, and Avril Metcalfe-Roach, who contributed to the work described herein; also Zachary Gingras, Allen Chiu, and Cicely Shillingford, who helped build the foundation.

I would also like to thank our collaborators from Bristol-Myers Squibb, including Heather Malakian, Michael Gosselin, Dr. Stanley Krystek, Dr. David Langley, and especially Dr. Daša Lipovšek.

This research was enabled in part by support provided by the Shared Hierarchical Academic Research Computing Network (SHARCNET; https://www.sharcnet.ca) and Compute Canada (https://www.computecanada.ca).

Dedication

To my parents, for encouraging independent thought and higher education. To Melissa, for her love and support. To Hanne, for the joy she brings into my life.

Table of Contents

List of Tables

List of Figures

Abbreviations

^{10}Fn3 10^{th} human fibronectin type III domain

AIC Akaike information criterion

APR aggregation-prone region

CAR chimeric antigen receptor

CARA computer aided resonance assignment

CDD conserved domain database

CGR chaos game representation

CSV comma separated values

DFT density functional theory

DLS dynamic light scattering

DSC differential scanning calorimetry

DSS 4,4-dimethyl-4-silapentane-1-sulfonic acid

DSSP dictionary of secondary structure of protein

ECM extracellular matrix

EDTA ethylenediaminetetraacetic acid

EGFR epidermal growth factor receptor

Fn3 fibronectin type III domain

FPT finite perturbation theory

FTIR Fourier transform infrared

GB1 B1 domain of protein G

GB3 B3 domain of protein G

GIAO gauge-independent atomic orbital

H/D hydrogeon/deuterium

HMQC heteronuclear multiple quantum coherence

HSQC heteronuclear single quantum coherence

IB inclusion body

IDP intrinsically disordered protein

IGF1R insulin-like growth factor-I receptor

IL-23 interleukin-23

IPTG isopropyl β-D-1-thiogalactopyranoside

mAb monoclonal antibody

MD molecular dynamics

MSA multiple sequence alignment

NESG northeast structural genomics

NMR nuclear magnetic resonance

NOESY nuclear Overhauser effect spectroscopy

ORF open reading frame

PAS proline/alanine/serine

PCM polarizable continuum model

PCSK9 proprotein convertase subtilisin/kexin type 9

PDB protein data bank

ppb parts per billion

ppm parts per million

PSSM position specific scoring matrix

pWT peudo wild-type

PXR human pregnane X receptor

RF random forest

RMSE root mean square deviation

RMSE root mean square error

SAP spatial aggregation propensity

SDS-PAGE sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SLS static light scattering

SSE sum of squared errors

SVM support vector machine

TIP3P transferable intermolecular potential with 3 points

TMS tetramethylsilane

TOCSY total correlation spectroscopy

tris tris(hydroxymethyl)aminomethane

VEGFR-2 vascular endothelial growth factor receptor-2

Part I

Introduction

Chapter 1

Proteins, Folding, and Aggregation

1.1 Amino Acids and Protein Folding

Four major classes of biological molecules are involved in the chemistry of life: proteins, nucleic acids, polysaccharides and lipids [1]. Of these, proteins may be the most adaptable, fulfilling structural, catalytic, cell signalling/transduction, and physical force-generating functions. This adaptability is observed despite the fact that, fundamentally, proteins are linear polymers composed of a relatively small number (~20) of different amino acid types. The functional diversity of proteins is enabled by complex three dimensional structure and dynamics, specified (primarily) by non-covalent residue-residue and residue-solvent interactions.

The physicochemical properties of the proteinogenic amino acid side chains vary widely, endowing each different sequence of amino acids with a unique character. For all but the shortest polypeptides, the number of possible conformations is vast [2]; yet, in most cases, evolved proteins fold into a well-determined three-dimensional conformation (or ensemble of closely related conformations) known as the native state, on a biologically accessible timescale. The sequences of most evolved proteins are therefore thought to be under selective pressure that favours those capable of an expeditious transition between the unfolded and natively folded states. Within the context of the energy landscape theory of protein folding [3], this is described by the 'folding funnel' hypothesis (Figure 1.1) [4]. Interesting exceptions to this general rule include intrinsically disordered proteins (IDPs), which may be described by a relatively flat (i.e., less funnelled) energy landscape [5, 6].

Like 'native state', the term 'unfolded' is understood to imply an ensemble of conformations - in this case, those that have in common a lack of well-defined structure. However, at physiologically relevant temperatures and in the absence of chemical denaturants, completely unfolded polypeptides are unlikely to persist; the hydrophobic effect [7] can drive the collapse of unfolded polypeptides into an ensemble of compact states [8, 9] (some of which may contain stabilizing secondary structure [10]), and partially synthesized

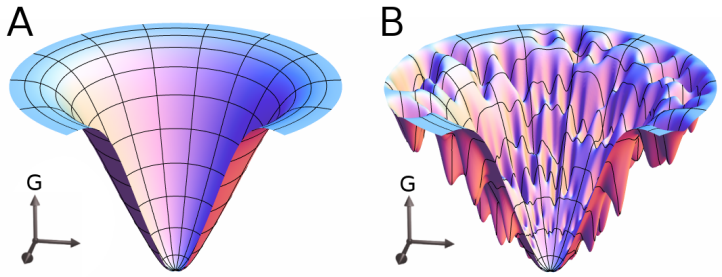


Figure 1.1: Free energy landscapes with Gibbs energy on the vertical axis, and a two-dimensional projection of conformational space on the horizontal axes; A: idealized folding funnel, B: rugged funnel with local minima.

polypeptides can begin to fold while still attached to the ribosome [11]. However, rates of folding to the native state may be limited by the height of one or more energy barriers that must be surmounted. As most proteins must be natively folded in order to function, there is evolutionary pressure to maintain a free energy difference between the native and unfolded states (ΔG) sufficient to ensure that a negligible fraction of protein is unfolded at equilibrium.

1.2 Solubility and Aggregation

The selective pressure for thermodynamically stable proteins that fold quickly is tempered by competing (e.g., functional) constraints on protein sequence and structure. This may manifest as deviations from the ideal (Fig. 1.1A) such as a rugged folding funnel with many local minima (Figure 1.1B), higher rate-limiting energy barriers (slower folding), a smaller free energy difference between the unfolded and natively folded states (lower global stability), or a native state defined by a broader minimum in the energy landscape (more dynamics/lower local stability). The exposure of protein segments that would ideally be sheltered from intermolecular association (e.g., in the native structure) can lead to protein aggregation [12]. Unlike monomer folding (which is unimolecular), aggregation is a minimally bimolecular, concentration-dependent process; nevertheless, aggregation can be considered a form of protein folding - one in which the formation of large, insoluble assemblies is driven by many of the same factors that promote the folding of soluble proteins (Table 1.1). Various types of aggregates can be differentiated by their internal structures and gross morphologies (e.g., fibrillar or amorphous), and study of the causes and effects of their formation has the potential to impact the fields of biology (including the study of disease), biotechnology, and biomaterials research. Protein aggregation is commonly encountered when high concentrations of proteins are produced in an expression system; in a cell-based expression system, these aggregates are frequently referred to as inclusion bodies (IBs).

Table 1.1: Types of intermolecular contacts

Intermolecular Interaction	**Characteristics**
Hydrophobic Side Chain Burial	Solvent exposure of hydrophobic side chains minimized through intermolecular interactions
Polar Contacts	Interactions between polar residues, possibly distributed over multiple contact patches
Intermolecular β-sheet	Secondary structure-promoting hydrogen bonds (also, burial of hydrophobic side chains; see above)
Covalent Bonding	Primarily disulfide bonds

The term 'inclusion bodies' encompasses many types of intracellular and periplasmic aggregates. Among the best studied IBs are those formed upon overexpression of a heterologous gene in *Escherichia coli* (*E. coli*). An indispensable workhorse of biotechnology since the birth of the field [13], *E. coli* remains one of the most widely used cell-based protein expression systems due to its status as a well-characterized model organism [14], and because it is relatively easy to grow cultures at high cell density [15].

Though the homogeneous composition of IBs may simplify the purification process (80-95% of a typical IB is made up of the overexpressed protein [16]), the structures adopted by the constituent protein may vary from near-native to distinctly non-native [17–21]; therefore, they must be 'refolded' to obtain soluble protein, a process that may be difficult and/or inefficient [22]. Alternatively, for applications in which solubility is not paramount, IBs that retain some degree of function may be used directly [23]. However, protein solubility is sometimes a very important consideration; for example, the aggregation of biopharmaceuticals such as Adnectins may trigger dangereous immune responses [24, 25]. In Chapters 4 and 5 we quantitate the propensities of Adnectins to form IBs when overexpressed in *E. coli* and treat these propensities as measures of solubility, but the complexity of the intracellular environment in which IB formation takes place must be acknowledged.

The intracellular concentration of proteins and other macromolecules may be as high as 300-400 g/L [26], endowing the cytoplasm with a gel-like consistency and giving rise to macromolecular crowding effects that may promote protein-protein association [27]. Also, many of these proteins are far from passive bystanders: proteases [28,29], chaperones [30,31] and molecular machines responsible for the transport of small aggregates to the poles of *E. coli* cells [32] are but a few examples of active intracellular macromolecules that may impact the expressed protein. Conditions such as temperature [20] and induction level [19] have also been shown to affect IB structure, as well as the partitioning of the expressed protein between the soluble and insoluble cellular fractions. Many aspects of intracellular complexity have been reviewed elsewhere [28–31].

Despite potentially confounding intracellular variables, IB formation may be driven by many of the same thermodynamic and kinetic factors that contribute to other forms of protein aggregation and folding. Relative to unfolded or denatured states, both native folding and aggregate formation are likely to be thermodynamically favoured under physiological conditions [33]; if the energy barriers to denaturation and disaggregation are sufficiently high, the fate of denatured proteins under folding conditions may be determined by a kinetic competition [34, 35]. Alternatively, marginal thermodynamic stability may promote aggregation by allowing a non-negligible equilibrium population of unfolded proteins [36], and aggregation initiated from native or native-like states (i.e., without surmounting the major energy barrier to global unfolding) has also been documented [37–39].

Historically, it was widely believed that energetically favourable interactions between the exposed hydrophobic side chains of unstructured protein regions were primarily responsible for protein aggregation [40]. The burial of hydrophobic residues remains a major factor in the stability of any water-solvated protein conformation, but more recent studies have emphasized intermolecular association mediated by self-complementary aggregation-prone regions [41–43]. The specificity of the interactions between such regions can be high; for example, even when two IB-forming proteins are expressed simultaneously, they may not appreciably co-aggregate (true co-aggregation can be distinguished from co-localization by Förster resonance energy transfer) [41, 44]. In theory, this self-complementarity could be attributed solely to hydrophobic side chain burial, but the enrichment of β-sheet structure frequently observed within aggregates [19, 20, 40] suggests that hydrogen bonding and other polar interactions also contribute (Table 1.1).

Intermolecular association driven by the burial of complementary hydrophobic surfaces has long been known to be a mechanism of protein complex formation [45]. The surfaces in question need not be exclusively hydrophobic; though the hydrophobic effect may be dominant, polar interactions at the interface can also play a stabilizing role [46], and either a single, contiguous hydrophobic patch or a number of smaller patches are possible [47]. The feature that differentiates this type of intermolecular association is the mean hydrophobicity of residues in the interface, which is greater than that of residues on the solvent-exposed exterior of the complex [48]. Hydrophobic surfaces that are safely buried in the monomeric native state may be exposed in fully or partially unstructured conformations. Aggregates stabilized primarily by non-specific hydrophobic interactions (i.e. between surfaces or segments that are not complementary) are also possible. However, this type of association does not provide a satisfactory explanation for the homogeneous composition of aggregates such as IBs [41, 44].

At the opposite end of the hydrophobicity spectrum, polar contacts, such as those commonly found in protein crystals, are intermolecular interfaces formed by the burial of predominantly polar surfaces. Crystals upon which X-ray crystallography is performed are typically composed of natively folded proteins. Under normal physiological conditions, these proteins are soluble, and the amino acid composition of the crystal-packing interface is often virtually indistinguishable from that of the solvent-exposed surface [49]. However, just as polar residues may be found in predominantly hydrophobic surfaces, non-polar ones

can be found in crystal-like contacts; the task of distinguishing 'biological' from 'crystal-packing' interfaces can be non-trivial [50]. Here we define polar contacts as those in which the intermolecular interaction is predominantly mediated by favourable interactions between polar and charged amino acid side chains, rather than the hydrophobic effect [51].

The intermolecular β-sheet form of protein-protein association frequently involves not only polar intermolecular contacts (hydrogen bonds), but also sequestration of hydrophobic amino acid side chains. It is unique in that segments from multiple polypeptides collectively form highly stabilizing secondary structure. This type of intermolecular association is found in the amyloid fibril spine [52], and is also commonly observed in domain-swapping oligomerization [53]. Evidence of non-native β-structure has also been found in amorphous aggregates such as some IBs [17, 19, 40, 42].

'Runaway' or open-ended domain-swapping is particularly interesting because a single mechanism of association may (conceptually) produce oligomers varying in size from dimers to large, insoluble aggregates (Figure 1.2). Domain swapping restricts translational and rotational degrees of (monomer) freedom, which is entropically unfavourable [53]; energetic compensation may result from the relief of strain (in the native monomer) or the formation of stabilizing structure in the intermolecular interface. In some cases, a protein segment must adopt the role of a 'hinge loop' linking the swapped domain to the rest of the monomer. If hinge loop strain in the native monomer is relieved in a more extended form, oligomerization may be energetically favoured; for example, strain induced by mutational hinge loop shortening has been observed to result in dimer formation [54]. The formation of domain-swapped oligomers can also be facilitated by longer, more flexible hinge loops capable of favourable self-association (β-sheet) in their extended form (Figure 1.2) [55].

As described in the next chapter (Chapter 2), proteins that adopt the fibronectin type III domain (Fn3) fold have been reported to form oligomers in which either the N or C terminal β-strand (A or G) may be swapped. Also discussed is the unfolding of human fibronectin domains, including the tenth human fibronectin type III domain (^{10}Fn3) from which the Adnectins are derived (Chapter 3), by cell traction forces as part of extracellular matrix assembly. As with strand-swapped Fn3 oligomers, assembly of multimeric protein complexes may be facilitated by unfolding of ^{10}Fn3 β-strands A and/or G.

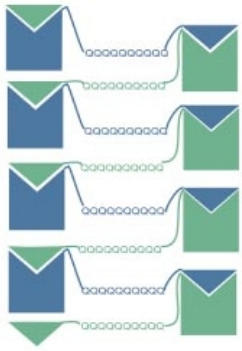


Figure 1.2: An illustration of open-ended, or 'runaway' domain swapping. Blue and green subunits swap domains, and the 'hinge loops' (middle) form stabilizing β-sheet structure. Reproduced from [55].

Chapter 2

Fibronectin Type III Domains

The Adnectins are derived from the tenth human fibronectin type III domain (^{10}Fn3), and therefore adopt a fibronectin type III domain (Fn3) fold. In this chapter we introduce ^{10}Fn3, describe the context (i.e., the multidomain human fibronectin protein; Fig. 2.1) in which it functions, and compare it to other well-studied examples of the Fn3 fold.

2.1 Fibronectin

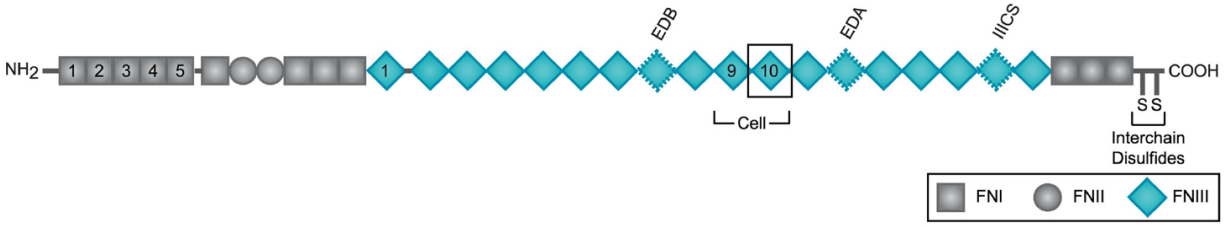


Figure 2.1: Multidomain human fibronectin with three domain types distinguished by different shapes. The EDA, EDB, and IIICS domains (labelled) may be subject to alternative splicing. The ninth and tenth type III domains are responsible for cell binding. Figure reproduced from [56].

Fibronectin is a multidomain glycoprotein, composed of repeats of three distinct domain types (I, II, and III), that is involved in cellular interactions with the extracellular matrix (ECM) [57]. The β-sandwich fold adopted by fibronectin type III domains (Fn3), considered to belong to the immunoglobulin superfamily, is by no means limited to fibronectin; by one estimate, this fold can be found in approximately 2% of all animal proteins [58]. The tenth human fibronectin type III domain (^{10}Fn3), illustrated in Fig. 2.2, is of particular interest because it has been used as a scaffold for the development of protein biotherapeutics called Adnectins, the solubility and dynamics of which are the subject of this thesis.

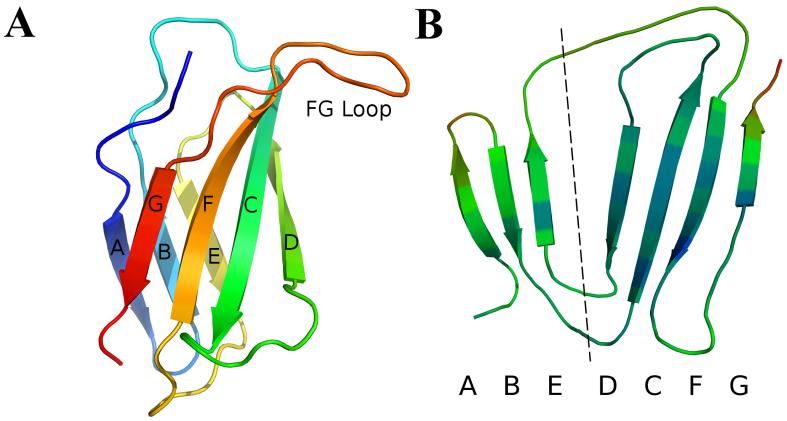


Figure 2.2: Cartoon representations of wild-type ^{10}Fn3 (derived from PDB structure 1FNF). A: coloured from N-terminus (blue) to C-terminus (red). B: flat representation coloured by crystallographic B-factor (which may be interpreted as a measure of uncertainty in the positions of the backbone atoms), from low (blue) to high (red); loops not to scale. The dashed line divides the two halves of the β-sandwich.

2.2 Mechanical Unfolding of Fn3 Domains

In its role as part of the ECM, fibronectin is assembled into elastic fibrils [59]. Studies of Chinese hamster ovary cells expressing green fluorescent protein-fibronectin chimeras have demonstrated that fibronectin fibrils can be extended up to four times their equilibrium length *in vivo* [60]. The extensibility of structural proteins such as fibronectin is a consequence of reversible unfolding of their constituent domains [61, 62]. This unfolding may expose 'cryptic' binding sites that promote the association of individual fibronectin molecules into fibrils [63–65]. The ^{10}Fn3 domain is thought to be among the first to unfold due to its low mechanical stability [66], and there is evidence that a cryptic binding site may be located in β-strand B [56].

Because the wild-type ^{10}Fn3 domain is believed to mediate fibronectin fibrillogenesis by unfolding under mechanical stress to reveal cryptic binding sites [56], ^{10}Fn3 unfolding has been studied extensively using both experimental and computational methods. Single-molecule force spectroscopy experiments [67] and steered molecular dynamics simulations [68–70] in which a tensile force was applied between the N and C-termini (similar to forces that might be applied to a ^{10}Fn3 domain through the ECM) independently concluded that at least two different unfolding intermediates are observed (as part of two distinct pathways): one in which β-strands A and B become detached and solvent-exposed, and another in which β-strand G becomes detached. The application of a tensile force between the N-terminus and the integrin-binding FG loop (a loading pattern that is physiologically relevant if cell-traction forces are responsible for ^{10}Fn3 unfolding) has also been simulated. These simulations consistently feature an unfolding intermediate in which β-strand A is

detached [70].

Given that the ^{10}Fn3 may have evolved to permit unfolding of β-strands A and/or G, it is unsurprising that these regions exhibit unusual structural dynamics even in the absence of mechanical stress. In the following section we describe a noteworthy lack of protection against amide hydrogen/deuterium (H/D) exchange in ^{10}Fn3 β-strands A and G, and contrast this with the protection of equivalent β-strands in an otherwise very similar Fn3 domain from human tenascin.

2.3 Structural Dynamics of Fn3 Domains

Like fibronectin, tenascin is a large, multi-domain protein found in the ECM. Approximately half of the domains in each of these proteins are classified as Fn3 folds [71]. A domain homologous to ^{10}Fn3, the third fibronectin type III domain of human tenascin (^{3}TnFn3), has also been shown to unfold under mechanical stress through a force-stabilized intermediate in which the A and/or G β-strands may be unstructured [72]. Despite the similarities between ^{10}Fn3 and ^{3}TnFn3, differences in structural dynamics are perceptible.

Amide (H/D) exchange experiments performed on ^{10}Fn3 demonstrated that, despite inter-β-strand hydrogen bonding (inferred from crystal structures), residues in β-strands A and G are not measurably protected from exchange at pD 7.0. In contrast, the corresponding residues in ^{3}TnFn3 are protected under the same conditions (despite the lower global stability of ^{3}TnFn3), which supports the hypothesis (explored in Chapters 4-6) that these regions of ^{10}Fn3 may be particularly dynamic [73].

2.4 Fn3 Folding and Transition State Structure

The folding of ^{10}Fn3 has been modelled as a three-state transition, with a folding intermediate apparent at low concentrations of denaturant [74]. Detailed ϕ-value analysis of ^{10}Fn3 has shown that β-strands C, D, E, and F are significantly structured in the transition state (though all of the ϕ-values are fractional) [75] (Figure 2.3). The corresponding β-strands are also highly structured in the transition state of ^{3}TnFn3 [76], and it has been proposed that these two proteins share a common folding nucleus [75] - one that excludes β-strands A, B, and G.

The A, B, and G strands are unambiguously less structured than the folding nucleus in the ^{3}TnFn3 transition state, but anomalously small changes in ^{10}Fn3 global stability upon mutation of residues in these strands (even those that appear to be deeply buried in the hydrophobic core of the native structure) complicate interpretation of the associated ϕ-values. This unusual accommodation of mutations without loss of stability has been attributed to the dynamic character of the native structure (Section 2.3) [73].

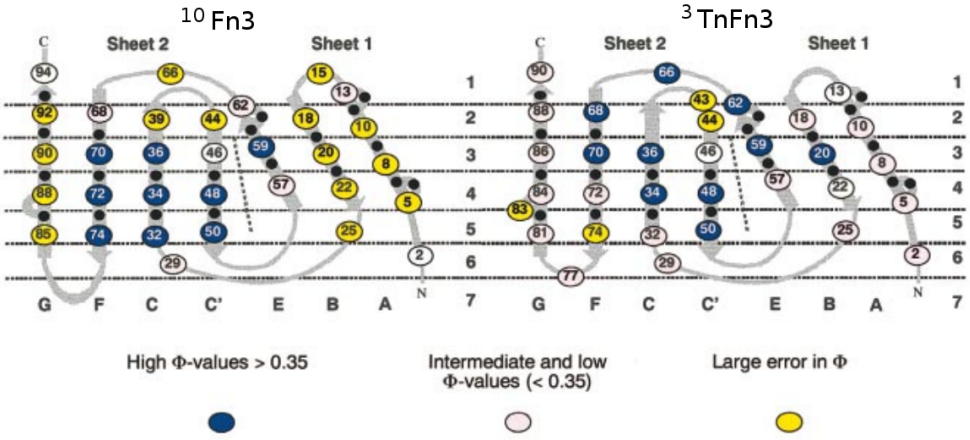


Figure 2.3: The results of ϕ-value analyses of wild-type ^{10}Fn3 (left) and ^{3}TnFn3 (right). Regions thought to be structured in the transition state are indicated by blue ovals. Reproduced from [75].

2.5 Multimerization of Fn3 Domains

Protein aggregates are often insoluble, precluding solution NMR-based characterization, and composed of structurally inhomogeneous or imperfectly ordered subunits, which complicates structure determination by X-ray crystallography (Chapter 1). However, in some cases ordered multimers suitable for X-ray diffraction may arise from the protein crystallization process, including examples composed of immunoglobulin-like domains [77–79]. In addition to the runaway domain-swapping mechanism of aggregation described in Section 1.2, swapping of β-strands can lead to the formation of highly ordered multimers. This phenomenon can be observed in crystal structures of several variants of Tencon, a consensus-designed protein of particular interest because it shares the same fold (Fn3) as the Adnectins [80, 81]. Fn3 hexamers composed of β-strand G swapped dimers are observed in several crystal structures (Fig. 2.4). This multimerization is not induced only by crystallization; analytical size exclusion chromatography (SEC) reveals that 4lpv (Fig. 2.4A) runs exclusively as a multimer, while two peaks (corresponding to a monomer and a multimer) are observed in the SEC profiles of 4lpw (Fig. 2.4B) and 4lpx (Fig. 2.4C).

Even more interesting, as it hints at plausible aggregation mechanisms for Adnectins, is the association of β-strand G swapped Tencon dimers into an open-ended (i.e., the number of monomers incorporated may be limited only by the size of the crystal) left-handed helical form (Fig 2.5). We note that β-strand A swapped Tencon dimers have also been reported [81]; the potential for strand swapping at both the N and C termini opens up a vast array of multimerization and aggregation possibilities.

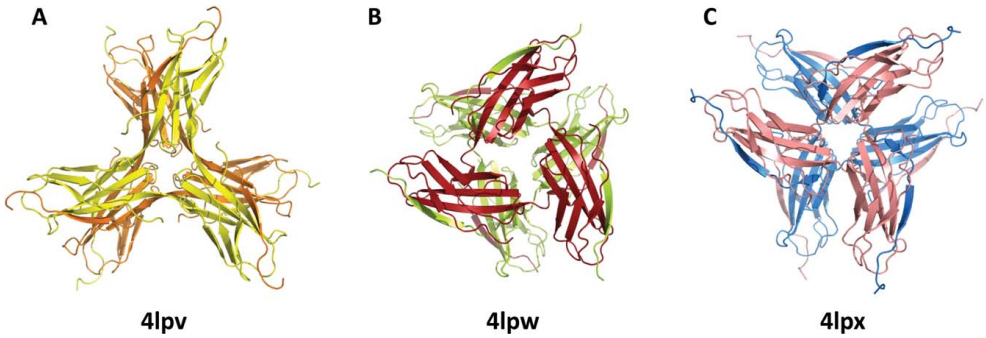


Figure 2.4: Fn3 (Tencon) hexamers formed by the association of β-strand G swapped dimers. A, B, and C: hexamers formed by dimers of different Tencon FG loop variants with the corresponding crystal structure PDB identifiers shown below. Figure reproduced from [80].

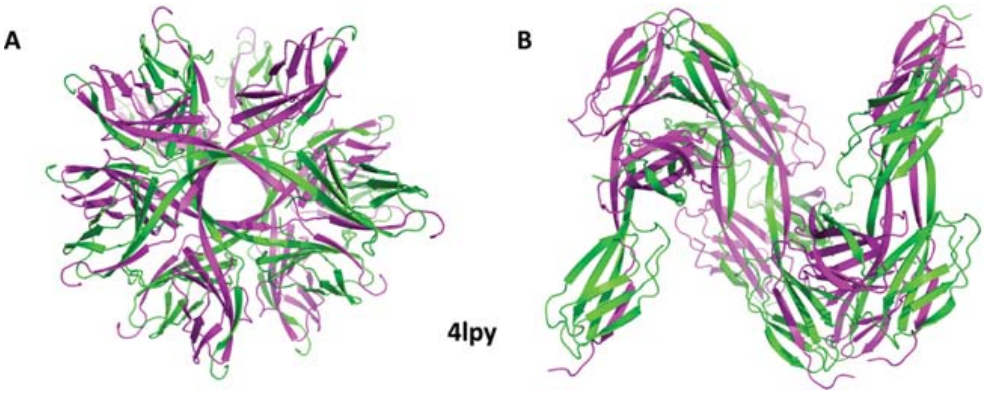


Figure 2.5: Association of β-strand G swapped Fn3 (Tencon) dimers to form a large (open-ended) helical multimer. A: Looking down the axis of symmetry; B: rotated 90° about an axis perpendicular to the axis of symmetry. The crystal structure PDB identifier is shown between the panels. Reproduced from [80].

Chapter 3

Adnectins

3.1 Engineered Binding Proteins Derived from ^{10}Fn3

The global market for monoclonal antibodies (mAbs) has grown faster each year since 2013, and exceeded US$98 billion in sales in 2017 [82]. This success has generated considerable interest in the development of alternative molecules that facilitate rapid discovery and manufacturing, while replicating the low toxicity/immunogenicity and tight, specific binding of mAbs. The fibronectin type III domain (Fn3) fold is structurally similar to the immunoglobulin fold, including solvent accessible loops resembling the V_H complementarity-determining regions H1, H2, and H3 of immunoglobulin. This, combined with favourable characteristics such as high thermostability, solubility, and expression level, as well as the absence of disulfide bonds or free cysteine residues, lead to the development of ^{10}Fn3-based binding proteins such as Adnectins [83–87]. Adnectins that bind with high affinity to diverse targets have been developed using *in vitro* evolution methods such as mRNA display, yeast display, and phage display [88]. Examples of targets for which crystal structures with Adnectins in complex have been published include the human pregnane X receptor (PXR) [89], proprotein convertase subtilisin/kexin type 9 (PCSK9) [90], epidermal growth factor receptor (EGFR) and interleukin-23 (IL-23) [91].

3.2 Bispecific Adnectins, Adnectin-Drug Conjugates, and Fusions

In contrast to antibodies, which are capable of multivalent binding via their variable domains as well as Fc-mediated activation of the immune system, most Adnectins are engineered for monovalent binding [87]. Individual (single domain) Adnectins that bind epidermal growth factor receptor (EGFR) and insulin-like growth factor-I receptor (IGF1R) have been combined to form a multidomain, bispecific binding protein that inhibited the

growth of human tumour xenografts [92]. This tandem Adnectin demonstrates the potential for Adnectin-based cancer therapies that target multiple growth factors. Similarly, bispecific Adnectins have been developed for antiviral applications. Joining an Adnectin targeting the glycoprotein 41 subunit of the HIV-1 envelope to a previously discovered human CD4-binding anti-HIV Adnectin [93] via a charged, relatively rigid linker resulted in a synergistic increase in antiviral potency [94].

Adnectins can also be conjugated to drugs, e.g., for the delivery of cytotoxic payloads to cancerous tumours [95]. Hydrogen/deuterium (H/D) exchange at pD 7.0 monitored by mass spectrometry has been used to demonstrate that conjugation to a drug did not result in significant protein conformational changes [96]. Likewise, antibodies can be used in cancer-targeting applications, but Adnectins are removed from bloodstream more quickly by the kidneys as a consequence of their smaller size, possibly reducing undesirable exposure of cancer-free tissues to the toxic payload [95].

In other contexts, the short *in vivo* half-lives of Adnectins may be a disadvantage. Fortunately, their pharmacokinetics can be modulated through the attachment of polyethylene glycol polymers (PEGylation) of varying molecular weights (which may also increase solubility and stability, and decrease immunogenicity [97]), as they were for an anti-angiogenic Adnectin targeting vascular endothelial growth factor receptor-2 (VEGFR-2) [98]. PASylation, genetic fusion of Pro/Ala/Ser (PAS) repeats 100-600 residues in length, has also been utilized to slow clearance of Adnectins from the bloodstream [99].

Large genetic fusions with high amino acid sequence and structural complexity are possible as well. An Adnectin has been used as the recognition domain of a chimeric antigen receptor (CAR), an artificial T-cell receptor that includes (in addition to the Adnectin) transmembrane, intracellular signalling, and costimulatory domains [100]. T-cells genetically engineered to express Adnectin-based CARs matched the ability of CARs based on single chain variable fragment recognition domains (derived from mAbs) to kill cancer cells both *in vitro* and in xenograft tumour-bearing mice.

3.3 Adnectin Solubility and Dynamics

Adnectins are known to vary in aggregation propensity, sometimes despite exceptionally high amino acid sequence identity, and have been used as a basis for protein aggregation/-solubility research [101–103]. Aggregation of therapeutic proteins can provoke a protein-specific immune response [24,25], and the solubility of Adnectins is therefore of immediate practical interest. The aggregation of proteins in general is a complicated and incompletely understood phenomenon [104], the study of which we aim to advance using Adnectins as a model system (Chapters 4-6). We also investigate protein dynamics (which can be related to protein aggregation, as in Chapter 6, but also has enormous impact on how we think about protein structure and function) through nuclear magnetic resonance (NMR) spectroscopic and computational study of Adnectins (Chapters 7-11).

Part II

Engineering Protein Solubility

Chapter 4

Solubility Prediction

4.1 Context

Aggregation can be thought of as a form of protein folding in which intermolecular associations lead to the formation of large, insoluble assemblies. Various types of aggregates can be differentiated by their internal structures and gross morphologies (e.g., fibrillar or amorphous), and the ability to accurately predict the likelihood of their formation by a given polypeptide is of great practical utility in the fields of biology (including the study of disease), biotechnology, and biomaterials research.

Solubility prediction is an essential part of our Adnectin solubility engineering strategy; therefore, here we review aggregation/solubility prediction methods, determine that methods from the 'sliding window' category (Section 4.2.2) meet our requirements, and evaluate promising methods by comparing their aggregation/solubility predictions for 41 different Adnectin sequences to experimentally determined propensities to form inclusion body (IB) intracellular aggregates. Of those evaluated, the CamSol method [105] performs best, and it is applied to predict the solubility of Adnectins and many other proteins that adopt the fibronectin type III domain (Fn3) fold in during the consensus sequence design for solubility described in Chapter 5.

4.1.1 Acknowledgements

As noted in Table 4.1, we previously published some of the data therein [102]. Zachary Gingras assisted with the collection of this previously published data. Much of the content in the review of aggregation/solubility prediction methods (Section 4.2) has also been published [104].

4.2 Introduction

In order to engineer solubility without prohibitive experimental labour, we must be able to predict it. Protein solubility is closely linked with aggregation, which we define as protein self-association that results in large, insoluble assemblies. Although aggregation is a complicated phenomenon, many fundamental aspects are understood - it occurs as a result of the same influences responsible for 'normal' protein folding (i.e., to the soluble native state) [40]. Thus, aggregation can be thought of as folding to an alternate state with features common to almost all thermodynamically stable protein conformations: relative compactness, substantial desolvation of hydrophobic side chains, and the satisfaction of many potential hydrogen bonds. A classic example is the amyloid fibril, a type of protein aggregate with a high degree of long-range order, extensive intermolecular (β-sheet) hydrogen bonding, and a well-packed core [106]. More generally, the degree of long-range order within aggregates may vary, and the conformations adopted by the constituent monomers may include native, native-like, and/or non-native secondary/tertiary structure. Gross aggregate morphologies range from fibrillar to amorphous, and a given polypeptide can form structurally distinct aggregates that may propagate in a prion-like fashion [107]. Furthermore, *in vivo* aggregation can involve additional complications such as macromolecular crowding and the activity of chaperones or proteases [16]; the impacts of such factors may be implicitly incorporated into predictive methods based on data from *in vivo* experiments.

Aggregation/solubility prediction is an area of active research owing to its significant practical utility in the production of recombinant proteins for research and biotechnological purposes [108], the formulation of biopharmaceuticals [109], and in the context of protein misfolding-related disease [110]. Over time, many methodological improvements have been driven by the accumulation of fundamental insights into the thermodynamics [111] and kinetics [12] of aggregation, the realization that short sequence segments may determine aggregation propensity [112,113], and advances in the modelling and simulation of dynamic aggregation-prone surface exposure [114–116]. Here we review recent developments in the field of aggregation/solubility prediction and evaluate the potential of several methods to aid in our solubility engineering endeavour by applying them to a reference set of 41 Adnectins for which aggregation propensities have been experimentally determined.

In this review, we group aggregation/solubility prediction methods into three broad categories: (a) statistical analyses and machine learning algorithms that abstract aggregation-related features from the amino acid sequences of proteins with known aggregation propensities, (b) 'sliding window' methods that can be used to inspect full-length (whole protein) amino acid sequences for aggregation-prone stretches as short as five residues, and (c) structure-based methods that consider the exposure of aggregation-prone regions, particularly in near-native conformations. Ultimately, we find that 'sliding window' methods suit our solubility engineering application best because they are computationally efficient, do not require accurate structural models, make no assumptions about accessibility of aggregation-prone regions (APRs), and several of them produce predictions that correlate well enough with experimentally measured Adnectin aggregation to be of use ($R^2 \geq 0.6$).

4.2.1 Sequence-based methods: amino acid composition and k-mer frequency

Many of the classification techniques reviewed in this section consist of machine learning algorithms, e.g., neural network, support vector machine (SVM), or random forest (RF) classifiers, trained using large protein solubility data sets. Amino acid composition, including amino acid or k-mer (dipeptide, tripeptide, etc.) frequencies, is often found to be the single feature that predicts solubility best [117–119], but many other features may be considered, and some of the classifiers have multi-layer architectures [117,120]. These measures of amino acid composition are mostly order-independent (i.e., the position of each amino acid in the sequence is not very important) - highly reductive, but perhaps necessary in order to generalize well to the wide diversity of proteins in the training and validation data sets. Possibly due to the difficulty of the prediction problem, these methods are often gauged by their abilities to classify proteins simply as either soluble or insoluble (i.e., binary classification), where the correct class has been determined by applying a cut-off to experimental solubility data.

Large protein solubility data sets facilitate classifier training/development by providing many examples from which generalizable aggregation/solubility-related features may be identified. For example, PROSO II [117] was trained using data from the expression of >82000 proteins in *Escherichia coli* (*E. coli*). However, the compilation of a large data set is challenging; annotation errors may creep in, biases towards extensively studied classes of proteins and/or proteins from common model organisms are inevitable, and experimental data may have been collected under varying conditions. Alternatively, smaller data sets obtained under uniform conditions can be used for training. For instance, the results of a study that quantified the soluble expression of 3173 *E. coli* proteins in a cell-free system (without chaperones) [121] have been used to train SVM and RF classifiers [122, 123]. Another valuable source of data is the high-throughput protein production platform of the Northeast Structural Genomics (NESG) consortium [124]. Statistical analysis of the expression level and solubility of 9644 proteins from the NESG pipeline revealed some interesting relationships [125]. Although mean hydrophobicity is negatively correlated with the expression level and solubility of the NESG proteins, it does not appear to be the dominant determinant. Instead, consistent with evidence for the inhibition of aggregation by high net (whole protein) charge and position-specific 'gatekeeper' residues [126–128], electrostatic charge seems to have the strongest influence. Accordingly, the fractional contents of aspartate, glutamate, and lysine are all positively correlated with the expression level and solubility of these proteins; however, in surprising contrast, the fractional content of arginine is negatively correlated.

Although solubility can be predicted to some degree from amino acid composition alone, potentially relevant information is discarded when the order of the amino acids in the sequence is disregarded. One way to salvage some of this information is by combining conventional measures of amino acid composition with factors that reflect some aspect of sequence order (pseudo-amino acid composition) [129]. Shannon entropy and parameters

derived from Chaos Game Representations (CGRs) of amino acid sequences are two such factors that have been reported to improve solubility predictions [130].

All of the methods and analyses reviewed to this point were trained on and/or produce binary classifications (soluble/insoluble) except for the NESG protein production analysis, in which a discrete six-point scale was employed [125]. Scales that recognize intermediate solubilities may be more useful than binary classifications, for example, in the context of mutagenesis for the purpose of increasing protein solubility. As the number of positions at which mutations are considered increases, the number of possible sequences undergoes a combinatorial explosion ($\sim 20^N$ for N positions), and the computational cost of analyzing all of them may become prohibitive. One can imagine an 'insolubility landscape' (Fig. 4.1) analogous to the energy landscapes of protein folding theory [131]. To whatever extent this landscape is 'funnelled' (Fig. 4.1B) it may be explored in a less-exhaustive fashion. If the effects of mutations on solubility are additive (even partially), some degree of funnelling is likely. Classification into more than two categories is one way to differentiate intermediate solubilities [120, 125]; as an alternative to discrete categories, aggregation rates or propensities can be predicted using parameterized equations such as the Chiti-Dobson equation [12], some derivatives of which have evolved into sliding window methods discussed in the next section.

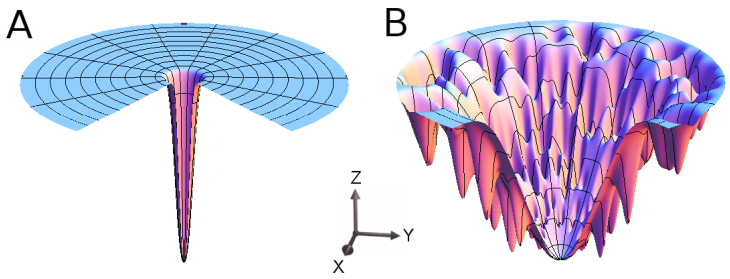


Figure 4.1: Hypothetical 'insolubility landscapes' in which multi-dimensional amino acid sequence space is simplified by projection onto the X and Y axes, and insolubility is plotted on the Z axis. Two examples are illustrated (others are also possible). A: a binary (soluble/insoluble) landscape with a flat, featureless plateau; B: a landscape in which intermediate insolubilities can be discerned and may facilitate identification of the most soluble sequence (the global insolubility minimum). Figure reproduced from [104]

4.2.2 Sequence-based methods: sliding windows and patterns

Many of the methods reviewed in this section make use of a sliding window to consider the aggregation propensity of amino acid sequence segments of various lengths; the order and locations of amino acids in the sequence are implicitly important. Superficially, this approach could be considered a logical extension of the strategy of relating k-mers (e.g.,

dipeptides), rather than individual amino acids, to solubility. In fact, these methods were inspired by theoretical and experimental results that suggest relatively short (5-7 residue) sequence segments can have a disproportionate influence on aggregation propensity [112]. Predictions of aggregation propensity produced by sliding window methods have been based on a variety of factors, including physicochemical characteristics (e.g., charge, hydrophobicity/hydrophilicity, and secondary structure propensity) [105, 132], statistical analysis of residue pairings between adjacent β-strands in known structures [133], and patterns of residue distribution in amyloidogenic hexapeptides [134]. The ESPRESSO method [135], which was developed using data from a genome-scale experiment in which 17739 human open reading frames (ORFs) were expressed in *E. coli* and 8850 ORFs were expressed in a wheat germ cell-free system [136], additionally makes use of sequence-based predictions of relative solvent accessible surface area, disordered regions, and transmembrane regions.

TANGO, among the earliest of the sliding window methods, estimates the population of various conformational states (including β-aggregate) according to a Boltzmann distribution [137]. TANGO remains in active use today, and was recently applied to engineer mutants of an Aβ42-green fluorescent protein fusion with differing aggregation rates [138]. The varying strength of evolutionary selection against *E. coli* expressing these aggregation-prone mutants was subsequently quantified. TANGO results were also used to target APRs in the sequences of human α-galactosidase (a deficiency of which causes Fabry disease) and the protective antigen of *Bacillus anthrax* (a component of recombinant anthrax vaccines) for solubilizing mutations [139]. The Solubis method [140] automates the identification of aggregation propensity-minimizing point mutations using a combination of TANGO predictions and thermodynamic stabilities calculated using the FoldX force field [141]. Computational prediction of protein thermodynamic stability is a challenging problem, particularly when the goal is to identify stabilizing mutations. Our recent work shows that combining the output of several stability prediction algorithms (meta-prediction) may improve performance on this type of task [142]. Furthermore, we find that individual stability prediction tools are often quite good at identifying mutations likely to have a large destabilizing effect, which bodes very well for combined aggregation propensity/thermodynamic stability approaches such as Solubis.

AGGRESCAN [143] aggregation propensity predictions, which are based on amino acid scores derived from measurements of intracellular aggregation by mutants of Aβ42, were recently shown to be correlated with both cell division defects and the resistance of purified Aβ42 inclusion bodies to chemical denaturation [144]. AGGRESCAN has also been adapted for aggregation propensity predictions that incorporate knowledge of tertiary structure [115]. Such predictions will be discussed further in the next section.

Sliding window aggregation predictions may be usefully combined with amino acid sequence-based predictions of other physicochemical properties. The Prediction of Amyloid STructure Aggregation (PASTA) sliding window algorithm [145] has been combined with ESpritz [146] intrinsic disorder scores (as well as hydrophobicity profiles and secondary structure propensities) to produce the SODA method [147]. SODA is capable of assessing all possible sequence substitutions as well as insertions and deletions.

Furthermore, the predictions generated by different methods are not always in agreement, and judicious combination may increase either the sensitivity or specificity of APR identification. Computational design of γD crystallin point mutants with increased aggregation resistance [148] was facilitated by a combination of AGGRESCAN, PASTA, and TANGO. AGGRESCAN, PASTA, TANGO, and Zyggregator [132] were used as part of a study of the conformational and colloidal stabilities of isolated constant domains from human immunoglobulin G [149]. AGGRESCAN, FoldAmyloid, TANGO, and Zyggregator were used to help identify APRs and potential Aβ42 binding sites in transthyretin [150]. Lastly, AGGRESCAN, PASTA, TANGO, Waltz [151], and the 3D Profile method (ZipperDB) [152] were used to help identify aggregation hotspots within and adjacent to CDR1 of a V_H antibody (single-domain) [153]. Negative charge-increasing mutations in these regions improved aggregation-resistance; however, the failure of similar (in location and impact on net charge) mutations to solubilize a different single-domain antibody suggests that electrostatic repulsion due to net charge does not always suffice to prevent aggregation [154].

In general, the sliding window methods recognize amino acid sequence features related to β-sheet (including β-aggregate) secondary structure propensity; however, β-aggregate structure is known to vary. For example, eight different classes of 'steric zippers' were identified among the structures of amyloid-forming peptides determined by X-ray microcrystallography [155]. The 3D Profile (ZipperDB) method predicts the aggregation propensity of amino acid sequence segments on the basis of their compatibility with the steric zipper structure adopted by one of these peptides (NNQQNY). This method was recently used to identify an APR in the amino acid sequence of the p53 tumour suppressor; a cell-penetrating peptide designed to inhibit aggregation by targeting this APR restored p53 function in ovarian cancer cells [156]. AmyloidMutants goes a step further than the other methods described in this section, predicting amyloid supersecondary structure (two-sheet β-solenoids with either parallel or antiparallel intra- and interchain interactions, or serpentine cross-β structure) [157]. The structural insights that underpin AmyloidMutants' predictions have also been applied to the assembly of complete models of amyloid-β structures using a tool called STITCHER [158].

In addition to aggregate structure, another important structural consideration is the protein conformation from which aggregation takes place. The methods reviewed in this section predict the aggregation propensity of amino acid sequence segments; however, some also calculate composite scores for full (whole protein) sequences, implicitly assuming that all sequence segments are equally exposed and available to participate in aggregation (as they might be, for example, in a fully unfolded polypeptide). Almost all open reading frames in the *E. coli*, *Homo sapiens*, and *Saccharomyces cerevisiae* genomes contain at least one region predicted to be an APR; however, in up to ~95% of known protein structures these regions are not solvent exposed [128, 159]. Mutations can modulate both the intrinsic aggregation propensity of APRs and the stability of the native fold relative to unfolded conformations; in some cases, APR aggregation propensity may correlate well with experimentally observed aggregation, while in other cases stability against unfolding

(leading to APR exposure) may be a better predictor. Even denatured polypeptides may have some degree of protective structure (residual or transient), and there is evidence that some proteins may aggregate from locally unfolded conformations (separated from the native state by a relatively low energy barrier) [102, 160]. The CamP method of predicting hydrogen exchange protection factors from amino acid sequence [161] has been used to adjust Zyggregator scores to account for predicted amino acid exposure [132]. CamSol [105] is also discussed in the next section because it can either be used in a Zyggregator-like 'sequence-only' mode, or incorporate a tertiary structure-based correction for APR exposure.

4.2.3 Tertiary/quaternary structure & simulation

None of the methods in the preceding sections explicitly consider the tertiary/quaternary structure of the native protein (though some do predict the compatibility of sequence with different types of secondary structure). This may be convenient, because no structure or structural model is required. Also, this may be consistent with a model of residue exposure in which the aggregating species is fully unfolded. However, even in these cases, it is possible for associations between natively folded proteins (prior to local or full unfolding) to impact the kinetics of aggregation [162], and aggregates may also be formed by partially structured proteins [102, 160, 163].

Zyggregator aggregation propensity scores have been combined with amino acid sequence-based predictions of residue accessibility [132], projected onto the known structures of natively folded proteins [164], projected onto the structure of an intermediate state identified by relaxation dispersion nuclear magnetic resonance (NMR) [165], and incorporated into a predictive model of local unfolding-based residue exposure and aggregation [102]. Also, the amino acid sequence-based solubility predictions of CamSol, which is partly derived from the Zyggregator method, can be projected onto a structure and smoothed on the basis of a surface patch size parameter and solvent exposure [105].

Like the 'structurally corrected' version of CamSol, AGGRESCAN3D projects amino-acid sequence-based predictions onto a structure [115]; coarse-grained simulations of near-native dynamics executed by the CABS-flex server [166], which has been found to produce results consistent with both all-atom MD simulations (at a fraction of the computational cost) and experimental (solution NMR) data [167], can also be integrated into the analysis. The CABS-flex web server is much simpler to use than typical MD software, giving a wider range of researchers the opportunity to explore near-native dynamics using computational methods.

Another popular method known as spatial aggregation propensity (SAP) measures the hydrophobicity of exposed surface patches, which may be averaged over 30 ns molecular dynamics (MD) simulations [114]. It has been used to engineer monoclonal antibodies with lower aggregation propensity [168], to investigate the influence of glycosylation on aggregation propensity [169], and has been incorporated into a 'developability index' designed to

facilitate the selection of candidate biotherapeutics [170]. SAP scores were also found correlate with the cytotoxicity of Aβ42 variants [171], and have been used to predict protein binding regions [172] and the ammonium sulfate concentrations at which biotherapeutics precipitate [173].

In addition to more conventional predictors of aggregation propensity such as surface hydrophobicity, hydrophilic/charged surface patches can also be considered. The AggScore method, of particular interest because it was trained using Adnectin IB formation data, surveys the distributions and relative orientations of various types of hydrophobic and hydrophilic surface patches [103]. Beyond the training set, AggScore has been used to rationalize the differing aggregation propensities of growth hormone proteins and β-amyloid mutants. It can also reliably discriminate between amyloidogenic and non-amyloidogenic hexapeptides on a strictly structural basis.

Among hydrophilic surface patches, those that are positively charged may be especially noteworthy, as they have been observed to decrease soluble expression (particularly when arginine residues are responsible for more of the charge than lysine residues) [174]. These findings are reminiscent of previous analyses in which negative surface charge was found to increase *in vitro* solubility more effectively than positive charge [175], and arginine content was found to be negatively correlated with soluble expression [125], but differ in that the spatial distribution of charge is analyzed and a specific mechanism involving nucleic acid binding proposed [174]. Extending the idea that the effect of arginine on solubility can be opposite to that of lysine, the ratio of lysine to arginine has been proposed as a novel, yet simple predictor of soluble expression [176].

The methods in the previous two sections may implicitly assume equal exposure of all residues in the sequence, and a variety of methods in this section consider native or native-like tertiary structure. This leaves a huge gap in our aggregation prediction tool kit - what about natively folded proteins that subsequently unfold (fully or locally) and aggregate? Ideally, this gap would be filled by a method that is both generally applicable and does not require experimental data. In theory, simulations of local dynamics, such as those included in the SAP and AGGRESCAN3D methods, could be extended to include more substantial unfolding; for now, however, the high computational cost of this approach renders it impractical. Simulations under non-physiological conditions that promote unfolding are often used to study the process on a shorter timescale; in a notable example, parameters recorded during relatively short (25 ns) thermal unfolding MD simulations conducted at 498 K were used to predict the soluble expression of recombinant proteins in *E. coli* using an SVM [116]. Although the number of proteins studied using this method was relatively small (approximately 15), the incorporation of simulated structural fluctuations outside of the native basin into aggregation/solubility predictions merits further study.

4.2.4 Solubility Engineering via Consensus Design

Not all of the methods reviewed above are equally well-suited for the application described in Chapter 5, consensus sequence design for solubility. The three types of methods (Sections 4.2.1-4.2.3) vary in their assumptions (e.g., regarding the importance of amino acid content vs. sequence order, or the exposure of different protein regions), requirements (e.g., for structural models), and computational demands. In Section 4.4, we introduce our aggregation/solubility reference data (from 41 Adnectins designed to bind insulin-like growth factor 1 receptor), narrow our focus to sliding window methods (Section 4.2.2) based on requirements imposed by consensus-based design, and use the reference data to evaluate methods that are candidates for use in Chapter 5.

4.3 Methods

4.3.1 Protein Expression

Adnectins were expressed in BL21 (DE3) pLysS cells (Edge Biosystems) transformed with pET-9d vectors into which the Adnectins had been cloned at the NcoI and BamHI sites. Cells were grown overnight at 37°C with shaking at 225 RPM in LB broth containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. The overnight culture was diluted 1:100 into fresh LB broth with the same antibiotics, and incubated at 37°C with shaking at 225 RPM to an A_{600} of 0.6-0.8. Expression was then induced using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

1 mL samples were taken 2, 4, 6, and 24 h after induction. Samples were centrifuged at 5000 g (room temperature) for 10 min and the pellets resuspended in 100 µL TEN buffer (20 mM Tris pH 8.0, 1 mM EDTA, 100 mM NaCl). The resuspended pellets were flash frozen in liquid nitrogen and stored at -80°C. Resuspended cell pellets were subjected to 5 cycles of freezing in liquid nitrogen and thawing in a 25°C water bath. 5 µL of 3 mg/mL DNase I was then added to each microcentrifuge tube, and mixed by gently inverting the tubes 30 times. After a 20 min incubation period, samples were subjected to an additional 5 freeze-thaw cycles. Soluble and insoluble fractions were separated by centrifugation at 16300 g (room temperature) for 15 min, and the supernatants were transferred to new tubes. The pellets were resuspended in TEN buffer and all tubes were flash frozen in liquid nitrogen and stored at -80°C.

4.3.2 Solubility Analysis

Samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and stained using Coomassie Blue. The gels were imaged using a BIS303PC

gel documentation system (DNR Bio-Imaging Systems) and pixel densities were quantitated using the TotalLab 100 software package (Nonlinear Dynamics). The background (an approximation of the pixel density of protein-free gel) was determined using the 'rolling ball' method and subtracted from the Adnectin bands. The percentage of Adnectin found in IBs was calculated as $100 \cdot (\text{Insoluble Integrated Pixel Density})/(\text{Soluble} + \text{Insoluble Integrated Pixel Density})$. More than one Adnectin band may be observed by SDS-PAGE as a consequence of non-enzymatic post-translational gluconoylation [102]. The total Adnectin integrated pixel density for each lane was taken to be the sum of the integrated pixel densities of the Adnectin bands.

4.3.3 Solubility Prediction

AGGRESCAN

Sequence segments that are known to be aggregation-prone are sometimes described as 'hot spots', and the relative aggregation propensities of individual amino acids have been characterized through mutational analysis of a model hot spot in Aβ42 [177]. Each Adnectin sequence (Appendix A) was submitted to the AGGRESCAN server (http://bioinf.uab.es/aggrescan), which assigns per-residue aggregation scores on the basis of the amino acid aggregation propensities observed in the model hot spot and averages them over a window seven residues in width.

FoldAmyloid

Strong hydrogen bonds formed between densely packed β-strands are known to be at the heart of the cross-β spine [106]. FoldAmyloid draws on amino acid packing density and hydrogen bond formation statistics captured from a database of protein structures to predict the amyloidogenicity of protein regions solely on the basis of primary sequence [178]. The Adnectin sequences (Appendix A) were submitted to the FoldAmyloid server (http://bioinfo.protres.ru/fold-amyloid) for analysis, with the following options selected: Scale, expected number of contacts 8 Å; averaging frame, 5; threshold, 21.4.

CamSol

CamSol assigns a solubility score to each amino acid in a protein based on hydrophobicity, charge, and secondary structure formation propensity [105]. These per-residue solubility scores are then averaged over a sliding window of seven residues, then combined with additional terms accounting for patterns of alternating hydrophobic/hydrophilic residues and the presence of charged 'gatekeeper' residues. The Adnectin sequences (Appendix A) were submitted to the CamSol server (http://www-mvsoftware.ch.cam.ac.uk/index.php/camsolintrinsic) with a pH of 7.0 selected.

4.4 Results & Discussion

We previously characterized the IB formation propensity of 31 Adnectins, engineered to bind insulin-like growth factor 1 receptor (IGF1R), with amino acid sequences that differ primarily in the FG loop [101, 102]. Here we enhance this aggregation/solubility data set by characterizing 10 additional IGF1R-binding Adnectins, increasing amino acid sequence diversity in the FG loop (Table 4.1, Fig. 4.2). As these Adnectins were all engineered to bind the same target, some lack of diversity is difficult to avoid; the RDY motif is notably well-represented in FG loop positions 3-5 (residue numbers 79-81).

Table 4.1: IGF1R-binding Adnectin expression data[1]

		% Adnectin in IB[3]			
Adnectin #	**FG Loop[2]**	**2 h**	**4 h**	**6 h**	**24 h**
1	GSRDYE	15 ± 3	20 ± 0	26 ± 6	58 ± 3
2	KMRDYR	17 ± 7	19 ± 3	25 ± 2	67 ± 5
3	RSRDYR	17 ± 7	21 ± 7	23 ± 3	57 ± 2
4	ERRDYR	20 ± 2	17 ± 1	19 ± 2	57 ± 4
5	SLRDYG	20 ± 1	22 ± 0	33 ± 0	83 ± 7
6	TQRDYG	21 ± 11	40 ± 6	37 ± 0	77 ± 8
7	DTRDYR	22 ± 5	20 ± 5	23 ± 2	58 ± 0
8	EMRDYG	22 ± 9	38 ± 2	44 ± 4	84 ± 1
9	NLRDYG	26 ± 4	24 ± 7	36 ± 3	78 ± 1
10	CRRDYG	28 ± 6	35 ± 2	39 ± 1	62 ± 3
11	KVRDYR	29 ± 7	25 ± 9	31 ± 5	81 ± 10
12	SRRDYG	29 ± 2	27 ± 2	27 ± 5	75 ± 0
13	EKNQRG	29 ± 9	39 ± 11	45 ± 5	79 ± 3
14	SLRDYA	29 ± 3	42 ± 1	56 ± 1	90 ± 3
15	RLRDYE	36 ± 3	47 ± 2	61 ± 1	89 ± 2
16	RIRDYG	36 ± 4	51 ± 2	62 ± 4	85 ± 7
17	HFRDYG	38 ± 3	44 ± 6	55 ± 10	80 ± 6
18	SLRDYV	40 ± 4	58 ± 3	64 ± 6	92 ± 3
19	NYDNDR	42 ± 1	74 ± 13	86 ± 1	95 ± 1
20	MSRDYG	46 ± 2	47 ± 3	61 ± 3	93 ± 2
21	DYRDYL	46 ± 2	59 ± 1	65 ± 2	89 ± 3
22	VLRDYR	47 ± 8	49 ± 3	55 ± 1	91 ± 5
23	KLRDYL	47 ± 5	53 ± 1	63 ± 3	90 ± 2
24	LLRDYG	47 ± 2	59 ± 0	67 ± 5	88 ± 11
25	LVRDYG	51 ± 7	65 ± 3	71 ± 1	No Data
26	LFRDYG	52 ± 5	54 ± 5	54 ± 7	70 ± 4
27	VCRDYR	54 ± 4	72 ± 4	75 ± 5	81 ± 8
28	TLRDYM	55 ± 9	70 ± 6	77 ± 7	94 ± 2
29	ALRDYV	57 ± 4	60 ± 0	67 ± 4	91 ± 4
30	QLRDYS	57 ± 4	64 ± 1	76 ± 7	93 ± 3
31	TWRDYL	61 ± 1	69 ± 3	72 ± 2	91 ± 5
32	YLRDYT	62 ± 7	73 ± 3	83 ± 1	96 ± 1

Table 4.1 (continued): IGF1R-binding Adnectin expression data[1]

Adnectin #	FG Loop[2]	% Adnectin in IB[3] 2 h	4 h	6 h	24 h
33	FIRDYG	63 ± 8	73 ± 7	76 ± 3	91 ± 2
34	LIRDYG	66 ± 6	74 ± 5	82 ± 9	92 ± 4
35	FVRDYF	66 ± 2	76 ± 7	75 ± 12	94 ± 0
36	LLRDYV	68 ± 2	76 ± 2	84 ± 3	95 ± 2
37	FQRDYQ	69 ± 3	90 ± 5	93 ± 1	91 ± 2
38	MFRNYG	72 ± 1	77 ± 6	80 ± 8	89 ± 1
39	ILRDYV	72 ± 2	92 ± 3	91 ± 0	92 ± 3
40	LIRNYG	74 ± 0	73 ± 6	73 ± 3	90 ± 3
41	MFWDYG	91 ± 6	92 ± 3	92 ± 6	93 ± 3

[1]Newly characterized Adnectins indicated by gray background; other data from [101, 102]

[2]Loop sequences colored according to the scheme: blue for R, K; red for D, E; yellow for polar, and green for hydrophobic residues

[3]Average of two experiments ± range

In our experience, amino acid composition-based methods may be useful for the statistical analysis of general trends, but methods that consider the structure of the aggregating polypeptide and/or of the aggregates themselves are more likely to produce insights useful for understanding and solving specific protein solubility problems. Some sliding window methods consider aggregate structure by including β-structure propensity as a predictive factor (in accordance with the observation that intermolecular β-sheet structure is frequently found in protein aggregates). Whole-sequence aggregation propensity scores are sometimes derived from the average of individual sliding window segment scores, a practice equivalent to assuming that all sequence segments of the aggregating polypeptide are equally exposed (as in an unfolded or denatured state). In contrast, higher order (secondary/tertiary) structure-based methods often consider amino acid side chain exposure in native or near-native conformations, explicitly assuming that the tendency of folded polypeptides to associate dominates aggregation propensity/solubility (Fig. 4.3).

We previously postulated that Adnectin aggregation involves some measure of unfolding [102]; in general, this might lead to uncertainty about which class of methods (higher order structure-based or sliding window amino acid sequence-based) is more suitable for this application. We resolve this uncertainty by noting that due to high sequence identity, differences in sliding window-based aggregation/solubility predictions for our reference set of IGF1R-binding Adnectins (Table 4.1) are attributable almost entirely to the amino acid sequence of the FG loop; thus, use of this class of methods does not constitute an assumption of aggregation from an unfolded or denatured state, instead placing focus on a region that is experimentally proven to influence aggregation propensity and will be the

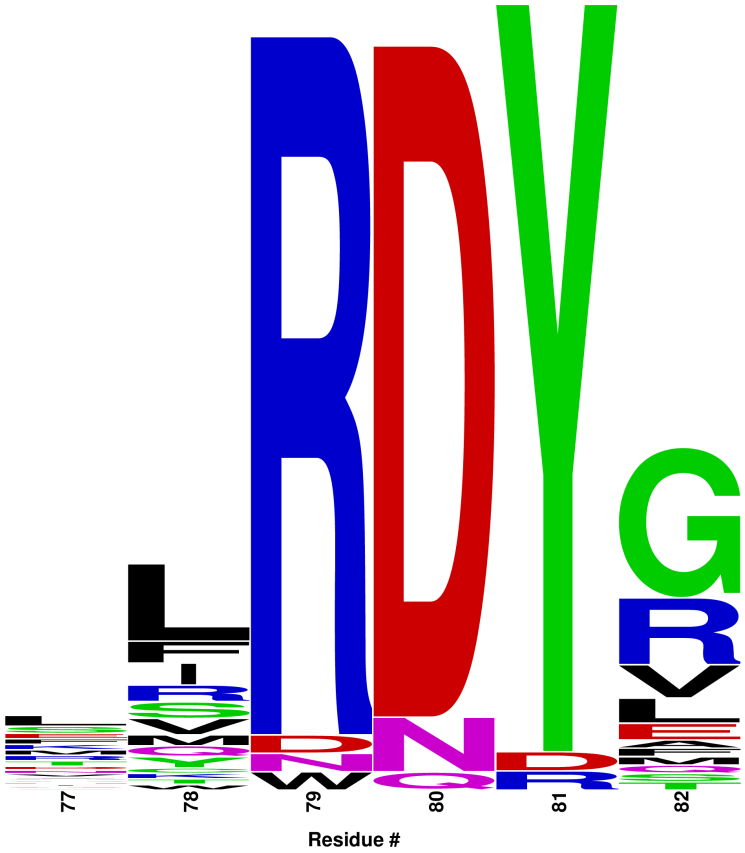


Figure 4.2: An amino acid sequence 'logo' [179] (created using WebLogo [180]) illustrating the degree of diversity at each position in the FG loop (within the set of IGF1R-binding Adnectins for which IB formation propensities have been experimentally determined). The height of each letter is proportional to the frequency of the corresponding amino acid at that position.

target of our solubility engineering efforts (Chapter 5). We will not consider higher order structure-based methods any further, as the required homology modelling adds little value in this particular application.

Some of the 10 newly characterized Adnectins (Table 4.1) have higher experimentally measured aggregation propensities than those previously described [101, 102]. After four or more hours of induction we observe that several Adnectins hit an aggregation 'ceiling'; once more than 90% of expressed protein is found in IB form, differences in aggregation propensity cannot be reliably distinguished. We find that after two hours of induction, better discrimination between Adnectins which have different propensities to form IBs is possible, and therefore compare these data to the amino acid sequence-based predictions of three sliding window methods: AGGRESCAN [143], FoldAmyloid [181], and CamSol [105] (Appendix A). AGGRESCAN and FoldAmyloid have previously been observed to produce predictions that correlate well with Adnectin data [101], while CamSol is the successor to

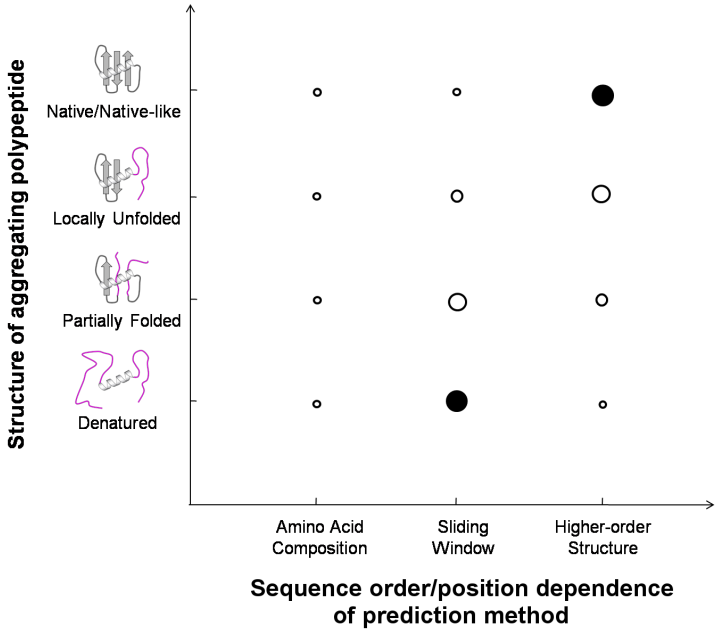


Figure 4.3: Aggregation/solubility prediction methods may (implicitly or explicitly) assume different degrees of structure in the aggregating polypeptide. If anything is known (or suspected) about the conformation of the aggregating polypeptide, the class of method best suited to the problem can be inferred. Increasing size of circles indicates increasing compatibility of assumptions built into the class of methods with the structure of the aggregating polypeptide. Filled circles indicate highest compatibility. Figure reproduced from [104]

Zyggregator [132], which has proven useful in Adnectin-related applications [102]. We find that CamSol produces predictions with the strongest correlation ($R^2 = 0.71$; Fig. 4.4) to our experimental results, outperforming AGGRESCAN ($R^2 = 0.59$) and FoldAmyloid ($R^2 = 0.61$). On this basis, we use CamSol to target and evaluate potential scaffold mutations in Chapter 5.

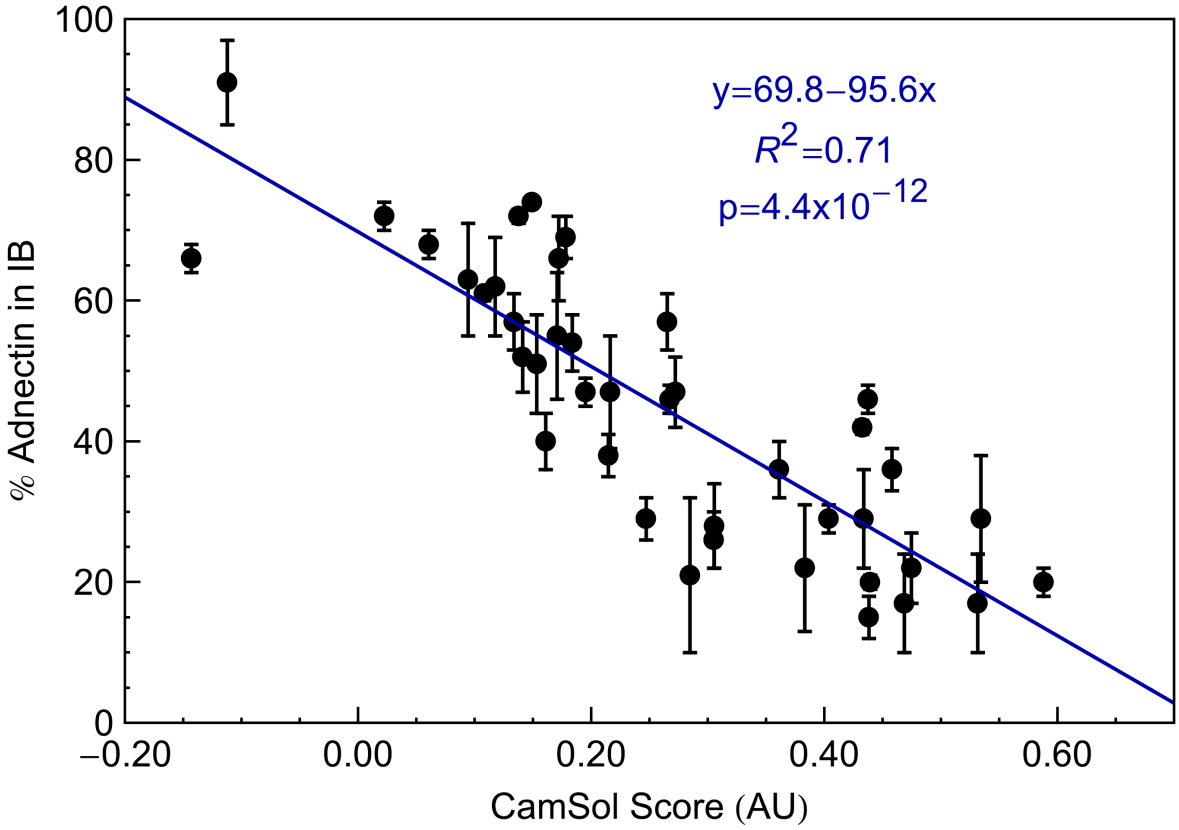


Figure 4.4: Percentages of Adnectins found in inclusion bodies (expressed in *E. coli*; 2 h post-induction) vs. CamSol solubility scores. Each point represents the average of two independent measurements, and the error bars indicate the observed range (average %IB ± (max-min)/2). The blue line shows the result of a linear regression through all of the points. The p-value is a measure of the probability of observing the same correlation in random data.

Chapter 5

Consensus Protein Design for Solubility

5.1 Context

We have established that the amino acid sequence of the FG loop modulates the degree of inclusion body (IB) formation, and thereby the yield of soluble protein, when Adnectins are expressed in *Escherichia coli* (*E. coli*) (Chapter 4). IB formation takes place in a complex intracellular environment and may be influenced by an array of factors (Chapter, 1, Section 1.2); nevertheless, it is a form of protein aggregation, and thus strongly dependent on intrinsic protein solubility. From these IB formation data it may be possible to accurately predict the solubility of as-yet uncharacterized FG loop variants. Such predictions are potentially useful, but Adnectins are engineered binding proteins, and the FG loop plays a prominent functional role. Here, guided by the principles of consensus design and CamSol amino acid sequence-based solubility predictions (Chapter 4), we implement changes to the 'scaffold' (i.e., mutations of β-strand residues) with the goal of increasing Adnectin solubility in a manner that does not place constraints on the composition of functionally important loops. We efficiently narrow the field of candidate mutations to one (V75R) that robustly increases the average solubility of Adnectins that bind insulin-like growth factor 1 receptor (IGF1R).

5.2 Introduction

Protein design can be thought of as an optimization problem for which, if we restrict ourselves to the 20 standard proteinogenic amino acids and disregard potential post-translational modifications or variation in the number of amino acids, the solution space for an N-residue protein consists of 20^N discrete possibilities. It is quite possible that many

of these 20^N sequences do not correspond to well-folded or otherwise useful (e.g., functional, but intrinsically disordered) proteins, but brute force experimental characterization is clearly not a practical optimization technique; though our understanding of protein folding and function is imperfect, we must make design choices that narrow the search to promising regions of the solution space in order to improve our chances of success.

Insights drawn from evolutionary exploration of the solution space (documented by DNA sequencing), and computational tools (physics-based and/or informed by decades of experimental protein characterization) can be used, separately or in combination, to constrain the optimization. For example, even point mutation of a template protein may constitute a use of evolutionary information, e.g., if the new sequence differs from that of an evolved protein, presumably selected for its desirable charactersistics, by a single amino acid (the smallest possible distance between points in the solution space). Computational modelling and prediction algorithms may be applied to identify promising point mutations or, conversely, to screen out those unlikely to meet design targets. It is a testament to the difficulty of the design problem that this apparently conservative strategy offers no guarantee of success [182].

The maturation of reliable and inexpensive DNA sequencing technologies spawned vast, ever-expanding databases of genetic information from diverse species [183]; in order to exploit these databases of DNA sequences in protein design, we operate on the assumption that the proteins encoded within have been selected for by evolutionary processes. This is the basis for techniques such as ancestral sequence reconstruction (inference of phylogenetic trees relating extant homologous proteins [184]) and consensus protein design (identification of conserved residues in the aligned amino acid sequences of homologous proteins [185]). Among many other applications, consensus design has been used to produce several proteins that adopt the same fold (fibronectin type III domain) as the Adnectins and feature remarkable thermodynamic stability [186, 187]. Inadequate thermodynamic stability is a common mode of failure among designed proteins, and must not be neglected; however, here our focus is on improving Adnectin solubility, a related but distinct characteristic - low stability may result in low solubility, but high stability does not necessarily confer high solubility [102, 104].

Approximately 50% of residues found to be conserved in multiple sequence alignments have been associated with thermodynamic stabilization, the remainder being stability neutral or destabilizing [185]. These latter outcomes may be rationalized as the result of evolutionary selection for beneficial traits aside from stability, including function, fast folding kinetics, and solubility. Accordingly, these traits, in addition to thermodynamic stability, may emerge from the consensus design process. On the other hand, just as not all conserved residues contribute to thermodynamic stability, neither are they equally important for solubility. This may be particularly true in the case of sequences homologous to the tenth human fibronectin type III domain (^{10}Fn3), which is believed to participate in a type of functional aggregation (fibrillization) within the extracellular matrix [56]. Here we apply CamSol [105], which produces solubility predictions that correlate with experimentally determined aggregation propensities for a set of IGF1R-binding Adnectins (Chapter 4),

to identify aggregation-prone target regions prior to the consensus design phase, and also to discriminate those conserved (in a multiple sequence alignment) residues most likely to contribute to solubility. We then experimentally test potentially solubilizing mutations suggested by consensus, singly and in combination, and find that one scaffold mutation in particular (V75R) increases the solubility of IGF1R-binding Adnectins with diverse FG loop sequences.

5.3 Methods

5.3.1 Amino Acid Sequence Construction

From the Conserved Domain Database (https://www.ncbi.nlm.nih.gov/cdd) [188] Fn3 multiple sequence alignment 'cd00063', sequences with β-strand G insertions longer than two amino acids were excluded. The remaining β-strand F sequences were extracted and used in the construction of hypothetical Fn3 sequences with the wild-type ^{10}Fn3 sequence at all non-β-strand F positions (Appendix B).

5.3.2 Solubility Prediction

The amino acid sequences constructed as described above (Appendix B) were submitted to the CamSol Intrinsic server (http://www-mvsoftware.ch.cam.ac.uk/index.php/camsolintrinsic) with a pH of 7.0 selected (as in Section 4.3).

5.3.3 Protein Expression

Adnectins were expressed in BL21 (DE3) pLysS cells (Edge Biosystems) transformed with pET-9d vectors into which the Adnectins had been cloned at the NcoI and BamHI sites. Cells were grown overnight at 37°C with shaking at 225 RPM in LB broth containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. The overnight culture was diluted 1:100 into fresh LB broth with the same antibiotics, and incubated at 37°C with shaking at 225 RPM to an A_{600} of 0.6-0.8. Expression was then induced using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

1 mL samples were taken 2, 4, 6, and 24 h after induction. Samples were centrifuged at 5000 g (room temperature) for 10 min and the pellets resuspended in 100 µL TEN buffer (20 mM Tris pH 8.0, 1 mM EDTA, 100 mM NaCl). The resuspended pellets were flash frozen in liquid nitrogen and stored at -80°C. Resuspended cell pellets were subjected to 5 cycles of freezing in liquid nitrogen and thawing in a 25°C water bath. 5 µL of 3 mg/mL DNase I was then added to each microcentrifuge tube, and mixed by gently inverting the tubes 30 times. After a 20 min incubation period, samples were subjected to an additional

5 freeze-thaw cycles. Soluble and insoluble fractions were separated by centrifugation at 16300 g (room temperature) for 15 min, and the supernatants were transferred to new tubes. The pellets were resuspended in 100 µL TEN buffer and all tubes were flash frozen in liquid nitrogen and stored at -80°C.

5.3.4 Solubility Analysis

Samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and stained using Coomassie Blue. The gels were imaged using a BIS303PC gel documentation system (DNR Bio-Imaging Systems) and pixel densities were quantitated using the TotalLab 100 software package (Nonlinear Dynamics). The background (an approximation of the pixel density of protein-free gel) was determined using the "rolling ball" method and subtracted from the Adnectin bands. The percentage of Adnectin found in IBs was calculated as $100 \cdot (\text{Insoluble Integrated Pixel Density})/(\text{Soluble} + \text{Insoluble Integrated Pixel Density})$. More than one Adnectin band may be observed by SDS-PAGE as a consequence of non-enzymatic post-translational gluconoylation [102]. The total Adnectin integrated pixel density for each lane was taken to be the sum of the integrated pixel densities of the Adnectin bands.

5.4 Results & Discussion

The ^{10}Fn3 domain has desirable characteristics that led to its selection as a scaffold for the development of protein biotherapeutics (Chapter 3). Changes to the amino acid sequence of this scaffold must be carefully considered, as each one runs the risk of inadvertently increasing immunogenicity or otherwise decreasing the suitability of the scaffold for biotherapeutic applications. Although consensus design methods may be applied to the development of whole-protein amino acid sequences, these circumstances dictate an approach that favours the fewest scaffold mutations. From experimentally determined aggregation propensity data for a set of IGF1R-binding Adnectins, we know that changes in FG loop amino acid sequence can have a substantial impact on solubility (Chapter 4, Table 4.1). The relative solubilities of FG loop mutants can be predicted using CamSol ($R^2 = 0.71$; Chapter 4, Fig. 4.4), and CamSol-generated per-residue solubility profiles illustrate the impact of the FG loop sequence, as well as showing that the residues of β strand F have the greatest potential to contribute to poor solubility of the scaffold (Fig. 5.1). We hypothesize that β-strand F may be made more accessible by transient unfolding (relative to the native state) of β-strand G, and here we apply the principles of consensus design to help select β-strand F mutations that will improve solubility.

The amino acid sequences of many domains that adopt the Fn3 fold are known. For our consensus design we draw on a carefully curated multiple sequence alignment (MSA) from the Conserved Domain Database (CDD). These MSAs are commonly used to generate

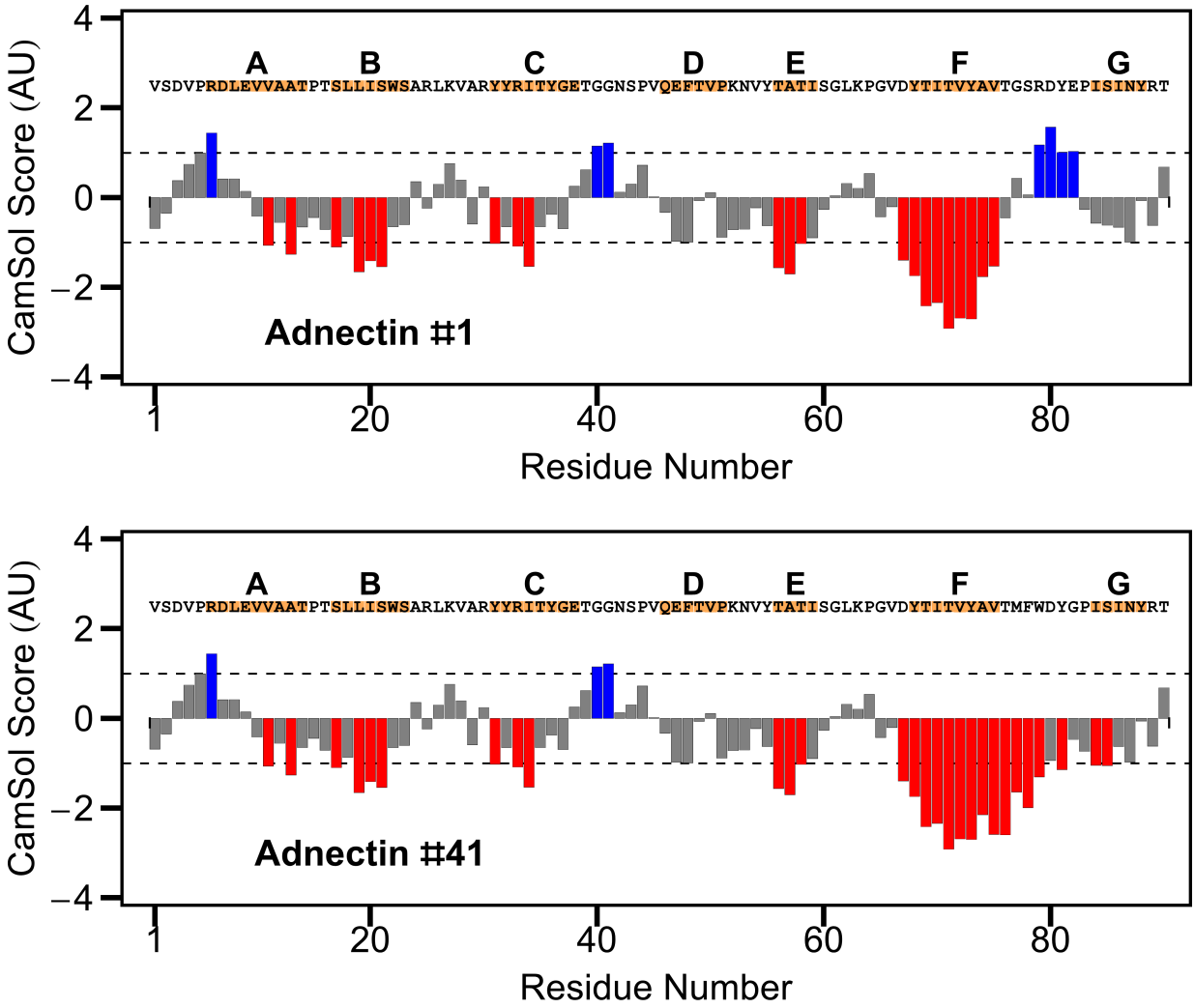


Figure 5.1: CamSol solubility profiles for Adnectins #1 and #41 (most and least soluble, respectively, based on Adnectin expression data presented in Chapter 4, Table 4.1). Per-residue solubility scores are based on a seven residue sliding window average. Scores above 1 indicate regions predicted to contribute to high solubility (blue bars), while scores below -1 indicate regions predicted to contribute to poor solubility (red bars). Amino acid sequences are shown, β-strands residues highlighted orange, and strands A-G labelled above.

position specific scoring matrices (PSSMs) for the prediction of domains within sequences of unknown structure [188], but here we employ one of them as the basis for a systematic study of variation in the predicted solubility of Fn3 domain β-strand F sequences. From several Fn3 domain models available in the CDD we select the MSA from 'cd00063' (curated by the National Center for Biotechnology Information), in which the quality of the alignment is very high - residues in β-strands A-F of the 741 sequences included in this MSA are aligned without gaps or insertions, while β-strand G appears to be less well-conserved. Exclusion of

sequences in which β-strand G has insertions longer than two residues relative to wild-type ^{10}Fn3 leaves 737 Fn3 sequences in total.

We use CamSol to predict the solubility of the CDD-derived β-strand F sequences substituted into the wild-type ^{10}Fn3 sequence (Appendix B), and visualize the amino acid preferences of the 100 sequences predicted to be least soluble and the 100 sequences predicted to be most soluble in the form of sequence logos (Fig. 5.2). The identities of β-strand F residues that have side chains buried in the hydrophobic core (Figs. 5.2 and 5.3, even numbers) are well-conserved across both low and high solubility sequences. In contrast, the identities of residues that have solvent-exposed side chains (Figs. 5.2 and 5.3, odd numbers) differ substantially between low and high solubility sequences. This suggests that solubility-increasing mutations need not target core residues - not an assumption we could safely have made *a priori*, given that the degree of unfolding (relative to the native state) involved in Fn3 aggregation is unknown.

Among low solubility sequences, a general preference for uncharged, but polar amino acids (e.g., threonine, serine, glutamine, and asparagine) is observed at positions 67, 69, and 71. It is quite common to find this class of amino acids in solvent-exposed positions; however, the preference for valine and amino acids with large hydrophobic side chains (e.g., phenylalanine and tyrosine) at positions 73 and 75, also solvent-exposed, is conspicuous. We note that some Fn3 domains are believed to contain 'cryptic' binding sites that may be exposed by mechanical unfolding, e.g., during fibronectin self-association as part of extracellular matrix assembly [189]. One such cryptic binding site is believed to be in β-strand B [56]; however, there is also experimental evidence of a mechanical unfolding pathway in which β-strand G is detached, exposing β-strand F [67]. If the high aggregation propensity of β-strand F is an assembly-related functional feature conserved in a subset of Fn3 domain sequences, the amino acids at positions 73 and 75 may be important contributors.

Among high solubility sequences, the trend at solvent-exposed positions is towards charged amino acids, with arginine and glutamate notably well-represented. Accordingly, we report the results of combining the binding (BC, DE, and FG) loop sequences from an IGFR1-binding Adnectin (#24 in Chapter 4, Table 4.1, hereafter referred to as the 'Parent' Adnectin) with moderately high aggregation propensity, allowing both increases and decreases in solubility to be resolved, with T71E/R, Y73E/R and V75E/R scaffold mutations (Table 5.1). At sites of mutation, careful selection of new codons approximately preserves the codon usage frequency of the Parent DNA sequence (Appendix B, Table B.3).

Mutation of Y73 to arginine or glutamate substantially decreases the solubility of the Parent (higher %IB), as does T71E. T71R produces a modest improvement in solubility (lower %IB), while either V75E or V75R produce a dramatic improvement in solubility. However, the solubility improvements resulting from V75E/R and T71R do not appear to be additive; in fact, either combination results in very poor solubility. From our original set of six scaffold mutations, we are left with two strong candidates at a single position (residue 75) that is poised to affect the solubility of both β-strand F and the FG loop (Fig. 5.4).

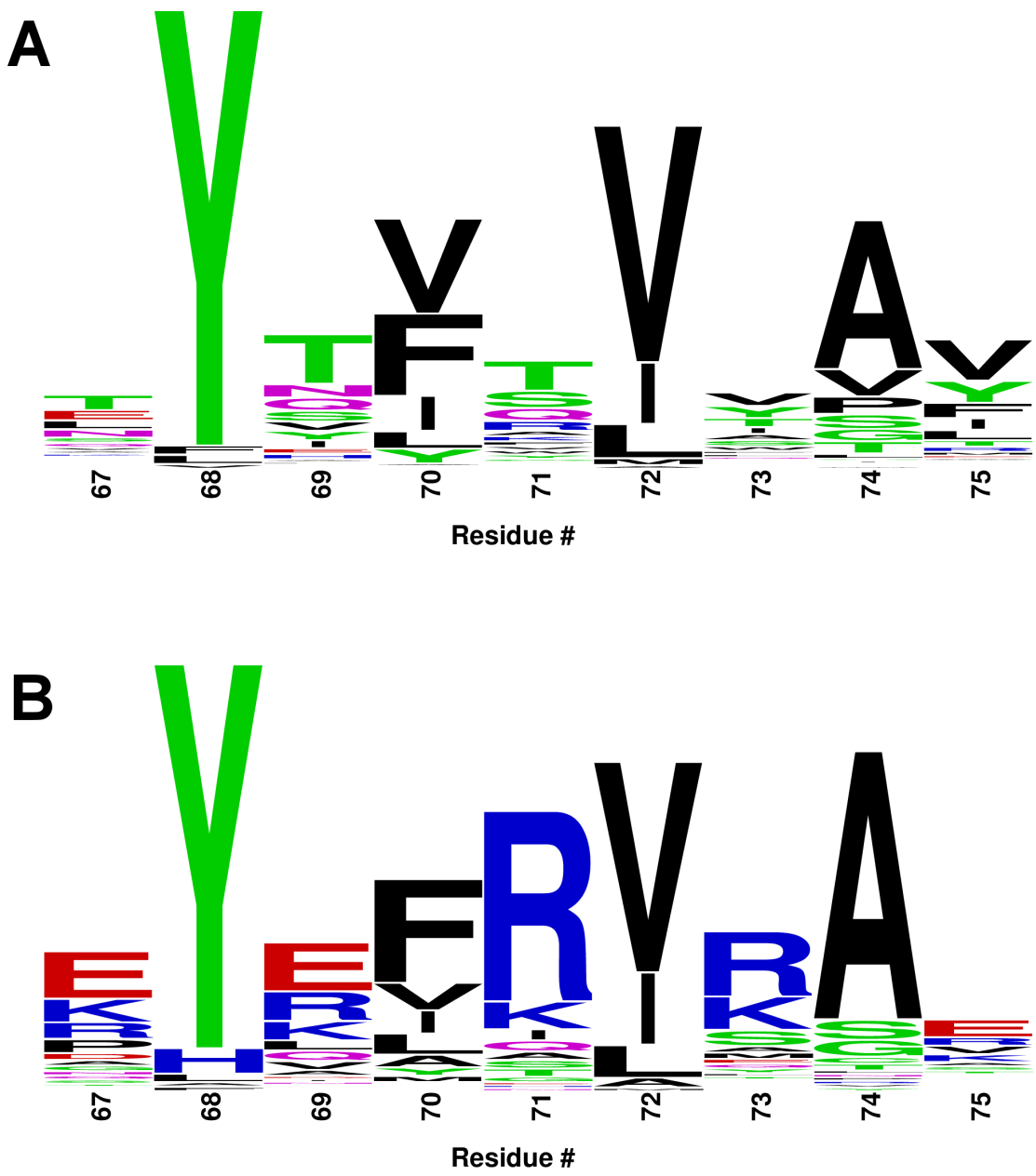


Figure 5.2: Amino acid sequence 'logos' [179] (created using WebLogo [180]) for soluble and insoluble Fn3 β-strand F sequences from the Conserved Domain Database multiple sequence alignment 'cd00063'. A: a logo constructed using the 100 least soluble sequences, as predicted by CamSol. B: a logo constructed using the 100 most soluble sequences, as predicted by CamSol. The height of each letter is proportional to the frequency of the corresponding amino acid at that position. Even residue numbers correspond to inward-facing (core) side chains, while odd residue numbers correspond to outward-facing (solvent-exposed) side chains.

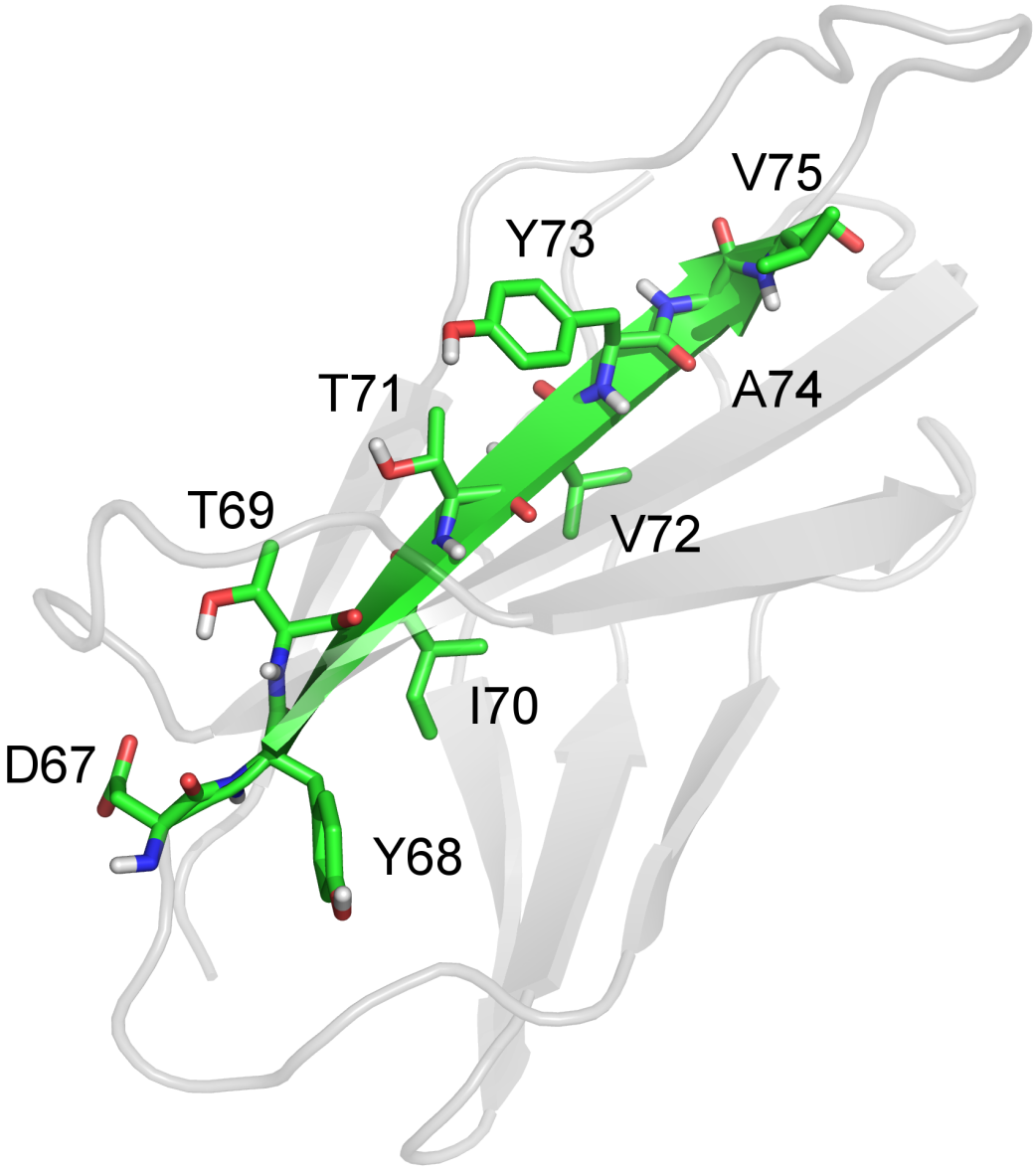


Figure 5.3: β-strand F residues (green; side chains visible) in the Fn3 domain (gray outside of β-strand F). Based on the ^{10}Fn3 domain from the 1FNF crystal structure. ^{10}Fn3/Adnectin β-strand F residue identities are labelled.

The determinants of protein solubility are numerous and complex; in general, we do not expect scaffold mutations to uniformly enhance the solubility of Adnectins with diverse loop sequences. As the amino acid composition of the FG loop is known to strongly influence Adnectin solubility, we assess the ability of our top candidates (V75E/V75R) to generalize by combining them with four additional FG loop sequences (from previously characterized Adnectins) that encompass variation from the Parent FG loop at all positions

Table 5.1: β-strand F scaffold mutations

		% Inclusion Body [4]			
Scaffold Mutations[1]	**Δ %IB[2,3]**	**2 h**	**4 h**	**6 h**	**24 h**
None (Parent)		47 ± 2	59 ± 0	67 ± 5	88 ± 11
T71E	⇑	85 ± 8	91 ± 4	91 ± 5	85 ± 8
T71R	⇓	42 ± 3	46 ± 1	49 ± 4	81 ± 5
Y73E	⇑	66 ± 8	82 ± 7	83 ± 2	97 ± 2
Y73R	⇑	76 ± 13	83 ± 10	93 ± 1	91 ± 2
V75E	⇓	25 ± 7	30 ± 6	35 ± 4	85 ± 2
V75R	⇓	27 ± 6	37 ± 8	44 ± 3	85 ± 5
T71R/V75E	⇑	81 ± 4	91 ± 1	96 ± 1	94 ± 1
T71R/V75R	⇑	82 ± 11	90 ± 4	93 ± 3	93 ± 1

[1]All loop sequences match those of the Parent
[2]Relative to the Parent
[3]Up arrow: higher % IB at 2, 4, and 6 h; Down arrow: lower % IB at 2, 4 and 6 h
[4]Average of two experiments ± range

outside of 'RDY' motif (Table 5.2). Despite showing initial promise, V75E proves to be a poor candidate; however, V75R continues to impress, producing %IB differences that vary (with FG loop sequence) from near-neutral to very solubilizing. In order to further increase diversity, we select 10 more FG loops for combination with the V75R scaffold mutation (for a total of 15; Table 5.3). A near-neutral effect on Adnectin solubility remains the worst case, with an average reduction of 14 %IB at two hours post-induction (Fig. 5.5).

All FG loop sequences tested with the V75R scaffold mutation share the 'RDY' motif; among those for which the change in solubility due to the V75R mutation is near-neutral, we note that additional (outside of the 'RDY' motif) charged residues are prevalent. We hypothesize that solubility enhancement by the V75R mutation is partially attributable to increased inter-Adnectin electrostatic repulsion, and that charged loop residues may confer a similar (evidently non-additive) benefit. However, the relationship between charge and solubility is not straightforward. We observe that both increases and decreases in Adnectin net charge due to FG loop residues may prevent solubility enhancement by the V75R scaffold mutation (Fig. 5.5). Furthermore, the substantial decreases in solubility observed when we mutate Y73 (to glutamate or arginine; Table 5.1) and the failure of the V75E scaffold mutation to replicate the success of V75R (Table 5.2) indicate that both position and the sign of the charge are relevant.

From a vast number of possible mutations, engineering protein solubility requires winnowing down to the relatively small number that can be experimentally evaluated. Taking advantage of available tools and information may help to avoid dead ends and needless labour; for example, the use of protein aggregation prediction algorithms to increase solubility is well-documented (Chapter 4). Here we apply the CamSol algorithm to good effect, but additionally demonstrate the value of incorporating evolutionary information, a resource that may be under-exploited in contemporary protein solubility engineering, into

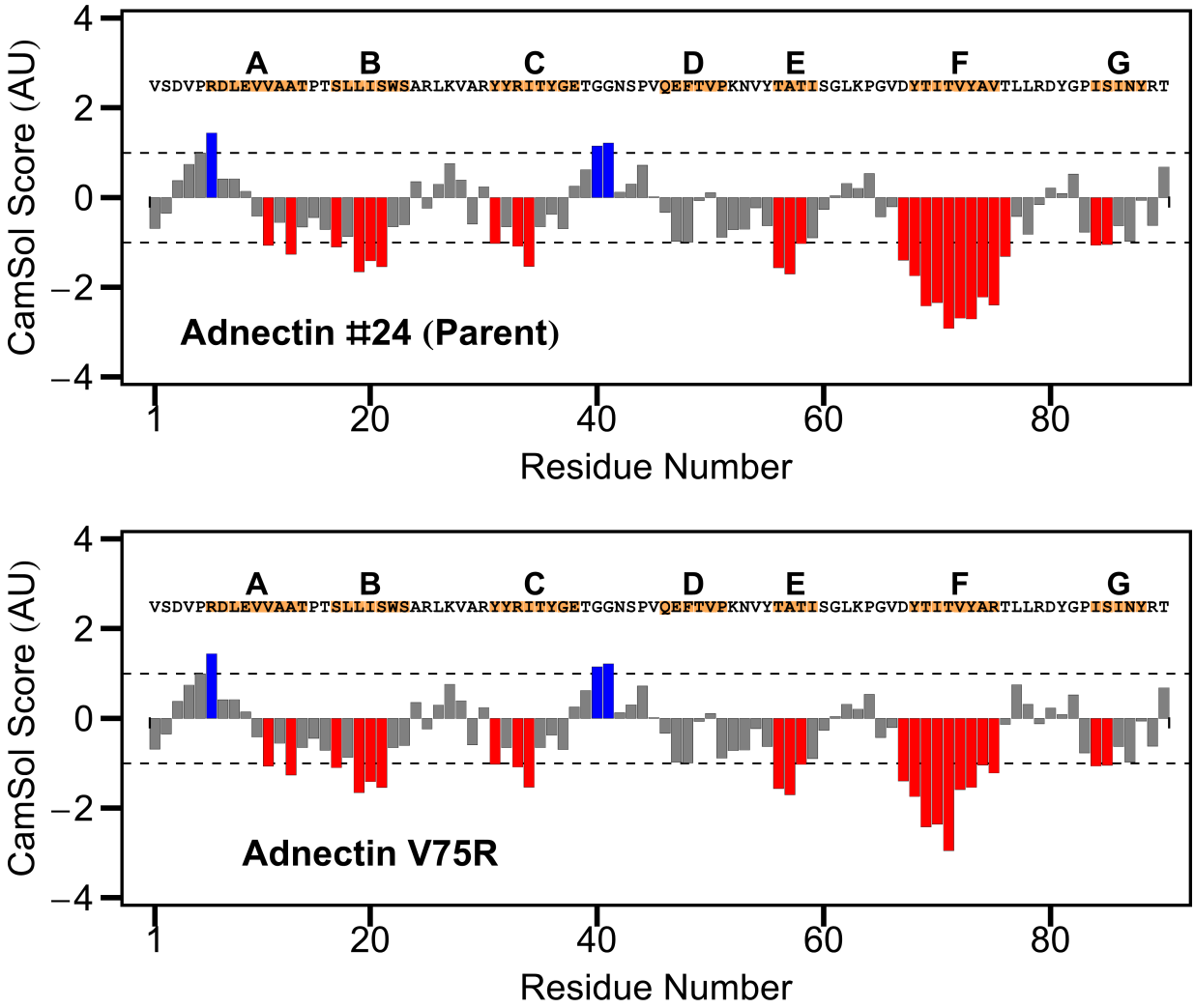


Figure 5.4: CamSol solubility profiles for Adnectins #24 (as numbered in Chapter 4, Table 4.1; also known as the 'Parent' because of its role as a scaffold mutation test case) and V75R (a point mutant of the Adnectin #24). Per-residue solubility scores are based on a seven residue sliding window average. Scores above 1 indicate regions predicted to contribute to high solubility (blue bars), while scores below -1 indicate regions predicted to contribute to poor solubility (red bars). Amino acid sequences are shown, β-strands residues highlighted orange, and strands A-G labelled above.

the decision-making process.

Table 5.2: V75E and V75R scaffold mutations with various FG loop sequences

Adnectin #[1]	FG Loop	Scaffold Mut.	Δ %IB[2,3]	% Inclusion Body[4] 2 h	4 h	6 h	24 h
21	DYRDYL	None		46 ± 2	59 ± 1	65 ± 2	89 ± 3
	DYRDYL	V75E	⇑	56 ± 10	66 ± 6	73 ± 2	94 ± 4
	DYRDYL	V75R		52 ± 4	55 ± 4	72 ± 4	95 ± 3
22	VLRDYR	None		47 ± 8	49 ± 3	55 ± 1	91 ± 5
	VLRDYR	V75E		40 ± 9	57 ± 1	64 ± 1	97 ± 1
	VLRDYR	V75R		49 ± 11	57 ± 6	54 ± 1	86 ± 6
24	LLRDYG	None		47 ± 2	59 ± 0	67 ± 5	88 ± 11
	LLRDYG	V75E	⇓	25 ± 7	30 ± 6	35 ± 4	85 ± 2
	LLRDYG	V75R	⇓	27 ± 6	37 ± 8	44 ± 3	85 ± 5
28	TLRDYM	None		55 ± 9	70 ± 6	77 ± 7	94 ± 2
	TLRDYM	V75E	⇑	76 ± 6	84 ± 0	92 ± 4	94 ± 0
	TLRDYM	V75R	⇓	37 ± 8	47 ± 2	52 ± 2	93 ± 1
36	LLRDYV	None		68 ± 2	76 ± 2	84 ± 3	95 ± 2
	LLRDYV	V75E	⇑	71 ± 5	77 ± 3	85 ± 3	89 ± 0
	LLRDYV	V75R	⇓	52 ± 1	58 ± 3	73 ± 3	92 ± 2

[1]Numbering matches Table 4.1
[2]Relative to the same sequence absent scaffold mutations (gray backgrounds)
[3]Up arrow: higher % IB at 2, 4, and 6 h; Down arrow: lower % IB at 2, 4 and 6 h
[4]Average of two experiments ± range

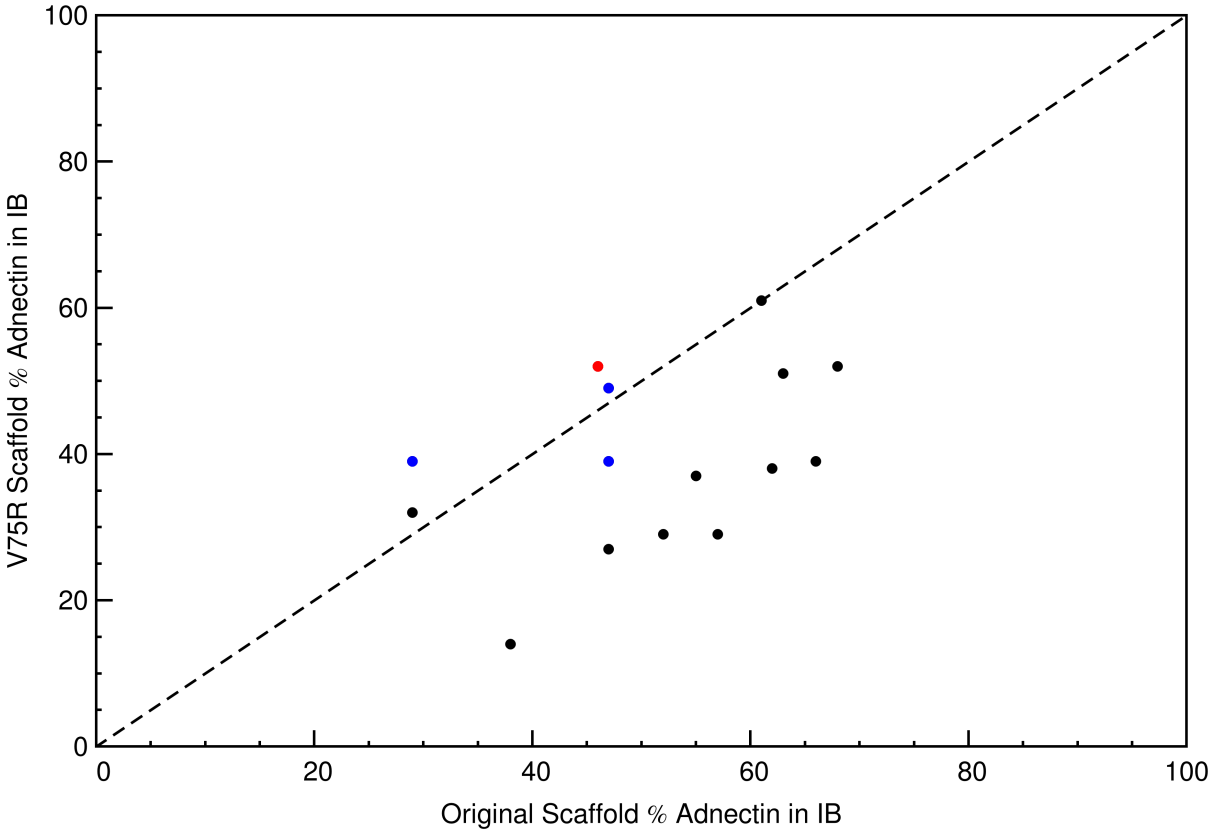


Figure 5.5: Solubility of the V75R scaffold vs. solubility of the original scaffold based on percentages of Adnectins with varying FG loop amino acid sequences found in inclusion bodies (expressed in *E. coli*; 2 h post-induction). Points (black: neutral FG loops; red: negatively charged; blue: positively charged) below the dashed line correspond to Adnectins that are more soluble with the V75R scaffold mutation, while points above the dashed line correspond to Adnectins that are less soluble with the V75R scaffold mutation.

Table 5.3: V75R Adnectins with diverse FG loop sequences

Adnectin #[1]	FG Loop	Scaffold Mut.	% Inclusion Body[2] 2 h	4 h	6 h	24 h
11	KVRDYR	None	29 ± 7	25 ± 9	31 ± 5	81 ± 10
	KVRDYR	V75R	39 ± 4	56 ± 2	56 ± 6	75 ± 11
		Δ Average % IB:	**+10**	**+31**	**+25**	**-7**
14	SLRDYA	None	29 ± 3	42 ± 1	56 ± 1	90 ± 3
	SLRDYA	V75R	32 ± 5	33 ± 1	43 ± 2	87 ± 9
		Δ Average % IB:	**+3**	**-10**	**-13**	**-3**
17	HFRDYG	None	38 ± 3	44 ± 6	55 ± 10	80 ± 6
	HFRDYG	V75R	14 ± 2	21 ± 3	23 ± 2	69 ± 3
		Δ Average % IB:	**-38**	**-44**	**-55**	**-80**
21	DYRDYL	None	46 ± 2	59 ± 1	65 ± 2	89 ± 3
	DYRDYL	V75R	52 ± 4	55 ± 4	72 ± 4	95 ± 3
		Δ Average % IB:	**+6**	**-4**	**+7**	**+6**
22	VLRDYR	None	47 ± 8	49 ± 3	55 ± 1	91 ± 5
	VLRDYR	V75R	49 ± 11	57 ± 6	54 ± 1	86 ± 6
		Δ Average % IB:	**+2**	**+8**	**-1**	**-5**
23	KLRDYL	None	47 ± 5	53 ± 1	63 ± 3	90 ± 2
	KLRDYL	V75R	39 ± 9	57 ± 7	56 ± 16	93 ± 5
		Δ Average % IB:	**-9**	**+4**	**-8**	**+3**
24	LLRDYG	None	47 ± 2	59 ± 0	67 ± 5	88 ± 11
	LLRDYG	V75R	27 ± 6	37 ± 8	44 ± 3	85 ± 5
		Δ Average % IB:	**-20**	**-22**	**-23**	**-3**
26	LFRDYG	None	52 ± 5	54 ± 5	54 ± 7	70 ± 4
	LFRDYG	V75R	29 ± 4	31 ± 5	35 ± 14	91 ± 3
		Δ Average % IB:	**-24**	**-24**	**-19**	**+21**
28	TLRDYM	None	55 ± 9	70 ± 6	77 ± 7	94 ± 2
	TLRDYM	V75R	37 ± 8	47 ± 2	52 ± 2	93 ± 1
		Δ Average % IB:	**-18**	**-23**	**-25**	**-1**
29	ALRDYV	None	57 ± 4	60 ± 0	67 ± 4	91 ± 4
	ALRDYV	V75R	29 ± 6	29 ± 4	32 ± 8	88 ± 1
		Δ Average % IB:	**-28**	**-28**	**-13**	**-4**
31	TWRDYL	None	61 ± 1	69 ± 3	72 ± 2	91 ± 5
	TWRDYL	V75R	61 ± 12	71 ± 13	81 ± 5	96 ± 0
		Δ Average % IB:	**0**	**-24**	**+9**	**+5**

[1]Numbering matches Table 4.1
[2]Average of two experiments ± range

Table 5.3 (continued): V75R Adnectins with diverse FG loop sequences

Adnectin #[1]	FG Loop	Scaffold Mut.	% Inclusion Body[2] 2 h	4 h	6 h	24 h
32	YLRDYT	None	62 ± 7	73 ± 3	83 ± 1	96 ± 1
	YLRDYT	V75R	38 ± 2	55 ± 1	63 ± 3	96 ± 0
		Δ Average % IB:	**-24**	**-19**	**-21**	**0**
33	FIRDYG	None	63 ± 8	73 ± 7	76 ± 3	91 ± 2
	FIRDYG	V75R	51 ± 12	68 ± 6	74 ± 4	96 ± 1
		Δ Average % IB:	**-12**	**-6**	**-3**	**+5**
34	LIRDYG	None	66 ± 6	74 ± 5	82 ± 9	92 ± 4
	LIRDYG	V75R	39 ± 4	51 ± 4	61 ± 5	90 ± 0
		Δ Average % IB:	**-27**	**-24**	**-21**	**-2**
36	LLRDYV	None	68 ± 2	76 ± 2	84 ± 3	95 ± 2
	LLRDYV	V75R	52 ± 1	58 ± 3	73 ± 3	92 ± 2
		Δ Average % IB:	**-16**	**-18**	**-11**	**-3**

[1]Numbering matches Table 4.1
[2]Average of two experiments ± range

Chapter 6

Protein Association-Induced Amide Hydrogen/Deuterium Exchange Curtailed by Design for Solubility

6.1 Context

Consistent with our hypothesis that transient separation of Fn3 β-strands F and G may promote aggregation, amide hydrogen/deuterium (H/D) exchange experiments at pD 7.0 show no measurable protection of wild-type ^{10}Fn3 β-strand G residues involved in inter-strand hydrogen bonds [73]. Here, taking advantage of slower exchange at pD 4.6, we quantify the protection of the corresponding pWT and Parent Adnectin residues. Using static and dynamic light scattering, we demonstrate that V75R (closely related to the Parent, but designed to be more soluble; Chapter 5) is less prone to transiently self-associate; furthermore, protection against amide H/D exchange of V75R β-strand G residues is strikingly higher. We propose that transient inter-Adnectin interactions lead to a state in which the amides of β-strand G are solvent-exposed, while their native state hydrogen bond partners in β-strand F remain protected within the protein-protein interface.

6.1.1 Acknowledgements

Colleen Doyle assisted with the collection of pWT and Parent Adnectin NOESY-HSQC and TOCSY-HSQC spectra, as well as analysis thereof for the purpose of assigning ^{1}H-^{15}N cross peaks to particular residues. Avril Metcalfe-Roach assisted with the collection of V75R amide H/D exchange data, conducted light scattering experiments, and analyzed data.

6.2 Introduction

Amide hydrogen/deuterium (H/D) exchange experiments are of significant interest because they report on the energetics of protein structural dynamics. Protonated protein dissolved in deuterium oxide (D_2O) exchanges amide protons for solvent deuterons over time. The pK_a values of protein backbone amides are quite high; consequently, the exchange of amide protons with solvent protons/deuterons is catalyzed only by the strongest acids and bases found in aqueous solution: H/D_3O^+ and OH/D^- ions [190]. Exchange rates, particularly those of amides sequestered from solvent interactions or protected from exchange by involvement in protein-protein hydrogen bonds, often fall within a range that can be monitored by mass spectrometry or nuclear magnetic resonance (NMR) spectroscopy. Furthermore, rates of exchange can be managed experimentally through control of the pH/pD (i.e., the concentration of catalyst). Above a pH/pD of approximately 3, base-catalyzed exchange is dominant, thus rates of exchange are reduced 10-fold with each drop of 1 pH/pD unit [191].

Amide H/D exchange is often modelled as a two-step process (Eq. 6.1): protective structure 'opens' and the amide adopts an exchange-competent state, then either the protective structure 'closes' up again, or the amide proton is exchanged for a solvent deuteron (an essentially irreversible process due to the vast excess of deuterons). The observed rate of exchange (k_{ex}) corresponding to this reaction scheme can be derived using a steady-state approximation (Eq. 6.2) [191]. Under so-called EX2 (bimolecular exchange) conditions, generally considered to prevail below a pD of approximately 8 (particularly if the protective structure is thermodynamically stable), the assumption that $k_{ch} << k_{cl}$ yields the simplification shown in Eq. 6.3. In this case, k_{ch} is so small that the protected and exchangeable conformations achieve pseudo-equilibrium (Eq. 6.1), and k_{ex} reports on the free energy difference between them.

$$\mathrm{N{-}H(protected)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \mathrm{N{-}H(exchangeable)} \xrightarrow{k_{ch}} \mathrm{N{-}D} \tag{6.1}$$

$$k_{ex} = \frac{k_{op} \cdot k_{ch}}{k_{cl} + k_{ch}} \tag{6.2}$$

$$k_{ex} = \frac{k_{op} \cdot k_{ch}}{k_{cl}} \tag{6.3}$$

In crystal and solution NMR structures of ^{10}Fn3 [192–194], backbone-backbone hydrogen bonds may be inferred between β-strands F and G; in particular, the amide nitrogens of three β-strand G residues (I88, I90, Y92) act as donors in hydrogen bonds with β-strand F carbonyl oxygen acceptors (Fig. 6.1). In contrast with the static picture offered by these structures, amide H/D exchange experiments conducted at pD 7.0 reveal that the three nominally hydrogen-bonded β-strand G amides are poorly protected from amide exchange

(no measurable protection in the cases of I90 and Y92) [73]. A qualitatively similar lack of β-strand G protection against H/D exchange (monitored by mass spectrometry) has been observed for an Adnectin that, like those studied here, has BC, DE, and FG loops that differ in amino acid sequence from wild-type ^{10}Fn3 [96]. These findings suggest the population of one or more alternate (non-native) states.

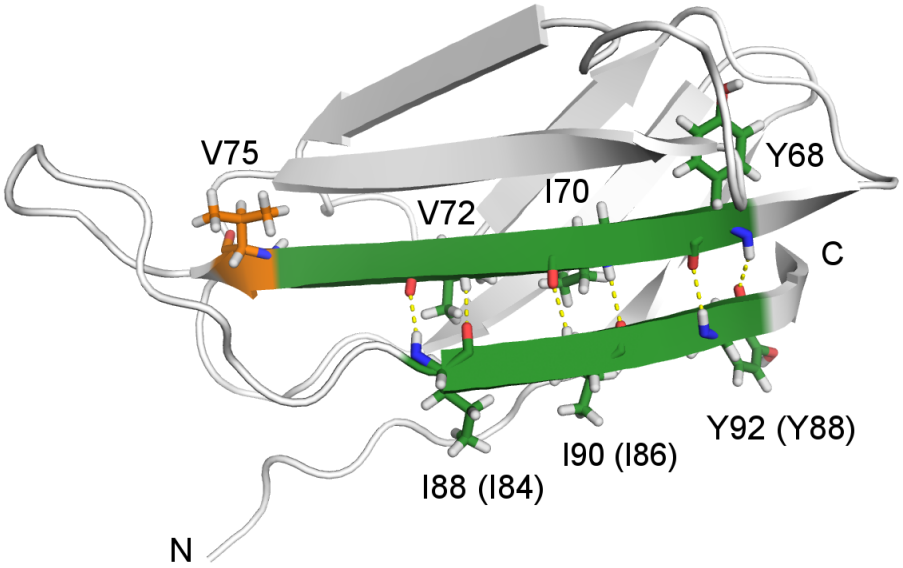


Figure 6.1: An illustration of the backbone-backbone hydrogen bonds joining β-strands F and G of ^{10}Fn3 (from PDB structure 1TTF). Where residue numbers differ between pWT and the other Adnectins due to FG loop length, residues are labelled 'pWT identifier (other Adnectin identifier)'. V75 is highlighted in orange, β-strands F (except for V75) and G are green, and backbone-backbone hydrogen bonds between β-strands F and G are indicated by dashed yellow lines. All atoms are shown for residues Y68, I70, V72, V75, I88 (I84), I90 (I86), and Y92 (Y88); carbon atoms are coloured to match the cartoon backbone, nitrogens are blue, oxygens are red, and hydrogens are white. N and C termini are labelled.

Here we report the results of amide H/D exchange experiments designed to shed light on these alternate states. We characterize fast H/D exchange by Adnectin amides, including those in β-strand G for which no protection could be measured in earlier experiments, by lowering the pD and utilizing the SOFAST HMQC pulse sequence [195] for rapid acquisition (<5 minutes) of ^{1}H-^{15}N correlation spectra. We find that the pWT and Parent Adnectins show patterns of β-strands F and G amide H/D exchange protection similar to those previously reported for wild-type ^{10}Fn3 [73]. We propose a model in which β-strand G residues occasionally adopt an exchange-competent state, while their native state hydrogen bond partner residues in β-strand F remain protected due to transient inter-Adnectin

interactions. A similarly surprising lack of protection against amide H/D exchange at pD 7.0 has been reported for wild-type ^{10}Fn3 β-strands A and B [73]; however, the hydrogen bonding between β-strands A and B is less regular, featuring few reciprocal bonds like those that facilitate comparisons of β-strand F and G amide H/D exchange (Fig. 6.1). We do not rule out N-terminal involvement in Adnectin self-association, but neither can we confirm it on the basis of these data.

In any case, the β-strand F/FG loop/β-strand G region of Adnectins is of particular interest, as it is the focus of our solubility characterization and engineering efforts (Chapters 4 and 5, respectively). We demonstrate that an Adnectin designed for increased solubility (V75R; Chapter 5) shows a lower propensity for transient self-association in solution (as measured by second virial coefficients from fits of static light scattering data to Eq. 6.9), and an unambiguous increase in the protection against amide H/D exchange of β-strand G residues involved in inter-strand hydrogen bonding. Based on these observations, we hypothesize that inter-Adnectin association is a prerequisite for the adoption of an exchange-competent state by these β-strand G residues; the V75R mutation discourages intermolecular associations and thus decreases rates of H/D exchange relative to the pWT and Parent Adnectins.

6.3 Methods

6.3.1 Expression & Purification

Expression and purification of ^{15}N-labelled protein were conducted as described in Chapter 7, Section 7.3.

6.3.2 ^{1}H-^{15}N Cross Peak Assignment

Cross peak assignment was conducted as described in Chapter 7, Section 7.3.

6.3.3 Amide Hydrogen/Deuterium Exchange

Each Adnectin NMR sample for amide H/D exchange was prepared in 40 mM sodium citrate pH 4.1 or 40 mM sodium phosphate pH 6.5 and concentrated to a volume of 450 µL. 50 µL of D_2O was added (final protein concentration ~30 mg/mL) and mixed by pipetting up and down, then the sample was transferred to a clean NE UL-5 NMR tube (New Era). Following reference ^{1}H-^{15}N SOFAST HMQC spectrum acquisition, the sample was lyophilized overnight in a FreeZone 4.5 freeze dry system (Labconco). The lyophilized sample was then dissolved in 500 µL of D_2O. Further ^{1}H-^{15}N SOFAST HMQC spectra were acquired in under five minutes each, for a minimum total acquisition time of 24 hours.

Subsequent ^{1}H-^{15}N SOFAST HMQC spectra were collected at varying intervals (2-7 days). All spectra were acquired at a temperature of 298 K using a Bruker AVANCE 700 MHz spectrometer. Following the final acquisition, the pD of the NMR sample was measured using an Accumet AB15 pH meter with an Orion ROSS Sure-Flow electrode (Thermo Fisher Scientific). As is common practice, we add a correction factor of 0.4 pD units to the value reported by the pH meter in order to account for the effect of D_2O on the glass electrode (Eq. 6.4, where pD^* is the uncorrected meter reading) [196]. Although there is some disagreement about the necessity of such a correction [197], we follow this convention in order that our data remain consistent and comparable with previously published protein and peptide H/D exchange data.

$$pD = pD^* + 0.4 \tag{6.4}$$

Exchange at pD 4.6

SOFAST HMQC (Bruker pulse program 'sfhmqcf3gpph' [195]; Section E.1) spectra were processed using Bruker TopSpin 3.5. The 'nmrglue' Python package [198] was used to read this processed data, and automatic peak area definition and peak volume integration algorithms were implemented in Python. Intensity thresholds that permitted resolution of peaks without apparent overlap in the first post-D_2O dissolution SOFAST HMQC spectrum were used to define peak areas (i.e., each peak area is a region of contiguous points above the threshold). Cross peak volumes in this and all subsequent SOFAST HMQC spectra were determined by numeric integration over these previously defined areas. Because of the pD and length of these experiments, no measurable H/D exchange was observed for many of the amide protons involved in backbone-backbone hydrogen bonds. One such amide proton (that of I59), selected for its high level of protection against exchange and the absence of any overlapping peaks in the SOFAST HMQC spectra, was used as an internal standard for the normalization of all peaks in each spectrum. Each individual peak was further normalized such that the peak volume in the first post-D_2O dissolution SOFAST HMQC spectrum was equal to 1.0 (Eq. 6.5; $\mathrm{I59}_{integrated}$: reference peak, volume in the spectrum from the current time point; $\mathrm{I59}_{initial}$: reference peak, volume in the first spectrum; $\mathrm{Peak\ Volume}_{integrated}$: peak of interest, volume in the spectrum from the current time point; $\mathrm{Peak\ Volume}_{initial}$: peak of interest, volume in the first spectrum).

$$\text{Normalized Peak Volume} = \frac{\mathrm{I59}_{initial}}{\mathrm{I59}_{integrated}} \cdot \frac{\mathrm{Peak\ Volume}_{integrated}}{\mathrm{Peak\ Volume}_{initial}} \tag{6.5}$$

Exchange at pD 7.0

In contrast to the pD 4.6 experiments, exchange was measurable for all amide protons. This precludes the use of the I59 peak as an internal standard, and the normalization was

instead carried out only with respect to the peak volume in the first post-D_2O dissolution SOFAST HMQC spectrum (Eq. 6.6; Peak Volume$_{integrated}$: peak of interest, volume in the spectrum from the current time point; Peak Volume$_{initial}$: peak of interest, volume in the first spectrum).

$$\text{Normalized Peak Volume} = \frac{\text{Peak Volume}_{integrated}}{\text{Peak Volume}_{initial}} \tag{6.6}$$

Data Fitting

Exchange constants for each residue were determined by nonlinear least squares fitting of peak volume vs. time data to a single exponential decay (Eq. 6.7) using the 'curve_fit' function from the SciPy Python library (Appendix C, Table C.1). For fast-exchanging amide protons the normalized peak volume (Eq. 6.7) rapidly approaches C, and later time points contribute little to the fit except error. In particular, after the initial 24 h period, acquisition of additional spectra required reinsertion of the sample into the spectrometer (along with the requisite matching, tuning, and shimming), and they were therefore collected under slightly variable conditions. To avoid fitting noise unnecessarily, some time series were truncated, i.e., for fast-exchanging amide protons only data up to the point where the slope of the tangent to the curve approaches zero were included in the fits.

Conversely, for very slowly exchanging amides, variability in normalized peak volume not attributable to amide H/D exchange by the residue of interest (e.g., due to incomplete T1 relaxation between experiments or partial overlap with a faster-exchanging peak) was sometimes evident near the beginning of the initial 24 h. Accordingly, in cases where the slope of the tangent to the curve does not approach zero (at the final data point), only data from 8 h and onward were included in the fits.

$$\text{Normalized Peak Volume} = A \cdot e^{-k_{ex} \cdot t} + C \tag{6.7}$$

Apparent Free Energy of Exchange

For reference, we calculate the apparent free energy of exchange for all residues (where possible) using Eq. 6.8 (Table 6.2; Appendix C, Table C.3). Use of this equation implies assumption of EX2 conditions [191, 199]; the validity of this assumption may vary among residues (discussion below). Intrinsic exchange rate constants (k_{int}) were determined from peptide data [200] using the SPHERE web server (https://protocol.fccc.edu/research/labs/roder/sphere) [201] (Appendix C, Table C.2); options selected include: poly D-L alanine reference data, exchange in D_2O, temperature of 25°C, and pH meter readings of 4.2 or 6.6 (for experiments at pD 4.6 and 7.0, respectively).

$$\Delta G_{ex}^{app} = -RT \cdot ln\left(\frac{k_{ex}}{k_{int}}\right) \tag{6.8}$$

As described in Section 6.2, in the EX2 limit ($k_{ch} << k_{cl}$) k_{ex} reports on the free energy difference between protected and exchangeable conformations. The assumption of EX2 conditions holds best at low pH/pD (low k_{ch}) and for thermodynamically stable exchange-protective protein structure such as intra-β-sheet hydrogen bonds (implying high k_{cl}). For many V75R residues we are able to calculate ΔG_{ex}^{app} at only one of pD 4.6 or pD 7.0; however, for two of three V75R residues with amides involved in intra-β-sheet hydrogen bonds, we calculate essentially the same ΔG_{ex}^{app} at pH 4.6 and 7.0 (Table 6.1). This is consistent with the amide H/D exchange literature [190], and supports validity of the EX2 assumption for the residues listed in Table 6.2 (upon which our conclusions depend), but not necessarily for residues with amides not involved in intra-β-sheet hydrogen bonds. We note that the ionizable side chain of D23 may account for its slightly larger pD 4.6/7.0 ΔG_{ex}^{app} discrepancy.

Table 6.1: Dependence of calculated ΔG_{ex}^{app} on pD

	ΔG_{ex}^{app} (kcal/mol)	
V75R Residue	**pD 4.6**	**pD 7.0**
S17	5.8	5.9
D23	3.9	5.6
I84	4.1	4.3

6.3.4 Light Scattering

Sample Preparation

All tubes and pipette tips were washed using 50% nitric acid to remove dust and debris, then washed with Milli-Q water (EMD Millipore) and air dried. Purified 10 mg/mL Adnectin in 40 mM sodium citrate buffer pH 4.0 and additional 40 mM sodium citrate buffer pH 4.0 were centrifuged at 20000 g for 15-30 minutes to sediment insoluble contaminants, and the supernatants were retained. Dilutions of the Adnectin samples were prepared using the additional buffer and filtered using Whatman Anotop 0.02 µm syringe filters (Sigma-Aldrich). After filtration the dilutions were thermally pre-equilibrated in a 298 K water bath, and light scattering measurements were conducted promptly in order to minimize the impact of protein aggregation on the results (because the intensity of scattered light depends nonlinearly on size, even a small number of aggregates or large oligomers may be unacceptable).

Dynamic & Static Light Scattering

All light scattering data were collected using a Zetasizer Nano ZS (Malvern). The size distribution of particles in solution was determined for each diluted sample by dynamic

light scattering (DLS). The size distribution peak (or fraction thereof) between 3-7 nm was designated the 'target peak'. Static light scattering (SLS) results from V75R Adnectin samples in which >75% of the scattered light intensity was attributable to the target peak are reported. For the Parent Adnectin, which is more aggregation-prone, we relax this restriction to >65% of the scattered light intensity in order to retain a sufficient number of data points; consequently, more error is associated with the Parent data. Debye plots (KC/R_θ vs. C, where K is a constant, R_θ is the Rayleigh ratio, and C is the protein concentration) of the data were created, and second virial coefficients (A_2), as described by the Rayleigh equation (Eq. 6.9, where MW is the molecular weight of the protein) were calculated via linear regression [202]. The constant K is defined in Eq. 6.10, where λ is the wavelength of the laser, N_A is the Avogadro constant, n_0 is the refractive index of the solvent, and dn/dC is the refractive index increment.

$$\frac{KC}{R_\theta} = \frac{1}{MW} + 2 \cdot A_2 \cdot C \tag{6.9}$$

$$K = \frac{4\pi^2}{\lambda^4 \cdot N_A}(n_0 \cdot \frac{dn}{dC})^2 \tag{6.10}$$

Final Adnectin Concentrations

After completion of the SLS measurements, the final (i.e., post-filtration) Adnectin concentration in each sample was determined by absorbance at 280 nm using extinction coefficients calculated by submitting the Adnectin amino acid sequence (Appendix D) to the ProtParam web server [203].

6.4 Results & Discussion

The pWT Adnectin amino acid sequence differs from that of the wild-type ^{10}Fn3 domain by a single aspartate to glutamate mutation in the FG loop (D80E; Fig. 2.2; Appendix D); our data, collected at pD 4.6, are complementary to ^{10}Fn3 amide H/D exchange data collected at pD 7.0 [73], as different exchange regimes (i.e., 'slow' vs. 'fast' exchanging amides) are probed due to the dependence of k_{ex} on pD (Section 6.2). The sequences of the three Adnectins studied here are identical outside of the BC, DE, and FG loops, and the k_{ex} values determined for residues in structurally equivalent positions outside of these loops can be directly compared. Comparisons between non-equivalent residues are facilitated by the calculation of ΔG_{ex}^{app} values, which also provide a useful estimate of the free energy gaps between protected and exchangeable states. The values of ΔG_{ex}^{app} determined for amides involved in hydrogen bonding between β-strands F and G are of particular interest (Table 6.2).

Table 6.2: Apparent free energies of exchange[1] in Adnectin β-strands F and G

			ΔG_{ex}^{app} (kcal/mol)			
			pD 4.6			pD 7.0
β-strand	**Residue[2]**	**HB Partner[2,3]**	**pWT**	**Parent**	**V75R**	**V75R**
F	Y68	Y92/Y88	High	High	High	7.8
F	I70	I90/I86	High	High	High	7.8
F	V72	I88/I84	High	High	High	8.3
G	I88/I84	V72	3.9	N/A	4.1	4.3
G	I90/I86	I70	5.3	4.9	High	6.7
G	Y92/Y88	Y68	5.2	5.0	High	7.9

[1]'High': less than one H-D exchange half-life is observed (k_{ex} cannot be determined with high accuracy, nor ΔG_{ex}^{app} calculated; 'N/A': missing assignment

[2][pWT residue]/[other Adnectins residue], where necessary

[3]Hydrogen bond partner (backbone carbonyl oxygen of the residue indicated)

Consistent with our findings, it has been reported that amide H/D exchange by ^{10}Fn3 residues 90 and 92 at pD 7.0 was too fast to measure, and calculated ΔG_{ex}^{app} values for residues 68, 70, and 72 fall into a range of 7.1 to 8.9 kcal/mol [73]. For the pWT and Parent Adnectins, the degree of protection against amide H/D exchange for β-strand G residues is substantially lower than that for β-strand F residues (Table 6.2) - a curious discrepancy considering the reciprocal hydrogen bonding arrangement (typical of antiparallel β-sheet structure) observed in the native structure (Fig. 6.1). This strongly suggests the transient population of a non-native conformation in which β-strand G residues access an exchange-competent state, while β-strand F residues remain protected. Given the contrast between pWT/Parent Adnectin and V75R results (Figs. 6.2-6.4), we propose that this mechanism of protection may involve intermolecular association between Adnectins.

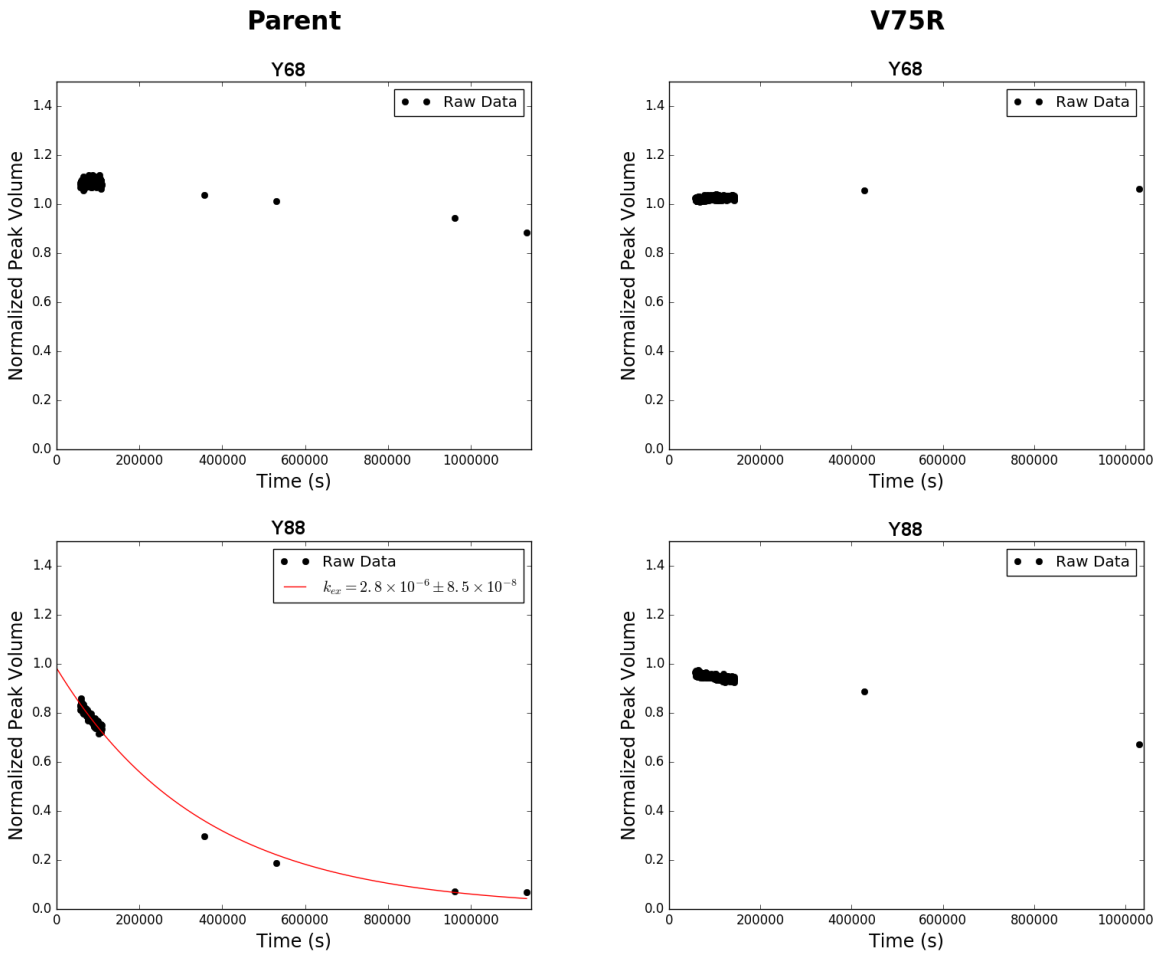


Figure 6.2: Adnectin amide H/D exchange at pD 4.6 monitored by ^{1}H-^{15}N SOFAST HMQC: a comparison of reciprocally hydrogen bonded residues Y68 and I88 in the Parent and V75R Adnectins. Black dots show integrated peak volumes as a fraction of the initial peak volume and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange over the course of the experiment). In each panel, the result of fitting the experimental data to an exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red if at least one H/D exchange half-life is captured.

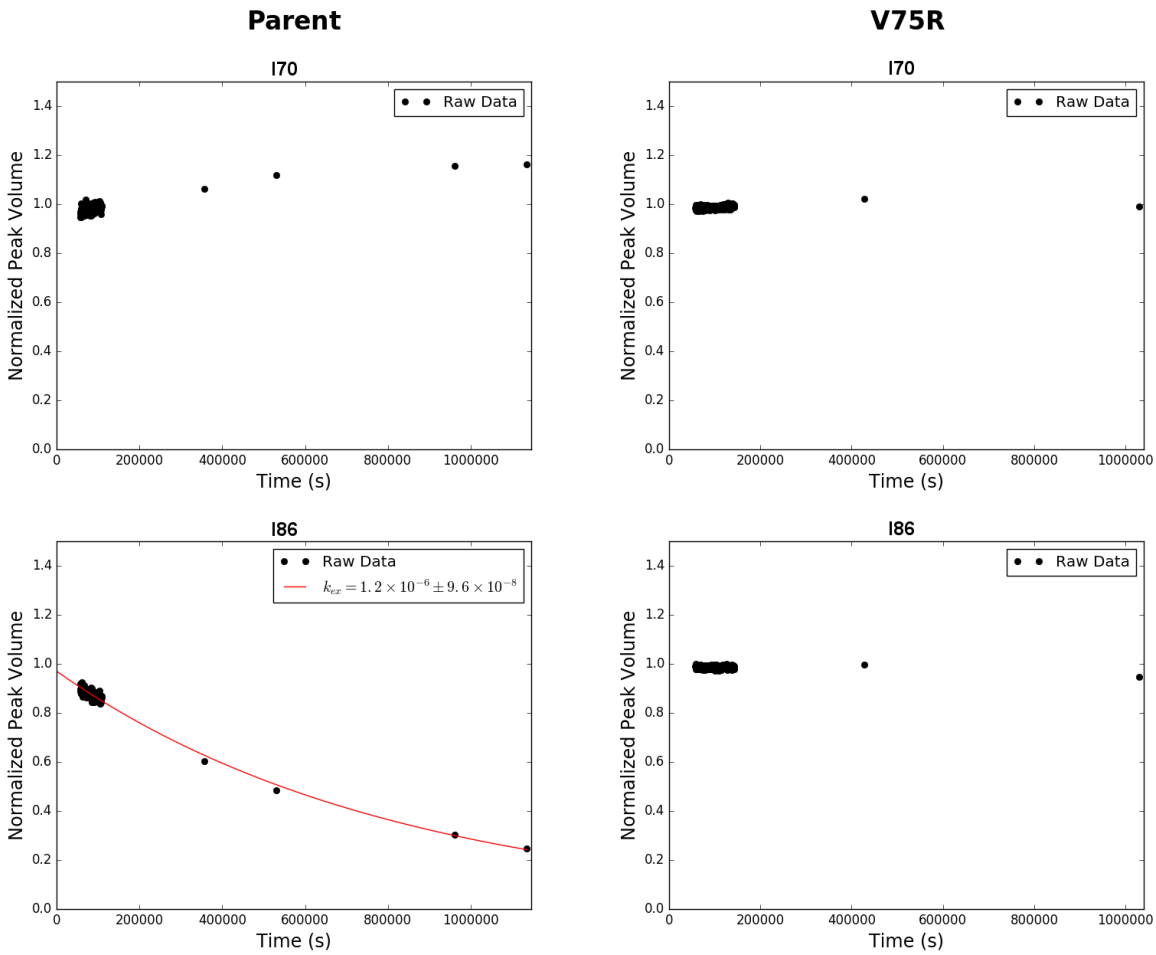


Figure 6.3: Adnectin amide H/D exchange at pD 4.6 monitored by ^{1}H-^{15}N SOFAST HMQC: a comparison of reciprocally hydrogen bonded residues I70 and I86 in the Parent and V75R Adnectins. Black dots show integrated peak volumes as a fraction of the initial peak volume and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange over the course of the experiment). In each panel, the result of fitting the experimental data to an exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red if at least one H/D exchange half-life is captured.

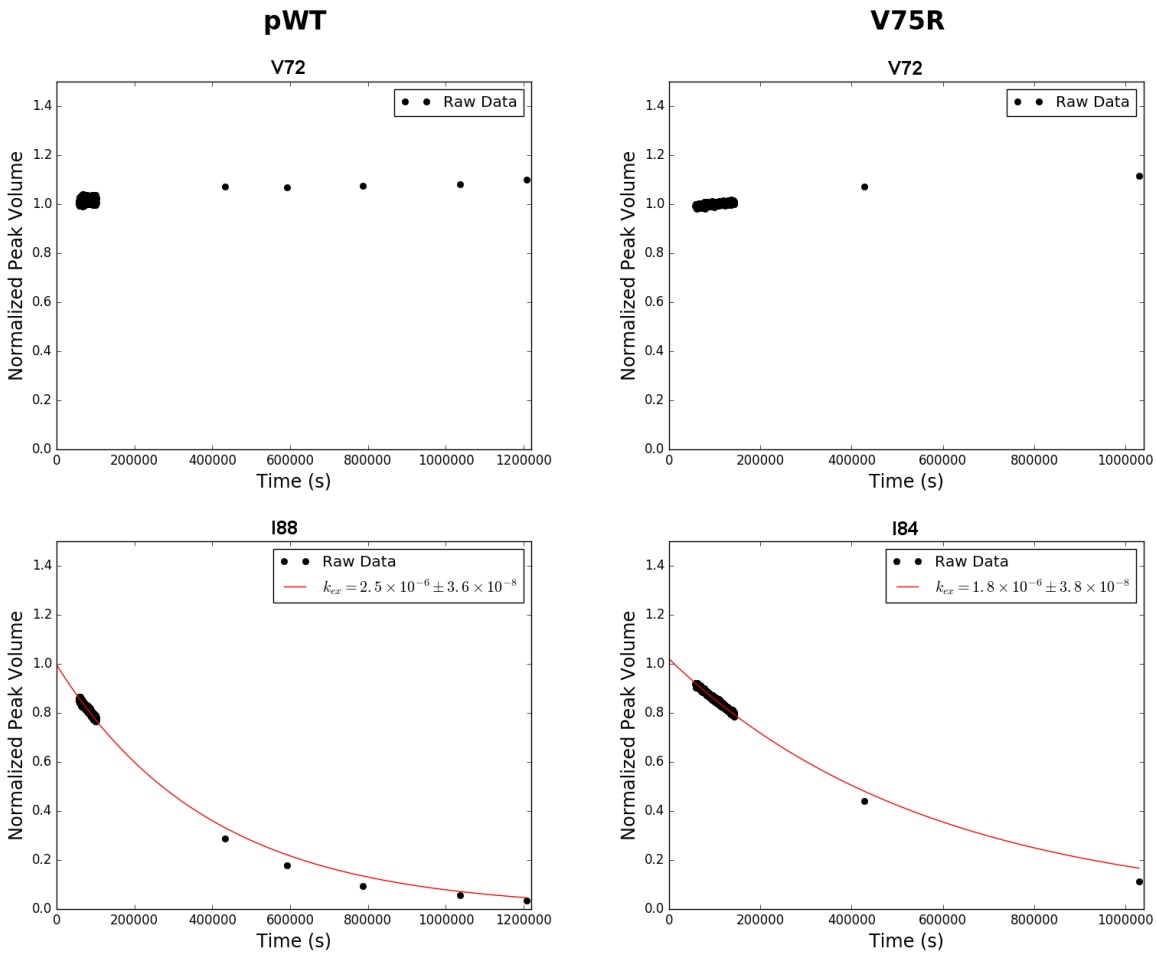


Figure 6.4: Adnectin amide H/D exchange at pD 4.6 monitored by ^{1}H-^{15}N SOFAST HMQC: a comparison of reciprocally hydrogen bonded residues V72 and I88 in pWT Adnectin with the structurally equivalent V72 and I84 in V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange over the course of the experiment). In each panel, the result of fitting the experimental data to an exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red if at least one H/D exchange half-life is captured.

V75R is a solubility-enhanced Adnectin developed using the consensus design strategy described in Chapter 5. Generally, soluble proteins may display a lower propensity for intermolecular association; the second virial coefficient calculated for V75R from SLS measurements (Fig. 6.5) confirms that interactions between V75R molecules in solution are predominantly repulsive (positive A_2) under buffer and temperature conditions that closely approximate those of our amide H/D exchange experiments (Section 6.3). In contrast, the interactions between Parent molecules are predominantly attractive (negative A_2), despite the fact that the amino acid sequences of these two Adnectins differ by a single point mutation. The rates of amide H/D exchange (k_{ex}) for V75R residues 86 and 88 are too slow to be reliably determined from our pD 4.6 amide H/D exchange data, but they are unambiguously lower than those observed for structurally equivalent residues in the pWT and Parent Adnectins (Figs. 6.2 and 6.3). In order to compare the ΔG_{ex}^{app} of V75R β-strand G amides with those of their hydrogen bond partner residues in β-strand F, we additionally present amide H/D exchange data collected at pD 7.0 (Fig. 6.6; Table 6.2).

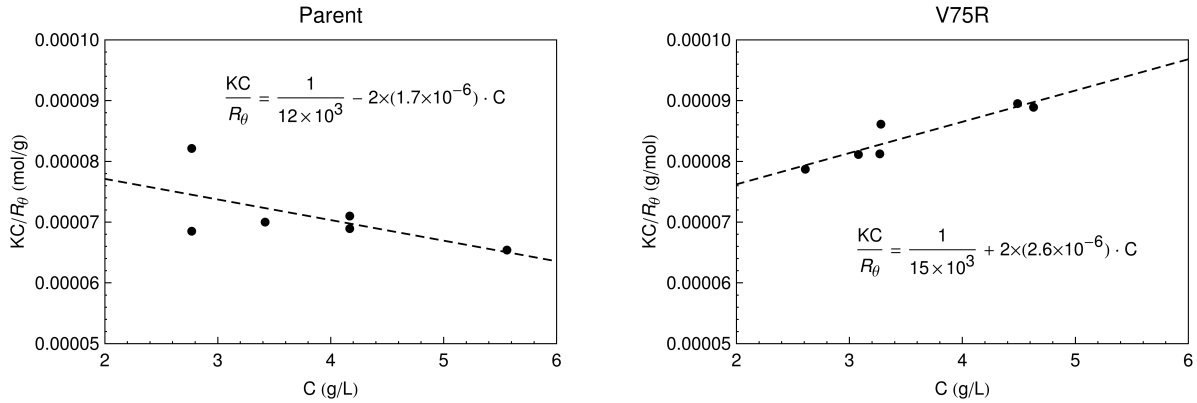


Figure 6.5: Debye plots of Parent (left) and V75R (right) static light scattering data. Within each panel is a dashed line showing the result of a linear regression including all of the points. The slope of the line corresponds to twice the second virial coefficient ($2 \cdot A_2$; Eq. 6.9), and the KC/R_θ intercept corresponds to inverse molecular weight.

The higher rate of base-catalyzed exchange at pD 7.0 allows us to determine that, relative to equivalent residues in the pWT and Parent Adnectins, the ΔG_{ex}^{app} values calculated for V75R residues 86 and 88 (β-strand G) are much closer to those of their hydrogen bond partners (residues 70 and 68, respectively, in β-strand F). Interestingly, the protection of V75R residue 84 is substantially lower than that of its hydrogen bond partner (residue 72), much like that of the structurally equivalent pWT residue 88 (Fig. 6.4); this suggests that residue 88/84 ([pWT #]/[Other Adnectins #]) amide H/D exchange may be dominated by a lower energy exchange-competent state that can be accessed independently of the other β-strand G residues.

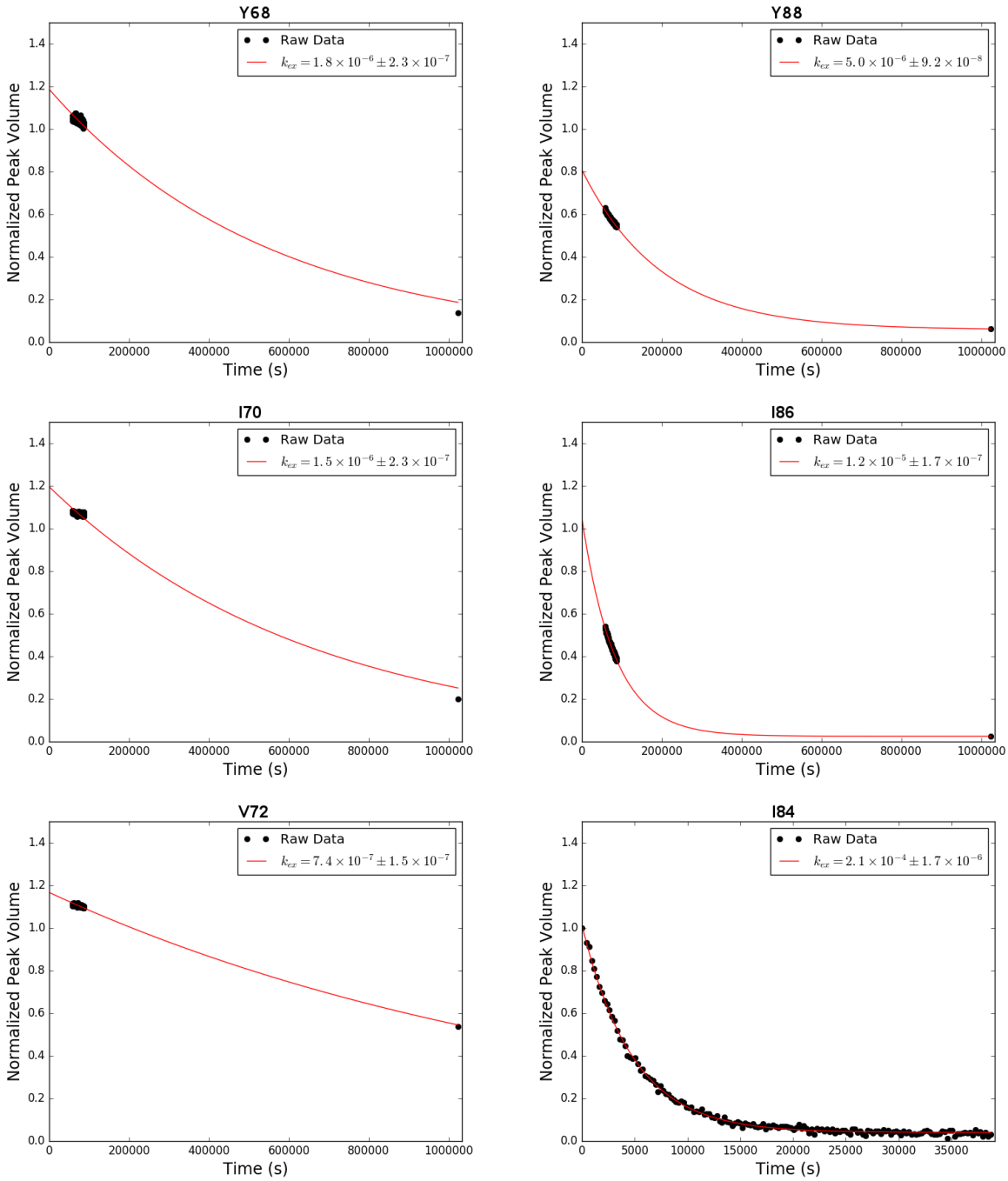


Figure 6.6: Adnectin amide H/D exchange at pD 7.0 monitored by ^{1}H-^{15}N SOFAST HMQC: a comparison of reciprocally hydrogen bonded residues Y68/Y88, I70/I86, and V72/I84 in the V75R Adnectin. Black dots show integrated peak volumes as a fraction of the initial peak volume. In each panel, the result of fitting the experimental data to an exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red. Note that the rates of exchange are amino acid sequence-dependent and therefore not directly comparable.

Considered jointly, the β-strand F/β-strand G protection discrepancy observed for the pWT/Parent Adnectins and the contrasting V75R results are best explained by a model in which inter-Adnectin association promotes a non-native state that protects β-strand F residues but not β-strand G residues from amide H/D exchange. Thus, lower propensity for inter-Adnectin association (e.g., by V75R) leads to lower rates of β-strand G amide H/D exchange. These data directly implicate β-strand F in transient intermolecular interactions, support our theory that this strand also plays an important role in Adnectin aggregation (Chapter 5), and illuminate the mechanism by which β-strand G may unfold (relative to the native state).

Part III

Protein Dynamics and Variable-Temperature NMR

Chapter 7

Temperature Dependence of NMR Chemical Shifts: Tracking and Statistical Analysis

7.1 Context

In this chapter we describe the experimental methods that we use to express and purify Adnectins, acquire variable-temperature nuclear magnetic resonance (VT-NMR) data, as well as computational methods for initial data analysis (detailed results and subsequent analyses are reported in Chapters 9-11). These computational methods include a fully automated tracking algorithm that is capable of propagating initial (single temperature) ^{1}H-^{15}N cross peak assignments to spectra collected over a range of temperatures. These sets of cross peaks are then decomposed into ^{1}H and ^{15}N chemical shift temperature dependences, which are fit to a linear model to yield linear temperature coefficients (further analyzed in Chapters 9 and 11). Also described here are methods for the detection of systematic, statistically significant deviation from linearity (curvature) in the temperature dependences of ^{1}H chemical shifts (further analyzed in Chapter 10). Resolution in the indirectly detected ^{15}N dimension is much lower, negatively impacting our ability to detect curvature; therefore, the ^{15}N chemical shift temperature dependences are not tested for deviations from linearity.

7.1.1 Acknowledgements

Colleen Doyle assisted with the collection of ^{1}H-^{15}N HSQC, NOESY-HSQC and TOCSY-HSQC pWT and Parent Adnectin spectra, the assignment of ^{1}H-^{15}N cross peaks to specific pWT and Parent Adnectin residues, as well as the collection of preliminary pWT and Parent VT-NMR data and identification of problematic hydrophobic interactions between

Adnectins and 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Colleen Doyle, Duncan MacKenzie, Dalia Naser, and Jeffrey Palumbo tested and provided valuable feedback on the ShiftTrack algorithm and the Shift-T web server using Adnectin, hisactophilin, and superoxide dismutase 1 data.

7.2 Introduction

Nuclear magnetic resonance (NMR) chemical shifts are sensitive probes of protein structure and dynamics. The chemical shifts of atoms such as C_α, C_β, C', and H_α have proven particularly useful because of their strong dependences on secondary structure and/or backbone dihedral angles [204, 205]. In contrast, the relationships between structure and amide proton/nitrogen chemical shifts are more complicated; disentangling the various through-bond and through-space contributions is a significant challenge. Paradoxically, measuring the temperature dependences of amide chemical shifts adds a dimension to the data, yet may simplify analysis if some contributions exhibit negligible temperature dependences (discussed in Chapter 9).

Empirically, the temperature dependences of both amide proton and amide nitrogen chemical shifts are frequently linear (Chapters 9 and 11). The linear temperature coefficients (i.e., slopes determined by fitting chemical shift vs. temperature data to a linear model) of amide protons have been used to probe the hydrogen bond status of individual amides, interpreted as a measure of temperature-dependent loss of structure, and grouped sequentially to distinguish between ordered and disordered protein regions (topics discussed in Chapter 9). No similarly straightforward interpretations of amide nitrogen temperature coefficients have been proposed, but we explore new avenues of investigation in Chapter 11. Systematic deviations from linearity are also of interest, as they may be explained by the two-state, fast exchange model described in Chapter 10; however, such curvature is often subtle, making it difficult to detect and validate.

This chapter is focused on methods for the collection and initial analysis of VT-NMR data, including automated tracking of ^{1}H-^{15}N cross peak movement over temperature, as well as curvature detection and validation. Results generated by application of these methods to VT-NMR data from experiments on pWT, Parent, L78I, V75R and L18V/Y88F Adnectins (Appendix D) are reported and interpreted in Chapters 9-11

7.3 Methods

7.3.1 Expression & Purification

BL21 (DE3) pLysS cells were transformed with pET-9d plasmids into which Adnectin genes were cloned at the NcoI and BamHI sites. These cells were streaked onto agar plates

with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. Cells from a single colony on each plate were transferred into test tubes containing 10 mL of sterile LB containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol, and grown overnight at 37°C with shaking at 225 RPM. The contents of each test tube were transferred into a 4 L Erlenmeyer flask containing 1 L of sterile M9 minimal media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 4 g/L glucose, and 0.5 g/L $^{15}NH_4Cl$) with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. These flasks were incubated at 37°C with shaking at 225 RPM until an A_{600} of 0.6-0.8 was reached. Expression was then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and the flasks were incubated at 25°C with shaking at 225 RPM. 24 h post-induction, cells were pelleted by centrifugation for 20 min at 5000 g and 4°C. The supernatant was poured off, and the cells from each flask were resuspended in 40 mL of buffer (50 mM sodium phosphate, pH 8.0, 0.5 M NaCl, and 25 mM imidazole) in a 50 mL conical tube. The cells in each conical tube were then lysed by sonication on ice using four 15 second pulses (60 W) separated by 10 second pauses, using a W-225R probe sonicator with a standard tapered microtip attached to a 1/2" disruptor horn (Heat Systems-Ultrasonics Inc.).

The cell lysate was centrifuged for 30 min at 20000 g to pellet insoluble material. Following the addition of 500 µL of 3 mg/mL DNase I, the supernatant was incubated for 20 min at room temperature, then syringe filtered (0.45 µm Supor® membrane, Pall Corporation). Adnectins, which have a 6-residue C-terminal polyhistidine tag, were purified from the supernatant by nickel affinity chromatography using Profinity IMAC resin (Bio Rad). The soluble fraction of the cell lysate was loaded onto the column and washed with a minimum of five column volumes of lysis buffer (50 mM sodium phosphate, pH 8.0, 0.5 M NaCl, and 25 mM imidazole), eluted by competition with excess imidazole (50 mM sodium phosphate, pH 8.0, 0.5 M NaCl, and 500 mM imidazole) and dialyzed into 20 mM acetic acid/sodium acetate pH 4.0 buffer. Initial protein concentrations were determined by absorbance at 280 nm using extinction coefficients calculated by submitting the Adnectin amino acid sequence (Appendix D) to the ProtParam web server [203]. The protein was concentrated to approximately 10 mg/mL using an Amicon Ultra-15 centrifugal filter (EMD Millipore) with a 3.5 kDa molecular weight cut-off. Aliquots of 0.5 mL and 1 mL were flash-frozen in liquid nitrogen and stored at -80°C.

7.3.2 NMR Sample Preparation

Adnectin NMR samples were prepared in 20 mM acetic acid/sodium acetate pH 4.0 buffer and concentrated to a volume of 450 µL using an Amicon Ultra-4 centrifugal filter (EMD Millipore). Following addition of 50 µL of D_2O (final protein concentration ~30 mg/mL) each sample was mixed by pipetting up and down, then transferred to a clean NE UL-5 NMR tube (New Era).

7.3.3 ^{1}H-^{15}N Cross Peak Assignment

Backbone amide nitrogen and proton resonance assignments were determined with the aid of ^{1}H-^{15}N HSQC (Bruker pulse program 'hsqcetfpf3gpsi' [206–209]; Section E.2) spectra, as well as ^{15}N-edited NOESY-HSQC (Bruker pulse program 'noesyhsqcetf3gp3d' [210]; Section E.3) and TOCSY-HSQC (Bruker pulse program 'dipsihsqcf3gpsi3d' [206–209]; Section E.4) spectra, all of which were acquired using either a Bruker AVANCE 600 MHz or a Bruker AVANCE 700 MHz spectrometer, processed using Bruker TopSpin 3.5 software and analyzed using Computer Aided Resonance Assignment (CARA) software [211].

7.3.4 Variable-Temperature NMR

Variable-temperature ^{1}H-^{15}N HSQC (Bruker pulse program 'hsqcetfpf3gpsi'; Section E.2) spectra were acquired using a Bruker AVANCE 600 MHz spectrometer. Reported amide proton chemical shifts are directly referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), while reported amide nitrogen chemical shifts are indirectly referenced to DSS using a ^{15}N/^{1}H Ξ ratio of 0.101329118 [212]. In order to avoid hydrophobic interactions between DSS and partially or fully unfolded protein [213], an NE-5-CIC coaxial inner cell (New Era) containing DSS was used (DSS was not added directly to Adnectin NMR samples).

Nominal temperatures ranging from 288 K to 328 K in 5 K increments were programmed into the temperature controller via Bruker TopSpin 1.3 spectrometer control software. Actual temperature differentials (relative to the lowest nominal temperature, 288 K) were calculated from the movement of the DSS peak (relative to the base frequency of the spectrometer) and the known temperature dependence of water (-11.9 ppb/K) [212] (Eq. 7.1). In this analysis, we ignore the slight temperature dependence of the DSS chemical shift, which is negligible relative to -11.9 ppb/K (a difference of two orders of magnitude) [214].

$$\Delta T = \frac{(\delta_{DSS}[T_{nom}] - \delta_{DSS}[288K])}{11.9\ ppb/K} \tag{7.1}$$

Calculation of absolute temperatures, as opposed to differentials (Eq. 7.1), is not required for accurate determination of temperature coefficients or curvature analysis, and would require a second standard (i.e., in addition to DSS) [214]. The H_2O resonance may be used for this purpose, but the chemical shift of water is sensitive to pH and salt concentration [212]; therefore, the accuracy of calculations based on published DSS-water temperature-dependent chemical shift differences may depend on these variables. Here we calculate only temperature differentials, and reference them to the lowest nominal temperature (288 K).

Data processing, including automatic peak picking (with parabolic interpolation) from unassigned ^{1}H-^{15}N spectra, was performed using Bruker TopSpin 3.5 software. Peak lists

were written to comma-separated variable (CSV) text files. The ShiftTrack and Curvalyzer algorithms described in the Results & Discussion were implemented in the Python programming language; full source code can be found in Appendix F, and both are available for use on the 'Shift-T: Automated Variable-Temperature Data Analysis' website, a beta version of which is currently available for testing at 'http://206.167.182.79/shiftt'.

7.4 Results & Discussion

7.4.1 Cross Peak Tracking and Temperature Coefficients

In order to combat slow drift in the strength of the field generated by the electromagnet, modern NMR spectrometers use a feedback control system (commonly referred to as the 'lock') that dynamically adjusts the strength of the magnetic field to maintain the resonant frequency of a particular nucleus at a fixed offset to the base frequency. In aqueous solutions such as those suitable for protein NMR, the resonance of deuterium nuclei from D_2O/HDO is monitored for this purpose. The chemical shift of water (including H_2O, D_2O, and HDO) is intrinsically temperature dependent [212, 214], but the feedback control system compensates. Thus, the chemical shift of water appears to be temperature-invariant, while all other peaks shift (because of change in the strength of the magnetic field) from their 'true' positions by an amount equal to the temperature dependence of water. To recover ^{1}H and ^{15}N temperature dependent chemical shifts unbiased by this 'deuterium lock artefact', spectra must be referenced to a standard with negligible temperature dependence, e.g., 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) [214, 215] (Fig. 7.1B). In spectra that have not been referenced to such a standard, the temperature dependence of water is often the largest contributor to the observed changes in chemical shifts (Fig. 7.1A).

Empirically, movement of cross peaks in the ^{1}H-^{15}N plane is often approximately linear with respect to temperature, even after referencing all chemical shifts to DSS. However, prior to such referencing a much greater degree of regularity is apparent (Fig. 7.1A); as temperature increases, cross peaks tend to move downfield in both dimensions in a highly linear fashion. From ^{1}H-^{15}N cross peak coordinates for a given amide, specified at a single temperature, our tracking algorithm (ShiftTrack) finds the set of cross peaks (one per temperature, selected from peaks picked automatically from unassigned spectra, e.g., by TopSpin; Section 7.3) with the smallest deviation from linearity (i.e., for which the simple linear regression has the lowest residual sum of squares) in the ^{1}H-^{15}N plane, subject to a weak constraint on the spacings between points. We prefer to apply this algorithm before referencing the spectra to DSS, turning artefact-induced regularity to our advantage.

ShiftTrack maintains a list of candidate 'lines' (sets of points). As the (unassigned) peak list from each temperature is processed, new candidate lines are constructed through extension of existing candidates by a single point (Fig. 7.2). In order to avoid a combinatorial explosion, new candidates are constructed using only points found within a user-defined

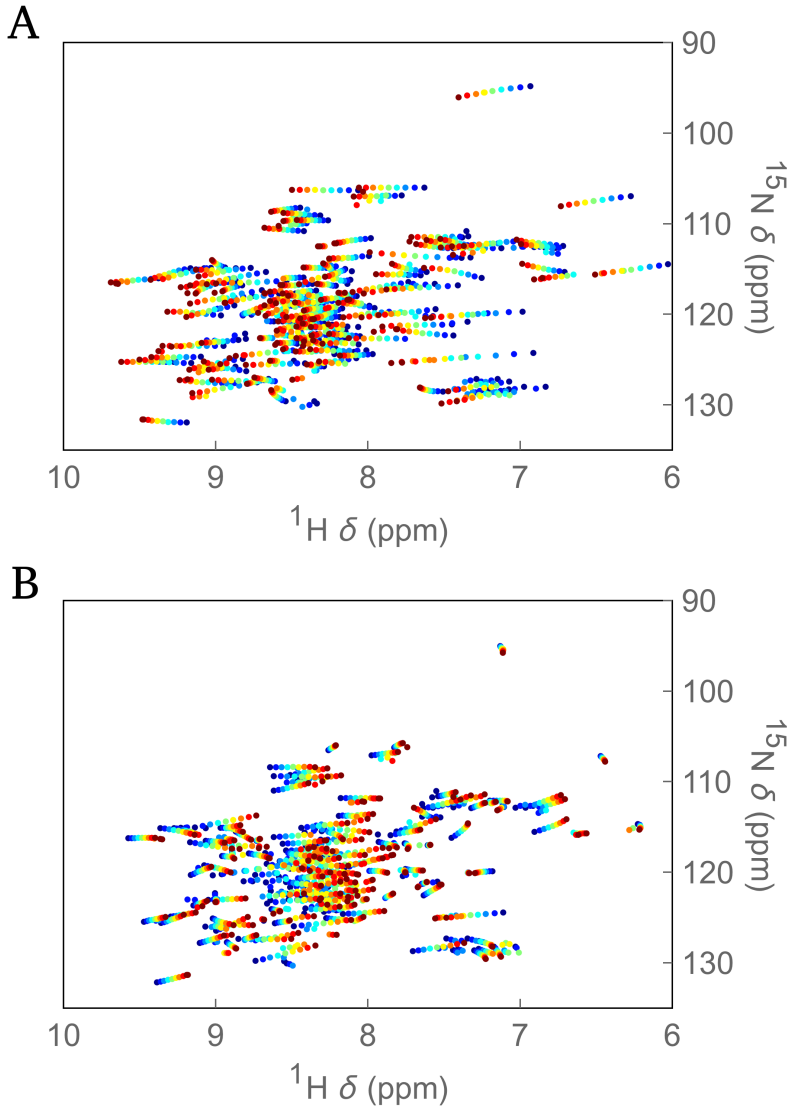


Figure 7.1: Overlaid variable-temperature ^{1}H-^{15}N peak positions picked from pWT Adnectin HSQC spectra: 288 K (nominal; dark blue) to 329 K (328 K nominal; dark red) in ~5 K increments (the distinction between nominal and calculated temperatures is explained in Section 7.3). A: peaks not yet referenced to DSS; observed differences between spectra are determined primarily by the temperature dependence of the chemical shift of water. B: peaks referenced to DSS; the intrinsic temperature dependence of amide ^{1}H and ^{15}N chemical shifts can be discerned.

radius of the cross peak from the previous temperature and filtered to remove candidates with pronounced nonlinearity from consideration. After peak lists from all temperatures have been processed, the list of candidate lines may contain entries with varying numbers of peaks (e.g. if no suitable points with which to extend a given candidate were found at one or more temperatures). Preference is shown for full-length (i.e., one peak per temperature) solutions. If no full-length solutions below the residual sum of squares linearity threshold are found, shorter solutions are considered. If these are also rejected on the basis of the residual sum of squares linearity criterion, no solution will be reported for the assignment in question. The output of ShiftTrack includes plots of chemical shifts vs. temperature, a line determined by simple linear regression, and regression residuals; these plots, particularly the residuals, facilitate verification of ShiftTrack solutions. At the discretion of the user, residuals more than two standard deviations from zero (the mean) may be flagged for manual review.

7.4.2 Detection and Statistical Validation of Curvature

Empirically, the temperature dependences of amide proton chemical shifts are predominantly linear (Chapter 9). Detectable deviations from linearity (hereafter referred to as 'curvature') may be attributed to temperature-dependent shifts in the population of distinct conformational states (Chapter 10). Here we focus first on strategies for the detection of curvature, which may be subtle, then on statistical validation that the curvature detected is likely to result from changes in the population of conformational states rather than a confluence of random errors. Collectively, these tests (which we refer to as Curvalyzer) are designed to be quite stringent. Temperature-dependent conformational changes that do not result in curvature are possible; absence of curvature is not particularly informative, therefore false negatives are unlikely to lead to incorrect conclusions. In contrast, false positives (inference of curvature where there is none) may result in serious errors.

Curvalyzer treats curvature detection as a 'model selection' problem. Two (nested) models are considered for each set of chemical shift vs. temperature data: linear and quadratic. There is no theoretical reason to believe that experimentally observed curvature should fit a quadratic model. However, as expected given our hypothesis that temperature-dependent shifts in the relative population of conformational states are causative, curvature generally manifests as a gentle curve with a single minimum or maximum (Fig. 7.3); thus, the quality of fit to a quadratic model is an adequate test of nonlinearity. Unless the temperature dependence is exactly linear, the quadratic model will have a lower sum of squared errors (SSE); however, if the improvement (relative to the linear model) is small, it may not be statistically significant. Curvalyzer uses an extra-sum-of-squares F test to quantify the statistical significance of the improvement [216]. The p-value resulting from this F test is reported (using a significance threshold of 0.01), where the null hypothesis is that the linear model is correct. To increase the stringency of this test, we ensure that nonlinearity introduced by a single outlier (and therefore unlikely to be attributable to the sampling of distinct conformational states) does not introduce false positives by requiring

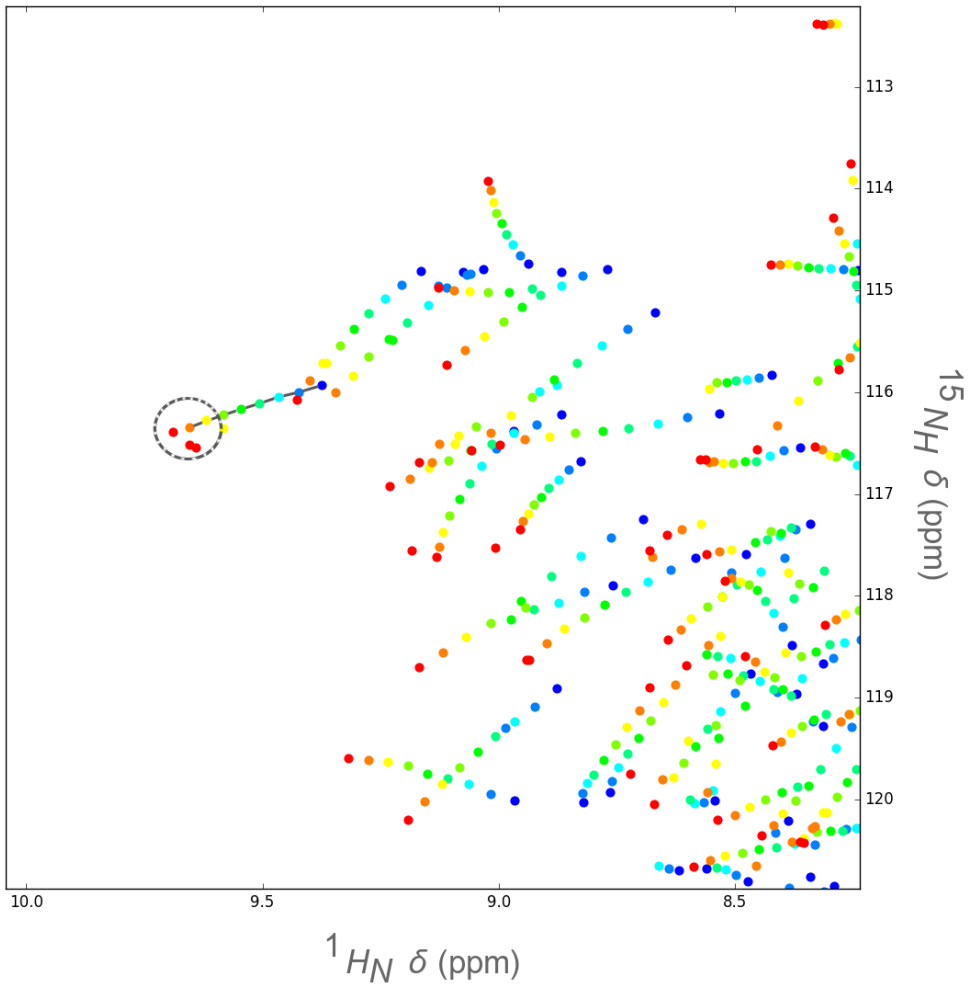


Figure 7.2: A ShiftTrack algorithm example. The coloured points represent peaks positions from pWT Adnectin ^{1}H-^{15}N HSQC spectra collected at temperatures between 288 K (nominal; dark blue) and 329 K (328 K nominal; dark red) in ~5 K increments (the distinction between nominal and calculated temperatures is explained in Section 7.3). Line segments join peaks in a candidate line from 288 K through 324.3 K (323 K nominal). The search radius (about the 324.3 K point) for 329 K (328 K nominal) points, indicated by the dashed circle, generates three new candidates. Ultimately, the candidate line that includes the left-most point will be selected as the solution most likely to be correct.

that, for each set of points generated by leave-one-out resampling without replacement (i.e., from a set of amide proton chemical shifts measured at N different temperatures, N subsets of N-1 points are possible), the quadratic fit must be significantly better (p-value less than 0.01, per the extra sum of squares F test) than that of the linear fit in order for the temperature dependence to be considered curved.

Once the existence of curvature has been established, we must consider its source. The F test referenced above implicitly considers the possibility that deviations from linearity are caused by errors in the data, but each set (one per amide proton) of chemical shifts is evaluated separately. For the set of all chemical shifts (at all temperatures) of amide protons for which curvature was not detected, we can calculate residual errors (residuals), i.e., the differences between measured chemical shifts and fits of the temperature dependences to a linear model. We find that distributions of residuals are bell-shaped with means near zero, but may have heavy tails that indicate deviation from normality (Fig. 7.4). We therefore fit residuals to a 't distribution', which includes the normal distribution as a special case, but allows for the possibility of heavier tails [216]. Plausible sets of residuals can be fabricated by drawing random values from this fitted distribution; by doing so, we calculate the probability of observing curvature (due to random errors) of a given magnitude. In our analysis of experimentally observed curvature, these probabilities are equivalent to p-values (using a significance threshold of 0.01) where the null hypothesis is that the curvature is due to measurement errors rather than shifts in the population of distinct conformational states. This statistical test complements the first by screening out cases where, although curvature was detected, there is a significant chance that it is a product of experimental error.

7.4.3 Conclusion

Manually tracking ^{1}H-^{15}N cross peak movement with temperature is a relatively straightforward, yet labour-intensive task. Here we implement an algorithm that automates the bulk of this task, freeing the user to focus on only the most ambiguous cases. For amide proton temperature dependences, we also automate the detection and statistical validation of curvature; stringent tests are required in order to prevent false (curvature) positives, which may lead to incorrect conclusions. Together, these methods form the foundation upon which our analyses of linear amide proton temperature coefficients (Chapter 9), curvature in the temperature dependence of amide proton chemical shifts (Chapter 10), and linear amide nitrogen temperature coefficients (Chapter 11) are built.

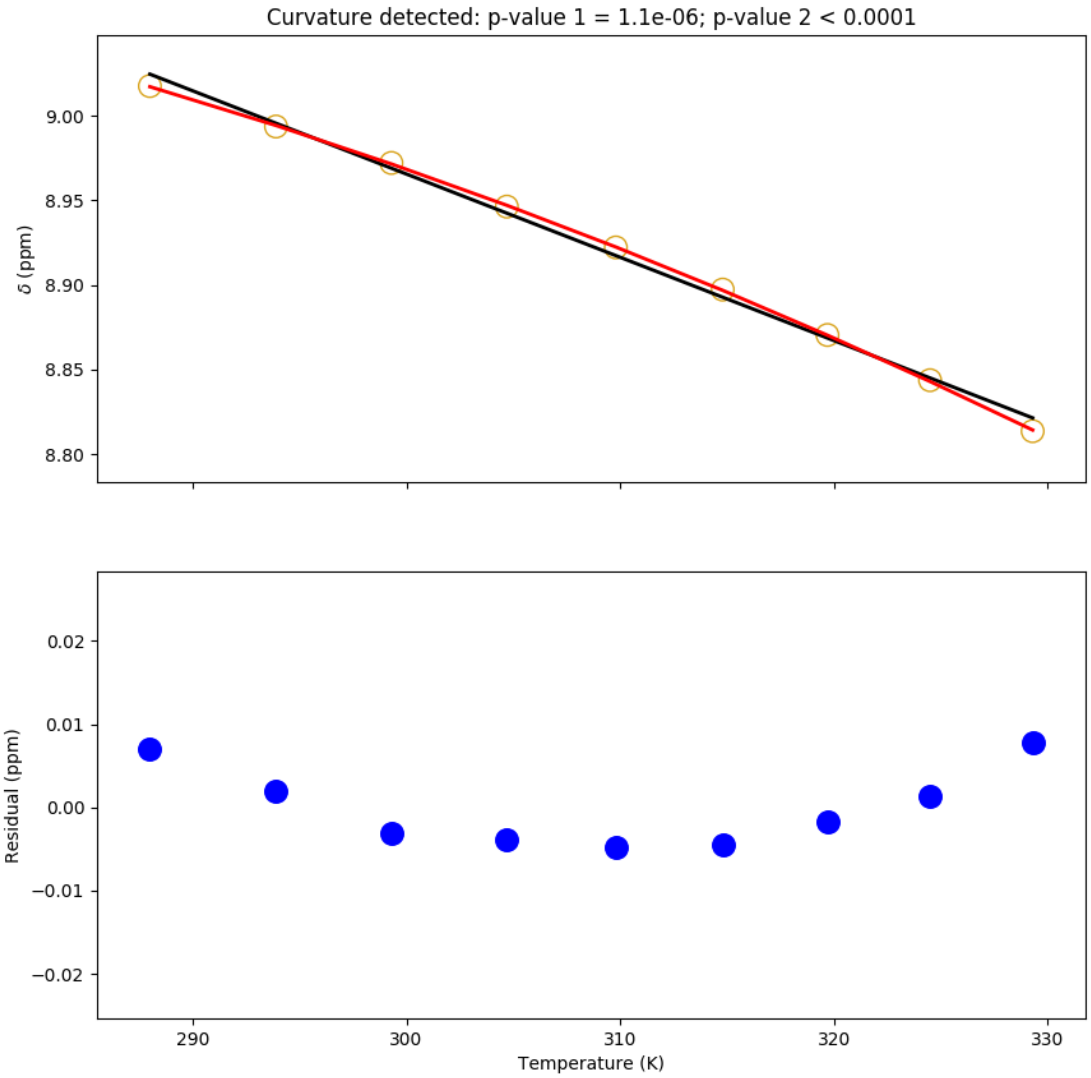


Figure 7.3: Curvature in the temperature dependence of amide proton chemical shifts. Chemical shifts (open circles, top) and residuals (filled circles, bottom) for residue 11 of Adnectin V75R over temperatures ranging from 288 (nominal) to 329.3 K (338 K nominal). The chemical shifts were fit to linear (black) and quadratic (red) models by linear regression. The probability that the linear, rather than the quadratic, model is correct is reported as p-value 1. The probability that the curvature observed is the result of random errors is reported as p-value 2.

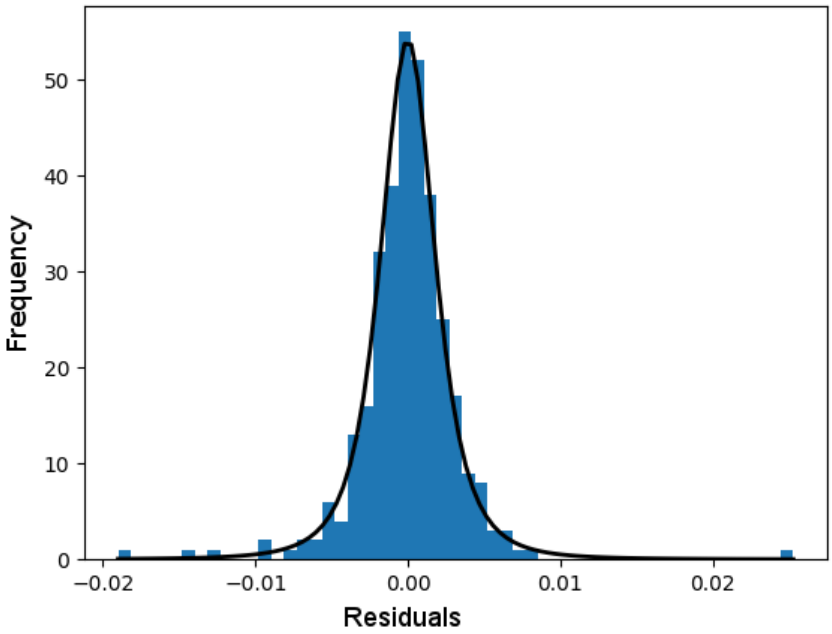


Figure 7.4: A histogram of pWT Adnectin VT-NMR amide proton chemical shift residuals (from all residues that show no curvature, at all temperatures) with a t distribution (black line) overlaid. 'Simulated' sets of residuals can be constructed using random values drawn from this distribution, allowing the probability of observing curvature due to random errors (rather than shifts in the population of distinct conformational states) to be calculated.

Chapter 8

Hydrogen Bond Dynamics and Computation of NMR Observables

8.1 Context

In Chapters 9-11, we discuss Adnectin backbone amide proton and nitrogen chemical shifts and temperature dependences thereof. For some types of nuclei, relationships between chemical shifts and protein conformation (e.g., for C_α, C_β, C', and H_α chemical shifts and secondary structure or backbone dihedral angles [204, 205]) are well-established; in contrast, disentangling the various contributions to amide chemical shifts (particularly those of amide protons) is more difficult [217]. In these later chapters, analysis of simulated dynamics informs some of our interpretations of Adnectin amide chemical shifts and their temperature dependences. However, because the determinants of amide chemical shifts are not entirely understood, in this chapter we validate our simulation methodology by using it to calculate trans-hydrogen bond ${}^{3h}J_{NC'}$ scalar couplings for human ubiquitin (the temperature dependences of which have been published [218]). These couplings depend on hydrogen bond geometry [219], which is also known to influence amide proton (Chapter 9) and nitrogen (Chapter 11) chemical shifts.

We replicate and extend published density functional theory (DFT) quantum chemical ${}^{3h}J_{NC'}$ coupling calculations [219] to encompass the wider range of hydrogen bond lengths we observe in molecular dynamics (MD) simulations and find that the previously derived formula (Eq. 8.2) still holds. We observe that the root-mean-square deviation (RMSD; relative to experiment) of MD average calculated ${}^{3h}J_{NC'}$ couplings decreases with increasing simulation length, suggesting that short simulations do not capture all of the relevant hydrogen bond dynamics. The RMSD of MD average couplings we calculate from 100 ns simulations (0.097 Hz) is ~1/3 lower than previously published results (based on shorter simulations) [220], suggesting that our simulations reproduce realistic protein dynamics; however, in part because the simulated temperature is only constant in a time-average

sense, reproducing the temperature dependence of these dynamics may require much longer trajectories.

8.2 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is perhaps the most powerful tool available for the measurement of protein dynamics in solution, but the price paid for sensitivity to dynamics is sometimes ease of interpretability. Each observable resonance is composed of contributions from structurally equivalent nuclei in a vast number of proteins, the conformations of which may both differ from each other at any particular instant and evolve over the timescale of an experiment. In order to make sense of such measurements, one may construct models constrained by both NMR data and physics (or an approximation thereof). A familiar example is protein structure determination, which can be facilitated by restrained (i.e., by NMR data) MD simulations. Weak restraints on protein structure in such simulations may result from inadequate experimental data; alternatively, sometimes an absence of strong restraints may be evidence of protein dynamics. Conversely, MD simulations unrestrained by experimental data can be used as a basis for the calculation of NMR observables such as scalar couplings, residual dipolar couplings, and relaxation order parameters [221], as well as chemical shifts [222, 223]; good agreement with experimental results may be be interpreted as evidence of both force field accuracy and the realism of the simulated dynamics.

There is a strong theoretical foundation for the hypothesis that ${}^{3h}J_{NC'}$ couplings are determined entirely by hydrogen bond geometry [219], which also influences amide chemical shifts (Chapter 9). We simulate human ubiquitin, the only protein (to our knowledge) for which the temperature dependences of ${}^{3h}J_{NC'}$ couplings have been published [218], and find that increasing the length of simulation (e.g., from 1 ns to 100 ns) captures slower dynamics and robustly decreases the RMSD (relative to experimental results) of calculated ${}^{3h}J_{NC'}$ couplings. Contrary to what has been published [220], we conclude that MD-average ${}^{3h}J_{NC'}$ calculations (based on simulations of sufficient length) may improve upon crystal structure-based calculations because the strength of ${}^{3h}J_{NC'}$ couplings varies nonlinearly with changes in hydrogen bond geometry, i.e., the average coupling cannot be calculated from the average positions of atoms.

8.2.1 ${}^{3h}\mathbf{J}_{\mathbf{NC'}}$ Measurement and Hydrogen Bond Geometry

Through-hydrogen bond scalar couplings have been measured for nucleic acids [224, 225] and proteins [226–229]; for a review of hydrogen bond scalar coupling mechanisms see [230]. Despite the fact that the inter-residue ${}^{3h}J_{NC'}$ couplings observed in proteins are relatively weak (with magnitudes less than 1 Hz), pulse sequences for the acquisition of spectra from which they can be determined are well-described [231], though proteins larger than approximately 10 kDa may present more of a challenge [231, 232].

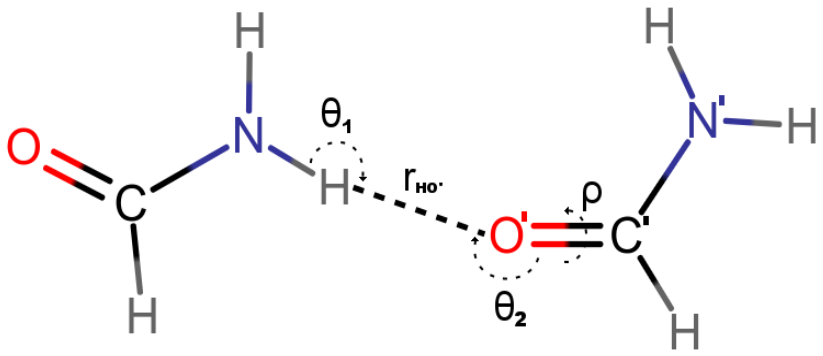


Figure 8.1: A formamide dimer model of hydrogen bond geometry. The H···O′ internuclear separation ($r_{HO'}$), the N-H···O angle (θ_1), the H···O′=C′angle (θ_2), and the H···O′=C′-N′ dihedral angle (ρ) are indicated.

A simple exponential dependence of ${}^{3h}J_{NC'}$ couplings on the distance between donor nitrogens and acceptor oxygens involved in intra-protein backbone hydrogen bonds was proposed on the basis of atomic coordinates from protein G crystal structures [233]. However, the influence of hydrogen bond geometry on ${}^{3h}J_{NC'}$ couplings in N-methylacetamide and formamide dimers (Fig. 8.1) has been studied using DFT and finite perturbation theory (FPT) computational methods [219,234,235], and DFT-computed ${}^{3h}J_{NC'}$ couplings can be accurately reproduced using a closed-form expression with an exponential dependence on the H···O' internuclear separation ($r_{HO'}$), a squared cosine dependence on the H···O′=C′ angle (θ_2), and weak dependence on the H···O′=C′-N′ dihedral angle (ρ) [219]. The form of this expression is shown in Eq. 8.1, which reduces to Eq. 8.2 if the dependence on ρ is neglected (which may be justifiable when θ_2 is close to 180°). The fully parameterized equations (corresponding to Eqs. 8.1 and 8.2) from [219] are reproduced here as Eqs. 8.3 and 8.4.

$$ {}^{3h}J_{NC'} = \Big(A \cdot cos^2(\theta_2) + [B \cdot cos^2(\rho) + C \cdot cos(\rho) + D]sin^2(\theta_2)\Big) \cdot e^{-3.2(r_{HO'}-1.76)} + E\ Hz \tag{8.1} $$

$$ {}^{3h}J_{NC'} = A \cdot cos^2(\theta_2) \cdot e^{-3.2(r_{HO'}-1.76)} + E\ Hz \tag{8.2} $$

$$ {}^{3h}J_{NC'} = \Big(-1.31cos^2(\theta_2) + [0.62cos^2(\rho) + 0.92cos(\rho) + 0.14]sin^2(\theta_2)\Big) \cdot e^{-3.2(r_{HO'}-1.76)}\ Hz \tag{8.3} $$

$$ {}^{3h}J_{NC'} = -1.29 \cdot cos^2(\theta_2)) \cdot e^{-3.2(r_{HO'}-1.76)} + 0.04\ Hz \tag{8.4} $$

8.2.2 Simulation of Hydrogen Bond Dynamics

Proteins in solution exhibit dynamic behaviour, and experimentally measured $^{3h}J_{NC'}$ couplings represent ensemble averages. Snapshots from MD trajectories can be used to generate ensembles of structures; if the the accuracy and length of the simulation are sufficient to capture the relevant dynamics, the average calculated $^{3h}J_{NC'}$ couplings may agree better with experimentally determined values than those based on the (static) starting structure, particularly if the resolution of that structure is not high [220, 236].

MD simulations have also been used to study the temperature dependence of $^{3h}J_{NC'}$ couplings by applying Eq. 8.3 to hydrogen bond geometries extracted from sixteen replica exchange trajectories of the B3 domain of protein G (GB3) spanning a temperature range of 278-331 K, but the correlations with experimentally-determined values were poor [237]. Here we show that the temperature dependence of the average (over all residues) $^{3h}J_{NC'}$ coupling calculated from 100 ns simulations of ubiquitin at five different temperatures follows the same trend as the experimentally measured values. It must be noted that the temperatures in our simulations are only constant in a time-averaged sense; the instantaneous temperatures are normally distributed with a standard deviation of approximately 3 K. As a consequence, we hypothesize that much longer simulations may be required to reproduce the experimentally observed temperature dependences of individual $^{3h}J_{NC'}$ couplings.

8.3 Methods

8.3.1 Quantum Chemical Calculations

All quantum chemical calculations were carried out via the GAUSSIAN 16 program [238] using DFT at the B3LYP/6-311++G(2d,2p) and B3PW91/6-311++G(2d,2p) levels (functional/basis set combinations similar to those selected for previously published calculations of this type [219, 220, 236]). The geometry of a formamide monomer was optimized and used to construct formamide dimers with various combinations of hydrogen bond geometrical parameters θ_1, θ_2, $r_{HO'}$, and ρ (Fig. 8.1). $^{3h}J_{NC'}$ coupling constants were calculated using finite perturbation theory (FPT) [239]; the total spin-spin couplings are reported here, though the Fermi contact contribution dominates (as previously described [219]). The isotropic magnetic shielding constants of hydrogen-bonded amide protons were calculated using the gauge-independent atomic orbital (GIAO) method [240]. These were then subtracted from a similarly calculated tetramethylsilane (TMS) proton isotropic magnetic shielding constant to give amide proton chemical shifts.

8.3.2 Molecular Dynamics

Explicit water MD simulations were performed using NAMD version 2.12 [241] with the CHARMM36 all-atom protein force field [242] and the TIP3P water model [243] (a force field/water model combination that has been validated using NMR data [221]). Hydrogen coordinates missing from crystal structures were inferred using PSFGEN version 1.6.5, a standalone program distributed with NAMD. Proteins were solvated in a water box that enclosed all protein atoms by a minimum of 8 Å, with Na^+ or Cl^- ions added when necessary (i.e., because of a net charge on the protein) to neutralize the system. The simulations used 1 fs time steps, periodic boundary conditions, and the particle mesh Ewald method of calculating long-range electrostatics. Constant temperatures were maintained using Langevin dynamics with a damping coefficient of 5 ps^{-1}, while a pressure of 1 atm was maintained via the Nosé-Hoover Langevin piston method. Production runs of 100 ns were preceded by 5000 steps of steepest-descent minimization, heating, and equilibration for 1 ns.

8.4 Results & Discussion

As described in Section 8.2.1, Barfield formulated an expression for ${}^{3h}J_{NC'}$ couplings based on hydrogen bond geometry (including an exponential dependence on $r_{HO'}$) that closely reproduces DFT-based calculations (Eq. 8.3) [219]. Perhaps informed by hydrogen bond geometries observed in protein crystal structures, these DFT calculations encompassed $r_{HO'}$ values between 1.8 and 2.2 Å. Though Eq. 8.3 has subsequently been applied to snapshots from MD simulations [220, 237], we find that dynamic (simulated) hydrogen bond geometries may deviate substantially from those found in crystal structures. In particular, although simulated θ_2 values most often fall within the bounds anticipated by Barfield, we must confirm the validity of Eq. 8.3 over a wider range of $r_{HO'}$ values (Fig. 8.2). Thus, here we reproduce and extend Barfield's DFT calculations using $r_{HO'}$ values between 1.5 and 3.0 Å, both with and without implicit (Polarizable Continuum Model) solvation (Table 8.1).

Barfield found it necessary to include the ρ angle (Fig. 8.1, Eq. 8.1) in order to maximize agreement between equation- and DFT-calculated results (in particular, to capture the effects of characteristic β-sheet and α-helix hydrogen bond geometries on ${}^{3h}J_{NC'}$ couplings). Here we observe that, in general, including the ρ parameter does not reduce the RMSD of MD-average ubiquitin ${}^{3h}J_{NC'}$ calculations (relative to experimentally measured couplings). We also find that use of the popular B3LYP functional produces results consistent with those from the B3PW91 functional employed by Barfield, that implicit solvation yields no improvement in RMSD, and, most importantly, that ${}^{3h}J_{NC'}$ values calculated using Eq. 8.3 continue to agree well with DFT results over our expanded $r_{HO'}$ range of 1.5 to 3 Å (justifying the use of this equation to calculate MD-average couplings).

In principle, because the strength of ${}^{3h}J_{NC'}$ couplings varies nonlinearly with changes in

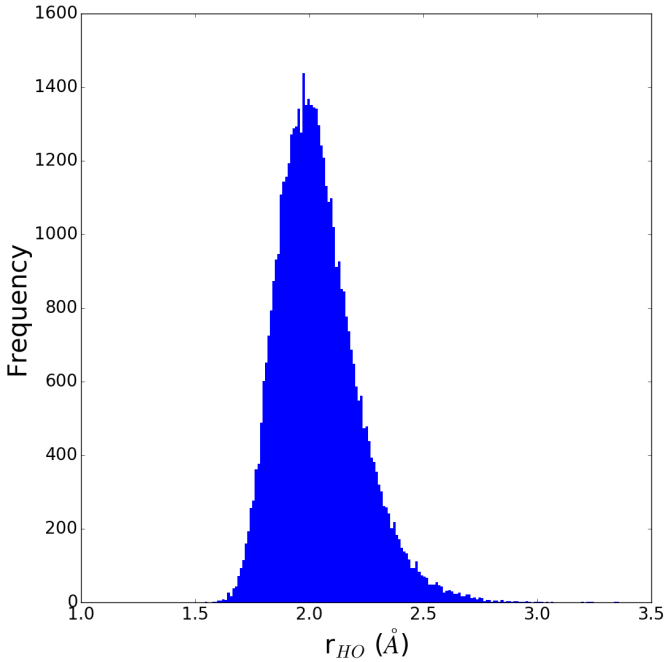


Figure 8.2: A histogram of $r_{HO'}$ distances observed in an MD simulation (at 298 K; Section 8.3) of human ubiquitin for a typical hydrogen bond (between the amide proton of residue 45 and the carbonyl oxygen of residue 48). For comparison, the corresponding $r_{HO'}$ observed in a static crystal structure (1UBQ) is 2.1 Å.

hydrogen bond geometry (i.e., the average coupling cannot be calculated from the average positions of atoms), it may be possible to calculate ${}^{3h}J_{NC'}$ values more accurately from MD trajectories than from high resolution crystal structures. In practice, previously published MD-based ${}^{3h}J_{NC'}$ calculations have only been found to improve upon calculations based on low resolution structures [220]. Here we demonstrate that MD-average ubiquitin ${}^{3h}J_{NC'}$ calculations substantially lower RMSD (relative to crystal structure-based calculations) in a timescale-dependent fashion (Figs. 8.3). We show that timescale, rather than simply the number of snapshots, is important by calculating the average couplings from 10 000 snapshots distributed evenly throughout the full length of the simulation; doing so gives the same RMSD (0.097) as calculating average couplings using all 500 000 snapshots. We hypothesize that the slower protein dynamics captured in longer simulations measurably impact the ${}^{3h}J_{NC'}$ couplings.

Relative to static structure-based calculations (1UBQ, 1.8 Å resolution), 100 ns MD-average ${}^{3h}J_{NC'}$ calculations yield a 34% improvement in RMSD (Fig. 8.4). Similarly, calculations of ${}^{3h}J_{NC'}$ couplings in the B1 immunoglobulin-binding domain of streptococcal protein G (1PGB, 1.9 Å resolution) based on longer simulations (>20 ns; Fig. G.1) are consistently more accurate than static structure-based calculations. As illustrated in Fig.

Table 8.1: DFT-based parameterizations of Eq. 8.1 and Eq. 8.2

	Parameters					
Parameterization	**A**	**B**	**C**	**D**	**E**	**RMSD[1]**
Barfield[2] Eq. 8.1	-1.31	0.62	0.92	0.14	0.00	0.098
Barfield[3] Eq. 8.2	-1.29	N/A	N/A	N/A	0.04	0.097
B3LYP Eq. 8.1	-1.39	0.40	0.05	0.18	-0.01	0.105
B3LYP Eq. 8.2	-1.31	N/A	N/A	N/A	0.04	0.096
B3PW91 Eq. 8.1	-1.31	0.36	0.05	0.18	0.00	0.094
B3PW91 Eq. 8.2	-1.24	N/A	N/A	N/A	0.04	0.100
B3LYP PCM[4] Eq. 8.1	-1.42	0.41	0.04	0.21	-0.02	0.114
B3LYP PCM[4] Eq. 8.2	-1.34	N/A	N/A	N/A	0.03	0.097
B3PW91 PCM[4] Eq. 8.1	-1.34	0.37	0.04	0.21	-0.01	0.098
B3PW91 PCM[4] Eq. 8.2	-1.26	N/A	N/A	N/A	0.04	0.097

[1]RMSD (relative to experiment) of MD-average ${}^{3h}J_{NC'}$ calculations (Hz)

[2]Parameterization from Eq. 12 in [219]

[3]Parameterization from Eq. 10 in [219]

[4]Implicit solvation (water) using the Polarizable Continuum Model

8.5, reductions in magnitude of deviation are unevenly distributed amongst the individual couplings in a simulation timescale-dependent manner. MD-average calculations of some couplings (e.g., donor residues 33, 56, and 61) actually result in greater deviation than static structure-based calculations, and show no improvement with increasing simulation length; from this we may infer that, in these few cases, realistic hydrogen bond dynamics are not simulated. Others (e.g., donor residues 17 and 33) show dramatic improvement with even short (0.2 ns) simulation, and modest improvement with increasing simulation length. Perhaps most interesting are those couplings that show the largest decreases in magnitude of deviation with an increase in simulation length from 2 ns to 100 ns (e.g., donor residues 3, 13, 26, 34, 45, and 50), suggesting that relatively slow hydrogen bond dynamics strongly influence experimental measurements of these particular couplings.

We indicate the positions of these hydrogen bonds in the ubiquitin structure (Fig. 8.6), and note that the $r_{HO'}$ distances corresponding to the couplings showing the largest decreases in magnitude of deviation are, in general, temporally uncorrelated (Table 8.2). This may indicate that local, rather than concerted global dynamics underlie the improvement in MD-average ubiquitin ${}^{3h}J_{NC'}$ couplings. In contrast, the ${}^{3h}J_{NC'}$ couplings of the B1 domain of protein G (GB1) that show the largest decreases in magnitude of deviation with increasing simulation length (Fig. G.2) show substantial correlations between several $r_{HO'}$ distances (Table G.1). Nevertheless, we confirm that, just as in our ubiquitin simulations, the $r_{HO'}$ distances of reciprocal hydrogen bonds in GB1 are not necessarily temporally correlated.

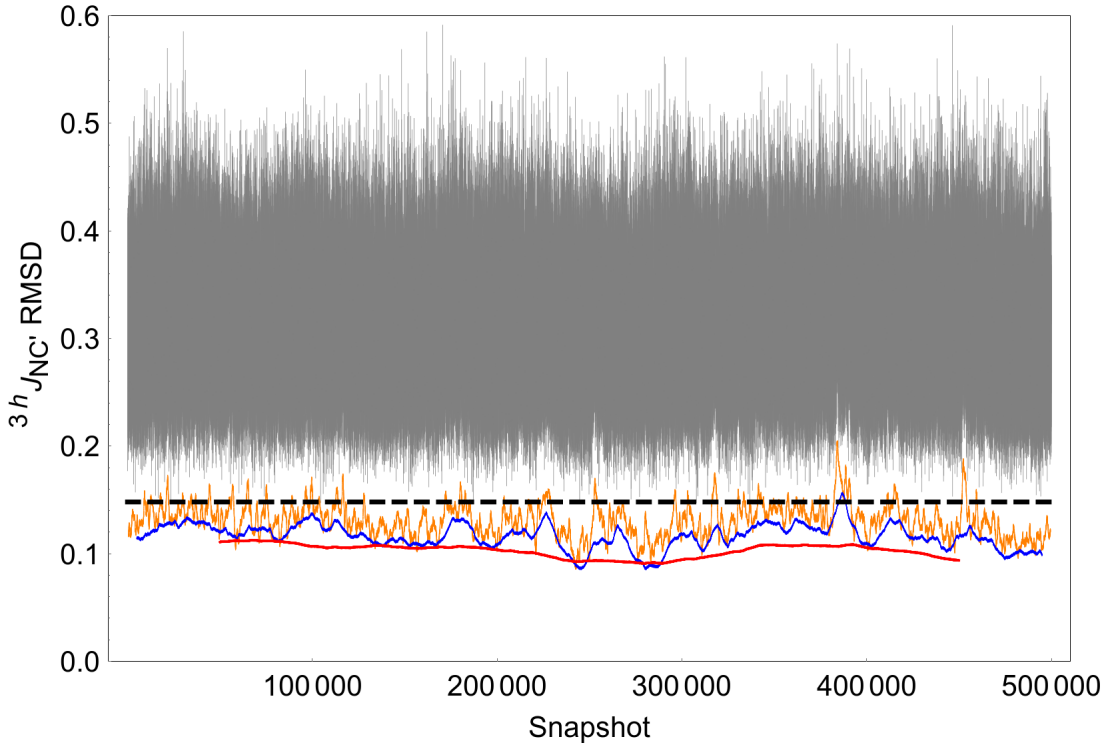


Figure 8.3: Sliding window averages of the RMSDs (relative to experimentally measured values) of ubiquitin $^{3h}J_{NC'}$ couplings calculated by applying Eq. 8.4 to hydrogen bond geometries from snapshots of a 100 ns MD simulation (one snapshot every 200 fs). Gray line: no averaging; orange line: 1 000 snapshots averaged; blue line: 10 000 snapshots averaged; red line: 100 000 snapshots averaged. For reference, the dashed black line shows the RMSD of ubiquitin $^{3h}J_{NC'}$ couplings calculated by applying Eq. 8.4 to hydrogen bond geometries extracted from the 1UBQ crystal structure.

Table 8.2: Linear (Pearson) correlations between simulated ubiquitin $r_{HO'}$ distances

Donor-Acceptor	**3 - 15**	**13 - 5**	**26 - 22**	**34 - 30**	**45 - 48**	**50 - 43**
3 - 15	1.00	0.01	-0.10	0.02	-0.01	0.02
13 - 5		1.00	0.01	-0.01	-0.01	0.01
26 - 22			1.00	-0.06	-0.03	0.01
34 - 30				1.00	0.02	0.01
45 - 48					1.00	0.14
50 - 43						1.00

[1]Correlation coefficients ≥ 0.1 indicated by yellow highlight

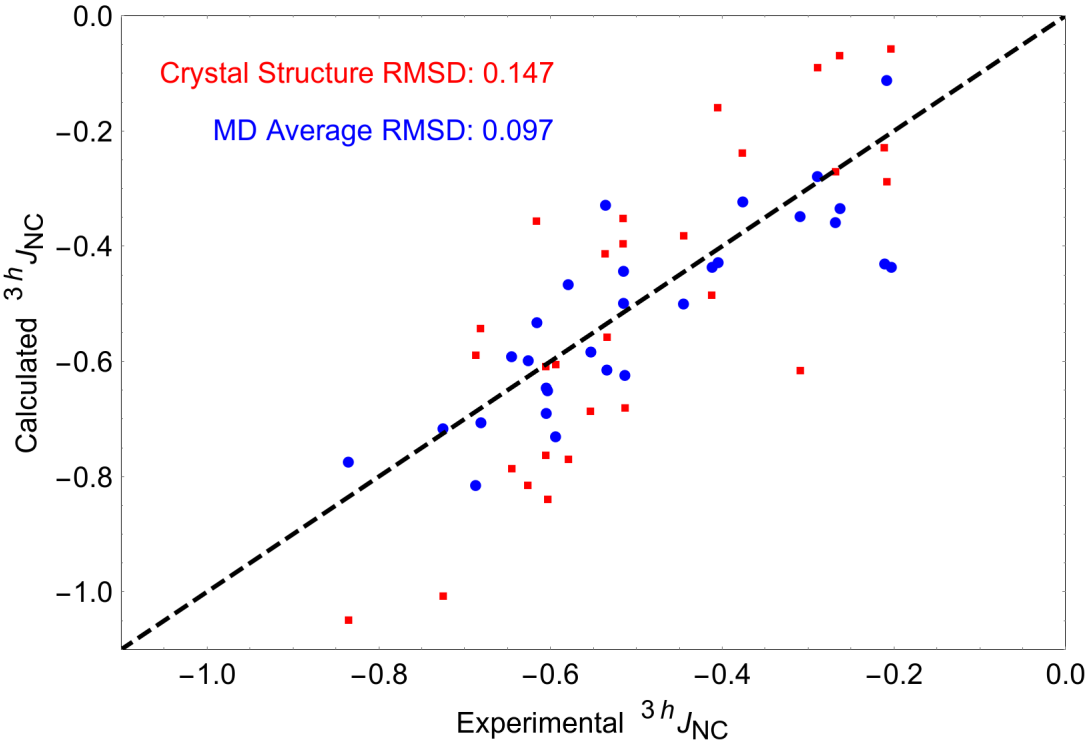


Figure 8.4: Static structure-based (1UBQ) and MD-average (100 ns, 298 K) ubiquitin $^{3h}J_{NC'}$ coupling calculations vs. experimental values. The dashed black line corresponds to perfect agreement.

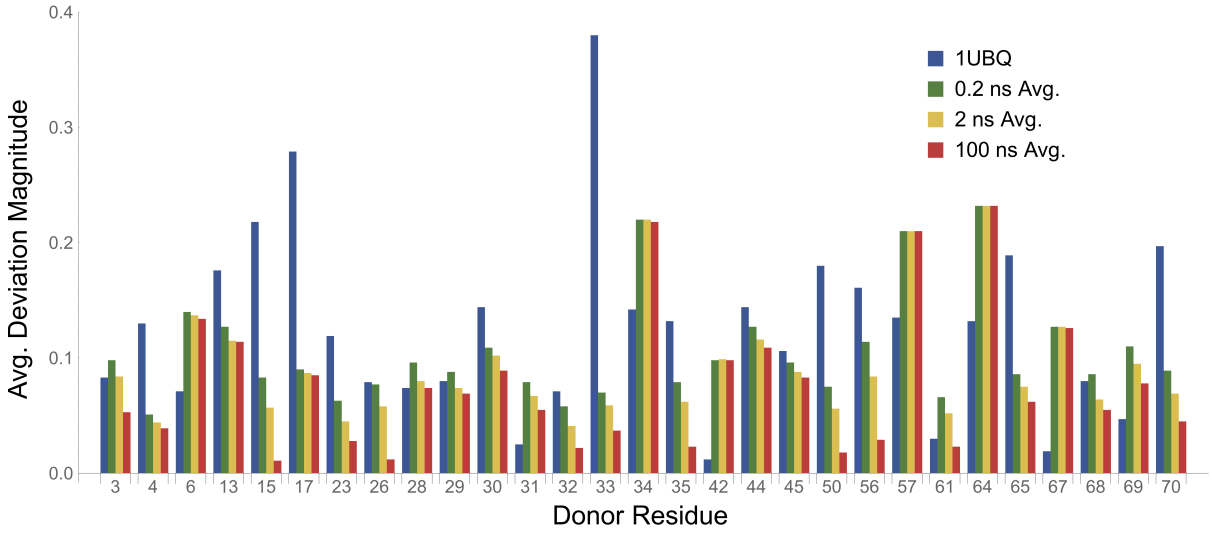


Figure 8.5: Calculated (Eq. 8.4) ubiquitin $^{3h}J_{NC'}$ deviation magnitudes decreasing with simulation length (static 1UBQ crystal structure, 0.2 ns, 2 ns, and 100 ns).

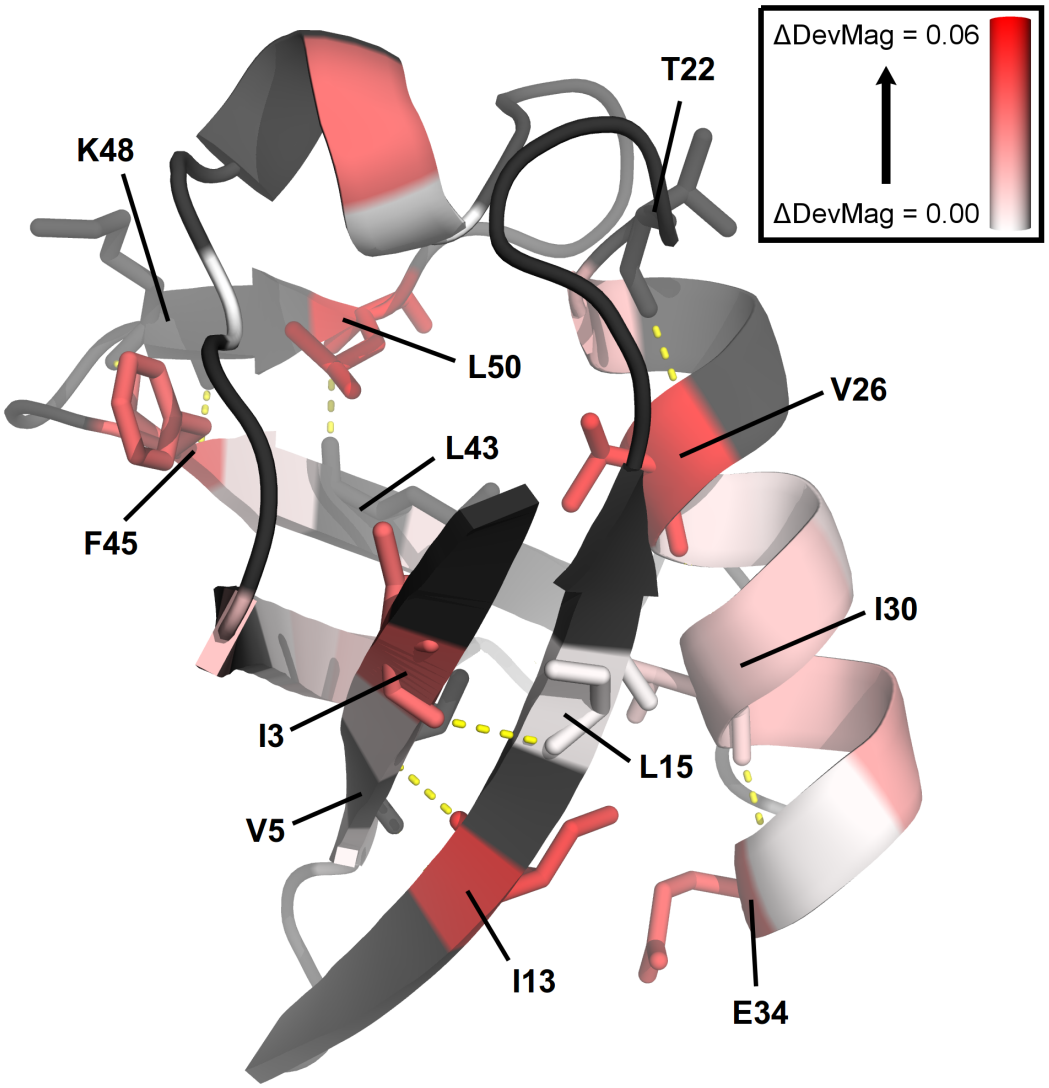


Figure 8.6: Changes in ubiquitin ${}^{3h}J_{NC'}$ deviation (relative to experiment) magnitudes resulting from averaging over longer (100 ns, rather than 2 ns) MD simulations projected onto the 1UBQ PDB structure (coloured by donor residue). This comparison emphasizes those hydrogen bonds for which slower dynamics may make the greatest contributions to experimentally measured ${}^{3h}J_{NC'}$ couplings. Black: not applicable (no hydrogen bond/measurable ${}^{3h}J_{NC'}$ coupling); white to red: increasing improvement (decreasing deviation magnitude) with simulation length. Hydrogen bonds (dashed yellow lines) corresponding to the six ${}^{3h}J_{NC'}$ averages showing the greatest RMSD improvement (>0.03; bright red) as well as the side chains of the donor/acceptor residues involved are shown.

We extend our study to the temperature dependence of ubiquitin ${}^{3h}J_{NC'}$ couplings by simulating at the same five temperatures for which experimental data have been published

[218]. Despite the unprecedented low RMSD of couplings (0.097 Hz, 34% lower than the ubiquitin 0.5 ns MD-average $^{3h}J_{NC'}$ RMSD previously published [220]) we calculate at a fixed temperature of 298 K, numerous discrepancies between the simulated (Table 8.3) and experimentally determined (Table 8.4) temperature dependences are apparent. Consistent with an overall trend toward weaker $^{3h}J_{NC'}$ coupling with increasing temperature, 19 of 31 experimentally measured $^{3h}J_{NC'}$ couplings show a monotonic decrease in magnitude (Table 8.4); in contrast, only 3 of 31 MD-average couplings show a monotonic decrease in magnitude (Table 8.3). We note that some of the couplings with the lowest magnitudes of deviation at 298 K (Fig. 8.5) are among those for which the temperature dependence is not accurately reproduced, suggesting that an element of randomness (e.g., the stochastic nature of conformational sampling in simulation) may contribute to our inability to resolve subtle temperature dependences. Interestingly, the average (mean of all couplings) trend is quite well reproduced (bottom rows of Tables 8.3 and 8.4).

We propose that the discrepancies between Table 8.3 and Table 8.4 may be explained by the length of the simulation (which is limited by practical considerations) and fluctuations in the instantaneous temperature (determined by the kinetic energies of the atoms). Experimental temperature measurements quantify the average kinetic energy of atoms in a vast number of molecules. In simulation, the temperature of the system (a single protein and the surrounding water molecules) is held constant only in a time-average sense; the instantaneous temperatures observed in simulation are normally distributed with a standard deviation of approximately 3 K (Fig. 8.7). Accordingly, the simulations required in order to calculate $^{3h}J_{NC'}$ coupling temperature dependences that match experiment may be extremely long. Other factors (e.g., force field accuracy) may also impact our ability to reproduce the temperature dependences of individual $^{3h}J_{NC'}$ couplings from simulation, but the match between calculated and experimental trends in the average over all couplings supports the hypothesis that simulation length (which facilitates coupling-specific averaging) is the primary obstacle.

NMR observables represent conformational averages; reproducing them from MD trajectories requires sampling the relevant protein conformations in the correct proportions. The low RMSD of the MD-average ubiquitin $^{3h}J_{NC'}$ couplings that we calculate at 298 K suggests that our simulations capture hydrogen bond dynamics that approximate those contributing to experimental measurements; however, conformational sampling sufficient for calculation of temperature dependences of individual $^{3h}J_{NC'}$ couplings that match experiment may require impractically long simulations. Having validated (and found the limitations of) our simulation methodology, we apply it in subsequent chapters to produce insights into Adnectin dynamics that facilitate our analysis of the temperature dependences of amide chemical shifts.

Table 8.3: Ubiquitin ${}^{3h}J_{NC'}$ couplings (Hz)[1] calculated (Eq. 8.2) from 100 ns MD simulations at five different temperatures.

		Temperature (K)				
Donor	**Acceptor**	**278**	**298**	**318**	**328**	**338**
3	15	-0.487	-0.498	-0.485	-0.482	-0.457
4	65	-0.669	-0.644	-0.630	-0.478	-0.550
6	67	-0.774	-0.728	-0.714	-0.743	-0.704
7	11	-0.487	-0.465	-0.463	-0.306	-0.440
13	5	-0.696	-0.714	-0.670	-0.635	-0.640
15	3	-0.570	-0.531	-0.534	-0.625	-0.579
17	1	-0.699	-0.660	-0.622	0.533	-0.523
23	54	-0.537	-0.581	-0.562	-0.511	-0.551
26	22	-0.306	-0.277	-0.246	-0.253	-0.271
27	23	-0.501	-0.441	-0.421	-0.342	-0.364
28	24	-0.322	-0.332	-0.365	-0.403	-0.366
29	25	-0.384	-0.357	-0.244	-0.324	-0.310
30	26	-0.363	-0.321	-0.233	-0.198	-0.213
31	27	-0.434	-0.427	-0.388	-0.303	-0.303
32	28	-0.368	-0.346	-0.276	-0.273	-0.268
33	29	-0.376	-0.429	-0.337	-0.393	-0.382
34	30	-0.748	-0.704	-0.740	-0.650	-0.632
35	31	-0.124	-0.110	-0.101	-0.085	-0.099
42	70	-0.650	-0.622	-0.649	-0.656	-0.610
44	68	-0.695	-0.688	-0.655	-0.719	-0.728
45	48	-0.530	-0.497	-0.561	-0.524	-0.500
50	43	-0.603	-0.597	-0.607	-0.653	-0.658
56	21	-0.356	-0.326	-0.349	-0.344	-0.318
57	19	-0.454	-0.435	-0.411	-0.434	-0.398
61	56	-0.460	-0.435	-0.426	-0.446	-0.405
64	2	-0.818	-0.773	-0.775	-0.713	-0.745
65	62	-0.062	-0.063	-0.053	-0.043	-0.047
67	4	-0.813	-0.813	-0.816	-0.770	-0.762
68	44	-0.619	-0.590	-0.602	-0.528	-0.506
69	6	-0.627	-0.612	-0.469	-0.570	-0.552
70	42	-0.612	-0.648	-0.536	-0.524	-0.525
	Mean:	-0.530	-0.515	-0.492	-0.479	-0.477

[1]Blue/red if the magnitude of the coupling decreased/increased relative to the previous (lower) temperature

Table 8.4: Ubiquitin $^{3h}J_{NC'}$ couplings (Hz)[1,2] measured at five different temperatures.

		Temperature (K)				
Donor	**Acceptor**	**278**	**298**	**318**	**328**	**338**
3	15	-0.436	-0.445	-0.444	-0.450	-0.496
4	65	-0.627	-0.605	-0.570	-0.558	-0.562
6	67	-0.604	-0.594	-0.587	-0.574	-0.562
7	11	-0.621	-0.579	-0.565	-0.559	-0.555
13	5	-0.753	-0.725	-0.699	-0.699	-0.661
15	3	-0.642	-0.616	-0.587	-0.573	-0.570
17	1	-0.598	N/D	-0.545	-0.524	-0.493
23	54	N/D	-0.553	-0.533	-0.514	N/D
26	22	N/D	-0.289	-0.268	-0.258	-0.250
27	23	-0.531	-0.515	-0.497	-0.492	-0.467
28	24	-0.261	-0.263	N/D	-0.248	-0.256
29	25	-0.259	-0.268	-0.263	-0.246	-0.247
30	26	-0.406	-0.376	-0.362	-0.348	-0.339
31	27	-0.406	-0.405	-0.399	-0.385	-0.365
32	28	-0.315	-0.309	-0.270	-0.232	N/D
33	29	N/D	-0.211	-0.173	-0.204	N/D
34	30	-0.725	-0.681	-0.666	-0.662	-0.645
35	31	-0.220	-0.208	-0.200	-0.181	-0.225
42	70	-0.472	-0.513	-0.514	-0.499	-0.486
44	68	-0.620	-0.605	-0.582	-0.571	-0.563
45	48	-0.504	-0.515	-0.479	-0.496	-0.462
50	43	N/D	-0.626	-0.589	-0.588	N/D
56	21	-0.549	-0.536	-0.477	N/D	N/D
57	19	N/D	-0.412	-0.369	-0.345	-0.380
61	56	-0.246	-0.203	-0.193	-0.198	-0.178
64	2	-0.890	-0.835	-0.781	-0.735	-0.723
65	62	N/D	N/D	-0.198	-0.182	N/D
67	4	-0.696	-0.687	-0.629	-0.619	-0.571
68	44	-0.663	-0.645	-0.605	-0.601	-0.592
69	6	-0.575	-0.534	-0.502	-0.496	-0.506
70	42	-0.637	-0.603	-0.582	-0.578	-0.595
	Mean[3]:	-0.536	-0.519	-0.510	-0.489	-0.483

[1]Blue/red if the magnitude of the coupling decreased/increased relative to the previous (lower) temperature

[2]Data from [218]

[3]Excluding rows with no data at one or more temperatures

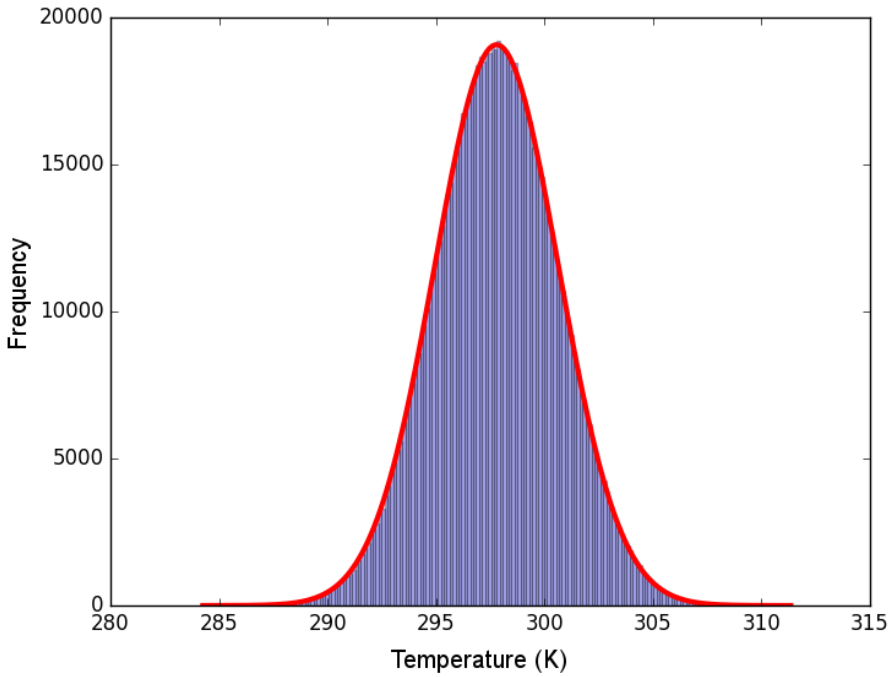


Figure 8.7: A histogram of instantaneous temperatures (based on the kinetic energy and degrees of freedom of the atoms in the system) from a nominally 'constant temperature' (298 K) 100 ns MD simulation of human ubiquitin. The temperature is controlled via Langevin dynamics with a damping coefficient of 5 ps^{-1}. The temperatures are normally distributed (fit shown in red) with a mean of 298 K and a standard deviation of 2.9 K.

Chapter 9

Temperature Dependence of β-sheet Hydrogen Bond Geometry

9.1 Context

The nuclear magnetic resonance (NMR) chemical shifts of proteins are known to be extremely sensitive to the local magnetic environments of nuclei and, consequently, to protein conformation and intermolecular (e.g., with ligands or solvent) interactions. In this chapter we analyze the temperature dependences of amide proton chemical shifts in structured proteins, for which the slopes of linear approximations (temperature coefficients) are often interpreted in terms of hydrogen bonding [244–246] and temperature-dependent loss of structure [247–251]. Deviations from linearity have also been observed [251–256], and we analyze curvature in the temperature dependences of Adnectin amide proton chemical shifts in Chapter 10. Though amide proton chemical shifts are not determined solely by hydrogen bond geometry, we leverage the temperature coefficients of amide protons involved in inter-β-strand hydrogen bonds to calculate bounds on thermal expansion of Adnectin β-sheet secondary structure consistent with experiment. We find that even the maximum temperature-dependent changes (over a 40 K range) in average Adnectin β-sheet hydrogen bond geometry consistent with amide proton temperature coefficients are very small (e.g., $\Delta r_{HO'} < 0.1$Å), affirming the extraordinary sensitivity of NMR spectroscopic measurements. Also, consistent with previous observations [247], we find no correlation between β-sheet hydrogen bond $r_{HO'}$ and amide proton temperature coefficients, suggesting that the strength of a β-sheet hydrogen bond may be a poor predictor of its thermal expansion. We note that protein stability is the net result of diverse contributions, and hydrogen bond geometry may be restrained by many factors aside from the strength of the bond itself.

9.1.1 Acknowledgements

Colleen Doyle assisted with the collection and analysis of NMR data (full details in Chapter 7, Section 7.1).

9.2 Introduction

The free energy difference between Boltzmann-distributed protein conformations is a function of temperature (Eq. 9.1), therefore the free energy landscape (including the native basin) of a protein is not fixed. No measurable fraction of protein is globally unfolded in the temperature range (288-328 K) over which we measure Adnectin chemical shifts; thus, in a broad sense, the temperature coefficients of these chemical shifts report on changes in the energy landscape of the native basin that reflect the increasing (with temperature) influence of entropy. In accordance with this interpretation, we find that Adnectin temperature coefficients are generally consistent with increasing dynamics and, particularly within stable secondary structure, weakening hydrogen bonds. However, we note that the energetics governing temperature-dependent conformational changes are those of the entire system (including solvent), while temperature coefficients are exclusively local reporters. In general, thermodynamic inference from individual temperature coefficients is not straightforward.

$$\Delta G(T) = \Delta H(T) - T \cdot \Delta S(T) \tag{9.1}$$

Though accurate prediction of protein chemical shifts is more difficult for amide protons than for other nuclei (e.g., from an experimentally determined structure) [217], many contributing factors have been studied; several known to be particularly important are summarized in Table 9.1. On the basis of what is known, we may interpret amide proton temperature coefficients as a consequence of specific types of conformational change.

To the extent that they change with temperature, the factors listed in Table 9.1 also determine the temperature dependence of the chemical shifts. Over the temperature range studied, those mediated by covalent bonds may vary little relative to those affecting amide proton shielding through space (e.g., hydrogen bond geometries, or the positions of amide protons relative to aromatic side chains). The relationship between hydrogen bonding and chemical shift is particularly well-studied, both computationally and experimentally.

As is the case for ${}^{3h}J_{NC'}$ couplings (Chapter 8), density functional theory (DFT) chemical shift calculations can be accurately reproduced as a function of hydrogen bond geometrical parameters (Fig. 8.1; Eq. 9.2 or, neglecting the weak dependence on ρ, Eq. 9.3) [219]. However, perhaps highlighting the importance of factors aside from hydrogen bonding (Table 9.1), the chemical shift predictions of Eqs. 9.2-9.3 (or the DFT calculations on which they are based) show only moderate correlation ($R^2 \sim 0.5$) with experimentally determined values [219].

Table 9.1: Factors determining amide proton chemical shifts

Factor	**Description**
Inductive Effects	Foremost, covalently mediated electron-withdrawing effects of the amide nitrogen and carbonyl; other inductive effects (e.g., through-bond influences of nearby side chains) may also contribute
Magnetic Anisotropy	Local magnetic fields induced by the circulation of π electrons, e.g., ring currents in aromatic side chains
Hydrogen Bonding	Amide proton deshielding, primarily as a function of $r_{HO'}$ and θ_2 (Fig. 8.1)
Conformational Averaging	Sufficiently fast exchange between states results in observation of population-weighted average chemical shifts

$$\delta_H = \Big(4.81cos^2(\theta_2) + [3.01cos^2(\rho) - 0.84cos(\rho) + 1.75]sin^2(\theta_2)\Big) \cdot e^{-2.0(r_{HO'}-1.76)} + 4.06\ ppm \tag{9.2}$$

$$\delta_H = 3.59cos^2(\theta_2) \cdot e^{-2.0(r_{HO'}-1.76)} + 5.32\ ppm \tag{9.3}$$

Experimentally determined temperature coefficients have been used to discriminate between amide protons involved in intramolecular hydrogen bonds and those hydrogen bonded to solvent [245] and, by the same token (because amides in random coil are unlikely to form stable intramolecular hydrogen bonds), as an indicator of disorder [257]. However, the chemical shift distributions of hydrogen bonded and non-hydrogen bonded amide protons (where hydrogen bond status is inferred from experimentally determined structures) show substantial overlap [245]. This overlap may be attributable to factors such as temperature-dependent changes in the relative positions of aromatic side chains and conformational averaging [249].

We distinguish two types of conformational averaging: exchange between experimentally indistinguishable microstates (e.g., like the oscillations in hydrogen bond length illustrated in Fig. 8.2), and exchange between distinct conformational states. The temperature dependence of the relative occupation of microstates is not easily probed (Chapter 8), and is neglected here in favour of average hydrogen bond distances and angles. On the other hand, exchange between (potentially) distinguishable macrostates is a very important consideration; where it occurs, the temperature dependence of the equilibrium between macrostates substantially complicates the interpretation of chemical shift changes. In Chapter 10 we

explicitly analyze curvature that may arise from such temperature-dependent equilibria, but even when no curvature is apparent, linear temperature coefficients may be determined primarily by conformational exchange [249].

Here, we first compare the temperature coefficients of five Adnectins that differ in one or more biophysical characteristics of interest; then, from those involved in intramolecular hydrogen bonds, we select the subset of pWT amide protons least likely to be measurably affected by conformational exchange between distinct states (i.e., those for which the relationship between temperature coefficients and hydrogen bond geometry may be more easily analyzed). In doing so, we consider experimentally determined wild-type ^{10}Fn3 structures, simulations thereof, amide hydrogen/deuterium (H/D) exchange data (Chapter 6), and the linearity of the temperature dependences of chemical shifts (Chapter 10). We analyze the temperature coefficients of this subset and place bounds on temperature-induced change in hydrogen bond geometries.

9.3 Methods

9.3.1 Expression & Purification

Expression and purification of ^{15}N-labelled protein, variable-temperature NMR experiments, and data analysis (including determination of temperature coefficients) were conducted as described in Chapter 7, Section sec:shiftt˙meth.

9.3.2 Molecular Dynamics

Molecular dynamics simulations were performed as described in Chapter 8, Section 8.3.2.

9.4 Results & Discussion

Amide proton temperature coefficients for pWT, Parent, L78I, V75R, and L18V/Y88F Adnectins are visualized in Fig. 9.1 (numerical data can be found in Appendix H). The L78I, V75R and L18V/Y88F Adnectins are named for their relationships to the Parent amino acid sequence, from which they differ by one or two point mutations. The pWT sequence differs from that of the Parent in three loops, one of which is longer by four residues. Specifically, the pWT BC loop (residues 23-29), DE loop (residues 52-55), and FG loop (residues 77-86) differ from equivalent loops in the other Adnectins (Fig. 2.2; Appendix D).

We find that, relative to the Parent Adnectin, L78I is substantially more aggregation-prone (Chapter 4), V75R is substantially less aggregation-prone (Chapter 5), while L18V/

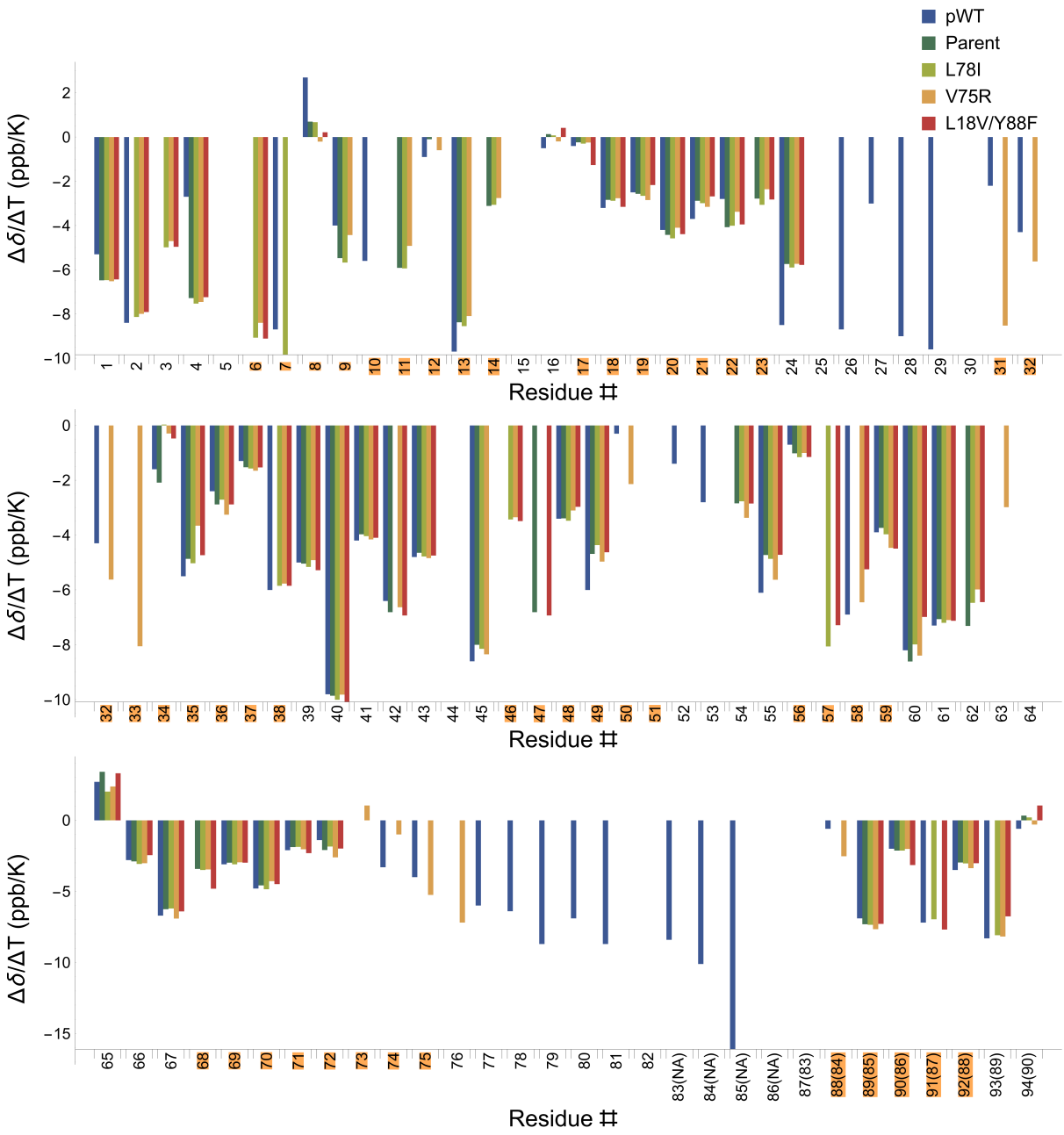


Figure 9.1: Adnectin amide proton linear temperature coefficients. Residue numbers shown as pWT #(other Adnectins #) where applicable. Residue numbers highlighted in orange are in the β-strands. Bars for residues missing assignments or for which the temperature dependence could not be tracked (distinguishable in Appendix H, Table H.1) are omitted.

Y88F has two mutations to side chains within the hydrophobic core. We see in Fig. 9.1 that, despite these biophysical differences, only minor variations in the temperature coefficients of equivalent residues are observed. The temperature coefficients of pWT are also generally similar, although we perceive differences (even in regions of identical sequence), such as a weak trend towards more positive pWT temperature coefficients in the N-terminal region and β-strand A.

In the N-terminal region, as well as others outside of stable secondary structure, we do not predict a simple relationship between the magnitudes of individual temperature coefficients and temperature-dependent changes in hydrogen bond geometry. As is the case for peptides [249], the temperature coefficients of less-structured residues may be dominated by conformational exchange (including, though not necessarily limited to, exchange between relatively structured and unstructured states, which may be described as temperature-dependent loss of structure). Though insights into the nature of this conformational exchange are certainly of great interest (Chapter 10), factors other than the relative populations of states (e.g., the difference between the temperature-dependent chemical shifts of the conformational states) contribute to experimentally measured temperature coefficients. Instead, we focus here on the Adnectin amide protons least likely to be affected by conformational exchange between distinct states and place bounds on the changes in hydrogen bond geometry that could explain chemical shift changes observed over our 288-328 K experimental temperature range.

In Figure 9.2 we show the variation in chemical shift with amide proton to acceptor distance ($r_{HO'}$) predicted by Eq. 9.3, which we favour over Eq. 9.2 for its greater simplicity, and because we found that including the ρ parameter did not generally improve our MD-average ${}^{3h}J_{NC'}$ calculations (Chapter 8, Table 8.1). The chemical shift of an amide proton with a temperature coefficient of -4.6 ppb/K (a minimum threshold widely used to discriminate hydrogen bonded amide protons [245]) changes by -0.18 ppm over a 40 K temperature range. Although chemical shift varies nonlinearly with $r_{HO'}$, deviation from linearity is negligible over the very small distances corresponding to a decrease of 0.18 ppm. Segments of the curve produced by varying θ_2 (with constant $r_{HO'}$) are also linear over ranges corresponding to plausible changes in chemical shift, except near $\theta_2 = 180°$. Interestingly, though they approach the (angular) maximum hydrogen bond strength, θ_2 values near 180° are rare (Table 9.2). We hypothesize that they may be entropically unfavourable (i.e., for a given $r_{HO'}$, there may be relatively few accessible configurations in which $\theta_2 = 180°$).

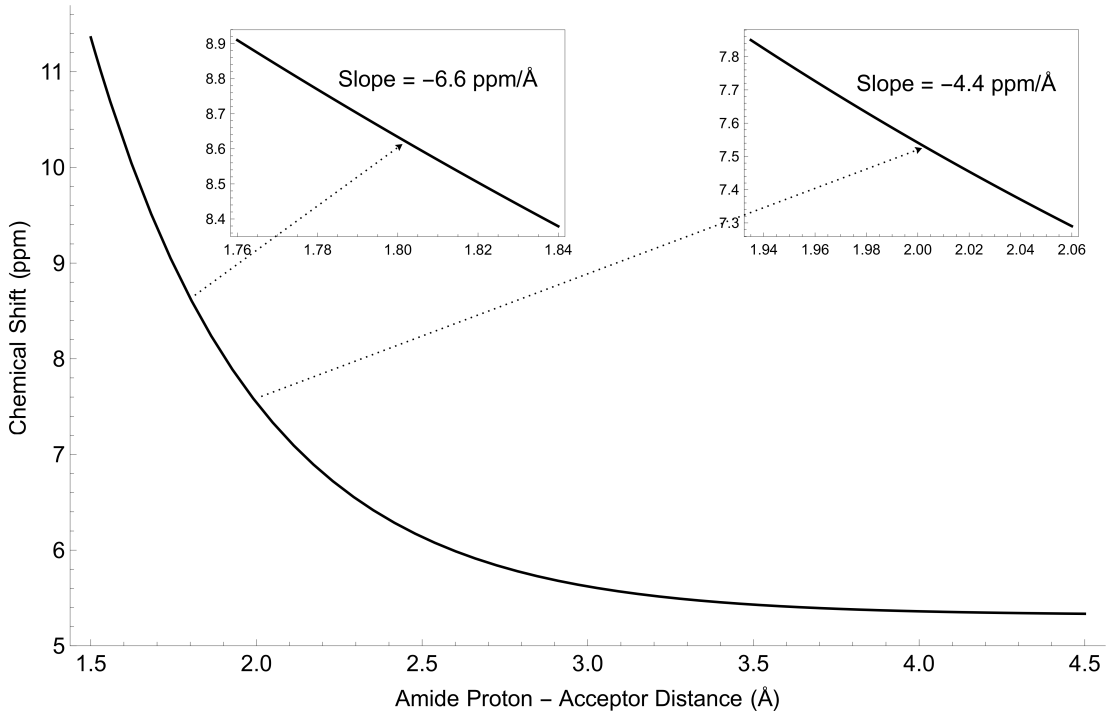


Figure 9.2: Changes in amide proton chemical shift with $r_{HO'}$ predicted by Eq. 9.3 (θ_2 fixed at 180°). Inset figures show the same curve over a narrower range of $r_{HO'}$ values; slopes of the tangents to the curve at the points indicated (black arrows) are labelled.

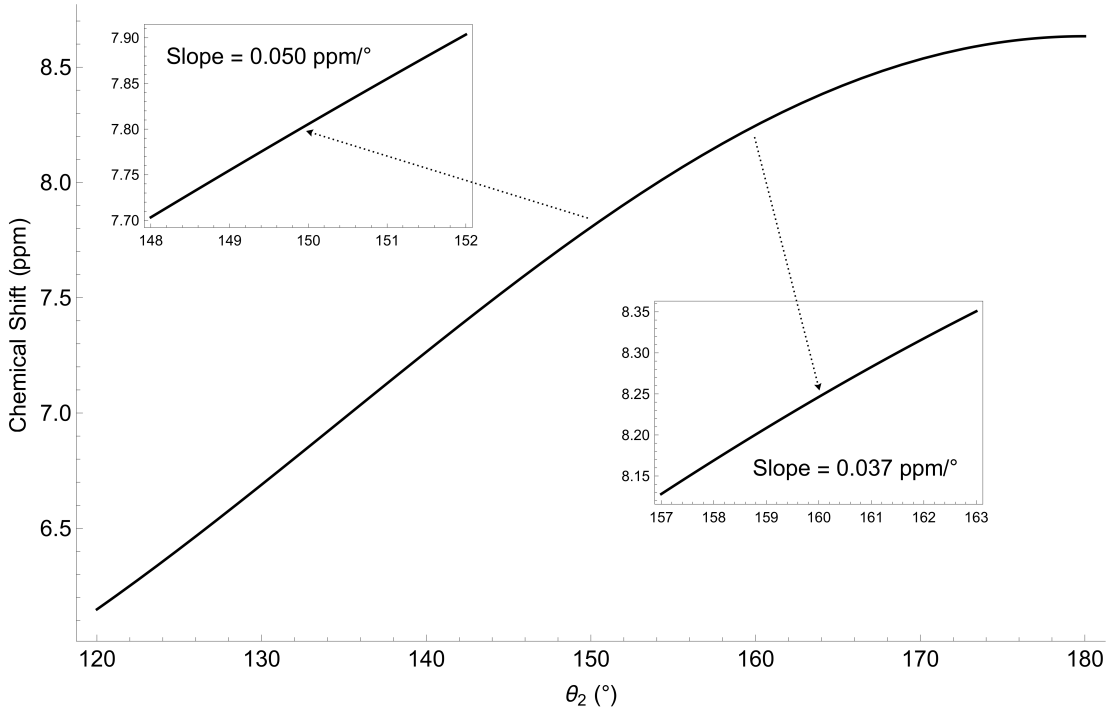


Figure 9.3: Changes in amide proton chemical shift with θ_2 predicted by Eq. 9.3 ($r_{HO'}$ fixed at 1.8 Å). Inset figures show the same curve over a narrower range of θ_2 values; slopes of the tangents to the curve at the points indicated (black arrows) are labelled.

Two crystal structures of wild-type ^{10}Fn3 are available from the Protein Data Bank (PDB): 1FNF (2.0 Å resolution) [194], and 1FNA (1.8 Å resolution) [193]. The C_α root mean square deviation (RMSD) between these structures is 1.7 Å including all residues, or 0.6 Å including only β-strand residues; in other words, allowing for loop flexibility, they align very well. We begin selection of the subset of amide protons least likely to have chemical shifts affected by conformational exchange by considering those that appear to be involved in intra-β-sheet hydrogen bonds, as determined by submission of the ^{10}Fn3 domain from the 1FNF PDB structure to the VADAR protein structural analysis web server [258]. Values of $r_{HO'}$ and θ_2 inferred from the 1FNF and 1FNA ^{10}Fn3 structures are listed in Table 9.2, as are the average values observed in a 100 ns MD simulation. Due to finite resolution and crystal packing artefacts on one hand, and timescale and force field accuracy limits on the other, all of these values must be regarded as approximate, but they may serve to define local slopes (as in Figs. 9.2 and 9.3).

We exclude from our analysis any amide protons found to exhibit curvature in the temperature dependence of their chemical shifts (Chapter 10; Table 9.2, gray rows). Despite the fact that no curvature was detected for the amide protons of residues 90 and 92, amide H/D exchange data suggest their involvement in conformational exchange (Chapter 6; Table 9.2, yellow rows). However, the apparent free energy of exchange (at 298 K) is greater

Table 9.2: pWT Adnectin intra-β-sheet hydrogen bond parameters and calculated changes with temperature

				1FNF[3]		1FNA[3]		MD Avg.[4]			
Donor[1]	β-β	Acceptor[1]	$\Delta\delta/\Delta T$[2]	$r_{HO'}$[5]	θ_2[6]	$r_{HO'}$[5]	θ_2[6]	$r_{HO'}$[5]	θ_2[6]	$\Delta r_{HO'}$[7]	$\Delta\theta_2$[8]
6	A-B	23	N/S	2.09	139	N/A	N/A	2.12	146		
9	A-B	21	-3.96	1.83	159	1.96	160	2.01	161		
11	A-B	19	N/S	1.82	157	2.07	150	1.95	153		
14	A-B	17	N/S	2.13	115	1.90	172	C/E	C/E		
17	B-A	14	-0.42	2.34	105	2.79	93	C/E	C/E		
18	B-E	59	-3.24	2.44	123	1.99	149	1.95	154	0.03	-3.7
19	B-A	12	-2.45	1.98	121	1.84	151	C/E	C/E		
20	B-E	57	-4.25	1.90	148	1.95	166	2.03	162	0.05	-6.9
21	B-A	9	-3.73	1.84	161	2.00	160	2.01	156	0.04	-4.9
22	B-E	55	-2.84	2.32	150	2.10	164	2.22	155	0.05	-5.5
23	B-A	6	N/S	1.69	147	N/A	N/A	C/E	C/E		
32	C-D	50	-4.31	1.83	143	1.82	159	1.87	160	0.04	-4.9
33	C-F	73	N/S	1.95	140	2.00	145	C/E	C/E		
34	C-D	48	-1.56	1.99	173	1.89	167	1.93	163	0.01	-2.4
35	C-F	71	-5.55	2.00	133	1.90	150	2.02	157	0.07	-7.4
36	C-D	46	-2.43	1.94	153	2.02	141	2.04	146		
37	C-F	69	-1.26	2.41	136	2.29	144	2.20	148	0.02	-2.1
46	D-C	36	N/S	1.97	157	1.83	156	C/E	C/E		
48	D-C	34	-3.37	2.26	149	2.16	148	2.04	152	0.04	-4.3
50	D-C	32	-0.26	2.07	144	2.01	145	1.94	145	0.00	-0.3
57	E-B	20	N/S	2.07	148	2.19	146	2.14	144		
59	E-B	18	-3.93	1.88	157	1.88	156	1.90	149		
68	F-G	92	N/S	2.78	125	1.82	140	1.98	155		
69	F-C	37	-3.06	1.81	167	1.93	166	1.94	161	0.03	-4.2
70	F-G	90	-4.83	2.01	146	1.82	165	1.93	160	0.04	-6.0
71	F-C	35	-2.13	1.84	153	1.91	156	1.89	158	0.02	-2.4
72	F-G	88	-1.36	1.98	158	1.92	153	1.95	162	0.01	-2.1
73	F-C	33	N/S	1.94	140	1.98	153	1.97	155		
75	F-C	31	-4.00	1.94	157	1.89	171	1.96	165	0.04	-6.4
88	G-F	72	-0.56	2.32	148	2.12	145	2.44	147		
90	G-F	70	-2.05	2.08	147	1.98	148	2.02	152	0.03	-2.6
92	G-F	68	-3.45	1.91	142	2.03	148	2.00	149	0.04	-3.9

[1] pWT Adnectin residue #; gray rows indicate donor amide proton chemical shifts with curved temperature dependences, and yellow rows indicate amide H/D exchange evidence of an alternate state (Chapter 6)
[2] Temperature coefficients ppb/K; N/S indicates that ShiftTrack (Chapter 7) found no solution
[3] PDB structure identifier; N/A indicates that no hydrogen bond could be inferred
[4] 100 ns simulation based on the ^{10}Fn3 domain from 1FNF; C/E indicates average parameters not calculated (due to evidence of conformational exchange between distinct states)
[5] Refer to Fig. 8.1 (Å)
[6] Refer to Fig. 8.1 (°)
[7] Maximum change over 278-328 K range calculated using Eq. 9.3 (Å)
[8] Maximum change over 278-338 K range calculated using Eq. 9.3 (°)

than 5 kcal/mol, thus the impact of the alternate state(s) on chemical shift observations may be very small; accordingly, though we have called attention to these amide protons, we include them in the analysis. Also excluded are any amide protons for which multimodal $r_{HO'}$ or θ_2 distributions (i.e., evidence of conformational exchange between distinct states) are observed in the MD simulation.

Interestingly, although our MD simulations are based on the ^{10}Fn3 domain from the 1FNF structure, many of the MD-average hydrogen bond geometries are closer to those observed in the slightly higher resolution 1FNA structure. Furthermore, where there are discrepancies between the geometries observed in the two crystal structures, the MD-average values often lie somewhere in between. Using MD-average $r_{HO'}$ and θ_2 values as our starting point, we apply Eq. 9.3 to estimate the maximum $\Delta r_{HO'}$ and $\Delta\theta_2$ by adjusting $r_{HO'}$ and θ_2 individually until the chemical shift difference matches that observed over the experimental (288-328 K) VT-NMR temperature range ($\Delta\delta/\Delta T \cdot 40\ K$). Thus, we find that changes in $r_{HO'}/\theta_2$ ranging from zero to $\Delta r_{HO'}/\Delta\theta_2$ are consistent with experiment. We also find that the $\Delta r_{HO'}$ and $\Delta\theta_2$ values each span a small range (over all amide protons), and are of magnitudes so small (e.g., distance changes on the order of hundredths of an angstrom) that detection by conventional structure determination methods, such as variable-temperature crystallography [259, 260], is unlikely.

It has been observed that short, strong hydrogen bonds sometimes have more negative temperature coefficients [246]. It is clear from Fig. 9.2 that the shorter the hydrogen bond, the steeper the $\Delta\delta/\Delta r_{HO'}$ gradient (a steeper $\Delta\delta/\Delta\theta_2$ gradient is also expected for shorter bonds); however, this only translates into a more negative temperature coefficient (than that of an amide proton involved in a longer hydrogen bond) for comparable temperature-dependent changes in hydrogen bond geometry ($\Delta r_{HO'}/\Delta T$ or $\Delta\theta_2/\Delta T$). Consistent with earlier reports [247], we report no correlation between inter-β-strand hydrogen bond length and temperature coefficients or $\Delta r_{HO'}$ (Fig. 9.4), suggesting that the strength of a hydrogen bond does not necessarily determine the extent of temperature-dependent changes in its geometry. Consistent with the results of experiments studying the sensitivity of hydrogen bonds to pressure and temperature [261], this lack of correlation suggests non-uniform β-strand thermostability. Reciprocally hydrogen bonded residues may be more likely to have comparable $\Delta r_{HO'}/\Delta T$ gradients; accordingly, the data in Table 9.2 show that the temperature coefficient of the amide proton involved in the shorter of two reciprocal hydrogen bonds is most often, though not always, more negative (examples include the residue pairs 20/57, 32/50, 37/69, and 70/90; counterexamples include 34/48; Table 9.2).

Plots of trans-hydrogen bond $^{3h}J_{NC'}$ couplings vs. amide proton chemical shift at multiple temperatures may be of interest because they require no assumptions about $\Delta r_{HO'}/\Delta T$ or $\Delta\theta_2/\Delta T$. Comparison of Eq. 8.4 ($^{3h}J_{NC'}$ coupling) and Eq. 9.3 (amide proton chemical shift), which share the general form of their nonlinear dependence on hydrogen bond geometry, but are parameterized differently, leads us to predict that shorter hydrogen bonds should display steeper $\Delta(^{3h}J_{NC'})/\Delta\delta$ slopes. Although Adnectin $^{3h}J_{NC'}$ couplings have not been measured, there is some evidence that this prediction holds for human ubiquitin (Fig. 9.5; stronger hydrogen bonds have more negative $^{3h}J_{NC'}$ couplings). These results

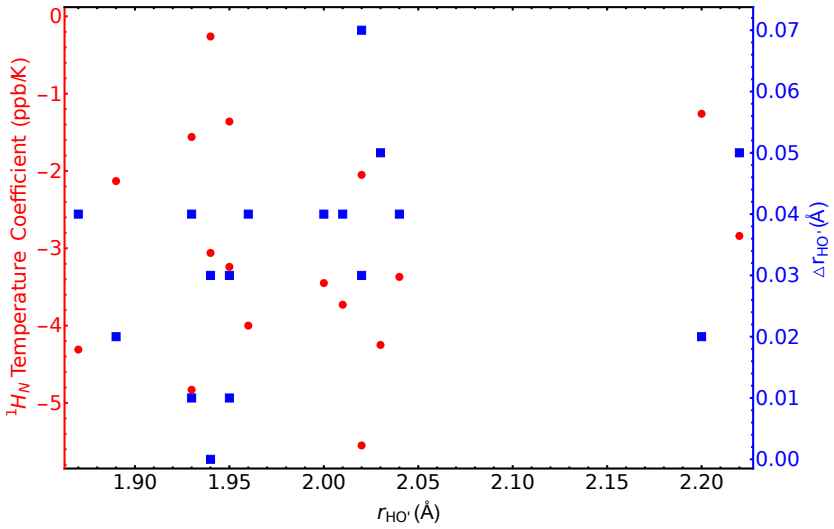


Figure 9.4: Temperature coefficients (red circles) and maximum changes in hydrogen bond length consistent with experiment (blue squares) vs. MD-average hydrogen bond length (data from Table 9.2). No correlations are apparent, suggesting non-uniform β-strand thermostability.

suggest that amide protons involved in stable intramolecular hydrogen bonds (i.e., those with measurable $^{3h}J_{NC'}$ couplings), generally have chemical shift temperature dependences attributable primarily to changes in hydrogen bond geometry, validating our analysis of select pWT Adnectin temperature coefficients (Table 9.2) in terms of $\Delta r_{HO'}$ and $\Delta\theta_2$.

The strong relationship between hydrogen bond geometry and amide proton temperature coefficients is well-documented (Section 9.1), and temperature coefficients are routinely presented in support of conclusions regarding the presence or absence of secondary structure. In the aggregate, this use may be justifiable; however, it is known that individual temperature coefficients are an error-prone indicator of hydrogen bond status, in part because of conformational averaging [249]. Here we include only the amide protons least likely to be involved in conformational exchange between distinct states in our analysis; however, acknowledging the possibility of temperature-dependent influences aside from hydrogen bonding (e.g., magnetic anisotropy), we present our results in terms of maximum changes (with temperature) in average hydrogen bond geometry that are consistent with experiment. Even the largest of these maximum changes may be too small to readily resolve by means other than NMR spectroscopy, affirming the phenomenal sensitivity of chemical shifts to thermal expansion of secondary structure. Our data and calculations confirm earlier reports that temperature coefficients (and implied changes in hydrogen bond geometry with temperature) are poorly correlated with hydrogen bond length. This observation may be rationalized by the fact that protein stability is the net result of diverse contributions;

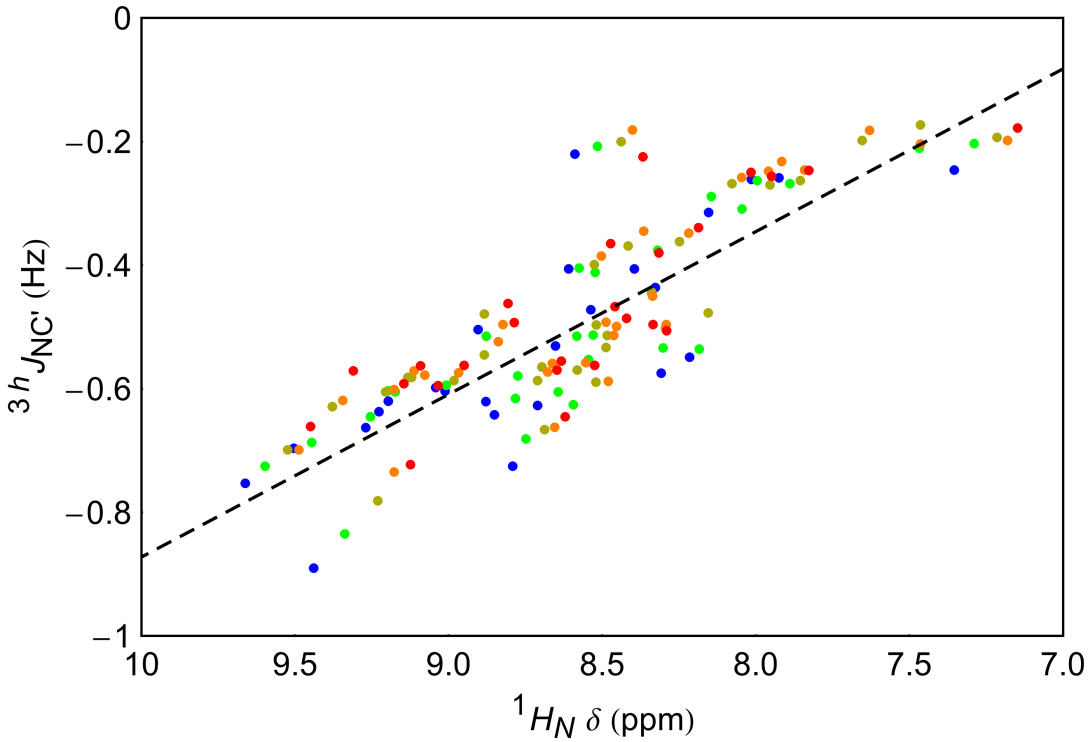


Figure 9.5: Ubiquitin $^{3h}J_{NC'}$ couplings vs. amide proton chemical shifts acquired at temperatures of 278 K, 298 K, 318 K, 328 K, and 338 K (blue to red; data from [218]). Strictly to provide a visual reference (slope), the dashed line shows the result of a linear regression including all points.

accordingly, hydrogen bond geometry may be restrained by numerous factors aside from the strength of the bond itself.

Chapter 10

Energetics of Conformational Exchange within the Native Ensemble

10.1 Context

The temperature dependences of amide proton chemical shifts are often well-approximated by a linear model (Chapter 9), and deviations from linearity (curvature) have been interpreted as evidence of fast exchange between distinct conformational states (Section 10.2). Here we find that Adnectin residues not involved in β-structure are more likely to exhibit measurable curvature, including loops (BC, DE, and FG; Fig. 2.2) that contribute to target-binding functionality. In what is, to our knowledge, the first quantitative thermodynamic analysis of curvature, we fit variable temperature nuclear magnetic resonance (VT-NMR) data to a two-state, fast exchange model. We compare model-based estimates of free energy differences between curvature-related states with the occupation of conformational states observed in molecular dynamics (MD) simulations, and find that it is possible to ascribe some instances of experimentally measured curvature to specific molecular mechanisms.

10.1.1 Acknowledgements

Colleen Doyle assisted with the collection and analysis of NMR data (full details in Chapter 7, Section 7.1).

10.2 Introduction

Curvature in the temperature dependences of amide proton chemical shifts (hereafter, simply 'curvature'), has been modelled as, and interpreted in terms of fast exchange between distinct conformations [251–256]. We introduced curvature in Chapter 7, where the process of detecting and validating this sometimes subtle phenomenon is described in detail. Here we undertake a quantitative analysis of experimentally measured curvature (to our knowledge, the first of its kind), beginning with the development of a model that describes curvature as the result of fast exchange between conformational ensembles with distinct chemical shifts (states).

We fit VT-NMR data to this model and reproduce experimental results with extremely high accuracy, i.e., with root-mean-square errors (RMSEs) lower than 0.0004 ppm in favourable cases. However, in part because the biophysical characteristics of Adnectins limit the temperature range over which they can be studied with no measurable fraction denatured, the experimental data are not sufficient to restrain the model parameters to a unique solution. Instead, the fitting exercise places bounds on the parameters; in the examples on which we focus, these bounds are tight enough to generate useful insights into the nature of the states between which fast exchange is postulated.

Furthermore, we detect states in MD simulations that may be related to curvature, and compare the observed occupancies of these states with thermodynamic predictions based on fits to our curvature model. The curvature of more than one residue may report on the same equilibrium, and we calculate very similar free energy differences between curvature-related states for such residues. Population of strongly disfavoured states is too low to impart measurable curvature; accordingly, most of the free energy differences ($\Delta G = G_{major} - G_{minor}$) estimated from simulation (at 298 K) range from -1.7 to -0.3 kcal/mol. However, we report one example in which curvature may report on the transient loss of a hydrogen bond, a conformational change disfavoured by 3.7 kcal/mol at 298 K.

10.2.1 Conformational Change and Chemical Shifts

Proteins are dynamic molecules, and this is reflected in isotropic chemical shifts measured by solution NMR. The impact of conformational change on chemical shifts depends on the rates of exchange between conformations and the separations between the corresponding resonant frequencies [262]. The temperature dependence of amide proton chemical shifts is due to differences in the ensemble of conformations sampled and their probabilities of occupation.

In the fast exchange limit, i.e., when the rate of exchange between sampled conformations is much greater than the separation between their resonant frequencies ($k_{ex} >> \Delta f$), the chemical shift observed for a given amide proton is a population-weighted average. In the slow exchange limit ($k_{ex} << \Delta f$), exchange is not predicted to contribute to the temperature dependence of chemical shifts (rather, we would expect separate peaks), while

at intermediate exchange rates ($k_{ex} \sim \Delta f$) significant line broadening would be expected. Thus, provided that such line broadening is not observed, we assume that the conformations contributing to each peak are in fast exchange. This assumption figures prominently in our model development below.

10.3 Methods

10.3.1 Expression & Purification

Expression and purification of ^{15}N-labelled protein, VT-NMR experiments, and data analysis (including statistical analysis of curvature in the temperature dependence of amide proton chemical shifts) were conducted as described in Chapter 7, Section 7.3.

10.3.2 Two-State, Fast Exchange Model

Fitting of experimental data to the two-state, fast exchange model of curvature developed herein (Section 10.4) was carried out using Mathematica 10.1 (Wolfram Research). States A and B are defined as the majority (lower energy) and minority (higher energy) state, respectively, at 288 K ($\Delta H = H_A - H_B$, $\Delta S = S_A - S_B$, and $\Delta G = G_A - G_B$).

For fitting to the $\Delta C_p = 0$ model, the values of ΔS and ΔH were swept from -0.3 to 0.3 kcal·mol^{-1}·K^{-1} in 0.001 kcal·mol^{-1}·K^{-1} increments and $-7.5 + (\Delta S \cdot 288\ K)$ to $-0.1 + (\Delta S \cdot 288\ K)$ kcal/mol in 0.1 kcal/mol increments (i.e., testing $\Delta H/\Delta S$ combinations that give ΔG at 288 K ranging between zero and the approximate stability against global unfolding of wild-type ^{10}Fn3), respectively, and ΔG was calculated using Eq. 10.13.

For fitting to the $\Delta C_p \geq 0$ model, the values of $\Delta S(T_{ref})$, $\Delta H(T_{ref})$, and ΔC_p were swept from -0.15 to 0.15 kcal·mol^{-1}·K^{-1} in 0.001 kcal·mol^{-1}·K^{-1} increments, $-7.5 + (\Delta S \cdot 288\ K)$ to $-0.1 + (\Delta S \cdot 288\ K)$ kcal/mol in 0.1 kcal/mol increments, and 0 to 0.1 kcal·mol^{-1}·K^{-1} (i.e., approximately up to the ΔC_p observed by differential scanning calorimetry for global unfolding of Adnectins [102]) in 0.01 kcal·mol^{-1}·K^{-1} increments, respectively, and ΔG was calculated using Eq. 10.12.

The root mean square error (RMSE) of points determined by Eq. 10.10 relative to experimental data is linearly dependent on the remaining parameters (m_A, m_B, $\delta_A(T_{ref})$, and $\delta_B(T_{ref})$; therefore, for each pair of ΔH and ΔS values ($\Delta C_p = 0$ model) or triplet of $\Delta H(T_{ref})$, $\Delta S(T_{ref})$, and ΔC_p values ($\Delta C_p \geq 0$ model), the optimal values of m_A, m_B, $\delta_A(T_{ref})$, and $\delta_B(T_{ref})$ were determined by linear regression.

10.3.3 Molecular Dynamics

Molecular dynamics simulations were performed as described in Chapter 8, Section 8.3.2.

10.4 Results & Discussion

We detect curvature in a minority of residues, e.g., less than 1/3 of non-proline pWT Adnectin residues (Fig. 10.1), of which 85 out of 86 have been assigned to HSQC peaks. Consistent with our proposal that residues in stable secondary structure may be less affected by exchange between distinct conformational states (Chapter 9), we find that residues with amide protons involved in intra-β-sheet hydrogen bonds are considerably less likely to show curvature (Table 10.1). Conversely, curvature in loop regions is common; we note that several Adnectin loops are implicated in binding functionality (BC, DE, and FG; Chapter 3; Fig. 2.2), and may sample related conformations even in the absence of ligands. However, it is important to note that, although curvature may be interpreted as evidence of fast exchange between distinct states, the absence of curvature is less informative. In general, curvature may be obscured by missing assignments, overlapping peaks, peak broadening, limits on the range of temperatures over which experimental data were collected, and fast exchange between different conformational states with similar chemical shifts/temperature coefficients.

Nevertheless, we hypothesize that linear temperature dependence of amide proton chemical shifts may be characteristic of fast exchange between conformations within a single well in the free energy landscape (hereafter considered a single 'state'). For example, differences between linear amide proton temperature coefficients have been explained, albeit imperfectly, in terms of changes in average hydrogen bond strength [244–246]; fast exchange amongst a shifting (with temperature) ensemble of hydrogen bond geometries is implied. In order to model curvature, we allow for fast exchange both within and between wells in the free energy landscape. The model developed here describes two states; in principle, a greater number of states may be sampled, but we find that the two-state model closely reproduces experimental results. Similar two-state models have been used to 'simulate' curvature [251, 253], but we are not aware of any previous attempts to fit experimental data to a curvature model.

Table 10.1: Breakdown of pWT curvature by amide proton hydrogen bond involvement

Amide Proton	Curvature	Total	Fraction
β-β H-bond	5	32	16%
Other	21	54	39%
All[1]	26	86	30%

[1]Non-proline residues

For Boltzmann-distributed conformational states A and B, the ratio of their temperature-dependent probabilities of occupation, $p_A(T)$ and $p_B(T)$, has an exponential dependence on the Gibbs free energy difference between the states, $\Delta G(T)$, where R represents the gas

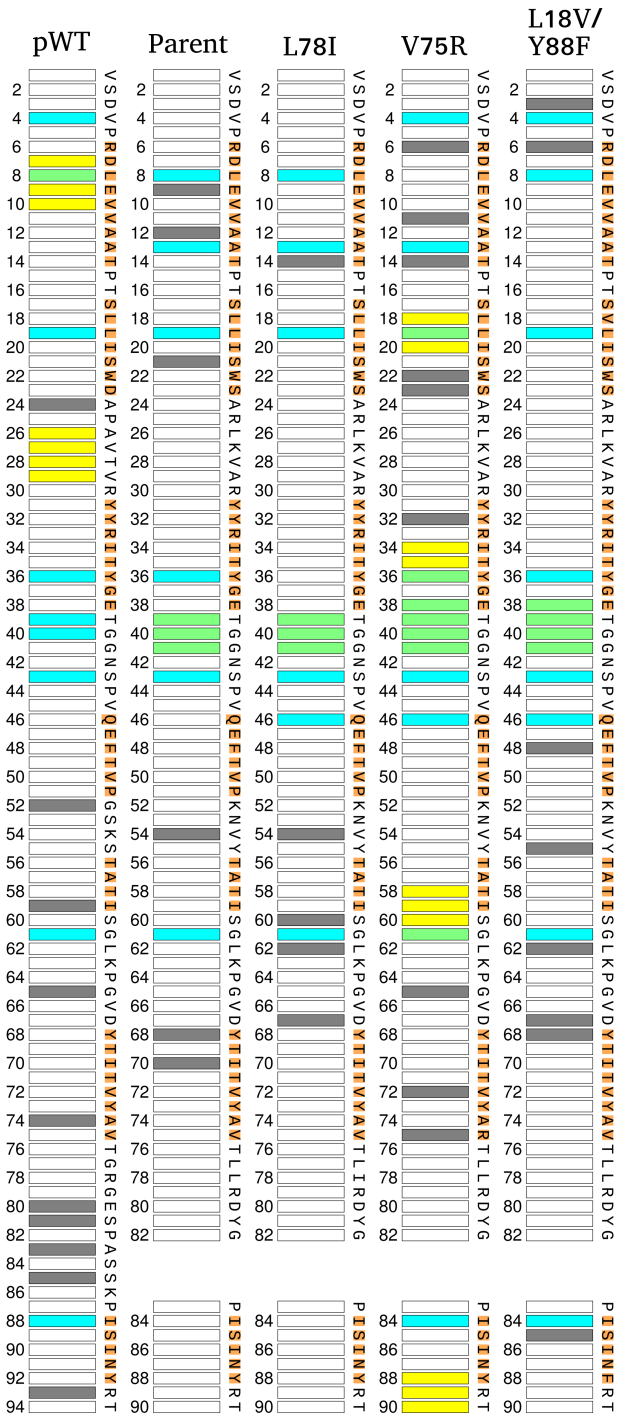


Figure 10.1: Curvature in the temperature dependence of amide proton chemical shifts. Grey bars indicate curvature, blue bars indicate curvature at the equivalent positions in three or more Adnectins, yellow bars indicate curvature at three or more consecutive positions in the same Adnectin, and green bars indicate that the criteria for both blue and yellow are met. The residues highlighted in orange form the β-sheets (for a more detailed mapping of curvature onto secondary structure, consult Table I.1)

constant ($0.001987\ kcal \cdot mol^{-1} \cdot K^{-1}$) and T is the temperature:

$$\frac{p_A(T)}{p_B(T)} = e^{-\Delta G(T)/RT} \tag{10.1}$$

$$p_A(T) = p_B(T) \cdot e^{-\Delta G(T)/RT} \tag{10.2}$$

If we consider an example with only two possible conformational states:

$$p_A(T) + p_B(T) = 1 \tag{10.3}$$

$$p_B(T) = 1 - p_A(T) \tag{10.4}$$

We can derive expressions for $p_A(T)$ and $p_B(T)$ by combining (10.2) and (10.4):

$$p_A(T) = (1 - p_A(T)) \cdot e^{-\Delta G(T)/RT} \tag{10.5}$$

$$p_A(T) = e^{-\Delta G(T)/RT} - p_A(T) \cdot e^{-\Delta G(T)/RT} \tag{10.6}$$

$$p_A(T) = \frac{e^{-\Delta G(T)/RT}}{1 + e^{-\Delta G(T)/RT}} \tag{10.7}$$

$$p_B(T) = 1 - p_A(T) = 1 - \frac{e^{-\Delta G(T)/RT}}{1 + e^{-\Delta G(T)/RT}} = \frac{1}{1 + e^{-\Delta G(T)/RT}} \tag{10.8}$$

In the fast exchange limit, the observed chemical shift, δ_{obs}, is the population-weighted average of the individual chemical shifts, δ_A and δ_B (Eq. 10.9), and into this equation we can substitute Eqs. 10.7 and 10.8:

$$\delta_{obs}(T) = p_A(T) \cdot \delta_A(T) + p_B(T) \cdot \delta_B(T) \tag{10.9}$$

$$\delta_{obs}(T) = \left(\frac{e^{-\Delta G(T)/RT}}{1 + e^{-\Delta G(T)/RT}}\right) \cdot \delta_A(T) + \left(\frac{1}{1 + e^{-\Delta G(T)/RT}}\right) \cdot \delta_B(T) \tag{10.10}$$

Eq. 10.10 embodies the most general two-state model, as it incorporates no assumptions about the temperature dependences of ΔG, δ_A, and δ_B. In order to proceed with our analysis, we now make such assumptions. Given the relationships between heat capacity at constant pressure (C_p), H, and S (Eq. 10.11), calorimetric data may suggest possible treatments of the temperature dependence of ΔG (Eq. 10.12). The pre-transition baselines from Adnectin differential scanning calorimetry (DSC) experiments show a slight positive slope [102], indicating that the excess (relative to aqueous solvent) heat capacity of folded protein gradually increases with temperature [102]. This temperature dependence of C_p is characteristic of folded globular proteins, but the magnitude is small; treating the C_p of folded protein as temperature-invariant is often a reasonable approximation [263]. Here we will compare the result of allowing small, positive values of ΔC_p (between curvature-related

states), implying temperature-dependent ΔH and ΔS (Eq. 10.12), with a simpler model in which $\Delta C_p = 0$, and Eq. 10.13 applies. We also assume that the chemical shifts of states A and B each have a linear temperature dependence (Eqs. 10.14 and 10.15, where m_A and m_B represent the linear temperature coefficients of states A and B, respectively), which is supported by extensive empirical observations (Chapter 9) and consistent with our hypothesis that this linearity is characteristic of fast exchange between conformations within a single well in the free energy landscape.

$$C_p = \frac{dH}{dT} = T\frac{dS}{dT} \tag{10.11}$$

$$\Delta G(T) = \Delta H(T) - T \cdot \Delta S(T) \tag{10.12}$$

$$\Delta G(T) = \Delta H - T \cdot \Delta S \tag{10.13}$$

$$\delta_A(T) = \delta_A(T_{ref}) + m_A \cdot (T - T_{ref}) \tag{10.14}$$

$$\delta_B(T) = \delta_B(T_{ref}) + m_B \cdot (T - T_{ref}) \tag{10.15}$$

Under the assumption that $\Delta C_p \approx 0$ between conformations in the native ensemble, ΔH and ΔS are approximately temperature-independent (Eq. 10.13). Altogether, the $\Delta C_p = 0$ model collectively described by Eqs. 10.10, 10.13, 10.14, and 10.15 has six parameters: ΔH, ΔS, $\delta_A(T_{ref})$, m_A, $\delta_B(T_{ref})$, and m_B; in general, a sigmoidal transition between the linear temperature dependences of states A and B is described (Fig. 10.2).

Fitting experimental data to this model presents two significant challenges: first, non-linear dependence of the model on ΔH and ΔS complicates parameter optimization, i.e., finding a global RMSE (between modelled and experimental chemical shifts) minimum is nontrivial; second, the sigmoidal transition described by the model typically takes place over a wide range of temperatures (Fig. 10.2). The relatively narrow temperature range over which we report chemical shifts (i.e., where the assumptions built into our model hold) may capture little of the sigmoidal curvature. To address the first challenge, we sweep ΔH and ΔS over ranges of fixed values, essentially converting one very difficult optimization problem into many (>45000; Section 10.3) straightforward linear regressions. Along with experimental error (Fig. 10.3), the narrowness of the range of experimental temperatures manifests as uncertainty in the values of the optimum parameters, i.e., various combinations of model parameters result in similar RMSEs. We report the degree of uncertainty as the percentage of all combinations of ΔH and ΔS for which the RMSE of the fit satisfies Eq. 10.16. Empirically, when this percentage is sufficiently small (<0.5%) we observe that fits with RMSE within 5% of the lowest RMSE (Eq. 10.16) describe a single minimum in the RMSE landscape (Fig. 10.4).

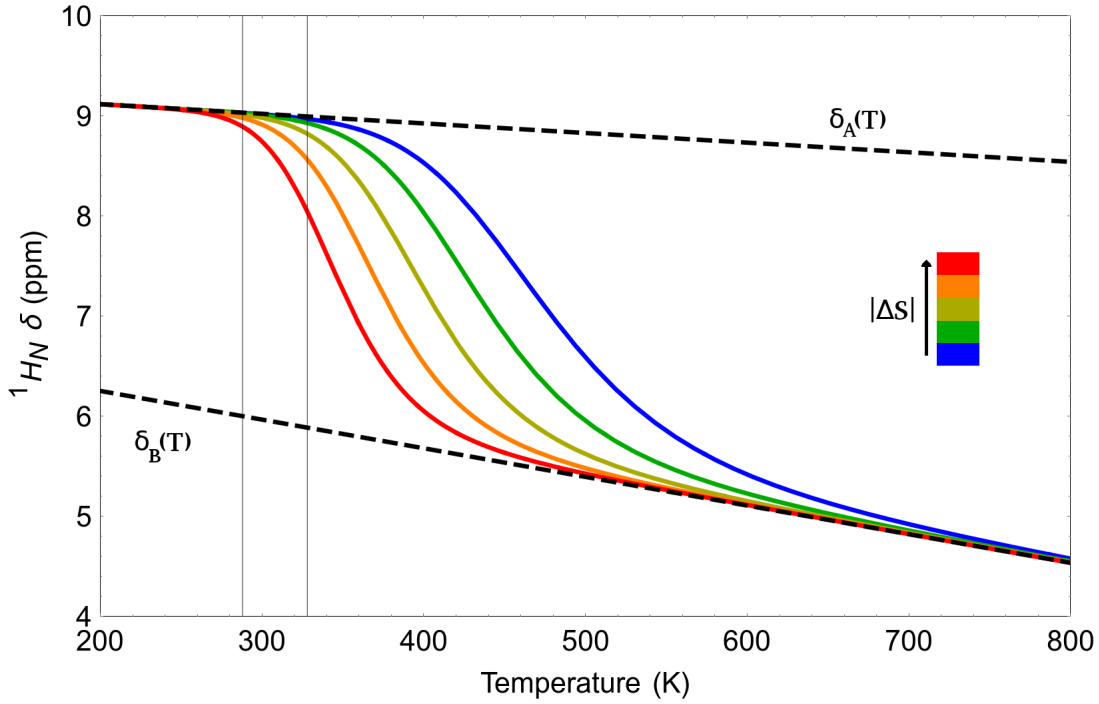


Figure 10.2: Sigmoidal temperature dependence of a hypothetical amide proton chemical shift according to the $\Delta C_p = 0$, two-state, fast-exchange model (Eqs. 10.10, 10.13, 10.14, and 10.15). The curve is plotted over an unrealistic temperature range to show the full breadth of the sigmoidal transition (the temperature range covered in our VT-NMR experiments, 288 to 328 K, falls between the two vertical lines). The effect of varying ΔS from -0.02 to -0.03 kcal·mol^{-1}·K^{-1} with ΔH fixed at -10 kcal·mol^{-1} (physically reasonable values; Appendix I) is illustrated.

$$RMSE < 1.05 \cdot RMSE_{min} \tag{10.16}$$

We also fit data to a model, again based on Eq. 10.10, that allows $\Delta C_p \geq 0$. In this case, formulae for $\Delta H(T)$ and $\Delta S(T)$ can be derived by rearranging Eq. 10.11 and integrating with respect to T (Eqs. 10.17 and 10.18) [264]. This model has one more parameter than the $\Delta C_p = 0$ model, and three parameters that must be swept rather than varying freely during optimization ($\Delta H(T_{ref})$, $\Delta S(T_{ref})$ and ΔC_p), substantially increasing the computational cost of the fitting procedure. In fits of the highest quality data (i.e., selected for unambiguous, smooth curvature; Table 10.2), increasing the number of parameters in order to allow non-zero values of ΔC_p results in only modest improvements over the simpler $\Delta C_p = 0$ model. However, we note that non-zero values of ΔC_p extend the family of curves described by Eq. 10.10 beyond the straightforward sigmoidal transition illustrated in Fig. 10.2. In exceptional cases, use of the more complicated model may be

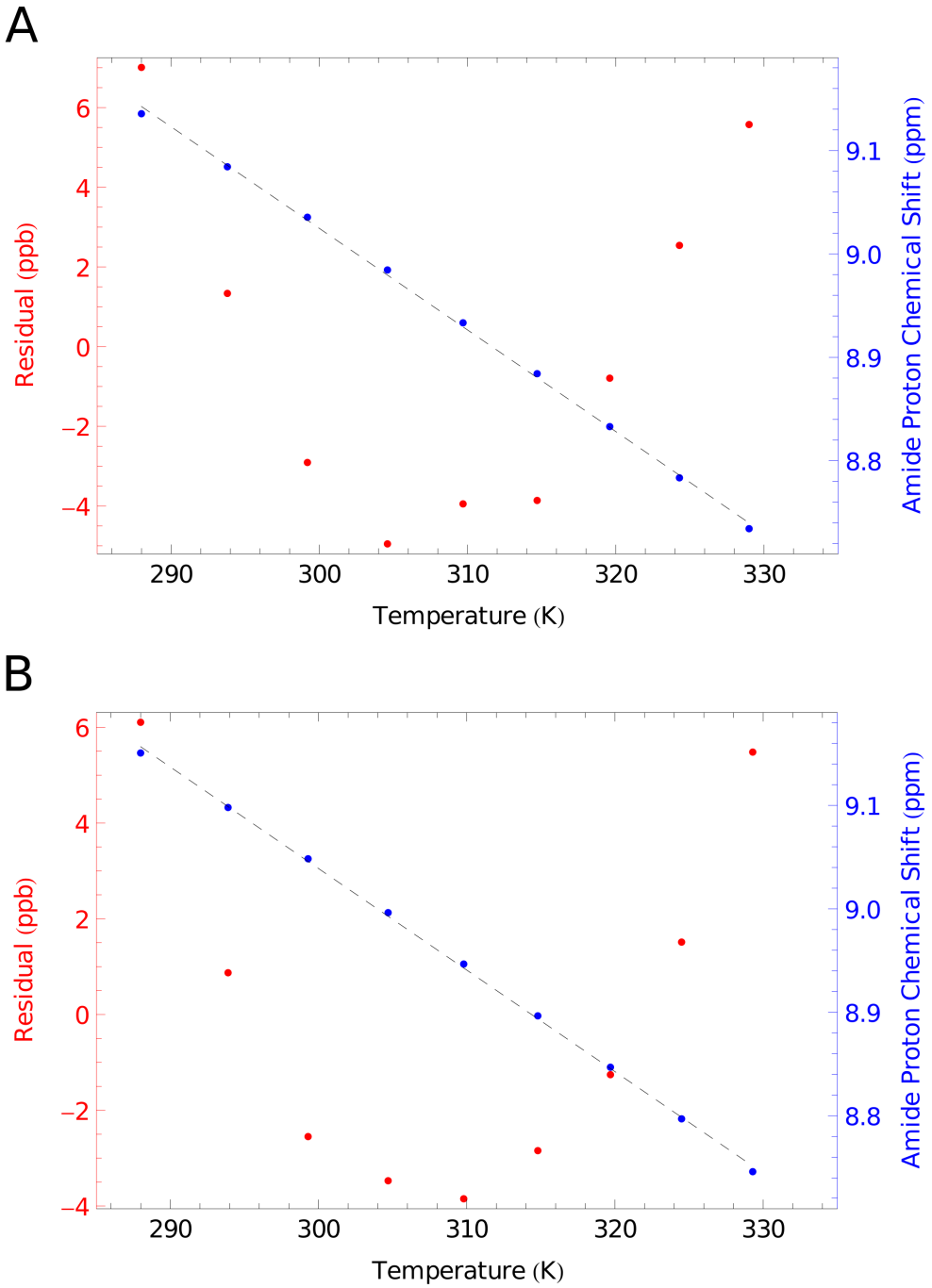


Figure 10.3: Amide proton chemical shifts (blue), linear fit (dashed line), and residuals (red; fit minus chemical shift). A: pWT Adnectin residue G40, and B: V75R residue G40. The V75R data describe a smoother curve; here we illustrate how smoothness of curvature can be disrupted by experimental error. The impact on the RMSE landscape is shown in Fig. 10.4.

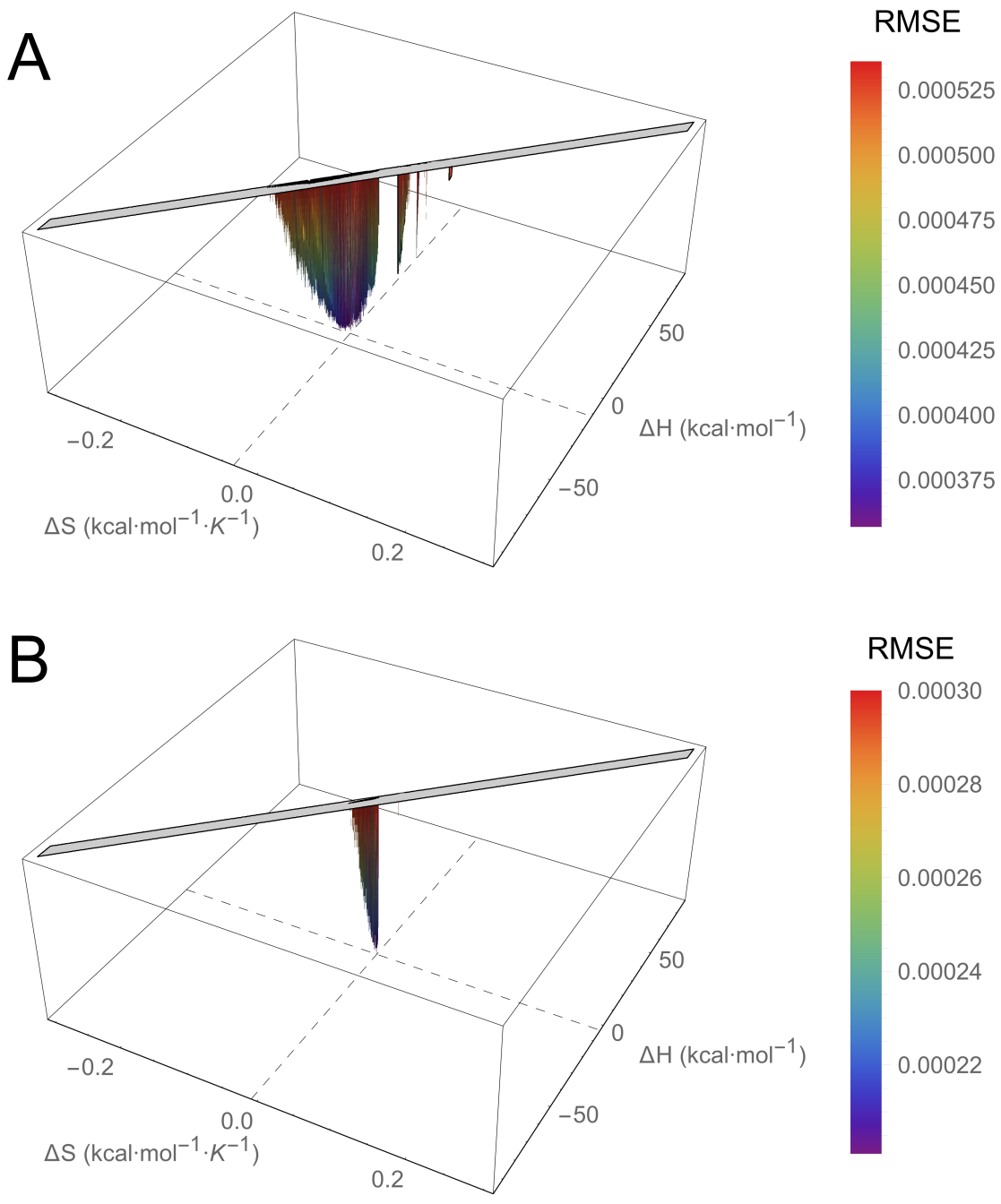


Figure 10.4: RMSE landscapes from fitting curvature data to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray. A: pWT residue G40; B: V75R residue G40. Smooth curvature (Fig. 10.3) narrows the breadth of the minimum in the RMSE landscape.

justified; here, model comparisons based on the Akaike information criterion (AIC) [265] indicate relatively high probabilities that the $\Delta C_p = 0$ model is sufficient (Table 10.2).

$$\Delta H(T) = \Delta H(T_{ref}) + \Delta C_p \cdot (T - T_{ref}) \tag{10.17}$$

$$\Delta S(T) = \Delta S(T_{ref}) + \Delta C_p \cdot ln(T/T_{ref}) \tag{10.18}$$

Table 10.2: Comparison of $\Delta C_p = 0$ and $\Delta C_p \geq 0$ models.

Adnectin	**Res.**	**RMSE$_{\Delta C_p=0}$[1]**	**RMSE$_{\Delta C_p\geq 0}$[1]**	**AIC$_{\Delta C_p=0}$[2]**	**AIC$_{\Delta C_p\geq 0}$[2]**	**Relative Likelihood[3]**
pWT	T28	0.00023	0.00023	-114.9	-115.1	0.9
pWT	S85	0.00042	0.00041	-104.6	-104.7	1.0
V75R	G40	0.00020	0.00019	-117.7	-118.4	0.7
V75R	R75	0.00041	0.00041	-104.8	-104.8	1.0
L18V-Y88F	G62	0.00023	0.00023	-115.2	-115.3	1.0

[1] Minimum RMSE (ppm)
[2] Akaike information criterion [265]
[3] Relative (to the $\Delta C_p \geq 0$ model) probability that the $\Delta C_p = 0$ model minimizes information loss: $e^{(AIC_{\Delta C_p\geq 0} - AIC_{\Delta C_P=0})/2}$

Of the fitted model parameters (Appendix I, Table I.2), we are particularly interested in ΔH and ΔS because they can be used to predict ΔG (or equivalently, the relative occupation of curvature-related states). Though we begin with little idea what conformations these states may entail, armed with ΔG predictions we examine pWT MD simulation trajectories seeking examples of conformational exchange that may measurably change the local magnetic environments of amide protons. In favourable cases, we find that states with relative occupations very close to those predicted can be discerned (Table 10.3).

Specifically, because the partial positive charge of amide protons allows them to participate in favourable electrostatic interactions (including, but not limited to hydrogen bonds), we consider electronegative nitrogen and oxygen atoms with multimodal distributions of distances to an amide proton (at least one mode being within 3.5 Å) candidates to perturb local magnetic environments; where only one candidate is apparent, we report ΔG determined from relative occupations (as illustrated in Fig. 10.5) per Eq. 10.1 (Table 10.3). If the free energy gap between curvature-related states is large, the higher energy state may not measurably impact population-weighted average amide proton chemical shifts; consistent with the theory that these MD-observed states are related to curvature, most of the ΔG values fall in a range between -1.7 and -0.3 kcal/mol. The ΔG calculated for L19 is of greater magnitude, and appears to be associated with the transient loss of an intramolecular hydrogen bond (Fig. 10.6). We hypothesize that transient hydrogen bond disruptions also affect other residues (as in Chapter 6), but the large magnitude of ΔG in such cases may often preclude observation of curvature.

Table 10.3: pWT Adnectin intra-β-sheet hydrogen bond parameters and calculated changes with temperature

				Curvature ΔG^4	
Res. #	**%[1]**	**Candidate Atom(s)[2]**	**MD ΔG^3**	**Min.**	**Max.**
4	0.16	D3 N, D3 $O_{\delta1}/O_{\delta2}$		-2.9	-1.8
7	0.36	R6 N, D7 $O_{\delta1}/O_{\delta2}$		-2.7	-1.6
8	0.18	D7 N	-1.0	-2.2	0.0
19	0.18	A12 O	-3.7	-5.3	-4.4
24	0.12	D23 $O_{\delta1}/O_{\delta2}$	-0.7	-1.8	-1.0
27	0.40	D3 $O_{\delta1}$	-0.6	-3.1	-1.8
28	0.03	T28 O_{γ}	-0.4	-0.6	-0.4
29	0.20	T28 O_{γ}	-0.5	-1.7	-0.7
39	0.30	T39 O_{γ}	-1.7	-2.5	-1.4
40	0.31	G40 O, T39 O_{γ}, N42 $O_{\delta1}$		-2.0	-0.6
74	0.25	S84 O_{γ}	-0.3	-0.4	0.0
83	0.19	T76 O, S81 O_{γ}, S81 O		-0.9	-0.3
85	0.02	S84 O_{γ}	-0.3	-0.7	-0.5
93	0.11	N/C		-0.7	-0.1

[1]Percentage of fits with RMSE within 5% of the minimum
[2]Oxygen/nitrogen atoms with multimodal distributions of distances to the amide proton; N/C: no candidate
[3]Calculated from occupation of states at 298 K using Eq. 10.1 (kcal/mol)
[4]Calculated at 298 K from ΔH and ΔS of fits with RMSE within 5% of the minimum (kcal/mol)

In some cases, an oxygen or nitrogen atom may contribute to curvature in two (or more) residues (e.g., T28/V29 or A74/S85; Table 10.3). If the MD ΔG values are similar, we infer that conformations in which the electronegative atom is near neither of the amide protons are unlikely. In the cases of T28 and V29 (Fig. 10.5), the curvature-related states are distinguished primarily by the T28 $\chi 1$ dihedral angle (though other conformational changes may be correlated with this side chain rotation). The majority state (at 298 K) corresponds to the *gauche*− rotamer, which places the T28 gamma oxygen near its own amide proton. The minority state corresponds to the *trans* rotamer, which places the T28 gamma oxygen near the V29 amide proton.

We are extending curvature analysis beyond the common binary test for conformational exchange, undertaking quantitative thermodynamic analysis of experimental data and exploring the underlying molecular mechanisms in atomistic detail. We find that curvature is more likely to be encountered in less rigidly structured regions, including Adnectin loops implicated in binding. Though few details about IGF1R-bound Adnectin conformations are available, we hypothesize that, in some cases, curvature-related states may have functional significance. In structured regions such as β-sheets, curvature attributable to side chain degrees of freedom may be distinguished from major structural reorganizations by the magnitudes of ΔG estimates from both fitting and simulation.

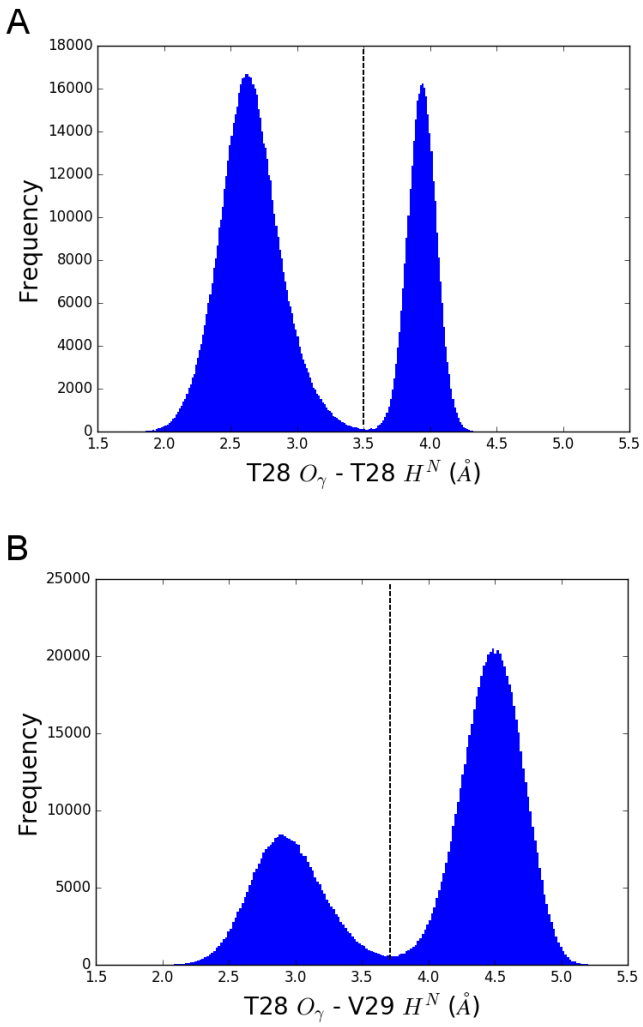


Figure 10.5: Histograms of A: T28 gamma oxygen to T28 amide proton, and B: T28 gamma oxygen to V29 amide proton distances from a 100 ns MD simulation of pWT Adnectin. The vertical dashed line indicates the distance cut-off distinguishing the two states.

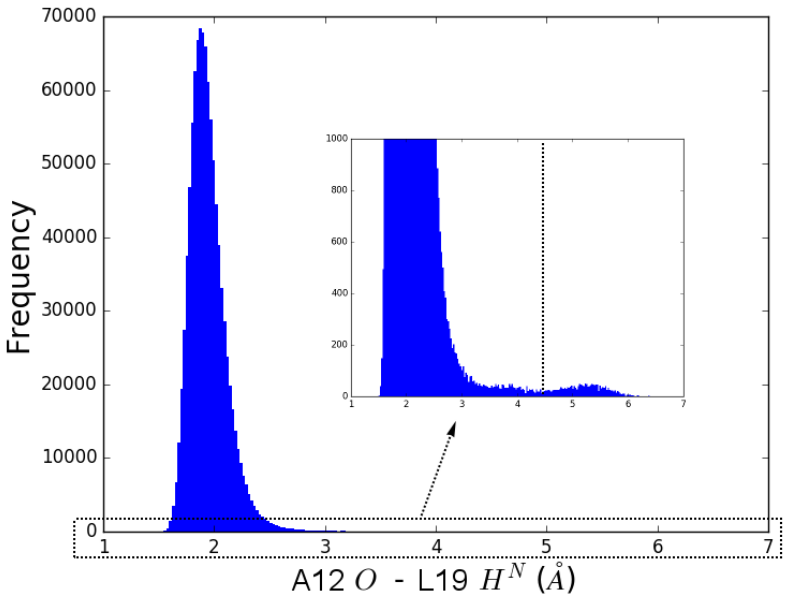


Figure 10.6: Histograms of A12 gamma oxygen to L19 amide proton distances from a 100 ns MD simulation of pWT Adnectin. The inset histogram shows the data on a vertical scale that facilitates perception of the minority state. The vertical dashed line indicates the distance cut-off distinguishing the two states.

Chapter 11

Amide Nitrogen Temperature Coefficients and Dynamic Protein Structure

11.1 Context

Isotropic chemical shifts measured by solution nuclear magnetic resonance (NMR) spectroscopy offer extensive insights into protein structure and dynamics [266]. Temperature dependences add a valuable dimension; notably, the temperature dependences of amide proton chemical shifts have been reported to probe hydrogen bonding, temperature-dependent loss of structure, and exchange between distinct protein conformations (Chapters 9 and 10). Despite extensive study of the determinants of amide nitrogen chemical shifts (Section 11.2.1), their temperature dependences have previously resisted interpretation [247]. We attack the problem on two fronts, comparing experimentally measured temperature coefficients to calculated random coil values, and resolving specific contributions from $\phi/\psi/\chi_1$ dihedral angle and hydrogen bond dynamics through a combination of chemical shift calculations and atomistic molecular dynamics (MD) simulations. We find that a transiently structured region of an RNA-binding intrinsically disordered protein (IDP) [267,268] is distinguished by amide nitrogen temperature coefficients that are substantially more positive, on average, than random coil temperature coefficients (which are invariably negative [269]). Adnectin amide nitrogen temperature coefficients that differ from random coil values are common throughout (including loop regions); although predominantly negative, positive temperature coefficients are not uncommon in β-strand regions. We hypothesize that positive temperature coefficients in β-strands may be rationalized by conformational exchange between distinct states, including, but not necessarily limited to, side chain rotamers.

11.1.1 Acknowledgements

Colleen Doyle assisted with the collection and analysis of NMR data (full details in Chapter 7, Section 7.1). Dr. Rodolfo Rasia provided dsRBD-1 domain (from *Arabidopsis thaliana* DCL1) NMR data, as well as the Python code used to generate Figs. 11.3 and 11.4.

11.2 Introduction

Empirically, as is the case for amide protons (Chapter 9), the temperature dependences of amide nitrogen chemical shifts are often well-approximated by a linear model [247]. In contrast to Adnectin amide proton temperature coefficients, which are almost uniformly negative (i.e., shielding of the nucleus increases with temperature) and distributed over a narrow range (Chapter 9, Fig. 9.1), we report amide nitrogen temperature coefficients that vary from -45 to 37 ppb/K (Appendix J, Table J.1). Increases in shielding (negative amide nitrogen temperature coefficients) are observed with greater frequency, but deshielding of similar magnitude is not uncommon.

Our approach to the previously unsolved puzzle of amide nitrogen temperature coefficients takes the form of two complementary lines of investigation. First, we develop a method for calculating random coil temperature coefficients using published model peptide data [269]. We demonstrate that differences between these calculated values and experimentally measured amide nitrogen temperature coefficients are extremely sensitive to deviations from random coil and may therefore serve as a measure of structure. Furthermore, referencing to random coil confers a degree of independence from amino acid sequence and greatly simplifies interpretation (relative to raw amide nitrogen temperature coefficients). Second, we combine the results of *ab initio* density functional theory (DFT) chemical shift calculations (for formamide dimer hydrogen bond models and short peptides with various fixed backbone and side chain dihedral angles) with observations of dihedral dynamics in classical MD simulations in order to elucidate the determinants of experimentally measured amide nitrogen temperature coefficients in highly structured (β-sheet) Adnectin regions. In this introductory section, we first consider some important influences on amide nitrogen chemical shifts (Section 11.2.1), then go on to discuss the referencing of chemical shifts and temperature coefficients to random coil values in order to facilitate comparisons on an approximately amino acid sequence-independent basis (Section 11.2.2).

11.2.1 Amide Nitrogen Chemical Shifts

The proteinogenic amino acids share (at a minimum) the same succession of 'backbone' atoms, but within a polypeptide the chemical shifts of the atoms in a particular amino acid may be strongly influenced by their local environment, including the characteristics of the attached and sequentially adjacent side chains [270]. Amide nitrogen chemical shifts

are known to be influenced by diverse factors, some of which contribute to amino acid sequence dependence (Table 11.1).

Table 11.1: Factors influencing nitrogen chemical shifts

Factor	**Description**
Inductive Effects	Covalently mediated, sequence-dependent electron withdrawal/donation
Backbone Dihedrals	Particularly ϕ_N and ψ_{N-1}
Side Chain Dihedrals	Particularly $\chi_{1,N}$
Conformational Averaging	Sufficiently fast exchange between states results in observation of population-weighted average chemical shifts
Primary Hydrogen Bonding	Deshielding by hydrogen bonds in which the N^{th} amide nitrogen acts as the donor
Secondary Hydrogen Bonding	Deshielding by hydrogen bonds in which the N-1^{th} carbonyl oxygen acts as the acceptor

Correlations between amide nitrogen chemical shifts and those of the C_α/C_β carbons from the preceding (N-1^{th}) residue have been reported [271, 272]. As the C_α chemical shift is believed to be strongly influenced by the electron withdrawing or donating nature of the side chain to which it is bonded, it has been proposed that amide nitrogen chemical shifts may be subject to similar covalently mediated effects (i.e., attributable to the N-1^{th} side chain) [271]. Furthermore, amide nitrogen chemical shifts are also sensitive to ψ_{N-1} and ϕ backbone dihedral angles [273], which change the positions of the N-1^{th} and N^{th} side chains, respectively, relative to the N^{th} amide nitrogen; thus, through-space effects may also contribute to sequence dependence.

Side chain dihedral angles may also play a role; in particular, the χ_1 angle is often important [273, 274], as some values position side chain γ atoms very near the amide nitrogen. Many side chains in proteins exhibit preferences for one of three non-eclipsed χ_1 angles ($\simeq -60/60/180°$). We find that in some cases these preferences are relatively weak (i.e., more than one rotamer may be sampled), thus experimentally measured chemical shifts may reflect conformational averaging of different rotamers.

As is also the case for amide protons (Chapter 9), hydrogen bonding deshields amide nitrogen atoms. Hydrogen bonds in which the N^{th} amide nitrogen acts as the donor (primary) or, to a lesser degree, in which the N-1^{th} carbonyl oxygen acts as the acceptor (secondary) may each have an impact [274].

11.2.2 Random Coil Chemical Shifts and Temperature Coefficients

The dependence of amide nitrogen chemical shifts on amino acid sequence complicates comparisons between the atoms of different residues. Random coil chemical shifts (measured from peptide models expected to exhibit random coil-like behaviour in solution [269, 270, 275–277]) similarly depend on amino acid sequence; these can be compared to protein chemical shifts to facilitate inferences regarding degree and/or type of structure. Crucially, the difference between a measured chemical shift (δ_{Obs}) and that expected for a random coil having the same amino acid sequence (δ_{RC}), known as the secondary chemical shift (δ_{Sec}; Eq. 11.1), is approximately independent of sequence.

$$\delta_{Sec} = \delta_{Obs} - \delta_{RC} \tag{11.1}$$

Structural considerations dominate secondary chemical shifts; they form the basis for chemical shift-based prediction of protein structure as well as the inverse problem of predicting chemical shifts from known structures. Here, by analogy with secondary chemical shifts, we define secondary temperature coefficients as the difference between observed and sequence-dependent random coil values (Eq. 11.2). Like secondary chemical shifts, secondary temperature coefficients are approximately independent of amino acid sequence.

$$\left(\frac{\Delta\delta}{\Delta T}\right)_{Sec} = \left(\frac{\Delta\delta}{\Delta T}\right)_{Obs} - \left(\frac{\Delta\delta}{\Delta T}\right)_{RC} \tag{11.2}$$

11.2.3 Quantum Chemical Calculations and Molecular Dynamics

The interpretation of amide nitrogen chemical shifts and temperature dependences thereof is challenging. Table 11.1 is not an exhaustive list of influences, and the relative importance of different factors is not well understood. Quantum calculations of chemical shifts may offer some insight, but they are frequently limited (by practical considerations such as finite computing power and time) to the analysis of fragments from experimentally determined protein structures [278–281] or short peptides with 'protein-like' dihedral angles [282, 283].

Here we systematically vary each of several model peptide degrees of freedom independently, including ψ_{N-1}, ϕ_N, ψ_N, $\chi_{1,N}$, and hydrogen bond parameters, in order to gauge their impacts on amide nitrogen chemical shifts. The chemical shifts of atoms within weakly structured protein regions may be affected by exchange between conformations incorporating diverse combinations of these parameters; therefore, this approach generates clearer insights when applied to stable secondary structure. Nevertheless, even within secondary structure we anticipate that protein dynamics may play an important role; accordingly, onto plots of amide nitrogen chemical shift vs. model peptide degrees of freedom we project histograms of dihedral angles observed via MD simulation at 298 K. Extrapolating from these plots, we hypothesize that conformational exchange (e.g., between side

chain rotamers) may contribute to the more positive temperature coefficients (on average) observed for Adnectin residues within β-strands.

11.3 Methods

11.3.1 Expression & Purification

Expression and purification of ^{15}N-labelled protein, variable-temperature NMR experiments, and data analysis (including determination of temperature coefficients) were conducted as described in Chapter 7, Section 7.3.

11.3.2 Random Coil Temperature Coefficients

Adopting assumptions described in Section 11.4, sequence-specific random coil amide nitrogen temperature coefficients were calculated using Eq. 11.5 and the data in Table 11.2.

11.3.3 Quantum Chemical Calculations

All quantum chemical calculations were carried out via the GAUSSIAN 16 program [238] using DFT at the B3LYP/6-311++G(2d,2p) level. Implicit solvation of peptides by water (ε=78.39) was accomplished using the polarizable continuum model (PCM). The FragBuilder Python library [284] was used to generate XG, and GX, and GXG dipeptide and tripeptide structures (where X is any of the standard proteinogenic amino acids except proline) with acetyl and N-methyl amide caps on the N termini and C termini, respectively. The structures underwent molecular mechanics optimization using the MMFF94 force field [285] to resolve potential steric clashes. Following quantum geometry optimization, the isotropic magnetic shielding constants of the amide nitrogens were calculated using the gauge-independent atomic orbital (GIAO) method [240]. These were then subtracted from a similarly calculated NH_3 nitrogen isotropic magnetic shielding constant (a gas phase calculation adjusted by -22.6 ppm to compensate for the liquid-association shift [286]) to give amide nitrogen chemical shifts.

The ψ_{N-1} and ϕ_N angles play a role in determining the positions of the N-1$^{\text{th}}$ and N$^{\text{th}}$ side chains relative to the N$^{\text{th}}$ amide nitrogen, and their impacts on amide nitrogen chemical shifts may therefore be inextricably sequence-dependent. Here we conduct DFT calculations in which we vary the ψ angles of the first residues in capped XG dipeptides (where X is any of the standard proteinogenic amino acids except proline) in 15° increments over typical β-sheet values ranging from -180° to -135°and 45° to 180°, and (separately) vary the ϕ angles of the second residues in capped GX dipeptides in 15° increments over typical β-sheet values ranging from -180° to -45° (Fig. 11.6; Appendix J, Section J.2). In

each case, we track the amide nitrogen chemical shift of the second residue in the dipeptide. When X is not glycine or alanine this variation of ψ and ϕ angles is repeated with the χ_1 angle of X set to -60°, 60°, and 180°; where applicable, χ angles aside from χ_1 are set to their single most probable value per the Dunbrack rotamer library [287]. For higher resolution insight into the influence of side chain dihedrals, we also investigate the effect of varying the χ_1 angle of the N^{th} residue in smaller increments (20° from -180 to 180°) on the chemical shift of the N^{th} amide nitrogen (Fig. 11.8). In β-structure, the $N+1^{th}$ side chain is relatively far from the N^{th} amide nitrogen; therefore, the impact of the ψ_N backbone dihedral is modelled in a sequence-independent fashion using a capped GG dipeptide (Fig. 11.7).

11.3.4 Molecular Dynamics

Molecular dynamics simulations were performed as described in Chapter 8, Section 8.3.2.

11.4 Results & Discussion

Experimentally measured amide nitrogen temperature coefficients (hereafter, simply 'temperature coefficients') for pWT, Parent, L78I, V75R, and L18V/Y88F Adnectins are visualized in Fig. 11.1 (numerical data can be found in Appendix J, Table J.1). The L78I, V75R and L18V/Y88F Adnectins are named for their relationships to the Parent amino acid sequence, from which they differ by one or two point mutations. The pWT sequence differs from that of the Parent in three loops, one of which is longer by four residues. Specifically, the pWT BC loop (residues 23-29), DE loop (residues 52-55), and FG loop (residues 77-86) differ from equivalent loops in the other Adnectins (Fig. 2.2, Appendix D).

The determinants of amide nitrogen chemical shifts have been studied extensively (Section 11.2). How amide nitrogen chemical shifts change with temperature is presumably attributable to changes in these same factors, but their relative importance (i.e., the magnitudes of their individual temperature dependences) is not well established. We see in Fig. 11.1 that negative temperature coefficients are somewhat more common than positive, and residues at equivalent positions in different Adnectins generally follow similar temperature coefficient trends. However, pWT Adnectin temperature coefficients differ noticeably in several instances, and sometimes even have the opposite sign, e.g., I34 and T49, residues centrally located within adjacent β-strands. The pWT amino acid sequence within these strands is identical to those of the other Adnectins, therefore the I34 and T49 temperature coefficients report on inter-Adnectin differences in structure and dynamics. We also observe sequence-dependent differences in temperature coefficients, e.g. for V18 and R89 in L18V-Y88F.

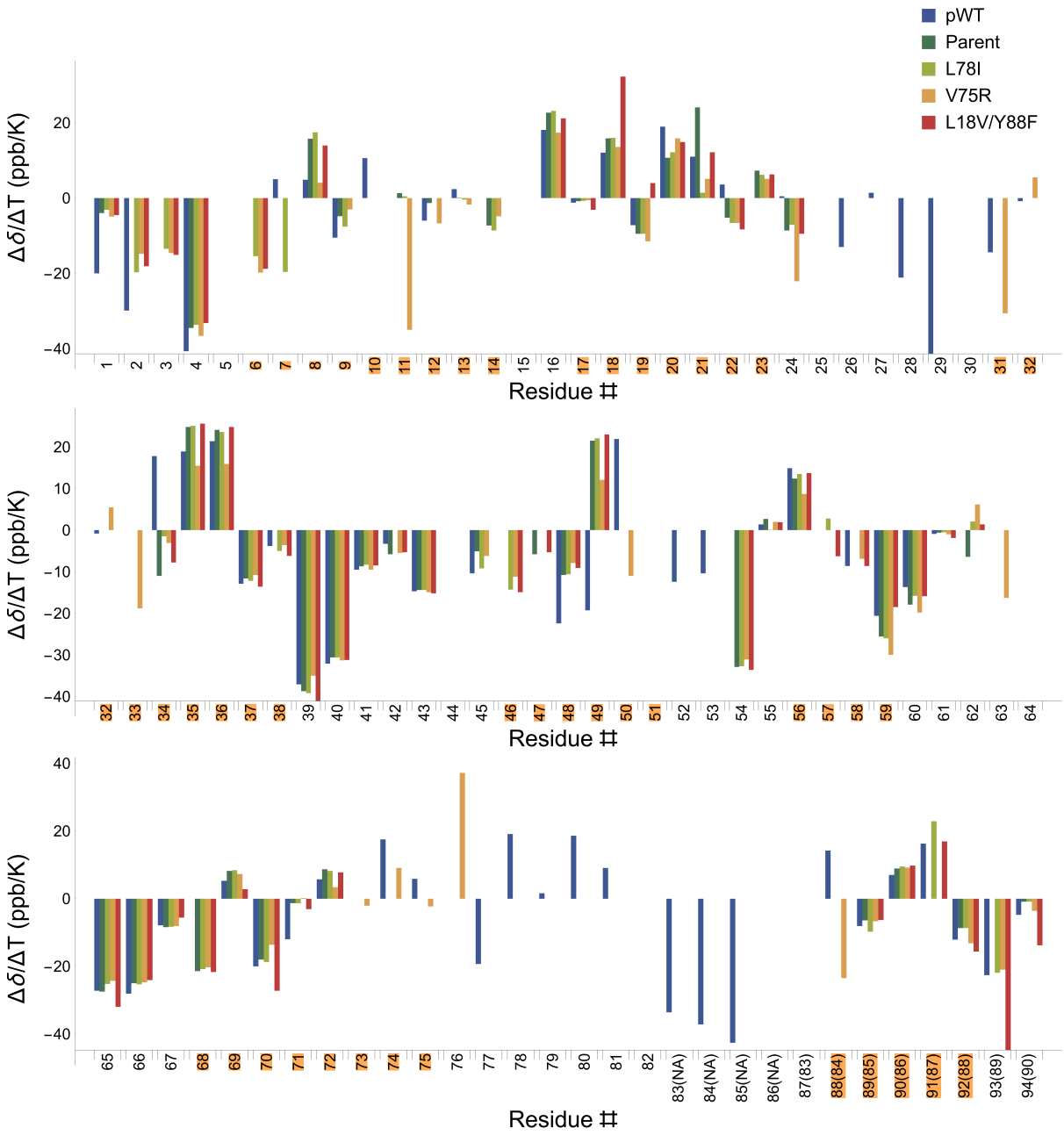


Figure 11.1: Adnectin amide nitrogen linear temperature coefficients. Residue numbers shown as pWT #(other Adnectins #) where applicable. Residue numbers highlighted in orange are in the β-strands. Bars for residues missing assignments or for which the temperature dependence could not be tracked (distinguishable in Appendix J, Table J.1) are omitted.

Temperature coefficients in regions of identical sequence are directly comparable, but without better understanding their determinants we cannot assign meaning to differences between them. Outside of these regions interpretation is even more difficult, as amide nitrogen temperature coefficients have a strong dependence on amino acid sequence [269]. Here we develop a method of calculating secondary temperature coefficients using published random coil model peptide data [269]. Secondary temperature coefficients are approximately independent of sequence and interpretable as a measure of how much the temperature-dependence of the local magnetic environment of an amide nitrogen nucleus differs from that expected in the absence of structure.

11.4.1 Secondary Temperature Coefficients

It has been demonstrated that not only amide nitrogen random coil chemical shifts themselves, but also their rates of change with temperature strongly depend on local amino acid sequence [269]. Accordingly, we propose a method for the calculation of random coil amide nitrogen temperature coefficients that accounts for the identity of the preceding residue and the class (glycine vs. non-glycine) of the following residue. Although we proceed directly from random coil model peptide temperature coefficients, the procedure is logically equivalent to calculating temperature-corrected random coil chemical shifts and determining the slope of the line ($\Delta\delta/\Delta T$) connecting them.

We note that several existing methods for calculating temperature-corrected random coil amide nitrogen chemical shifts implicitly (and incorrectly) assume that the temperature dependences of amide nitrogen chemical shifts are sequence-independent or negligible [277, 288, 289]; in accordance with [269], we avoid this pitfall. Our method facilitates calculation of temperature-corrected secondary chemical shifts, and therefore has wide-ranging applications including chemical shift-based prediction of protein structure, the inverse problem of chemical shift prediction from known structures, and the analysis of IDP chemical shifts; here we interpret experimentally measured temperature coefficients by comparison to calculated random coil values.

The number of random coil peptides that must be characterized in order to determine how interactions between sequentially adjacent side chains might contribute to observed chemical shifts (and temperature coefficients) increases exponentially with the number of amino acids considered. Fortunately, it has been shown that the identities of the attached (N^{th}) and previous ($N\text{-}1^{th}$) side chains have the greatest influences on the observed amide nitrogen random coil chemical shift [277, 290], which may justify the simplification allowed by consideration of just two amino acids at a time (as in [269, 275, 290]). We also continue the common practice of treating the contributions of each side chain as approximately independent [269, 275, 277, 290].

From the data in Table 11.2, there are two subtly different ways to model the contributions of the N^{th} and $N\text{-}1^{th}$ side chains to amide nitrogen temperature coefficients. We can take a value from the X3 column (i.e., the effect of the N^{th} side chain) as our starting point;

Table 11.2: Random coil amide ^{15}N temperature coefficients.

	X3 (ppb/K)[1,2]	**A4 (ppb/K)[2]**	**GX (ppb/K)[1,3]**	**XA (ppb/K)[1,4]**
Ala (A)	-9.0	-15.6	-6.8	-17.8
Cys (C)	-12.2	-14	-10.0	-16.2
Asp (D)	-6.8	-14.7	-4.6	-16.9
Glu (E)	-8.1	-16.2	-5.9	-18.4
Phe (F)	-10.5	-15.3	-8.3	-17.5
Gly (G)	-7.6	-6.8	-5.4	-9.0
His (H)	-6.7	-12.1	-4.5	-14.3
Ile (I)	-17.1	-26.5	-14.9	-28.7
Lys (K)	-11.5	-17.2	-9.3	-19.4
Leu (L)	-5.1	-17.7	-2.9	-19.9
Met (M)	-8.3	-17.3	-6.1	-19.5
Asn (N)	-8.9	-11.7	-6.7	-13.9
Pro (P)		-22.2		
Gln (Q)	-7.2	-18.3	-5	-20.5
Arg (R)	-12.2	-16.5	-10	-18.7
Ser (S)	-10.1	-10.9	-7.9	-13.1
Thr (T)	-15.4	-17.3	-13.2	-19.5
Val (V)	-19.0	-25.5	-16.8	-27.7
Trp (W)	-9.6	-21.9	-7.4	-24.1
Tyr (Y)	-12.4	-22.0	-10.2	-24.2

[1] X represents any amino acid (determined by row)
[2] Measured from GGXAGG peptides; data reproduced from [269]
[3] Calculated using Eq. 11.4 for comparison with X3 column
[4] Calculated using Eq. 11.3 for comparison with A4 column

X3 is preceded by a glycine in all of the experimentally characterized peptide sequences, but we can model the effect of the N-1$^{\text{th}}$ side chain as the difference between two rows from the A4 column (Eq. 11.3, where $AA_{\#}$ is the amino acid type of the specified residue and $\frac{\Delta\delta_{\text{Col}}}{\Delta T}[Row]$ is a value selected from Table 11.2). Though more direct, this approach produces the same result as calculating sequence and temperature corrected amide nitrogen chemical shifts as described in [269] at more than one temperature, then determining the slope of the line joining them.

$$\frac{\Delta\delta_{\text{RC}}}{\Delta T} = \frac{\Delta\delta_{\text{X3}}}{\Delta T}[AA_N] + \left(\frac{\Delta\delta_{\text{A4}}}{\Delta T}[AA_{N-1}] - \frac{\Delta\delta_{\text{A4}}}{\Delta T}[Gly]\right) \tag{11.3}$$

Alternatively, we can begin with a value from the A4 column (i.e., the effect of the N-1$^{\text{th}}$ side chain), then model the effect of the N$^{\text{th}}$ side chain as the difference between two rows from the X3 column (Eq. 11.4).

$$\frac{\Delta\delta_{\text{RC}}}{\Delta T} = \frac{\Delta\delta_{\text{A4}}}{\Delta T}[AA_{N-1}] + \left(\frac{\Delta\delta_{\text{X3}}}{\Delta T}[AA_N] - \frac{\Delta\delta_{\text{X3}}}{\Delta T}[Ala]\right) \tag{11.4}$$

Ideally, use of either Eq. 11.3 or Eq. 11.4 would give the same answer (for a given amino acid sequence), but we find that there are small discrepancies. Temperature coefficients for the A (alanine) in XA (where X is any amino acid except proline) can be calculated using Eq. 11.3 and directly compared to the A4 column of Table 11.2. Similarly, temperature coefficients for X in GX can be calculated using Eq. 11.4 for comparison to column X3 of Table 11.2. The temperature coefficients calculated using Eq. 11.3 are systematically more negative than the measured values (A4), while the temperature coefficients calculated using Eq. 11.4 are systematically less negative than the measured (X3) values, revealing modest errors (approximately 2 ppb/K) introduced by our imperfect assumptions.

The errors described above can be traced to a difference (-9.0 ppb/K vs. -6.8 ppb/K) in the temperature coefficients of A3 in the GGAAGG peptide and A4 in the GGGAGG peptide (AA_{N-1}=G and AA_N=A in both cases). We hypothesize that the less negative temperature coefficient in the latter case may be attributable to the presence of a glycine in the N+1$^{\text{th}}$ position, and that the temperature coefficients of residues with non-glycine amino acids in the N+1$^{\text{th}}$ position may be better approximated by Eq. 11.3. Thus, we further attempt to limit discrepancies between calculated and experimental values by applying Eq. 11.4 when the N+1$^{\text{th}}$ amino acid is a glycine, and Eq. 11.3 otherwise, giving the composite Eq. 11.5 (usage illustrated in Fig 11.2).

$$\frac{\Delta\delta_{\text{RC}}}{\Delta T} = \begin{cases} \frac{\Delta\delta_{\text{A4}}}{\Delta T}[AA_{N-1}] + \left(\frac{\Delta\delta_{\text{X3}}}{\Delta T}[AA_N] - \frac{\Delta\delta_{\text{X3}}}{\Delta T}[Ala]\right), \text{if } AA_{N+1} = Gly \\ \frac{\Delta\delta_{\text{X3}}}{\Delta T}[AA_N] + \left(\frac{\Delta\delta_{\text{A4}}}{\Delta T}[AA_{N-1}] - \frac{\Delta\delta_{\text{A4}}}{\Delta T}[Gly]\right), \text{otherwise} \end{cases} \tag{11.5}$$

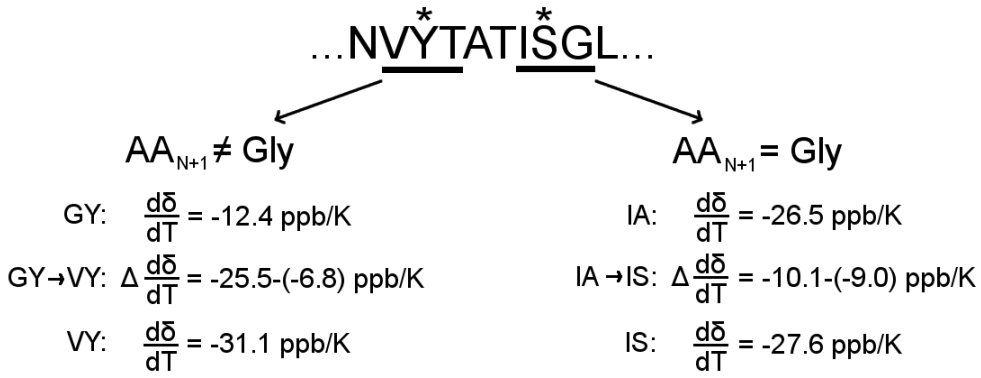


Figure 11.2: Example calculations of random coil amide nitrogen temperature coefficients (Eq. 11.5) for the residues identified by an asterisk. Values for experimentally measured GX and XA temperature coefficients are taken from columns X3 and A4, respectively, of Table 11.2. The random coil temperature coefficient difference expected upon substitution of a single amino acid side chain is modelled as $\frac{\Delta\delta_{\mathrm{A4}}}{\Delta T}[AA_{N-1}] - \frac{\Delta\delta_{\mathrm{A4}}}{\Delta T}[Gly]$ (left, N-1$^{\mathrm{th}}$ side chain substitution) or $\frac{\Delta\delta_{\mathrm{X3}}}{\Delta T}[AA_{N}] - \frac{\Delta\delta_{\mathrm{X3}}}{\Delta T}[Ala]$ (right, N$^{\mathrm{th}}$ side chain substitution).

We validate our method by calculating secondary amide nitrogen temperature coefficients for the intrinsically disordered dsRBD-1 domain from *Arabidopsis thaliana* DCL1. DCL1 is a ribonuclease, and the dsRBD-1 domain folds upon binding RNA; though generally disordered in the absence of RNA, chemical shift changes upon titration with urea suggest that the domain may transiently sample folded conformations that are destabilized by the denaturant [267, 268]. In Fig. 11.3 we compare these chemical shift changes with secondary temperature coefficients calculated using Eq. 11.5. We find that over most of the domain the secondary temperature coefficients are near zero (consistent with random coil-like chemical shift temperature dependences); however, in the H2 helix, where chemical shift changes upon titration with urea are most prominent, large positive secondary temperature coefficients support the hypothesis that this region is transiently structured in the absence of denaturant.

Strictly, negative secondary temperature coefficients also suggest deviations from random coil; however, as we see for both the dsRBD-1 domain and pWT Adnectin (Fig. 11.4), positive secondary temperature coefficients are far more common in regions of secondary structure. In contrast to the dsRBD-1 domain, pWT Adnectin secondary temperature coefficients suggestive of behaviour substantially unlike random coil are common throughout the protein; this confirms our intuition that, within well-structured proteins, even loops may be distinguishable from random coil.

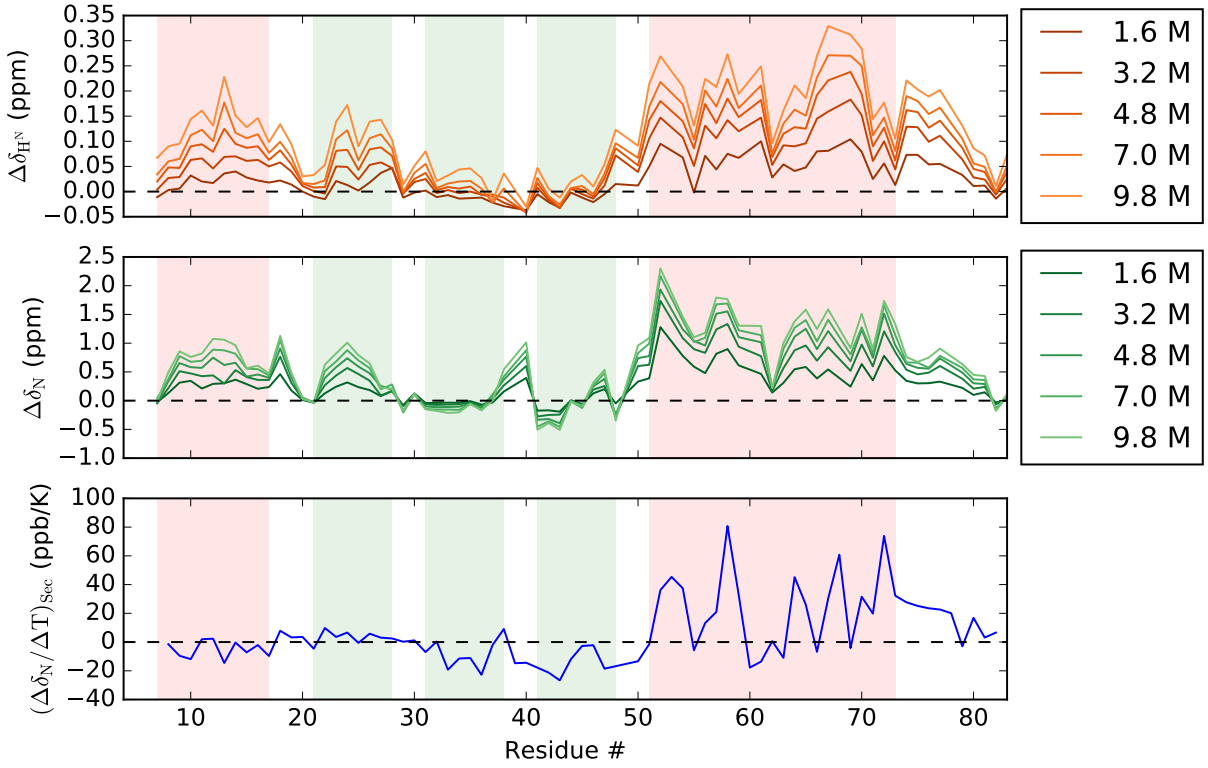


Figure 11.3: Chemical shift perturbations in the intrinsically disordered dsRBD-1 domain from DCL1 due to changes in urea concentration or temperature (data from [268]). Amide proton (top) and amide nitrogen (middle) chemical shift changes due to titration with urea. Secondary amide nitrogen temperature coefficients (bottom) are near-zero (random coil-like) towards the N-terminus, but large positive values are apparent in the H2 helix (residues 51-73). Structure induced in RNA-bound conformations is indicated: α-helix (light red); β-strand (light blue).

11.4.2 Amide Nitrogen Chemical Shift Dependence on Dihedral Angles and Hydrogen Bonding

In the previous section, we propose secondary temperature coefficients as a measure of similarity to random coil. Although of practical value (e.g., for the identification of transiently structured regions within IDPs), this approach generates little insight into the specifics of Adnectin structure and dynamics; in particular, establishing a structural basis for the interpretation of experimentally measured amide nitrogen temperature coefficients requires detailed knowledge of the relative contributions of various chemical shift determinants. In this section, we report that the simulated dynamics of many residues in Adnectin β-sheet secondary structure are restricted enough that direct analysis, via DFT calculations of chemical shifts in model peptides with backbone dihedral angles (Fig. 11.5) typical of

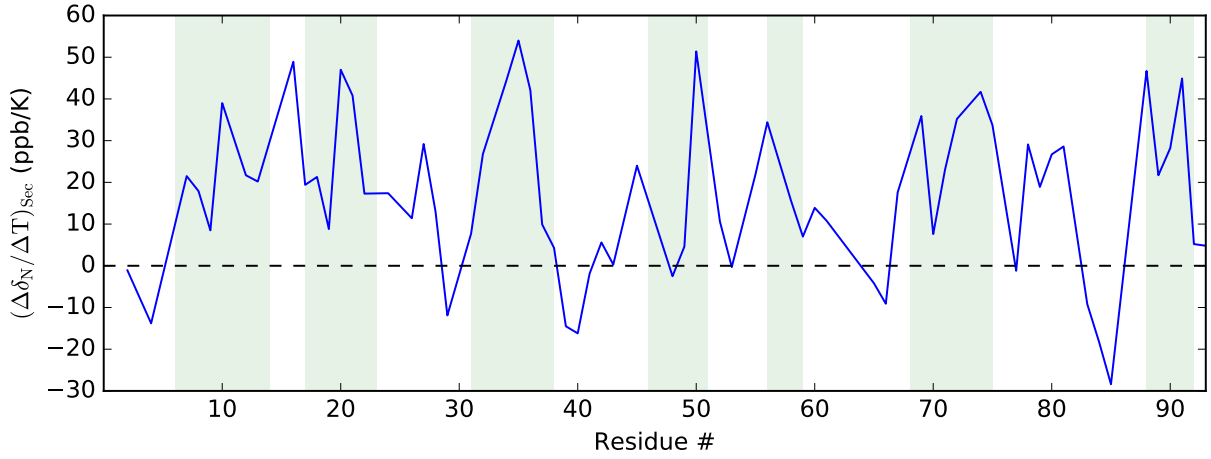


Figure 11.4: Secondary amide nitrogen temperature coefficients for pWT Adnectin. Temperature dependences substantially unlike random coil are apparent throughout the protein (including loop regions), but especially in the β-strands (light blue background).

β-structure, is feasible. We calculate the impacts on amide nitrogen chemical shifts of independently varying several backbone/side chain dihedral angles and hydrogen bond parameters, which, combined with the dynamics observed in simulation, provides the context within which we interpret our experimental data. Due to the approximations inherent in such an approach, we restrict ourselves to pWT Adnectin simulations (for which experimentally determined structures are available as a starting point) and interpret the results with caution; nevertheless, we find potentially explanatory points of contrast between computational results for amide nitrogens with positive and negative experimentally measured temperature coefficients.

Residues in flexible loops may sample broad ranges of backbone dihedral angles, producing distributions consistent with a complex free energy landscape. On the other hand, in MD simulation at 298 K we find that pWT Adnectin residues in β-sheet secondary structure often sample backbone dihedrals with distributions that are relatively narrow and approximately symmetric (Fig. 11.6). Because the ψ_{N-1} and ϕ_N backbone dihedrals help determine the position of the N-1$^{\text{th}}$ and N$^{\text{th}}$ side chains relative to the N$^{\text{th}}$ amide nitrogen (Fig. 11.5), we calculate the effect of varying these angles using XG and GX dipeptides where the identity of residue X is determined by pWT Adnectin sequence, and the chemical shift of the second residue is monitored. In contrast, we treat the effect of varying ψ_N (which may affect the relative position of the more distant N+1$^{\text{th}}$ side chain) as approximately independent of sequence, i.e., due mainly to the positions of backbone atoms relative to the N$^{\text{th}}$ amide nitrogen (Fig. 11.7).

Side chain χ_1 angles are generally constrained by steric considerations to regions near 180°, 60°, and -60° [292]; in our sequence-dependent analysis of ψ_{N-1} and ϕ_N variation (in

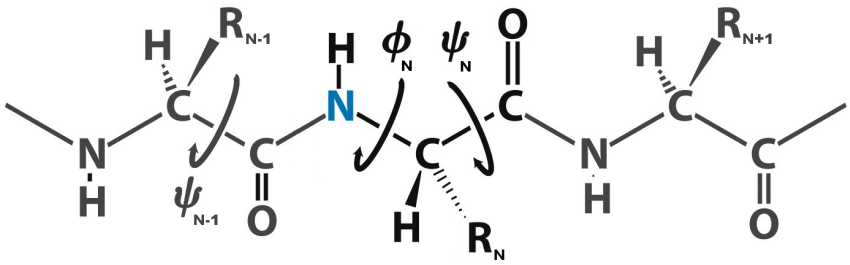


Figure 11.5: Protein backbone dihedral angles ψ_{N-1} and ϕ_N affect the positions of the N-1$^{\text{th}}$ and N$^{\text{th}}$ side chains relative to the N$^{\text{th}}$ amide nitrogen (blue). The N+1$^{\text{th}}$ side chain is relatively distant (in extended β-structure), therefore we account only for changes in the positions of backbone atoms (relative to the N$^{\text{th}}$ amide nitrogen) in our treatment of ψ_N variation. Figure adapted from [291].

model peptides), we fix each of $\chi_{1,N-1}$ and $\chi_{1,N}$ to one of these three angles (whichever is closest to the corresponding χ_1 angle in the 1FNF crystal structure). The $\chi_{1,N}$ dihedral angle, which determines the positions of side chain γ atoms relative to the N$^{\text{th}}$ amide nitrogen, is particularly likely affect the chemical shift [274]; therefore, we also examine this degree of freedom independently and at higher resolution (20° increments from -180 to 180°; Fig. 11.8). We extrapolate from these results in an effort to decipher experimentally measured amide nitrogen temperature coefficients.

Random coil amide nitrogen temperature coefficients, both measured (Table 11.2) and calculated (Eq. 11.5) are invariably negative, indicating an increase in the shielding of the nitrogen atom with rising temperature. In contrast, the temperature coefficients of structured proteins vary over a wide range that includes positive values (Fig. 11.1; Appendix J, Table J.1), though negative values remain more common. On average, the temperature coefficients of residues involved in stable secondary structure are more positive than those that are not (Fig. 11.1), but on a per-residue basis there are many exceptions to this trend. Of particular interest are those residues with a positive temperature coefficient, indicating an overall decrease in shielding of the amide nitrogen atom as temperature increases; given the prevalence of negative temperature coefficients, we hypothesize that protein degrees of freedom that affect the amide nitrogen chemical shift often make negative contributions to the temperature dependence. Where positive contributions are identified, we may infer that these same factors play a role in moderating the negativity (relative to random coil) of Adnectin temperature coefficients in general.

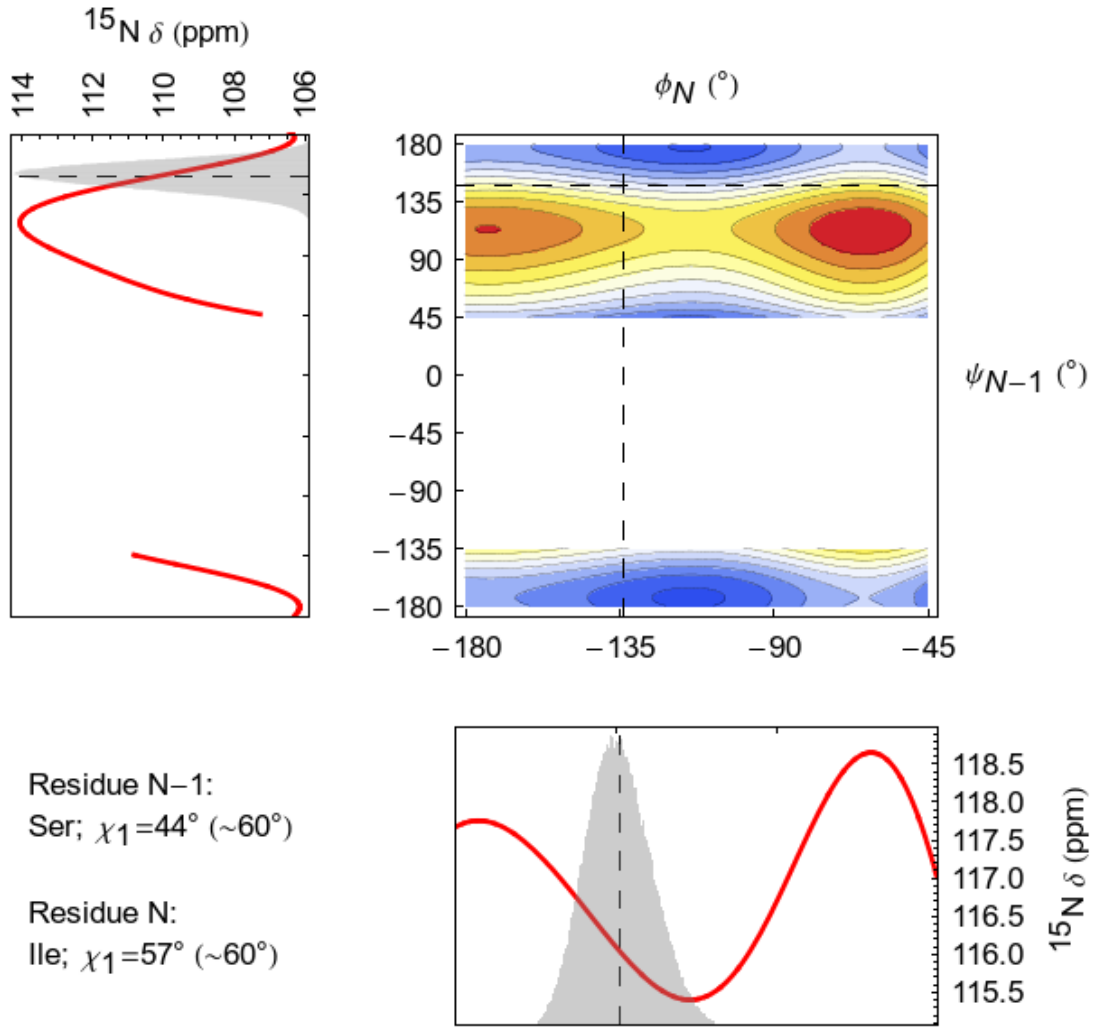


Figure 11.6: Smooth curves (red) representing DFT calculations (Appendix J, Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues S89 and I90 (i.e., N=90) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

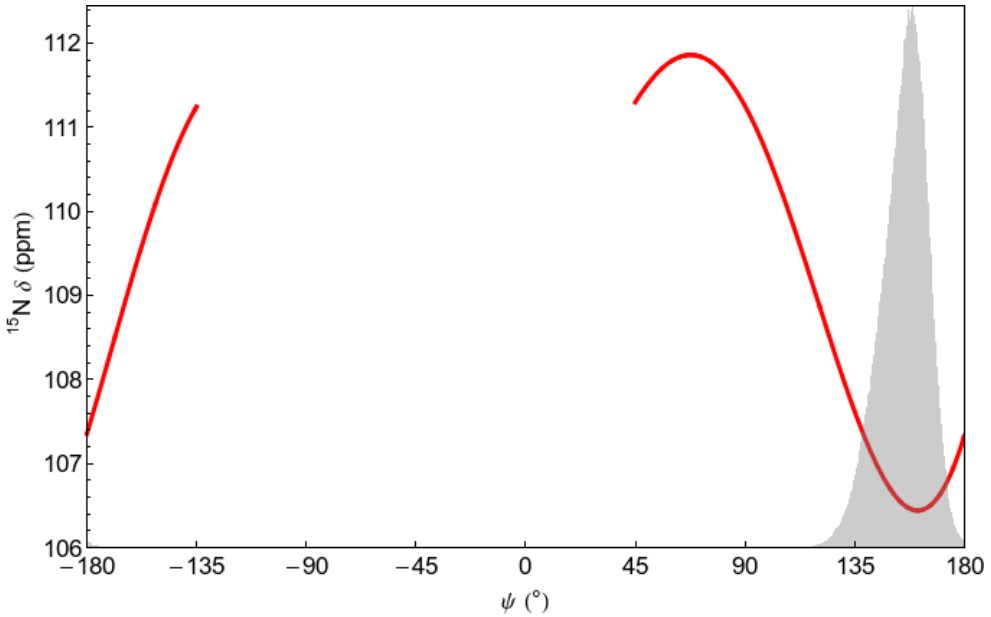


Figure 11.7: A smooth curve (red) representing DFT calculations (Appendix J, Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue I90 (i.e., N=90) in a 100 ns MD simulation.

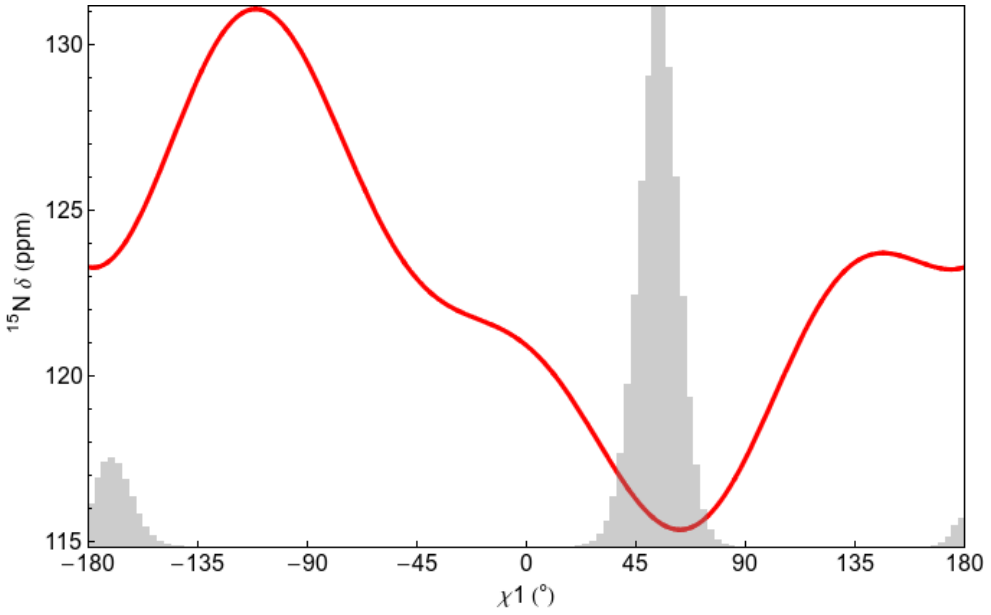


Figure 11.8: A smooth curve (red) representing DFT calculations (Appendix J, Section J.2) of isoleucine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue I90 (i.e., N=90) in a 100 ns MD simulation.

We therefore examine residues with consistently (in all five Adnectins studied here; Appendix J, Table J.1) positive amide nitrogen temperature coefficients (Table 11.3, top). We note that β-branched residues are over-represented among those in the top part of Table 11.3, suggesting that side chain degrees of freedom may be relevant, and, importantly, the amide nitrogens of all but two (L8, T56) act as donors in inter-β-strand hydrogen bonds (inferred from the 1FNF crystal structure; Appendix J, Table J.1). Furthermore, L8 and T56 have unusually positive amide proton temperature coefficients (especially for protons that are not known to be involved in intramolecular hydrogen bonds). DFT calculations using a formamide dimer hydrogen bond model (Chapter 8, Fig. 8.1) show that both primary (N^{th} amide nitrogen donor) and secondary (N-1^{th} carbonyl oxygen acceptor) hydrogen bonds deshield the N^{th} amide nitrogen (Fig. 11.9). Therefore, like amide proton temperature coefficients (Chapter 9), amide nitrogen temperature coefficients may be less negative when thermal expansion of hydrogen bonds is subject to restraints (as in β-structure), creating an opportunity for positive contributions (e.g., from dihedral angle degrees of freedom) to determine the overall sign.

Table 11.3: pWT Adnectin residues with positive (top) and negative (bottom) amide nitrogen temperature coefficients

Res.	$\boldsymbol{\Delta\delta_N/\Delta T}$[1]	$\boldsymbol{\Delta\delta_H/\Delta T}$[2]	**Curved**[3]	$\boldsymbol{\psi_{N-1}}$[4]	$\boldsymbol{\phi_N}$[4]	$\boldsymbol{\psi_N}$[4]	$\boldsymbol{\chi_{1,N}}$[4]	**Ex.**[5]
L8	4.9	2.75	✓	+	-	N	+	
L18	12.1	-3.24		A	A	+	-	✓
I20	19.0	-4.25		-	+	+	+	✓
S21	11.0	-3.73		N	N	+	A	✓
T35	18.9	-5.55		-	+	+	A	✓
Y36	21.4	-2.43	✓	N	N	+	+	
T56	14.9	-0.74		A	A	+	A	✓
T69	5.3	-3.06		N	N	N	A	✓
V72	5.7	-1.36		-	N	N	N	✓
I90	7.0	-2.05		N	+	+	+	✓
I59	-20.6	-3.93	✓	-	A	+	-	✓
I70	-20.0	-4.83		-	+	N	N	

[1] Amide nitrogen temperature coefficient (ppb/K)
[2] Amide proton temperature coefficient (ppb/K)
[3] Amide proton curvature detected
[4] +/-/N/A: amide nitrogen chemical shift increasing/decreasing/neutral/ambiguous with temperature
[5] $\chi_{1,N}$ conformational exchange between rotamers

The combination of $\phi/\psi/\chi_1$ dihedral angle distributions sampled in pWT Adnectin MD simulation with quantum chemical calculations of amide nitrogen chemical shifts may provide clues as to which of these degrees of freedom contribute to positive amide nitrogen temperature coefficients (Table 11.3, top). We contrast these results with those for select residues that have β-branched side chains, participate in intramolecular hydrogen bonds,

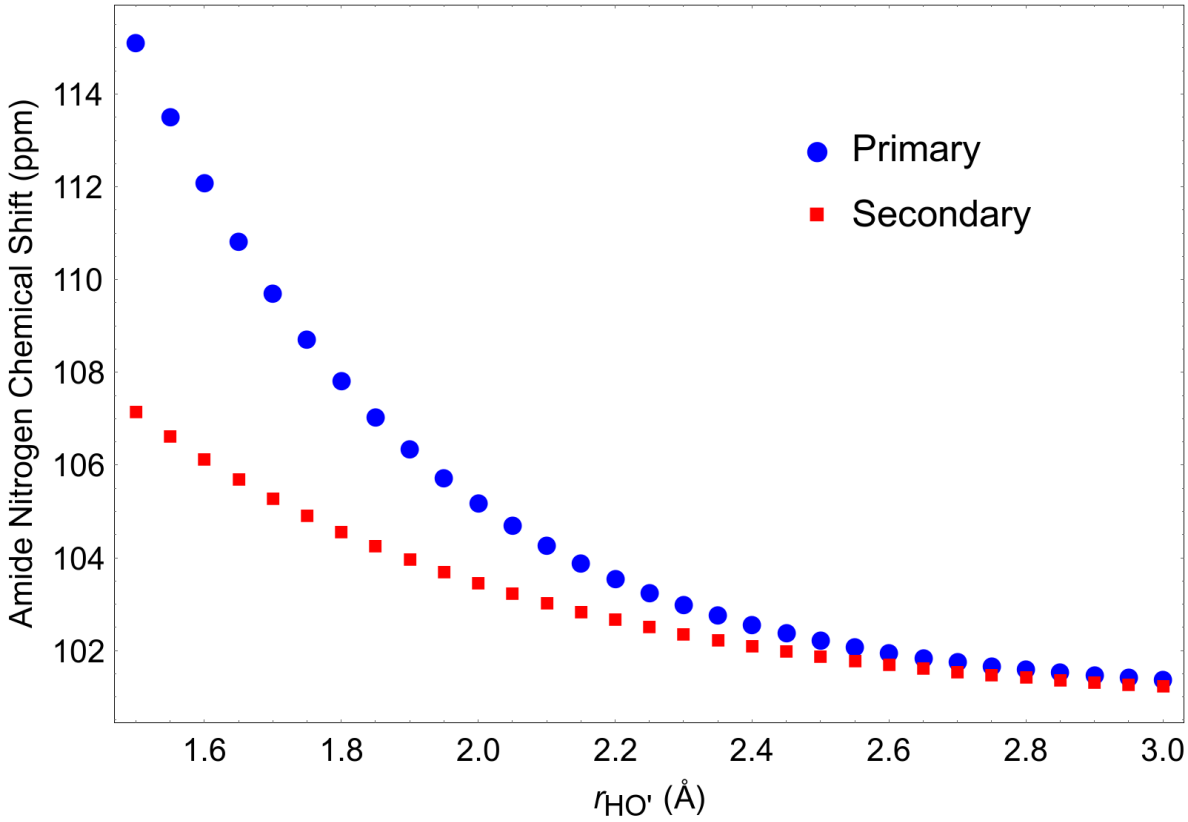


Figure 11.9: Calculated (DFT) N^{th} amide nitrogen chemical shift vs. primary (N^{th} amide nitrogen donor) and secondary (N-1^{th} carbonyl oxygen acceptor) hydrogen bond length (proton to oxygen distance).

yet have negative amide nitrogen temperature coefficients (Table 11.3, bottom). Direct simulation of the temperature dependence of protein dynamics presents difficulties (Chapter 8). Here, based on simulation at a single temperature (298 K), we hypothesize that the increasing (with temperature) influence of entropy has two likely consequences. First, consistent with standard deviations obtained by parameterizing a Karplus-style equation (modified to account for Gaussian distributions of dihedral angles arising from harmonic motion) using NMR data collected at different temperatures [293], we propose that distributions of Adnectin dihedral angles contributing to conformational averaging may broaden. Second, where evidence of exchange between distinct conformations is observed, we predict temperature-dependent changes in the equilibria between states. In particular, for side chains that are not confined to a single rotamer by tertiary protein structure, we expect that alternative (minority) rotamers may be increasingly populated at higher temperatures (as has been experimentally observed for leucine side chains [294]).

Entries in the ψ_{N-1}, ϕ_N, ψ_N, and $\chi_{1,N}$ columns of Table 11.3 reflect the assumption that broadening (with increasing temperature) of an approximately symmetric dihedral

distribution near a chemical shift maximum decreases the average chemical shift, broadening near a minimum increases the average chemical shift, and broadening in a linear region may have little impact (neutral). Also, exchange between different rotamers is evident in many χ_1 angle distributions (Fig. 11.8; Appendix J, Section J.3); in some cases it may be possible to infer the direction of chemical shift change with temperature from the observation that increasing population of minority rotamer population(s) is likely to be entropically favoured [294]. The picture that emerges in Table 11.3 is not straightforward, but detailed inspection does reveal some interesting trends.

First, despite the fact that temperature-dependent changes in ψ_{N-1}, ϕ_N, and especially ψ_N distributions characteristic of β-sheet secondary structure all have the potential to make positive contributions to the N^{th} amide nitrogen temperature coefficient, they are not good predictors of the overall sign. Second, temperature-dependent changes in equilibria between states (e.g., side chain rotamers) may have a larger impact than broadening of distributions; out of four examples (Table 11.3) in which the likely (i.e., according to our assumptions) direction of chemical shift change due to the temperature-dependence of equilibria between rotamers is clear (L18, I20, I90, I59), the sign of the temperature coefficient matches for all but L18 (Figs. 11.8, 11.10, 11.11). We note that L18 has multimodal distributions of ψ_{N-1} and ϕ_N angles (Appendix J, Fig. J.40), thus conformational exchange other than between rotameric states may also impact temperature coefficients.

Random coil amide nitrogen temperature coefficients are invariably negative (Table 11.2); deviations from random coil values, particularly in the positive direction, may be diagnostic of structure (Section 11.4.1). Examination of Adnectin residues with positive amide nitrogen temperature coefficients suggests that side chain ($\chi_{1,N}$) and hydrogen bond degrees of freedom make important contributions. The temperature dependences of ubiquitin ${}^{3h}J_{NC'}$ couplings [218] suggest that hydrogen bonds generally weaken with increasing temperature. Most Adnectin amide *proton* temperature coefficients are consistent with this hypothesis (i.e., they are negative; Chapter 9), including those for which the corresponding amide *nitrogen* temperature coefficients are positive (Table 11.3, top). DFT chemical shift calculations show that hydrogen bonds deshield the amide nitrogen (Fig. 11.9); therefore, we generally expect the temperature dependence of hydrogen bond geometry to make negative contributions to amide nitrogen temperature coefficients. However, hydrogen bond geometries in structured regions may change less (Chapter 9), creating an opportunity for other factors to push amide nitrogen temperature coefficients into positive territory. These other factors may include those we investigate here via quantum chemical calculations and classical MD: ψ_{N-1}, ϕ_N, ψ_N, and $\chi_{1,N}$ dihedral angles. We find that conformational exchange between distinct states evident in dihedral angle distributions is a promising candidate to rationalize the signs of amide nitrogen temperature coefficients; examples in which these states are distinguished by χ_1 angles are readily identified, but other differentiators (e.g., backbone dihedrals) are also possible.

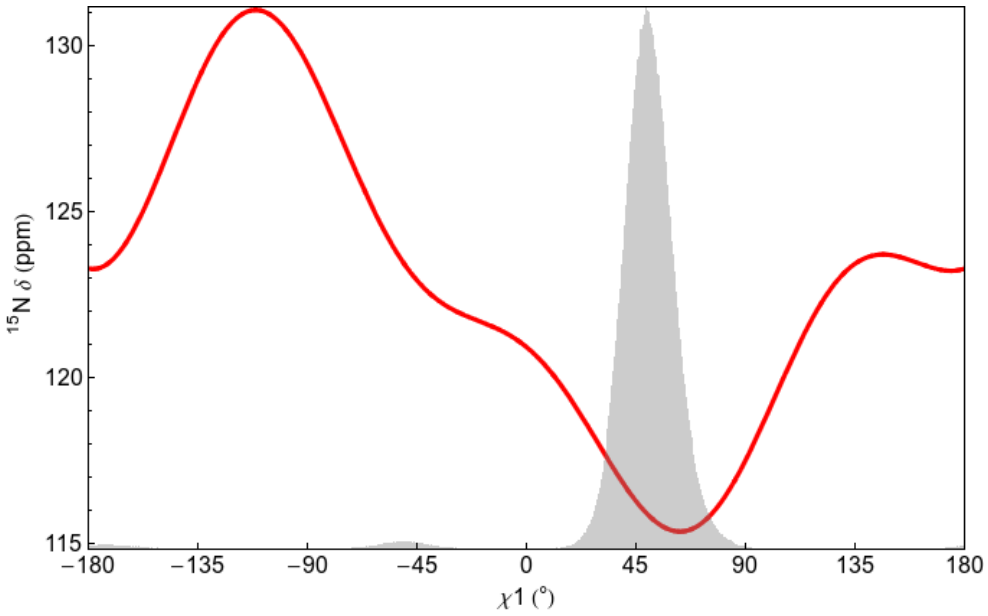


Figure 11.10: A smooth curve (red) representing DFT calculations (Appendix J, Section J.2) of isoleucine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue I20 (i.e., N=20) in a 100 ns MD simulation. The direction of chemical shift change that would result from increased sampling (at higher temperature) of minority rotamers matches the sign (positive) of the amide nitrogen temperature coefficient (Table 11.3).

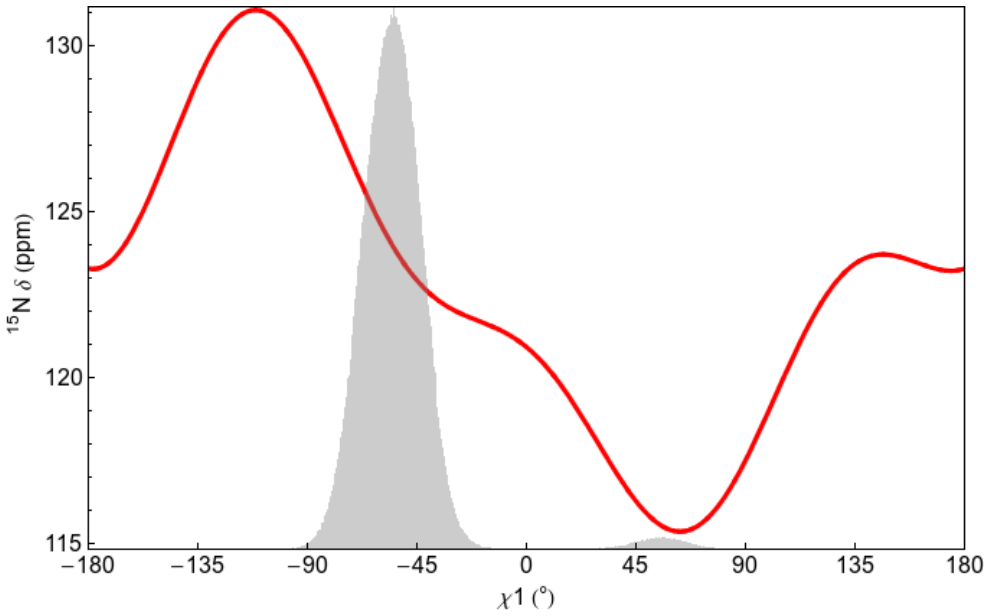


Figure 11.11: A smooth curve (red) representing DFT calculations (Appendix J, Section J.2) of isoleucine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue I59 (i.e., N=59) in a 100 ns MD simulation. The direction of chemical shift change that would result from increased sampling (at higher temperatures) of the minority rotamer matches the sign (negative) of the amide nitrogen temperature coefficient (Table 11.3).

Part IV

Conclusion

Chapter 12

Discussion & Future Research

Like the main body of this thesis (Chapters 4-11), the content of this final chapter is divided into two sections. In the first, we discuss prospects for future research into protein solubility and aggregation, a broad category of which the solubility engineering covered in Part II of this thesis represents only a sliver. In the second, we tackle topics related to Part III, such as variable-temperature nuclear magnetic resonance (VT-NMR) spectroscopy and hybrid quantum/classical approaches to the computation of NMR observables, as well as estimation of thermodynamic parameters from either VT-NMR data or simulation.

12.1 Protein Solubility and Aggregation

In Chapters 4-6, we address protein solubility/aggregation propensity in some detail, and even postulate the existence of an alternate state in which Adnectins transiently associate (Chapter 6). The proverbial elephant in the room is the nature of the aggregates themselves, about which we say little aside from noting the possibilities arising from domain-swapping oligomerization at either or both termini of Fn3 domains (Section 2.5). Here, we first discuss transient association as a possible precursor to aggregation, then move on to characterization of the aggregates themselves.

Under some conditions (e.g., a pH near the isoelectric point of the Parent Adnectin), the aforementioned transiently-associated state may promote aggregation; accordingly, observing the evolution of chemical shifts over time under such conditions might be informative. This state, distinguished by the susceptibility to amide H/D exchange of several residues in the C-terminal β-strand, may be described as 'invisible', i.e., a higher-energy (less populated) state detected primarily by indirect means. Beyond amide H/D exchange, other NMR-based techniques for the elucidation of invisible states include chemical exchange saturation transfer (CEST), Carr-Purcell-Meiboom-Gill relaxation dispersion and paramagnetic relaxation enhancement [295, 296]. Furthermore, when invisible states arise from intermolecular association (Chapter 6), peaks may narrow or broaden, and move upfield or downfield in a concentration-dependent manner.

High-resolution structural characterization of protein aggregates is, in general, very difficult. Although β-amyloid has been studied by X-ray crystallography [52, 106, 297] and solid-state NMR [298], these methods require a degree of long-range order that may be lacking in other types of aggregates. Direct study of aggregates by solution NMR is ruled out by their size and insolubility; however, dissolution of aggregates in a denaturing aprotic solvent following quenched amide H/D exchange allows spectroscopic study of selectively labelled monomer, revealing the relative exposure of protein regions within the aggregate [299].

12.2 Protein Dynamics, NMR, and Computational Methods

Molecular dynamics (MD) and nuclear magnetic resonance (NMR) spectroscopy may be a perfect match. NMR observables are exquisitely sensitive probes of protein structure and dynamics, but they are subject to averaging both over time and many individual molecules; fortunately, context for their interpretation may be found in simulation. On the other hand, the realism of MD simulations may be validated by comparison with NMR observables, or enhanced through the incorporation of NMR-based constraints.

Although classical physics-based calculations may be sufficient for the simulation of protein dynamics (exclusive of making or breaking covalent bonds) that are consistent with experiment [300], the magnetic resonance of nuclei is an inherently quantum phenomenon [262]. Quantum chemical calculations can bridge this gap, but their computational demands may be high, particularly when the object is to rationalize NMR spectroscopic measurements of nuclei throughout a molecule (in contrast with multiscale modelling, where quantum calculations may be restricted to a particular region of interest [301]).

In Chapters 8-11, we combine classical MD simulations of protein dynamics with NMR chemical shifts and couplings derived from density functional theory (DFT) quantum chemical calculations. In Chapter 8, we make a case for the validity of the approximations inherent in such an approach by calculating ${}^{3h}J_{NC'}$ couplings (which have the advantage of a relatively well-understood structural basis) in human ubiquitin that match experiment with unprecedented low root mean square error; future measurement of Adnectin ${}^{3h}J_{NC'}$ couplings may allow additional validation of the realism of simulated Adnectin dynamics reported in Chapters 9-11. Furthermore, the measurement of such couplings as a function of temperature (to our knowledge, something that has only previously been reported for human ubiquitin [218]) would, along with the amide proton temperature coefficients reported here, facilitate investigation of the temperature dependence of Adnectin hydrogen bond geometries.

Temperature coefficients are often reported for amide protons, but amide nitrogen temperature coefficients also shed light on protein structure and dynamics (Chapter 11). Beyond amide protons and nitrogens, the temperature dependences of other nuclei with

chemical shifts known to report on protein structure (e.g., C_α, H_α, and C_β) may be an underexploited source of insight into the dynamics thereof. Furthermore, like those we describe in Chapter 7, the experiments required to track the temperature dependences of C_α, H_α, and C_β atoms are of relatively low complexity (2D HSQCs). Curvature (Chapter 10), in the temperature dependences of nuclei other than protons may also be of interest, but increasing resolution in the indirectly detected dimension greatly expands the time required to collect each spectrum; use of ^{15}N and ^{13}C directly-detected experiments might sidestep this problem.

Curvature in the temperature dependences of amide proton chemical shifts may result from conformational exchange; we propose in Chapter 10 that the curvature detected in Adnectin binding loop regions (BC, DE, and FG loops; Fig. 2.2) could result from the sampling of binding-related conformations, despite the absence of ligand. Accordingly, some instances of curvature might exhibit a dependence on ligand concentration, and comparisons of the energetics of binding (e.g., determined by isothermal titration calorimetry) with estimates of the ΔG between curvature-related states (determined by fitting VT-NMR data to our curvature model and/or single-temperature MD simulation) might be informative.

Accurate reproduction of temperature-dependent protein dynamics through MD simulation at multiple temperatures may require much better conformational sampling than we achieve in 100 ns trajectories. To address this problem, various techniques for the enhancement of conformational sampling [302] are worth considering; however, if calculated temperature dependences of NMR chemical shifts and couplings are to match experimentally determined values, the method of sampling enhancement must not disturb the equilibria between fast-exchanging conformational states.

12.3 Conclusion

The future experiments proposed above all have the potential to help draw together what we have learned about Adnectin dynamics and solubility into a more cohesive whole. Those that probe the boundaries between protein dynamics, function, and aggregation may be of particular interest. Accordingly, we most strongly advocate further research into 'invisible states' and aggregate structure (which may be related), as well as curvature and binding.

References

[1] Voet D, Voet J (2010) *Biochemistry, 4th Edition.* (John Wiley & Sons).

[2] Levinthal C (1969) How to fold graciously in *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois.* (University of Illinois Press), pp. 22–24.

[3] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics* 21(3):167–195.

[4] Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences of the United States of America* 89(18):8721–8725.

[5] Fisher CK, Stultz CM (2011) Constructing ensembles for intrinsically disordered proteins. *Current Opinion in Structural Biology* 21(3):426–431.

[6] Jensen MR, Zweckstetter M, Huang Jr, Blackledge M (2014) Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chemical Reviews* 114(13):6632–6660.

[7] Southall NT, Dill KA, Haymet A (2002) A view of the hydrophobic effect. *The Journal of Physical Chemistry B* 106(3):521–533.

[8] Sadqi M, Lapidus LJ, Muñoz V (2003) How fast is protein hydrophobic collapse? *Proceedings of the National Academy of Sciences of the United States of America* 100(21):12117–12122.

[9] Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proceedings of the National Academy of Sciences of the United States of America* 104(8):2655–2660.

[10] Gilmanshin R, Williams S, Callender RH, Woodruff WH, Dyer RB (1997) Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proceedings of the National Academy of Sciences of the United States of America* 94(8):3709–3713.

[11] Fedorov AN, Baldwin TO (1997) Cotranslational protein folding. *Journal of Biological Chemistry* 272(52):32715–32718.

[12] Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424(6950):805–808.

[13] Cohen SN, Chang AC, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 70(11):3240–3244.

[14] Baneyx F (1999) Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology* 10(5):411–421.

[15] Lee SY (1996) High cell-density culture of Escherichia coli. *Trends in Biotechnology* 14(3):98–105.

[16] Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in Escherichia coli. *Nature Biotechnology* 22(11):1399–1408.

[17] Oberg K, Chrunyk BA, Wetzel R, Fink AL (1994) Native-like secondary structure in interleukin-1ß. inclusion bodies by attenuated total reflectance FTIR. *Biochemistry* 33(9):2628–2634.

[18] Przybycien TM, Dunn JP, Valax P, Georglou G (1994) Secondary structure characterization of ß-lactamase inclusion bodies. *Protein Engineering* 7(1):131–136.

[19] Ami D, Natalello A, Taylor G, Tonon G, Doglia SM (2006) Structural analysis of protein inclusion bodies by fourier transform infrared microspectroscopy. *Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics* 1764(4):793–799.

[20] Ami D, Natalello A, Gatti-Lafranconi P, Lotti M, Doglia SM (2005) Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. *FEBS Letters* 579(16):3433–3436.

[21] García-Fruitós E, González-Montalbán N, Morell M, Vera A, Ferraz RM, Arís A, Ventura S, Villaverde A (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microbial Cell Factories* 4(1):27.

[22] Singh SM, Panda AK (2005) Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering* 99(4):303–310.

[23] García-Fruitós E, Vázquez E, Díez-Gil C, Corchero JL, Seras-Franzoso J, Ratera I, Veciana J, Villaverde A (2012) Bacterial inclusion bodies: making gold from waste. *Trends in Biotechnology* 30(2):65–70.

[24] Ratanji KD, Derrick JP, Dearman RJ, Kimber I (2014) Immunogenicity of therapeutic proteins: influence of aggregation. *Journal of mmunotoxicology* 11(2):99–109.

[25] Moussa EM, Panchal JP, Moorthy BS, Blum JS, Joubert MK, Narhi LO, Topp EM (2016) Immunogenicity of therapeutic protein aggregates. *Journal of Pharmaceutical Sciences* 105(2):417–430.

[26] Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *Journal of Molecular Biology* 222(3):599–620.

[27] Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences* 26(10):597–604.

[28] Enfors SO (1992) Control of in vivo proteolysis in the production of recombinant proteins. *Trends in Biotechnology* 10:310–315.

[29] Corchero J, Cubarsi R, Enfors S, Villaverde A (1997) Limited in vivo proteolysis of aggregated proteins. *Biochemical and Biophysical Research Communications* 237(2):325–330.

[30] Carrió M, Villaverde A (2003) Role of molecular chaperones in inclusion body formation. *FEBS Letters* 537(1):215–221.

[31] Martínez-Alonso M, Vera A, Villaverde A (2007) Role of the chaperone DnaK in protein solubility and conformational quality in inclusion body-forming Escherichia coli cells. *FEMS Microbiology Letters* 273(2):187–195.

[32] Rokney A, Shagan M, Kessel M, Smith Y, Rosenshine I, Oppenheim AB (2009) E. coli transports aggregated proteins to the poles by a specific and energy-dependent process. *Journal of Molecular Biology* 392(3):589–601.

[33] Jahn TR, Radford SE (2008) Folding versus aggregation: Polypeptide conformations on competing pathways. *Archives of Biochemistry and Biophysics* 469(1):100–117.

[34] Kiefhaber T, Rudolph R, Kohler HH, Buchner J (1991) Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. *Nature Biotechnology* 9(9):825–829.

[35] Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G, Dobson CM (2002) Kinetic partitioning of protein folding and aggregation. *Nature Structural & Molecular Biology* 9(2):137–143.

[36] Chi EY, Krishnan S, Randolph TW, Carpenter JF (2003) Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharmaceutical Research* 20(9):1325–1336.

[37] Marcon G, Plakoutsi G, Chiti F (2006) Protein aggregation starting from the native globular state. *Methods in Enzymology* 413:75–91.

[38] Bemporad F, Chiti F (2009) Native-like aggregation of the acylphosphatase from Sulfolobus solfataricus and its biological implications. *FEBS Letters* 583(16):2630–2638.

[39] Chiti F, Dobson CM (2008) Amyloid formation by globular proteins under native conditions. *Nature Chemical Biology* 5(1):15–22.

[40] Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and Design* 3(1):R9–R23.

[41] Morell M, Bravo R, Espargaró A, Sisquella X, Avilés FX, Fernàndez-Busquets X, Ventura S (2008) Inclusion bodies: specificity in their aggregation process and amyloid-like structure. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1783(10):1815–1825.

[42] Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R (2008) Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biology* 6(8):e195.

[43] de Groot NS, Sabate R, Ventura S (2009) Amyloids in bacterial inclusion bodies. *Trends in Biochemical Sciences* 34(8):408–416.

[44] Rajan RS, Illing ME, Bence NF, Kopito RR (2001) Specificity in intracellular protein aggregation and inclusion body formation. *Proceedings of the National Academy of Sciences of the United States of America* 98(23):13060–13065.

[45] Chothia C, Janin J (1975) Principles of protein-protein recognition. *Nature* 256(5520):705–708.

[46] Tsai CJ, Lin SL, Wolfson HJ, Nussinov R (1997) Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Science* 6(1):53–64.

[47] Larsen TA, Olson AJ, Goodsell DS (1998) Morphology of protein–protein interfaces. *Structure* 6(4):421–427.

[48] Jones S, Thornton JM (1996) Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* 93(1):13–20.

[49] Carugo O, Argos P (1997) Protein-protein crystal-packing contacts. *Protein Science* 6(10):2261–2263.

[50] Kobe B, Guncar G, Buchholz R, Huber T, Maco B, Cowieson N, Martin J, Marfori M, Forwood J (2008) Crystallography and protein-protein interactions: biological interfaces and crystal contacts. *Biochemical Society Transactions* 36:1438–1441.

[51] Dasgupta S, Iyer GH, Bryant SH, Lawrence CE, Bell JA (1997) Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins: Structure, Function, and Bioinformatics* 28(4):494–514.

[52] Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB, Blake CC (1997) Common core structure of amyloid fibrils by synchrotron x-ray diffraction. *Journal of Molecular Biology* 273(3):729–739.

[53] Bennett MJ, Schlunegger MP, Eisenberg D (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Science* 4(12):2455–2468.

[54] Rousseau F, Schymkowitz JW, Itzhaki LS (2003) The unfolding story of three-dimensional domain swapping. *Structure* 11(3):243–251.

[55] Sambashivan S, Liu Y, Sawaya MR, Gingery M, Eisenberg D (2005) Amyloid-like fibrils of ribonuclease a with three-dimensional domain-swapped and native-like structure. *Nature* 437(7056):266–269.

[56] Gee EP, Yüksel D, Stultz CM, Ingber DE (2013) SLLISWD sequence in the 10FNIII domain initiates fibronectin fibrillogenesis. *Journal of Biological Chemistry* 288(29):21329–21340.

[57] Pankov R, Yamada KM (2002) Fibronectin at a glance. *Journal of Cell Science* 115(20):3861–3863.

[58] Bork P, Doolittle RF (1992) Proposed acquisition of an animal protein domain by bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 89(19):8990–8994.

[59] Baneyx G, Vogel V (1999) Self-assembly of fibronectin into fibrillar networks underneath dipalmitoyl phosphatidylcholine monolayers: role of lipid matrix and tensile forces. *Proceedings of the National Academy of Sciences of the United States of America* 96(22):12518–12523.

[60] Ohashi T, Kiehart DP, Erickson HP (1999) Dynamics and elasticity of the fibronectin matrix in living cell culture visualized by fibronectin–green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* 96(5):2153–2158.

[61] Erickson HP (1994) Reversible unfolding of fibronectin type III and immunoglobulin domains provides the structural basis for stretch and elasticity of titin and fibronectin. *Proceedings of the National Academy of Sciences of the United States of America* 91(21):10114–10118.

[62] Baneyx G, Baugh L, Vogel V (2002) Fibronectin extension and unfolding within cell matrix fibrils controlled by cytoskeletal tension. *Proceedings of the National Academy of Sciences of the United States of America* 99(8):5139–5143.

[63] Hocking DC, Smith RK, McKeown-Longo PJ (1996) A novel role for the integrin-binding III-10 module in fibronectin matrix assembly. *The Journal of Cell BBiology* 133(2):431–444.

[64] Ingham KC, Brew SA, Huff S, Litvinovich SV (1997) Cryptic self-association sites in type III modules of fibronectin. *Journal of Biological Chemistry* 272(3):1718–1724.

[65] Zhong C, Chrzanowska-Wodnicka M, Brown J, Shaub A, Belkin AM, Burridge K (1998) Rho-mediated contractility exposes a cryptic site in fibronectin and induces fibronectin matrix assembly. *The Journal of Cell Biology* 141(2):539–551.

[66] Oberhauser AF, Badilla-Fernandez C, Carrion-Vazquez M, Fernandez JM (2002) The mechanical hierarchies of fibronectin observed with single-molecule AFM. *Journal of Molecular Biology* 319(2):433–447.

[67] Li L, Huang HHL, Badilla CL, Fernandez JM (2005) Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *Journal of Molecular Biology* 345(4):817–826.

[68] Paci E, Karplus M (1999) Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *Journal of Molecular Biology* 288(3):441–459.

[69] Gao M, Craig D, Vogel V, Schulten K (2002) Identifying unfolding intermediates of FN-III 10 by steered molecular dynamics. *Journal of Molecular Biology* 323(5):939–950.

[70] Gee EP, Ingber DE, Stultz CM (2008) Fibronectin unfolding revisited: modeling cell traction-mediated unfolding of the tenth type-III repeat. *PLoS One* 3(6):e2373.

[71] Chiquet-Ehrismann R (1990) What distinguishes tenascin from fibronectin? *The FASEB Journal* 4(9):2598–2604.

[72] Ng SP, Rounsevell RW, Steward A, Geierhaas CD, Williams PM, Paci E, Clarke J (2005) Mechanical unfolding of TNfn3: the unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *Journal of Molecular Biology* 350(4):776–789.

[73] Cota E, Hamill SJ, Fowler SB, Clarke J (2000) Two proteins with the same structure respond very differently to mutation: the role of plasticity in protein stability. *Journal of Molecular Biology* 302(3):713–725.

[74] Cota E, Clarke J (2000) Folding of beta-sandwich proteins: Three-state transition of a fibronectin type III module. *Protein Science* 9(1):112–120.

[75] Cota E, Steward A, Fowler SB, Clarke J (2001) The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. *Journal of Molecular Biology* 305(5):1185–1194.

[76] Hamill SJ, Steward A, Clarke J (2000) The folding of an immunoglobulin-like greek key protein is defined by a common-core nucleus and regions constrained by topology. *Journal of Molecular Biology* 297(1):165–178.

[77] Ultsch MH, Wiesmann C, Simmons LC, Henrich J, Yang M, Reilly D, Bass SH, de Vos AM (1999) Crystal structures of the neurotrophin-binding domain of TrkA, TrkB and TrkC. *Journal of Molecular Biology* 290(1):149–159.

[78] Abrams CS, Cines DB (2002) Platelet glycoprotein IIb/IIIa inhibitors and thrombocytopenia: possible link between platelet activation, autoimmunity and thrombosis. *Thrombosis and Haemostasis* 88(12):888–889.

[79] Spinelli S, Desmyter A, Frenken L, Verrips T, Tegoni M, Cambillau C (2004) Domain swapping of a llama VHH domain builds a crystal-wide β-sheet structure. *FEBS Letters* 564(1-2):35–40.

[80] Teplyakov A, Obmolova G, Malia TJ, Luo J, Jacobs SA, Chan W, Domingo D, Baker A, O'Neil KT, Gilliland GL (2014) C-terminal β-strand swapping in a consensus-derived fibronectin type III scaffold. *Proteins: Structure, Function, and Bioinformatics* 82(7):1359–1369.

[81] Luo J, Teplyakov A, Obmolova G, Malia TJ, Chan W, Jacobs SA, O'Neil KT, Gilliland GL (2014) N-terminal β-strand swapping in a consensus-derived alternative scaffold driven by stabilizing hydrophobic interactions. *Proteins: Structure, Function, and Bioinformatics* 82(7):1527–1533.

[82] Grilo AL, Mantalaris A (2019) The increasingly human and profitable monoclonal antibody market. *Trends in Biotechnology* 37(1):9–16.

[83] Koide A, Bailey CW, Huang X, Koide S (1998) The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of Molecular Biology* 284(4):1141–1151.

[84] Xu L, Aha P, Gu K, Kuimelis RG, Kurz M, Lam T, Lim AC, Liu H, Lohse PA, Sun L, Weng S, Wagner RW, Lipovšek D (2002) Directed evolution of high-affinity antibody mimics using mrna display. *Chemistry & Biology* 9(8):933–942.

[85] Parker M, Chen Y, Danehy F, Dufu K, Ekstrom J, Getmanova E, Gokemeijer J, Xu L, Lipovsek D (2005) Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two. *Protein Engineering Design and Selection* 18(9):435–444.

[86] Lipovšek D, Lippow SM, Hackel BJ, Gregson MW, Cheng P, Kapila A, Wittrup KD (2007) Evolution of an interloop disulfide bond in high-affinity antibody mimics based on fibronectin type III domain and selected by yeast surface display: molecular convergence with single-domain camelid and shark antibodies. *Journal of Molecular Biology* 368(4):1024–1041.

[87] Lipovšek D (2011) Adnectins: engineered target-binding protein therapeutics. *Protein Engineering Design and Selection* 24(1-2):3–9.

[88] Koide S, Koide A, Lipovšek D (2012) Target-binding proteins based on the 10th human fibronectin type III domain (10Fn3) in *Methods in Enzymology.* (Elsevier) Vol. 503, pp. 135–156.

[89] Khan JA, Camac DM, Low S, Tebben AJ, Wensel DL, Wright MC, Su J, Jenny V, Gupta RD, Ruzanov M, Russo KA, Bell A, Yongmi A, Bryson JW, Gao M, Gambhire P, Baldwin ET, Gardner D, Cavallaro CJ, Duncia JV, Hynes Jr J (2015) Developing Adnectins that target SRC co-activator binding to PXR: a structural approach toward understanding promiscuity of PXR. *Journal of Molecular Biology* 427(4):924–942.

[90] Mitchell T, Chao G, Sitkoff D, Lo F, Monshizadegan H, Meyers D, Low S, Russo K, DiBella R, Denhez F, Gao M, Myers J, Duke G, Witmer M, Miao B, Ho SP, Khan J, Parker RA (2014) Pharmacologic profile of the Adnectin BMS-962476, a small protein biologic alternative to PCSK9 antibodies for low-density lipoprotein lowering. *Journal of Pharmacology and Experimental Therapeutics* 350(2):412–424.

[91] Ramamurthy V, Krystek Jr SR, Bush A, Wei A, Emanuel SL, Gupta RD, Janjua A, Cheng L, Murdock M, Abramczyk B, Cohen D, Lin Z, Morin P, Davis JH, Dabritz M, McLaughlin DC, Russo KA, Chao G, Wright MC, Jenny VA, Engle LJ, Furfine E, Sheriff S (2012) Structures of adnectin/protein complexes reveal an expanded binding footprint. *Structure* 20(2):259–269.

[92] Emanuel SL, Engle LJ, Chao G, Zhu RR, Cao C, Lin Z, Yamniuk AP, Hosbach J, Brown J, Fitzpatrick E, Gokemeijer J, Morin P, Morse BA, Carvajal IM, Fabrizio D, Wright MC, Das Gupta R, Gosselin M, Cataldo D, Ryseck RP, Doyle ML, Wong TW, Camphausen RT, Cload ST, Marsh HN, Gottardis MM, Furfine ES (2011) A fibronectin scaffold approach to bispecific inhibitors of epidermal growth factor receptor and insulin-like growth factor-I receptor. *mAbs* 3(1):38–48.

[93] Wensel D, Sun Y, Li Z, Zhang S, Picarillo C, McDonagh T, Fabrizio D, Cockett M, Krystal M, Davis J (2017) Discovery and characterization of a novel CD4-binding adnectin with potent anti-HIV activity. *Antimicrobial Agents and Chemotherapy* 61(8):e00508–17.

[94] Wensel D, Sun Y, Davis J, Li Z, Zhang S, McDonagh T, Fabrizio D, Cockett M, Krystal M (2018) A novel gp41-binding Adnectin with potent anti-HIV activity

is highly synergistic when linked to a CD4-binding adnectin. *Journal of Virology* 92(14):e00421–18.

[95] Lipovšek D, Carvajal I, Allentoff AJ, Barros Jr A, Brailsford J, Cong Q, Cotter P, Gangwar S, Hollander C, Lafont V, , et al. (2018) Adnectin–drug conjugates for Glypican-3-specific delivery of a cytotoxic payload to tumors. *Protein Engineering, Design and Selection* 31(5):159–171.

[96] Huang RYC, ONeil SR, Lipovšek D, Chen G (2018) Conformational assessment of adnectin and Adnectin-drug conjugate by hydrogen/deuterium exchange mass spectrometry. *Journal of The American Society for Mass Spectrometry* 29(7):1524–1531.

[97] Palm T, Esfandiary R, Gandhi R (2011) The effect of PEGylation on the stability of small therapeutic proteins. *Pharmaceutical Development and Technology* 16(5):441–448.

[98] Mamluk R, Carvajal IM, Morse BA, Wong HK, Abramowitz J, Aslanian S, Lim AC, Gokemeijer J, Storek MJ, Lee J, Gosselin M, Wright MC, Camphausen RT, Wang J, Chen Y, Miller K, Sanders K, Short S, Sperinde J, Prasad G, Williams S, Kerbel R, Ebos J, Mutsaers A, Mendlein JD, Harris AS, Furfine ES (2010) Anti-tumor effect of CT-322 as an adnectin inhibitor of vascular endothelial growth factor receptor-2. *mAbs* 2(2):199–208.

[99] Aghaabdollahian S, Cohan RA, Norouzian D, Davami F, Karam MRA, Torkashvand F, Vaseghi G, Moazzami R, Dizaji SL (2019) Enhancing bioactivity, physicochemical, and pharmacokinetic properties of a nano-sized, anti-vegfr2 adnectin, through pasylation technology. *Scientific Reports* 9(1):2978.

[100] Han X, Cinay GE, Zhao Y, Guo Y, Zhang X, Wang P (2017) Adnectin-based design of chimeric antigen receptor for T cell engineering. *Molecular Therapy* 25(11):2466–2476.

[101] Trainor K (2015) Master's thesis: Inclusion body formation by mutants of the tenth human fibronectin type III domain (University of Waterloo).

[102] Trainor K, Gingras Z, Shillingford C, Malakian H, Gosselin M, Lipovšek D, Meiering EM (2016) Ensemble modeling and intracellular aggregation of an engineered immunoglobulin-like domain. *Journal of Molecular Biology* 428(6):1365–1374.

[103] Sankar K, Krystek Jr SR, Carl SM, Day T, Maier JK (2018) AggScore: Prediction of aggregation-prone regions in proteins based on the distribution of surface patches. *Proteins: Structure, Function, and Bioinformatics* 86(11):1147–1156.

[104] Trainor K, Broom A, Meiering EM (2017) Exploring the relationships between protein sequence, structure and solubility. *Current Opinion in Structural Biology* 42:136–146.

[105] Sormanni P, Aprile FA, Vendruscolo M (2015) The CamSol method of rational design of protein mutants with enhanced solubility. *Journal of Molecular Biology* 427(2):478–490.

[106] Nelson R, Sawaya MR, Balbirnie M, Madsen AØ, Riekel C, Grothe R, Eisenberg D (2005) Structure of the cross-β spine of amyloid-like fibrils. *Nature* 435(7043):773–778.

[107] Collinge J, Clarke AR (2007) A general model of prion strains and their pathogenicity. *Science* 318(5852):930–936.

[108] Hamodrakas SJ (2011) Protein aggregation and amyloid fibril formation prediction software from primary sequence: towards controlling the formation of bacterial inclusion bodies. *FEBS Journal* 278(14):2428–2435.

[109] Agrawal NJ, Kumar S, Wang X, Helk B, Singh SK, Trout BL (2011) Aggregation in protein-based biotherapeutics: Computational studies and tools to identify aggregation-prone regions. *Journal of Pharmaceutical Sciences* 100(12):5081–5095.

[110] Belli M, Ramazzotti M, Chiti F (2011) Prediction of amyloid aggregation in vivo. *EMBO Reports* 12(7):657–663.

[111] Baldwin AJ, Knowles TP, Tartaglia GG, Fitzpatrick AW, Devlin GL, Shammas SL, Waudby CA, Mossuto MF, Meehan S, Gras SL, Christodoulou J, Anthony-Cahill SJ, Barker PD, Vendruscolo M, Dobson CM (2011) Metastability of native proteins and the phenomenon of amyloid formation. *Journal of the American Chemical Society* 133(36):14160–14163.

[112] Esteras-Chopo A, Serrano L, de la Paz ML (2005) The amyloid stretch hypothesis: Recruiting proteins toward the dark side. *Proceedings of the National Academy of Sciences of the United States of America* 102(46):16672–16677.

[113] Teng PK, Eisenberg D (2009) Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Engineering, Design & Selection* 22(8):531–536.

[114] Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL (2009) Design of therapeutic proteins with enhanced stability. *Proceedings of the National Academy of Sciences of the United States of America* 106(29):11937–11942.

[115] Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S (2015) AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Research* p. gkv359.

[116] Schaller A, Connors NK, Oelmeier SA, Hubbuch J, Middelberg AP (2015) Predicting recombinant protein expression experiments using molecular dynamics simulation. *Chemical Engineering Science* 121:340–350.

[117] Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D (2012) PROSO II–a new method for protein solubility prediction. *FEBS Journal* 279(12):2192–2200.

[118] Huang HL, Charoenkwan P, Kao TF, Lee HC, Chang FL, Huang WL, Ho SJ, Shu LS, Chen WL, Ho SY (2012) Prediction and analysis of protein solubility using a novel scoring card method with dipeptide composition. *BMC Bioinformatics* 13(17):1.

[119] Khurana S, Rawi R, Kunji K, Chuang GY, Bensmail H, Mall R (2018) DeepSol: a deep learning framework for sequence-based protein solubility prediction. *Bioinformatics* 34(15):2605–2613.

[120] Yang Y, Niroula A, Shen B, Vihinen M (2016) PON-Sol: prediction of effects of amino acid substitutions on protein solubility. *Bioinformatics* p. btw066.

[121] Niwa T, Ying BW, Saito K, Jin W, Takada S, Ueda T, Taguchi H (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proceedings of the National Academy of Sciences of the United States of America* 106(11):4201–4206.

[122] Samak T, Gunter D, Wang Z (2012) Prediction of protein solubility in E. coli. *E-Science (e-Science), 2012 IEEE 8th International Conference on* pp. 1–8.

[123] Agostini F, Vendruscolo M, Tartaglia GG (2012) Sequence-based prediction of protein solubility. *Journal of Molecular Biology* 421(2):237–241.

[124] Xiao R, Anderson S, Aramini J, Belote R, Buchwald WA, Ciccosanti C, Conover K, Everett JK, Hamilton K, Huang YJ, Janjua H, Jiang M, Kornhaber GJ, Lee DY, Locke JY, Ma LC, Maglaqui M, Mao L, Mitra S, Patel D, Rossi P, Sahdev S, Shastry R, Swapna GVT, Tong SN, Wang D, Wang H, Zhao L, Montelione GT, Acton TB (2010) The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *Journal of Structural Biology* 172(1):21–33.

[125] Price WN, Handelman SK, Everett JK, Tong SN, Bracic A, Luff JD, Naumov V, Acton T, Manor P, Xiao R, Rost B, Montelione GT, Hunt JF (2011) Large-scale experimental studies show unexpected amino acid effects on protein expression and solubility in vivo in E. coli. *Microbial Informatics and Experimentation* 1(1):1.

[126] Otzen DE, Kristensen O, Oliveberg M (2000) Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. *Proceedings of the National Academy of Sciences of the United States of America* 97:9907–9912.

[127] Richardson JS, Richardson DC (2002) Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proceedings of the National Academy of Sciences of the United States of America* 99(5):2754–2759.

[128] Rousseau F, Serrano L, Schymkowitz JW (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. *Journal of Molecular Biology* 355:1037–1047.

[129] Chou KC (2001) Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins: Structure, Function, and Bioinformatics* 43(3):246–255.

[130] Niu X, Shi F, Hu X, Xia J, Li N (2014) Predicting the protein solubility by integrating chaos games representation and entropy in information theory. *Expert Systems with Applications* 41(4):1672–1679.

[131] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics* 21(3):167–195.

[132] Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, Vendruscolo M (2008) Prediction of aggregation-prone regions in structured proteins. *Journal of Molecular Biology* 380(2):425–436.

[133] Walsh I, Seno F, Tosatto SC, Trovato A (2014) PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Research* 42(W1):W301–W307.

[134] Gasior P, Kotulska M (2014) FISH Amyloid–a new method for finding amyloidogenic segments in proteins based on site specific co-occurence of aminoacids. *BMC Bioinformatics* 15(1):1.

[135] Hirose S, Noguchi T (2013) ESPRESSO: a system for estimating protein expression and solubility in protein expression systems. *Proteomics* 13(9):1444–1456.

[136] Hirose S, Kawamura Y, Yokota K, Kuroita T, Natsume T, Komiya K, Tsutsumi T, Suwa Y, Isogai T, Goshima N, Noguchi T (2011) Statistical analysis of features associated with protein expression/solubility in an in vivo Escherichia coli expression system and a wheat germ cell-free expression system. *Journal of Biochemistry* 150(1):73–81.

[137] Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnology* 22(10):1302–1306.

[138] Navarro S, Villar-Piqué A, Ventura S (2014) Selection against toxic aggregation-prone protein sequences in bacteria. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1843(5):866–874.

[139] Ganesan A, Siekierska A, Beerten J, Brams M, Van Durme J, De Baets G, Van der Kant R, Gallardo R, Ramakers M, Langenberg T, Wilkinson H, De Smet F, Ulens C, Rousseau F, Schymkowitz J (2016) Structural hot spots for the solubility of globular proteins. *Nature Communications* 7.

[140] De Baets G, Van Durme J, van der Kant R, Schymkowitz J, Rousseau F (2015) Solubis: optimize your protein. *Bioinformatics* 31(15):2580–2582.

[141] Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005) The FoldX web server: an online force field. *Nucleic Acids Research* 33:W382–W388.

[142] Broom A, Jacobi Z, Trainor K, Meiering EM (2017) Computational tools help improve protein stability but with a solubility tradeoff. *Journal of Biological Chemistry* 292(35):14349–14361.

[143] Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* 8(1):1.

[144] Villar-Pique A, de Groot NS, Sabaté R, Acebrón SP, Celaya G, Fernàndez-Busquets X, Muga A, Ventura S (2012) The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity. *Journal of Molecular Biology* 421(2):270–281.

[145] Trovato A, Seno F, Tosatto SC (2007) The PASTA server for protein aggregation prediction. *Protein Engineering Design and Selection* 20(10):521–523.

[146] Walsh I, Martin AJ, Di Domenico T, Tosatto SC (2011) Espritz: accurate and fast prediction of protein disorder. *Bioinformatics* 28(4):503–509.

[147] Paladin L, Piovesan D, Tosatto SC (2017) SODA: prediction of protein solubility from disorder and aggregation propensity. *Nucleic acids research* 45(W1):W236–W240.

[148] Sahin E, Jordan JL, Spatara ML, Naranjo A, Costanzo JA, Weiss IV WF, Robinson AS, Fernandez EJ, Roberts CJ (2011) Computational design and biophysical characterization of aggregation-resistant point mutations for γD crystallin illustrate a balance of conformational stability and intrinsic aggregation propensity. *Biochemistry* 50(5):628–639.

[149] Yageta S, Lauer TM, Trout BL, Honda S (2015) Conformational and colloidal stabilities of isolated constant domains of human immunoglobulin G and their impact on antibody aggregation under acidic conditions. *Molecular Pharmaceutics* 12(5):1443–1455.

[150] Du J, Cho PY, Yang DT, Murphy RM (2012) Identification of beta-amyloid-binding sites on transthyretin. *Protein Engineering Design and Selection* 25(7):337–345.

[151] Maurer-Stroh S, Debulpaep M, Kuemmerer N, de la Paz ML, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, Schymkowitz JWH, Rousseau F (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nature Methods* 7(3):237–242.

[152] Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, Eisenberg D (2006) The 3D profile method for identifying fibril-forming segments of proteins. *Proceedings of the National Academy of Sciences of the United States of America* 103(11):4074–4078.

[153] Perchiacca JM, Bhattacharya M, Tessier PM (2011) Mutational analysis of domain antibodies reveals aggregation hotspots within and near the complementarity determining regions. *Proteins: Structure, Function, and Bioinformatics* 79(9):2637–2647.

[154] Perchiacca JM, Ladiwala ARA, Bhattacharya M, Tessier PM (2012) Aggregation-resistant domain antibodies engineered with charged mutations near the edges of the complementarity-determining regions. *Protein Engineering Design and Selection* 25(10):591–602.

[155] Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJ, McFarlane HT, Madsen A, Riekel C, Eisenberg D (2007) Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* 447(7143):453–457.

[156] Soragni A, Janzen DM, Johnson LM, Lindgren AG, Nguyen AT, Tiourin E, Soriaga AB, Lu J, Jiang L, Faull KF, Pellegrini M, Memarzedeh S, Eisenberg D (2016) A designed inhibitor of p53 aggregation rescues p53 tumor suppression in ovarian carcinomas. *Cancer Cell* 29(1):90–103.

[157] O'Donnell CW, Waldispühl J, Lis M, Halfmann R, Devadas S, Lindquist S, Berger B (2011) A method for probing the mutational landscape of amyloid structure. *Bioinformatics* 27(13):i34–i42.

[158] Bryan AW, O'Donnell CW, Menke M, Cowen LJ, Lindquist S, Berger B (2012) STITCHER: Dynamic assembly of likely amyloid and prion β-structures from secondary structure predictions. *Proteins: Structure, Function, and Bioinformatics* 80(2):410–420.

[159] Goldschmidt L, Teng PK, Riek R, Eisenberg D (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proceedings of the National Academy of Sciences of the United States of America* 107(8):3487–3492.

[160] Chiti F, Dobson CM (2009) Amyloid formation by globular proteins under native conditions. *Nature Chemical Biology* 5(1):15–22.

[161] Tartaglia GG, Cavalli A, Vendruscolo M (2007) Prediction of local structural stabilities of proteins from their amino acid sequences. *Structure* 15(2):139–143.

[162] Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS (2010) Stability of protein pharmaceuticals: an update. *Pharmaceutical Research* 27(4):544–575.

[163] Wetzel R (1994) Mutations and off-pathway aggregation of proteins. *Trends in Biotechnology* 12(5):193–198.

[164] Pechmann S, Levy ED, Tartaglia GG, Vendruscolo M (2009) Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proceedings of the National Academy of Sciences of the United States of America* 106(25):10159–10164.

[165] Neudecker P, Robustelli P, Cavalli A, Walsh P, Lundström P, Zarrine-Afsar A, Sharpe S, Vendruscolo M, Kay LE (2012) Structure of an intermediate state in protein folding and aggregation. *Science* 336(6079):362–366.

[166] Jamroz M, Kolinski A, Kmiecik S (2013) CABS-flex: server for fast simulation of protein structure fluctuations. *Nucleic Acids Research* 41(W1):W427–W431.

[167] Jamroz M, Kolinski A, Kmiecik S (2014) CABS-flex predictions of protein flexibility compared with NMR ensembles. *Bioinformatics* p. btu184.

[168] Courtois F, Schneider CP, Agrawal NJ, Trout BL (2015) Rational design of biobetters with enhanced stability. *Journal of Pharmaceutical Sciences* 104(8):2433–2440.

[169] Kayser V, Chennamsetty N, Voynov V, Forrer K, Helk B, Trout BL (2011) Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. *Biotechnology Journal* 6(1):38–44.

[170] Lauer TM, Agrawal NJ, Chennamsetty N, Egodage K, Helk B, Trout BL (2012) Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. *Journal of Pharmaceutical Sciences* 101(1):102–115.

[171] Tiwari MK, Kepp KP (2015) Modeling the aggregation propensity and toxicity of amyloid-β variants. *Journal of Alzheimer's Disease* 47(1):215–229.

[172] Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL (2011) Prediction of protein binding regions. *Proteins: Structure, Function, and Bioinformatics* 79(3):888–897.

[173] Yamniuk AP, Ditto N, Patel M, Dai J, Sejwal P, Stetsko P, Doyle ML (2013) Application of a kosmotrope-based solubility assay to multiple protein therapeutic classes indicates broad use as a high-throughput screen for protein therapeutic aggregation propensity. *Journal of Pharmaceutical Sciences* 102(8):2424–2439.

[174] Chan P, Curtis RA, Warwicker J (2013) Soluble expression of proteins correlates with a lack of positively-charged surface. *Scientific Reports* 3.

[175] Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM (2012) Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophysical Journal* 102(8):1907–1915.

[176] Warwicker J, Charonis S, Curtis RA (2014) Lysine and arginine content of proteins: computational analysis suggests a new tool for solubility design. *Molecular Pharmaceutics* 11(1):294–303.

[177] de Groot NS, Pallarés I, Avilés FX, Vendrell J, Ventura S (2005) Prediction of "hot spots" of aggregation in disease-linked polypeptides. *BMC Structural Biology* 5(1):18.

[178] Garbuzynskiy SO, Lobanov MY, Galzitskaya OV (2010) FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* 26(3):326–332.

[179] Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* 18(20):6097–6100.

[180] Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Research* 14(6):1188–1190.

[181] Garbuzynskiy SO, Lobanov MY, Galzitskaya OV (2010) FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* 26(3):326–332.

[182] Broom RA (2016) Ph.D. thesis: Computational design of protein structure and prediction of ligand binding (University of Waterloo).

[183] Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH (2017) DNA sequencing at 40: past, present and future. *Nature* 550(7676):345.

[184] Merkl R, Sterner R (2016) Ancestral protein reconstruction: techniques and applications. *Biological Chemistry* 397(1):1–21.

[185] Porebski BT, Buckle AM (2016) Consensus protein design. *Protein Engineering, Design and Selection* 29(7):245–251.

[186] Jacobs SA, Diem MD, Luo J, Teplyakov A, Obmolova G, Malia T, Gilliland GL, O'Neil KT (2012) Design of novel FN3 domains with high stability by a consensus sequence approach. *Protein Engineering Design and Selection* p. gzr064.

[187] Porebski BT, Nickson AA, Hoke DE, Hunter MR, Zhu L, McGowan S, Webb GI, Buckle AM (2015) Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain. *Protein Engineering, Design & Selection* 28(3):67–78.

[188] Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, , et al. (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Research* 43(D1):D222–D226.

[189] Singh P, Carraher C, Schwarzbauer JE (2010) Assembly of fibronectin extracellular matrix. *Annual Review of Cell and Developmental Biology* 26:397.

[190] Krishna MM, Hoang L, Lin Y, Englander SW (2004) Hydrogen exchange methods to study protein folding. *Methods* 34(1):51–64.

[191] Hvidt A, Nielsen SO (1966) Hydrogen exchange in proteins in *Advances in Protein Chemistry*. (Elsevier) Vol. 21, pp. 287–386.

[192] Main AL, Harvey TS, Baron M, Boyd J, Campbell ID (1992) The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell* 71(4):671–678.

[193] Dickinson CD, Veerapandian B, Dai XP, Hamlin RC, Xuong NH, Ruoslahti E, Ely KR (1994) Crystal structure of the tenth type III cell adhesion module of human fibronectin. *Journal of Molecular Biology* 236(4):1079–1092.

[194] Leahy DJ, Aukhil I, Erickson HP (1996) 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* 84(1):155–164.

[195] Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society* 127(22):8014–8015.

[196] Glasoe PK, Long F (1960) Use of glass electrodes to measure acidities in deuterium oxide[1, 2]. *The Journal of Physical Chemistry* 64(1):188–190.

[197] Rubinson KA (2017) Practical corrections for p(H,D) measurements in mixed h_2o/d_2o biological buffers. *Analytical Methods* 9(18):2744–2750.

[198] Helmus JJ, Jaroniec CP (2013) Nmrglue: an open source python package for the analysis of multidimensional NMR data. *Journal of Biomolecular NMR* 55(4):355–367.

[199] Linderstrøm-Lang K (1958) Deuterium exchange and protein structure in *Symposium on Protein Structure*. (Methuen London), pp. 23–34.

[200] Bai Y, Milne JS, Mayne L, Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, and Bioinformatics* 17(1):75–86.

[201] Zhang YZ (1995) Ph.D. thesis: Protein and peptide structure and interactions studied by hydrogen exchange and NMR (University of Pennsylvania).

[202] Malvern Instruments (2004) *Zetasizer Nano Series User Manual*.

[203] Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server in *The Proteomics Protocols Handbook*. (Springer), pp. 571–607.

[204] Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts. *Journal of Biomolecular NMR* 26(3):215–240.

[205] Wishart DS (2011) Interpreting protein chemical shift data. *Progress in Nuclear Magnetic Resonance Spectroscopy* 58(1-2):62.

[206] Palmer III AG, Cavanagh J, Wright PE, Rance M (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance (1969)* 93(1):151–170.

[207] Kay L, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society* 114(26):10663–10665.

[208] Grzesiek S, Bax A (1993) The importance of not saturating water in protein NMR. application to sensitivity enhancement and NOE measurements. *Journal of the American Chemical Society* 115(26):12593–12594.

[209] Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletzky O, Glaser SJ, Sørensen OW, Griesinger C (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *Journal of Biomolecular NMR* 4(2):301–306.

[210] Davis AL, Keeler J, Laue ED, Moskau D (1992) Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients. *Journal of Magnetic Resonance* 98:207–216.

[211] Keller R (2004) The computer aided resonance assignment tutorial.

[212] Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *Journal of Biomolecular NMR* 6(2):135–140.

[213] Shimizu A, Ikeguchi M, Sugai S (1994) Appropriateness of DSS and TSP as internal references for 1H NMR studies of molten globule proteins in aqueous media. *Journal of Biomolecular NMR* 4(6):859–862.

[214] Hoffman RE (2006) Standardization of chemical shifts of TMS and solvent signals in NMR solvents. *Magnetic Resonance in Chemistry* 44(6):606–616.

[215] Cross BP, Schleich T (1977) Temperature dependence of the chemical shifts of commonly employed proton NMR reference compounds. *Organic Magnetic Resonance* 10(1):82–85.

[216] DeGroot MH, Schervish MJ (2012) *Probability and statistics*. (Pearson Education).

[217] Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. *Journal of Biomolecular NMR* 50(1):43.

[218] Cordier F, Grzesiek S (2002) Temperature-dependence of protein hydrogen bond properties as studied by high-resolution NMR. *Journal of Molecular Biology* 317(5):739–752.

[219] Barfield M (2002) Structural dependencies of interresidue scalar coupling $^{h3}J_{NC}$ and donor ^{1}H chemical shifts in the hydrogen bonding regions of proteins. *Journal of the American Chemical Society* 124(15):4158–4168.

[220] Sass HJ, Schmid FFF, Grzesiek S (2007) Correlation of protein structure and dynamics to scalar couplings across hydrogen bonds. *Journal of the American Chemical Society* 129(18):5898–5903.

[221] Huang J, MacKerell Jr AD (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* 34(25):2135–2145.

[222] Markwick PR, Cervantes CF, Abel BL, Komives EA, Blackledge M, McCammon JA (2010) Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. *Journal of the American Chemical Society* 132(4):1220–1221.

[223] Robustelli P, Stafford KA, Palmer III AG (2012) Interpreting protein structural dynamics from NMR chemical shifts. *Journal of the American Chemical Society* 134(14):6365–6374.

[224] Dingley AJ, Grzesiek S (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^{2}J_{NN}$ couplings. *Journal of the American Chemical Society* 120(33):8293–8297.

[225] Pervushin K, Ono A, Fernández C, Szyperski T, Kainosho M, Wüthrich K (1998) NMR scalar couplings across Watson–Crick base pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* 95(24):14147–14151.

[226] Blake PR, Park JB, Adams MW, Summers MF (1992) Novel observation of NH–S (cys) hydrogen-bond-mediated scalar coupling in cadmium-113 substituted rubredoxin from Pyrococcus furiosus. *Journal of the American Chemical Society* 114(12):4931–4933.

[227] Blake PR, Lee B, Summers MF, Adams MW, Park JB, Zhou ZH, Bax A (1992) Quantitative measurement of small through-hydrogen-bond and through-space1 H-113 Cd and 1 H-199 Hg J couplings in metal-substituted rubredoxin from Pyrococcus furiosus. *Journal of Biomolecular NMR* 2(5):527–533.

[228] Cordier F, Grzesiek S (1999) Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC}$ scalar couplings. *Journal of the American Chemical Society* 121(7):1601–1602.

[229] Cornilescu G, Hu JS, Bax A (1999) Identification of the hydrogen bonding network in a protein by scalar couplings. *Journal of the American Chemical Society* 121(12):2949–2950.

[230] Dingley AJ, Cordier F, Grzesiek S (2001) An introduction to hydrogen bond scalar couplings. *Concepts in Magnetic Resonance: An Educational Journal* 13(2):103–127.

[231] Cordier F, Nisius L, Dingley AJ, Grzesiek S (2008) Direct detection of N-H$\cdots$O=C hydrogen bonds in biomolecules by NMR spectroscopy. *Nature Protocols* 3(2):235.

[232] Wang YX, Jacob J, Cordier F, Wingfield P, Stahl SJ, Lee-Huang S, Torchia D, Grzesiek S, Bax A (1999) Measurement of $^{3h}J_{NC}$ connectivities across hydrogen bonds in a 30 kDa protein. *Journal of Biomolecular NMR* 14(2):181–184.

[233] Cornilescu G, Ramirez BE, Frank MK, Clore GM, Gronenborn AM, Bax A (1999) Correlation between $^{3h}j_{NC}$ and hydrogen bond length in proteins. *Journal of the American Chemical Society* 121(26):6275–6279.

[234] Scheurer C, Brüschweiler R (1999) Quantum-chemical characterization of nuclear spin- spin couplings across hydrogen bonds. *Journal of the American Chemical Society* 121(37):8661–8662.

[235] Bagno A (2000) Quantum chemical modeling of through-hydrogen bond spin-spin coupling in amides and ubiquitin. *Chemistry–A European Journal* 6(16):2925–2930.

[236] Markwick PR, Sprangers R, Sattler M (2003) Dynamic effects on J-couplings across hydrogen bonds in proteins. *Journal of the American Chemical Society* 125(3):644–645.

[237] Hong J, Jing Q, Yao L (2013) The protein amide $^{1}H^{N}$ chemical shift temperature coefficient reflects thermal expansion of the N-H$\cdots$O=C hydrogen bond. *Journal of Biomolecular NMR* 55(1):71.

[238] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J,

Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery, Jr. JA, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ (2016) Gaussian16 Revision B.01. Gaussian Inc. Wallingford CT.

[239] Pople J, McIver Jr J, Ostlund N (1968) Self-consistent perturbation theory. I. Finite perturbation methods. *The Journal of Chemical Physics* 49(7):2960–2964.

[240] Ditchfield R (1974) Self-consistent perturbation theory of diamagnetism: I. a gauge-invariant LCAO method for NMR chemical shifts. *Molecular Physics* 27(4):789–807.

[241] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26(16):1781–1802.

[242] Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, MacKerell Jr AD (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *Journal of Chemical Theory and Computation* 8(9):3257–3273.

[243] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79(2):926–935.

[244] Baxter NJ, Williamson MP (1997) Temperature dependence of 1H chemical shifts in proteins. *Journal of Biomolecular NMR* 9(4):359–369.

[245] Cierpicki T, Otlewski J (2001) Amide proton temperature coefficients as hydrogen bond indicators in proteins. *Journal of Biomolecular NMR* 21(3):249–261.

[246] Cierpicki T, Zhukov I, Byrd RA, Otlewski J (2002) Hydrogen bonds in human ubiquitin reflected in temperature coefficients of amide protons. *Journal of Magnetic Resonance* 157(2):178–180.

[247] Tomlinson JH, Williamson MP (2012) Amide temperature coefficients in the protein G B1 domain. *Journal of Biomolecular NMR* 52(1):57–64.

[248] Andersen NH, Chen C, Marschner TM, Krystek Jr SR, Bassolino DA (1992) Conformational isomerism of endothelin in acidic aqueous media: a quantitative NOESY analysis. *Biochemistry* 31(5):1280–1295.

[249] Andersen NH, Neidigh JW, Harris SM, Lee GM, Liu Z, Tong H (1997) Extracting information from the temperature gradients of polypeptide NH chemical shifts. 1. the importance of conformational averaging. *Journal of the American Chemical Society* 119(36):8547–8561.

[250] Bouvignies G, Vallurupalli P, Cordes MH, Hansen DF, Kay LE (2011) Measuring 1HN temperature coefficients in invisible protein states by relaxation dispersion NMR spectroscopy. *Journal of Biomolecular NMR* 50(1):13–18.

[251] Doyle CM, Rumfeldt JA, Broom HR, Sekhar A, Kay LE, Meiering EM (2016) Concurrent increases and decreases in local stability and conformational heterogeneity in Cu, Zn superoxide dismutase variants revealed by temperature-dependence of amide chemical shifts. *Biochemistry* 55(9):1346–1361.

[252] Baxter NJ, Hosszu LL, Waltho JP, Williamson MP (1998) Characterisation of low free-energy excited states of folded proteins. *Journal of Molecular Biology* 284(5):1625–1639.

[253] Williamson MP (2003) Many residues in cytochrome c populate alternative states under equilibrium conditions. *Proteins: Structure, Function, and Bioinformatics* 53(3):731–739.

[254] Krishna Mohan P, Mukherjee S, Chary KV (2008) Differential native state ruggedness of the two Ca2+-binding domains in a Ca2+ sensor protein. *Proteins: Structure, Function, and Bioinformatics* 70(4):1147–1153.

[255] Chandra K, Sharma Y, Chary K (2011) Characterization of low-energy excited states in the native state ensemble of non-myristoylated and myristoylated neuronal calcium sensor-1. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1814(2):334–344.

[256] Bari KJ, Sharma S, Chary KV (2019) On identifying low energy conformational excited states with differential ruggedness in human γs-crystallin promoting severe infantile cataracts. *Biochemical and Biophysical Research Communications* 514(3):946–952.

[257] Okazaki H, Matsuo N, Tenno T, Goda N, Shigemitsu Y, Ota M, Hiroaki H (2018) Using $^{1}h^{N}$ amide temperature coefficients to define intrinsically disordered regions: An alternative NMR method. *Protein Science*.

[258] Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS (2003) VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Research* 31(13):3316–3319.

[259] Tilton Jr RF, Dewan JC, Petsko GA (1992) Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320K. *Biochemistry* 31(9):2469–2481.

[260] Young AC, Tilton RF, Dewan JC (1994) Thermal expansion of hen egg-white lysozyme: comparison of the 1· 9 å resolution structures of the tetragonal form of the enzyme at 100 K and 298 K. *Journal of Molecular Biology* 235(1):302–317.

[261] Nisius L, Grzesiek S (2012) Key stabilizing elements of protein structure identified through pressure and temperature perturbation of its hydrogen bond network. *Nature Chemistry* 4(9):711.

[262] Keeler J (2011) *Understanding NMR spectroscopy*. (John Wiley & Sons).

[263] Privalov PL, Gill SJ (1988) Stability of protein structure and hydrophobic interaction in *Advances in Protein Chemistry*. (Elsevier) Vol. 39, pp. 191–234.

[264] Prabhu NV, Sharp KA (2005) Heat capacity in proteins. *Annual Reviews of Physical Chemistry* 56:521–548.

[265] Akaike H (1974) A new look at the statistical model identification in *Selected Papers of Hirotugu Akaike*. (Springer), pp. 215–222.

[266] Berjanskii MV, Wishart DS (2017) Unraveling the meaning of chemical shifts in protein NMR. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1865(11):1564–1576.

[267] Suárez IP, Burdisso P, Benoit MP, Boisbouvier J, Rasia RM (2015) Induced folding in RNA recognition by arabidopsis thaliana DCL1. *Nucleic Acids Research* 43(13):6607–6619.

[268] Suárez IP, Gauto DF, Hails G, Mascali FC, Crespo R, Zhao L, Wang J, Rasia RM (2018) Conformational sampling of the intrinsically disordered dsRBD-1 domain from arabidopsis thaliana DCL1. *Physical Chemistry Chemical Physics* 20(16):11237–11246.

[269] Lam SL, Hsu VL (2003) NMR identification of left-handed polyproline type II helices. *Biopolymers* 69(2):270–281.

[270] Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. investigations of nearest-neighbor effects. *Journal of Biomolecular NMR* 5(1):67–81.

[271] Wang Y, Jardetzky O (2002) Investigation of the neighboring residue effects on protein chemical shifts. *Journal of the American Chemical Society* 124(47):14075–14084.

[272] Wang L, Markley JL (2009) Empirical correlation between protein backbone ^{15}N and ^{13}C secondary chemical shifts and its application to nitrogen chemical shift re-referencing. *Journal of Biomolecular NMR* 44(2):95–99.

[273] Wang Y, Jardetzky O (2004) Predicting ^{15}N chemical shifts in proteins using the preceding residue-specific individual shielding surfaces from φ, ψ_{i-1}, and χ_1 torsion angles. *Journal of Biomolecular NMR* 28(4):327–340.

[274] De Dios AC, Pearson JG, Oldfield E (1993) Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. *Science* 260(5113):1491–1496.

[275] Glushka J, Lee M, Coffin S, Cowburn D (1989) Nitrogen-15 chemical shifts of backbone amides in bovine pancreatic trypsin inhibitor and apamin. *Journal of the American Chemical Society* 111(20):7716–7722.

[276] Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ (2001) Sequence-dependent correction of random coil NMR chemical shifts. *Journal of the American Chemical Society* 123(13):2970–2978.

[277] Kjaergaard M, Brander S, Poulsen FM (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *Journal of Biomolecular NMR* 49(2):139–149.

[278] Frank A, Onila I, Möller HM, Exner TE (2011) Toward the quantum chemical calculation of nuclear magnetic resonance chemical shifts of proteins. *Proteins: Structure, Function, and Bioinformatics* 79(7):2189–2202.

[279] Frank A, Möller HM, Exner TE (2012) Toward the quantum chemical calculation of NMR chemical shifts of proteins. 2. level of theory, basis set, and solvents model dependence. *Journal of Chemical Theory and Computation* 8(4):1480–1492.

[280] Zhu T, He X, Zhang JZ (2012) Fragment density functional theory calculation of NMR chemical shifts for proteins with implicit solvation. *Physical Chemistry Chemical Physics* 14(21):7837–7845.

[281] Zhu T, Zhang JZ, He X (2013) Automated fragmentation QM/MM calculation of amide proton chemical shifts in proteins with explicit solvent model. *Journal of Chemical Theory and Computation* 9(4):2104–2114.

[282] Xu XP, Case DA (2001) Automated prediction of ^{15}n, ^{13}c$^{\alpha}$, ^{13}C$^{\beta}$ and ^{13}C chemical shifts in proteins using a density functional database. *Journal of Biomolecular NMR* 21(4):321–333.

[283] Xu XP, Case DA (2002) Probing multiple effects on ^{15}N, ^{13}Cα, ^{13}Cβ, and ^{13}C chemical shifts in peptides using density functional theory. *Biopolymers: Original Research on Biomolecules* 65(6):408–423.

[284] Christensen AS, Hamelryck T, Jensen JH (2014) FragBuilder: an efficient Python library to setup quantum chemistry calculations on peptides models. *PeerJ* 2:e277.

[285] Halgren TA (1996) Merck molecular force field. I. basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* 17(5-6):490–519.

[286] Alei Jr M, Florin AE, Litchman WM, O'Brien JF (1971) A study of Nitrogen-15 nuclear magnetic resonance shifts in pure methylamines and pure $ch_3c^{15}n$. *The Journal of Physical Chemistry* 75(7):932–938.

[287] Dunbrack Jr RL, Cohen FE (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science* 6(8):1661–1681.

[288] Tamiola K, Acar B, Mulder FA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *Journal of the American Chemical Society* 132(51):18000–18003.

[289] Nielsen JT, Mulder FA (2018) POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. *Journal of Biomolecular NMR* 70(3):141–165.

[290] Braun D, Wider G, Wuethrich K (1994) Sequence-corrected ^{15}N "random coil" chemical shifts. *Journal of the American Chemical Society* 116(19):8466–8469.

[291] Berg J, Tymoczko J, Stryer L (2012) *Biochemistry*. (WH Freeman, Basingstoke).

[292] Dunbrack Jr RL (2002) Rotamer libraries in the 21st century. *Current Opinion in Structural Biology* 12(4):431–440.

[293] Brüschweiler R, Case D (1994) Adding harmonic motion to the Karplus relation for spin-spin coupling. *Journal of the American Chemical Society* 116(24):11199–11200.

[294] Mulder FA (2009) Leucine side-chain conformation and dynamics in proteins from 13C NMR chemical shifts. *ChemBioChem* 10(9):1477–1479.

[295] Baldwin AJ, Kay LE (2009) NMR spectroscopy brings invisible protein states into focus. *Nature Chemical Biology* 5(11):808–814.

[296] Sekhar A, Kay LE (2013) NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proceedings of the National Academy of Sciences of the United States of America* 110(32):12867–12874.

[297] Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC (2005) Molecular basis for amyloid fibril formation and stability. *Proceedings of the National Academy of Sciences of the United States of America* 102(2):315–320.

[298] Jaroniec CP, MacPhee CE, Bajaj VS, McMahon MT, Dobson CM, Griffin RG (2004) High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* 101(3):711–716.

[299] Budyak IL, Krishnan B, Marcelino-Cruz AM, Ferrolino MC, Zhuravleva A, Gierasch LM (2013) Early folding events protect aggregation-prone regions of a β-rich protein. *Structure* 21(3):476–485.

[300] Karplus M (2014) Development of multiscale models for complex chemical systems: from H+H2 to biomolecules (Nobel lecture). *Angewandte Chemie International Edition* 53(38):9992–10005.

[301] Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition* 48(7):1198–1229.

[302] Bernardi RC, Melo MC, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1850(5):872–877.

APPENDICES

Appendix A

Solubility Prediction: Supplemental Information

A.1 Full Amino Acid Sequences of IGF1R-binding Adnectins

Table A.1: Full amino acid sequences (as encoded by the DNA constructs cloned into pET-9d expression vectors) of the IGF1R-binding Adnectins in Table 4.1.

Seq. #[1]	Amino Acid Sequence[2]
1	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTGSRDYEPISINYRTEIDKPSQHHHHHH
2	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTKMRDYRPISINYRTEIDKPSQHHHHHH
3	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTRSRDYRPISINYRTEIDKPSQHHHHHH
4	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTERRDYRPISINYRTEIDKPSQHHHHHH
5	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KGKYTATISGLKPGVDYTITVYAVTSLRDYGPISINYRTEIDKPSQHHHHHH
6	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTTQRDYGPISINYRTEIDKPSQHHHHHH
7	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTDTRDYRPISINYRTEIDKPSQHHHHHH
8	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTEMRDYGPISINYRTEIDKPSQHHHHHH
9	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTNLRDYGPISINYRTEIDKPSQHHHHHH

Seq. #[1]	Amino Acid Sequence[2]
10	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTCRRDYGPISINYRTEIDKPSQHHHHHH
11	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTKVRDYRPISINYRTEIDKPSQHHHHHH
12	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTSRRDYGPISINYRTEIDKPSQHHHHHH
13	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTEKNQRGPISINYRTEIDKPSQHHHHHH
14	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTSLRDYAPISINYRTEIDKPSQHHHHHH
15	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTRLRDYEPISINYRTEIDKPSQHHHHHH
16	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTRIRDYGPISINYRTEIDKPSQHHHHHH
17	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTHFRDYGPISINYRTEIDKPSQHHHHHH
18	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTSLRDYVPISINYRTEIDKPSQHHHHHH
19	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTNYDNDRPISINYRTEIDKPSQHHHHHH
20	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KDRYTATISGLKPGVDYTITVYAVTMSRDYGPISINYRTEIDKPSQHHHHHH
21	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTDYRDYLPISINYRTEIDKPSQHHHHHH
22	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTVLRDYRPISINYRTEIDKPSQHHHHHH
23	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTKLRDYLPISINYRTEIDKPSQHHHHHH
24	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTLLRDYGPISINYRTEIDKPSQHHHHHH
25	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTLVRDYGPISINYRTEIDKPSQHHHHHH
26	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTLFRDYGPISINYRTEIDKPSQHHHHHH
27	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTVCRDYRPISINYRTEIDKPSQHHHHHH
28	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTTLRDYMPISINYRTEIDKPSQHHHHHH
29	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTALRDYVPISINYRTEIDKPSQHHHHHH
30	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTQLRDYSPISINYRTEIDKPSQHHHHHH

Seq. #[1]	Amino Acid Sequence[2]
31	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTTWRDYLPISINYRTEIDKPSQHHHHHH
32	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTYLRDYTPISINYRTEIDKPSQHHHHHH
33	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTFIRDYGPISINYRTEIDKPSQHHHHHH
34	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTLIRDYGPISINYRTEIDKPSQHHHHHH
35	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTFVRDYFPISINYRTEIDKPSQHHHHHH
36	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTLLRDYVPISINYRTEIDKPSQHHHHHH
37	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTFQRDYQPISINYRTEIDKPSQHHHHHH
38	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTMFRNYGPISINYRTEIDKPSQHHHHHH
39	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTILRDYVPISINYRTEIDKPSQHHHHHH
40	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTLIRNYGPISINYRTEIDKPSQHHHHHH
41	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTMFWDYGPISINYRTEIDKPSQHHHHHH

[1]Numbering matches Table 4.1
[2]Single letter amino acid code; FG loop highlighted

A.2 Adnectin Aggregation/Solubility Predictions

Table A.2: Adnectin aggregation/solubility predictions

			Aggregation/Solubility Scores (AU)		
Adnectin #	**FG Loop[1]**	**% IB, 2 h[2]**	**AGGRESCAN**	**FoldAmyloid**	**CamSol**
1	GSRDYE	15	-1.4	20.91	0.438
2	KMRDYR	17	-0.5	21.01	0.469
3	RSRDYR	17	-1.9	20.98	0.532
4	ERRDYR	20	-3.0	20.98	0.588
5	SLRDYG	20	-0.4	20.91	0.439
6	TQRDYG	21	-1.1	20.94	0.285
7	DTRDYR	22	-2.4	20.96	0.475
8	EMRDYG	22	-0.3	20.98	0.383
9	NLRDYG	26	0.3	20.99	0.305
10	CRRDYG	28	-0.4	21.00	0.306

Table A.2: Adnectin aggregation/solubility predictions

			Aggregation/Solubility Scores (AU)		
Adnectin #	**FG Loop[1]**	**% IB, 2 h[2]**	**AGGRESCAN**	**FoldAmyloid**	**CamSol**
11	KVRDYR	29	0.2	21.01	0.434
12	SRRDYG	29	-1.3	20.95	0.404
13	EKNQRG	29	-3.8	20.85	0.534
14	SLRDYA	29	1.7	21.01	0.247
15	RLRDYE	36	-0.5	21.02	0.458
16	RIRDYG	36	0.8	21.02	0.361
17	HFRDYG	38	0.9	21.04	0.215
18	SLRDYV	40	3.3	21.05	0.161
19	NYDNDR	42	-3.5	20.93	0.432
20	MSRDYG	46	-2.4	20.94	0.437
21	DYRDYL	46	1.4	21.06	0.267
22	VLRDYR	47	2.4	21.08	0.216
23	KLRDYL	47	2.5	21.06	0.272
24	LLRDYG	47	2.8	21.06	0.195
25	LVRDYG	51	3.0	21.04	0.153
26	LFRDYG	52	3.2	21.07	0.141
27	VCRDYR	54	1.6	21.06	0.184
28	TLRDYM	55	2.7	21.08	0.171
29	ALRDYV	57	3.5	21.07	0.134
30	QLRDYS	57	0.6	21.01	0.265
31	TWRDYL	61	2.9	21.11	0.108
32	YLRDYT	62	3.0	21.09	0.118
33	FIRDYG	63	3.6	21.08	0.094
34	LIRDYG	66	3.3	21.06	0.172
35	FVRDYF	66	5.6	21.16	-0.143
36	LLRDYV	68	4.9	21.12	0.061
37	FQRDYQ	69	0.0	21.03	0.178
38	MFRNYG	72	3.2	21.08	0.138
39	ILRDYV	72	5.3	21.12	0.022
40	LIRNYG	74	3.8	21.07	0.149
41	MFWDYG	91	4.9	21.14	-0.112
	R^2 with % IB, 2 h:		0.59	0.61	0.71

[1]Loop sequences colored according to the scheme: blue for R, K; red for D, E; yellow for polar, and green for hydrophobic residues

[2]Percentage of Adnectin found in inclusion bodies 2 h post-induction (Chapter 4)

Appendix B

Consensus Protein Design for Solubility: Supplemental Information

B.1 β-strand F Substitutions

Table B.1: Diverse β-strand F amino acid sequences from the cd00063 Fn3 domain model substituted into the wild-type ^{10}Fn3 sequence

Seq. #[1]	Amino Acid Sequence[2]
1	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE FTVPKNVYTATISGLKPGVGLEDIVNNETGRGESPASSKPISINYRT
2	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE FTVPKNVYTATISGLKPGVDYEIRVRSRTGRGESPASSKPISINYRT
3	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE FTVPKNVYTATISGLKPGVEYEVRVRTRTGRGESPASSKPISINYRT
4	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE FTVPKNVYTATISGLKPGVRYRLQLRARTGRGESPASSKPISINYRT
5	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE FTVPKNVYTATISGLKPGVEYQFRVRAETGRGESPASSKPISINYRT
6	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE FTVPKNVYTATISGLKPGVEYQFRVRAETGRGESPASSKPISINYRT
7	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE FTVPKNVYTATISGLKPGVGHKVKVRAGTGRGESPASSKPISINYRT
8	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE FTVPKNVYTATISGLKPGVRYRFRVRAETGRGESPASSKPISINYRT
9	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE FTVPKNVYTATISGLKPGVQYRFRVRAETGRGESPASSKPISINYRT
10	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVEYQFRVKAETGRGESPASSKPISINYRT
11	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYRFRIRAETGRGESPASSKPISINYRT
12	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQFRVKAQTGRGESPASSKPISINYRT
13	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEVQFQVRAETGRGESPASSKPISINYRT
14	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQFRIKAETGRGESPASSKPISINYRT
15	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVKAKTGRGESPASSKPISINYRT
16	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRHVVQLRAQTGRGESPASSKPISINYRT
17	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVKCETGRGESPASSKPISINYRT
18	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYRFRVRADTGRGESPASSKPISINYRT
19	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYEVQVRGKTGRGESPASSKPISINYRT
20	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVSAETGRGESPASSKPISINYRT
21	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYDLQVRAHTGRGESPASSKPISINYRT
22	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRIKAKTGRGESPASSKPISINYRT
23	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYDIQVSAHTGRGESPASSKPISINYRT
24	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVSAQTGRGESPASSKPISINYRT
25	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYEFRVRAETGRGESPASSKPISINYRT
26	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYEFRVSAETGRGESPASSKPISINYRT
27	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYTFKVRSRTGRGESPASSKPISINYRT
28	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKHVVQVRGKTGRGESPASSKPISINYRT
29	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYEFRIRAKTGRGESPASSKPISINYRT
30	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVGYQIRIEAETGRGESPASSKPISINYRT
31	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVKHSVKIRAATGRGESPASSKPISINYRT
32	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYEFRIRAKTGRGESPASSKPISINYRT
33	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYNVQVAAETGRGESPASSKPISINYRT
34	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYEIRVQAETGRGESPASSKPISINYRT
35	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSFRVRAQTGRGESPASSKPISINYRT
36	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYQFRVRAETGRGESPASSKPISINYRT
37	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPHAVRVSARTGRGESPASSKPISINYRT
38	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYDVKVRAHTGRGESPASSKPISINYRT
39	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYHFRVSAETGRGESPASSKPISINYRT
40	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQFRVEANTGRGESPASSKPISINYRT
41	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFNVEAQTGRGESPASSKPISINYRT
42	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQWNFEVQAATGRGESPASSKPISINYRT
43	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYRFEVQASTGRGESPASSKPISINYRT
44	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKFRVRAKTGRGESPASSKPISINYRT
45	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYRFQLQATTGRGESPASSKPISINYRT
46	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYALAIEADTGRGESPASSKPISINYRT
47	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFQISSKTGRGESPASSKPISINYRT
48	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYQMKMTATTGRGESPASSKPISINYRT
49	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYKFDIQTTTGRGESPASSKPISINYRT
50	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEHRVTAITGRGESPASSKPISINYRT
51	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYNARLQAMTGRGESPASSKPISINYRT
52	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVEYQFRISARTGRGESPASSKPISINYRT
53	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQLQVRAKTGRGESPASSKPISINYRT
54	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEFRVRACTGRGESPASSKPISINYRT
55	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTVHVRAHTGRGESPASSKPISINYRT
56	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYRFRVKAETGRGESPASSKPISINYRT
57	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYDVKIRAHTGRGESPASSKPISINYRT
58	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYEIKARPFTGRGESPASSKPISINYRT
59	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVGYDINIKGTTGRGESPASSKPISINYRT
60	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEIHLFKETGRGESPASSKPISINYRT
61	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEYRVYAETGRGESPASSKPISINYRT
62	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSFRVSSETGRGESPASSKPISINYRT
63	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYNFRFAAQTGRGESPASSKPISINYRT
64	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVHYKVEVRATTGRGESPASSKPISINYRT
65	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVAGSTGRGESPASSKPISINYRT
66	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKFRVRAATGRGESPASSKPISINYRT
67	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPAEVHIVRNTGRGESPASSKPISINYRT
68	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYRVQLWGRTGRGESPASSKPISINYRT
69	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYRARVRAWTGRGESPASSKPISINYRT
70	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYQFRAIARTGRGESPASSKPISINYRT
71	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPHELKICASTGRGESPASSKPISINYRT
72	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVCAQTGRGESPASSKPISINYRT
73	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVPYKFKVQARTGRGESPASSKPISINYRT
74	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYRFRIKAATGRGESPASSKPISINYRT
75	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYRLALRASTGRGESPASSKPISINYRT
76	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYTVQVRADTGRGESPASSKPISINYRT
77	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYEMTIRSATGRGESPASSKPISINYRT
78	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTAHLHGLTGRGESPASSKPISINYRT
79	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVMAETGRGESPASSKPISINYRT
80	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYQFRVYAETGRGESPASSKPISINYRT
81	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYRFSLSARTGRGESPASSKPISINYRT
82	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYDFQVSAHTGRGESPASSKPISINYRT
83	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTFRVSAETGRGESPASSKPISINYRT
84	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSFRVSAQTGRGESPASSKPISINYRT
85	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQFRICAETGRGESPASSKPISINYRT
86	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTFRVSAETGRGESPASSKPISINYRT
87	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYRYTVKARTGRGESPASSKPISINYRT
88	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKFRVAARTGRGESPASSKPISINYRT
89	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFHVMAETGRGESPASSKPISINYRT
90	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYRFRVYSETGRGESPASSKPISINYRT
91	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYTYRVKANTGRGESPASSKPISINYRT
92	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRIMAETGRGESPASSKPISINYRT
93	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFTVEARTGRGESPASSKPISINYRT
94	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVAYKVRVGAETGRGESPASSKPISINYRT
95	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFKIEACTGRGESPASSKPISINYRT
96	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYTFHLSARTGRGESPASSKPISINYRT
97	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYMARVRCATGRGESPASSKPISINYRT
98	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYRIRVSASTGRGESPASSKPISINYRT
99	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYSVRFRATTGRGESPASSKPISINYRT
100	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYWARVRVRTGRGESPASSKPISINYRT
101	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYSVRIRARTGRGESPASSKPISINYRT
102	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQFKIQVATRTGRGESPASSKPISINYRT
103	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFTVQARTGRGESPASSKPISINYRT
104	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQMRVIARTGRGESPASSKPISINYRT
105	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYRVWVRAHTGRGESPASSKPISINYRT
106	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYNVELQAVTGRGESPASSKPISINYRT
107	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYWLQLRSETGRGESPASSKPISINYRT
108	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSFRFAARTGRGESPASSKPISINYRT
109	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYTFAVRARTGRGESPASSKPISINYRT
110	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYRITVTAHTGRGESPASSKPISINYRT
111	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYLFRVRAETGRGESPASSKPISINYRT
112	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYNISVSCKTGRGESPASSKPISINYRT
113	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYEVRVRARTGRGESPASSKPISINYRT
114	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYSLRVRCATGRGESPASSKPISINYRT
115	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVKYRVGVAAKTGRGESPASSKPISINYRT
116	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYSVRVQCSTGRGESPASSKPISINYRT
117	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKVSVRASTGRGESPASSKPISINYRT
118	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEYRLTAVTGRGESPASSKPISINYRT
119	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYRLALPSTTGRGESPASSKPISINYRT
120	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVFAETGRGESPASSKPISINYRT
121	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYCLQTEAQTGRGESPASSKPISINYRT
122	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYRFRVRAATGRGESPASSKPISINYRT
123	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTATVRGLTGRGESPASSKPISINYRT
124	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVFRVRAETGRGESPASSKPISINYRT
125	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYNVKVRAYTGRGESPASSKPISINYRT
126	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYNARVKAFTGRGESPASSKPISINYRT
127	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYRINVTEVTGRGESPASSKPISINYRT
128	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVYYRIELKAHTGRGESPASSKPISINYRT
129	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYRYQVIAETGRGESPASSKPISINYRT
130	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYEFRVRAVTGRGESPASSKPISINYRT
131	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYVVQVRARTGRGESPASSKPISINYRT
132	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYDVNVYAQTGRGESPASSKPISINYRT
133	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIFRVRAETGRGESPASSKPISINYRT
134	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIFRVRAETGRGESPASSKPISINYRT
135	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYIFRIHAETGRGESPASSKPISINYRT
136	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVEYQFRVRAVTGRGESPASSKPISINYRT
137	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYKIRVAARTGRGESPASSKPISINYRT
138	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYDVKVSAVTGRGESPASSKPISINYRT
139	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVGYDVSVAGTTGRGESPASSKPISINYRT
140	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYFFRVRAETGRGESPASSKPISINYRT
141	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVKAVTGRGESPASSKPISINYRT
142	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYEFRVFARTGRGESPASSKPISINYRT
143	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYTVRVSARTGRGESPASSKPISINYRT
144	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYRFRVLAETGRGESPASSKPISINYRT
145	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEFRVLASTGRGESPASSKPISINYRT
146	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYSVEVAASTGRGESPASSKPISINYRT
147	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYKFYFNAQTGRGESPASSKPISINYRT
148	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYVVSLRARTGRGESPASSKPISINYRT
149	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYTYTVRARTGRGESPASSKPISINYRT
150	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYSYSVKARTGRGESPASSKPISINYRT
151	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYLFRIRAQTGRGESPASSKPISINYRT
152	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYEFRVIARTGRGESPASSKPISINYRT
153	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYNVLLTAETGRGESPASSKPISINYRT
154	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQVKVQVQTGRGESPASSKPISINYRT
155	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYIVRARVNTGRGESPASSKPISINYRT
156	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYKFRVKAVTGRGESPASSKPISINYRT
157	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVKNNARILVYTGRGESPASSKPISINYRT
158	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEMKVCAYTGRGESPASSKPISINYRT
159	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYEFRIIAETGRGESPASSKPISINYRT
160	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEFTVRSATGRGESPASSKPISINYRT
161	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEISLNSVTGRGESPASSKPISINYRT
162	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEFKIAAKTGRGESPASSKPISINYRT
163	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYEFRVRAVTGRGESPASSKPISINYRT
164	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYSIDVSSCTGRGESPASSKPISINYRT
165	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQFRVKAVTGRGESPASSKPISINYRT
166	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYEISAWAKTGRGESPASSKPISINYRT
167	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEVSLISRTGRGESPASSKPISINYRT
168	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRLLFRVRAHTGRGESPASSKPISINYRT
169	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYEFRVQTKTGRGESPASSKPISINYRT
170	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQFRTYALTGRGESPASSKPISINYRT
171	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKFRVKALTGRGESPASSKPISINYRT
172	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYKLEILPGTGRGESPASSKPISINYRT
173	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYDIKVRAWTGRGESPASSKPISINYRT
174	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYLYRVSAETGRGESPASSKPISINYRT
175	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYEFRIIARTGRGESPASSKPISINYRT
176	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYKLRVSAVTGRGESPASSKPISINYRT
177	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYQIRVRAITGRGESPASSKPISINYRT
178	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVTYQFQVQSWTGRGESPASSKPISINYRT
179	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQFRIIAQTGRGESPASSKPISINYRT
180	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYSFRVRAVTGRGESPASSKPISINYRT
181	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVLTLQSHTGRGESPASSKPISINYRT
182	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYTITVEATTGRGESPASSKPISINYRT
183	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKFRVKAVTGRGESPASSKPISINYRT
184	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKFRVKAVTGRGESPASSKPISINYRT
185	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTVQLHGVTGRGESPASSKPISINYRT
186	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYTYAVRAKTGRGESPASSKPISINYRT
187	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYSIQMIAETGRGESPASSKPISINYRT
188	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYQFTVRTVTGRGESPASSKPISINYRT
189	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYQFRVTAVTGRGESPASSKPISINYRT
190	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFEIDAVTGRGESPASSKPISINYRT
191	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYTFTVRARTGRGESPASSKPISINYRT
192	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVHYVLKVRAKTGRGESPASSKPISINYRT
193	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVYYTARVRVRTGRGESPASSKPISINYRT
194	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYARILAYNTGRGESPASSKPISINYRT
195	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYRVEVHGVTGRGESPASSKPISINYRT
196	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYRFRVKAVTGRGESPASSKPISINYRT
197	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYLFRVSARTGRGESPASSKPISINYRT
198	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEVRVVAETGRGESPASSKPISINYRT
199	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVEYEFRVVASTGRGESPASSKPISINYRT
200	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYLVQVRARTGRGESPASSKPISINYRT
201	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYKFRVKAVTGRGESPASSKPISINYRT
202	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVHYTATMYATTGRGESPASSKPISINYRT
203	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYEVQVQAVTGRGESPASSKPISINYRT
204	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYDVRVAALTGRGESPASSKPISINYRT
205	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYQCRLLLGTGRGESPASSKPISINYRT
206	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQVTVRALTGRGESPASSKPISINYRT
207	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYTFRVRSRTGRGESPASSKPISINYRT
208	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYYVQVAAQTGRGESPASSKPISINYRT
209	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYEVSVIAETGRGESPASSKPISINYRT
210	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSFRLNAFTGRGESPASSKPISINYRT
211	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKFRVIAKTGRGESPASSKPISINYRT
212	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYQYRVCAVTGRGESPASSKPISINYRT
213	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYRLTVRAVTGRGESPASSKPISINYRT
214	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYKIRISATTGRGESPASSKPISINYRT
215	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYQVKVKAFTGRGESPASSKPISINYRT
216	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYTFSVRAKTGRGESPASSKPISINYRT
217	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYFFKIQARTGRGESPASSKPISINYRT
218	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYDVKVQAITGRGESPASSKPISINYRT
219	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYHVEVHGVTGRGESPASSKPISINYRT
220	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVAYRVQFVGRTGRGESPASSKPISINYRT
221	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYQMQVALCTGRGESPASSKPISINYRT
222	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYYIRVRANTGRGESPASSKPISINYRT
223	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYLFRVRASTGRGESPASSKPISINYRT
224	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYHFSLRALTGRGESPASSKPISINYRT
225	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYRFQLAARTGRGESPASSKPISINYRT
226	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYHIDVSAFTGRGESPASSKPISINYRT
227	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYHLEVQAFTGRGESPASSKPISINYRT
228	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSFRVRAVTGRGESPASSKPISINYRT
229	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYNFRVAALTGRGESPASSKPISINYRT
230	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYHFRVRGVTGRGESPASSKPISINYRT
231	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYELRVMAETGRGESPASSKPISINYRT
232	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEVTVSTLTGRGESPASSKPISINYRT
233	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTLRVHAVTGRGESPASSKPISINYRT
234	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQVKIAAMTGRGESPASSKPISINYRT
235	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYLLRIKTTTGRGESPASSKPISINYRT
236	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQFRVSAVTGRGESPASSKPISINYRT
237	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYDFSLVLDTGRGESPASSKPISINYRT
238	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYTFKAYFHTGRGESPASSKPISINYRT
239	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYELSLHGVTGRGESPASSKPISINYRT
240	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYHVSVRAYTGRGESPASSKPISINYRT
241	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVEYVFRLAARTGRGESPASSKPISINYRT
242	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYTVKLQALTGRGESPASSKPISINYRT
243	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYLFRVAAETGRGESPASSKPISINYRT
244	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFAVRAVTGRGESPASSKPISINYRT
245	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYDIKMQCFTGRGESPASSKPISINYRT
246	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYAVAIAARTGRGESPASSKPISINYRT
247	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYRISITATTGRGESPASSKPISINYRT
248	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQIGVQAITGRGESPASSKPISINYRT
249	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYYFKVQARTGRGESPASSKPISINYRT
250	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYNISISAGTGRGESPASSKPISINYRT
251	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYVFRVRAQTGRGESPASSKPISINYRT
252	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYEIGVKSVTGRGESPASSKPISINYRT
253	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEFRVFAETGRGESPASSKPISINYRT
254	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVTAITGRGESPASSKPISINYRT
255	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVGYSLWVQAHTGRGESPASSKPISINYRT
256	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYTFEVQAVTGRGESPASSKPISINYRT
257	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYEIKVQALTGRGESPASSKPISINYRT
258	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYSVQIIPRTGRGESPASSKPISINYRT
259	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVHYIFRIQAKTGRGESPASSKPISINYRT
260	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYYFRVCAETGRGESPASSKPISINYRT
261	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVFYRVRISGKTGRGESPASSKPISINYRT
262	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVSYQFRVIAETGRGESPASSKPISINYRT
263	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYQFRAMAITGRGESPASSKPISINYRT
264	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEFRVLAKTGRGESPASSKPISINYRT
265	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEFEFRIIAKTGRGESPASSKPISINYRT
266	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYVVKVQARTGRGESPASSKPISINYRT
267	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIFRVAGKTGRGESPASSKPISINYRT
268	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVYYEAIVQAKTGRGESPASSKPISINYRT
269	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVGVTVRTLSATGRGESPASSKPISINYRT
270	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVGYSVRVQLSTGRGESPASSKPISINYRT
271	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYDYRFVATTGRGESPASSKPISINYRT
272	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIFRVAAETGRGESPASSKPISINYRT
273	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQFRVRAVTGRGESPASSKPISINYRT
274	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYVFHVRARTGRGESPASSKPISINYRT
275	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYLVQIRCKTGRGESPASSKPISINYRT
276	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTVKLQAITGRGESPASSKPISINYRT
277	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYTFEIQAVTGRGESPASSKPISINYRT
278	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYTVLLQAATGRGESPASSKPISINYRT
279	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYEFRVIAETGRGESPASSKPISINYRT
280	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYRIIIRAKTGRGESPASSKPISINYRT
281	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYQFNVKTVTGRGESPASSKPISINYRT
282	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYTLRVYSRTGRGESPASSKPISINYRT
283	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVSYTFTIKAKTGRGESPASSKPISINYRT
284	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYHFEIIPRTGRGESPASSKPISINYRT
285	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYRFILRAHTGRGESPASSKPISINYRT
286	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKFRVSAVTGRGESPASSKPISINYRT
287	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYEIKVQAVTGRGESPASSKPISINYRT
288	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVGYYFRVSAQTGRGESPASSKPISINYRT
289	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYSFQVSAMTGRGESPASSKPISINYRT
290	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKHFIRVQAVTGRGESPASSKPISINYRT
291	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYEITVAASTGRGESPASSKPISINYRT
292	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYTVRVRAKTGRGESPASSKPISINYRT
293	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYSFQVTATTGRGESPASSKPISINYRT
294	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYEVEVIVETGRGESPASSKPISINYRT
295	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQFRVKALTGRGESPASSKPISINYRT
296	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRIIATTGRGESPASSKPISINYRT
297	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTVTVWAETGRGESPASSKPISINYRT
298	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKMNLYGLTGRGESPASSKPISINYRT
299	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEIELYGVTGRGESPASSKPISINYRT
300	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYLVTATANTGRGESPASSKPISINYRT
301	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKFRVSAITGRGESPASSKPISINYRT
302	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYQFRVTAQTGRGESPASSKPISINYRT
303	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKVHVRVKAITGRGESPASSKPISINYRT
304	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVTYIFRVRARTGRGESPASSKPISINYRT
305	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYVFQIRARTGRGESPASSKPISINYRT
306	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEYRVSAVTGRGESPASSKPISINYRT
307	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYNVTVSAATGRGESPASSKPISINYRT
308	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYHIRVACTTGRGESPASSKPISINYRT
309	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYQVQAYSVTGRGESPASSKPISINYRT
310	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTFRLRIRSVTGRGESPASSKPISINYRT
311	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYEVKVTAHTGRGESPASSKPISINYRT
312	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYDFRIYGLTGRGESPASSKPISINYRT
313	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYLINIAAKTGRGESPASSKPISINYRT
314	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEIHVVAHTGRGESPASSKPISINYRT
315	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKFRVAAVTGRGESPASSKPISINYRT
316	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYAFRVKAVTGRGESPASSKPISINYRT
317	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQFRVKAVTGRGESPASSKPISINYRT
318	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYSIRVMATTGRGESPASSKPISINYRT
319	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYTFTVRATTGRGESPASSKPISINYRT
320	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYNISVQTMTGRGESPASSKPISINYRT
321	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQFRVKAVTGRGESPASSKPISINYRT
322	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVYQVTATTGRGESPASSKPISINYRT
323	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYRFRVLACTGRGESPASSKPISINYRT
324	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYSVRVTAITGRGESPASSKPISINYRT
325	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVNYSLSVQAVTGRGESPASSKPISINYRT
326	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYVFQIRARTGRGESPASSKPISINYRT
327	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYCLQVQAQTGRGESPASSKPISINYRT
328	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSFVLMNRTGRGESPASSKPISINYRT
329	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYSFTIRASTGRGESPASSKPISINYRT
330	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQVRIVAKTGRGESPASSKPISINYRT
331	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYSFVLTNRTGRGESPASSKPISINYRT
332	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYEFRVIAKTGRGESPASSKPISINYRT
333	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTISLVAETGRGESPASSKPISINYRT
334	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQINVYAMTGRGESPASSKPISINYRT
335	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYVVQVRCTTGRGESPASSKPISINYRT
336	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFSVEALTGRGESPASSKPISINYRT
337	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKMNLYGFTGRGESPASSKPISINYRT
338	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPFEVKIKAYTGRGESPASSKPISINYRT
339	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYEVSVSAVTGRGESPASSKPISINYRT
340	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYNITISPETGRGESPASSKPISINYRT
341	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPFQIRVQAVTGRGESPASSKPISINYRT
342	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYLATLTVRTGRGESPASSKPISINYRT
343	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMHYFRVLPETGRGESPASSKPISINYRT
344	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEFRVAAYTGRGESPASSKPISINYRT
345	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYNFRISAITGRGESPASSKPISINYRT
346	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVEYVVTVTAETGRGESPASSKPISINYRT
347	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIIQVAAKTGRGESPASSKPISINYRT
348	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYVFRLSARTGRGESPASSKPISINYRT
349	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEIELFGITGRGESPASSKPISINYRT
350	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKICIDIPTGRGESPASSKPISINYRT
351	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVHYDLAISAFTGRGESPASSKPISINYRT
352	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYAFSVRAETGRGESPASSKPISINYRT
353	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYLFKVCARTGRGESPASSKPISINYRT
354	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKFWVTATTGRGESPASSKPISINYRT
355	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEVYVVAETGRGESPASSKPISINYRT
356	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYKLEMVVTTGRGESPASSKPISINYRT
357	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFHVKAITGRGESPASSKPISINYRT
358	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYLFSVRARTGRGESPASSKPISINYRT
359	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKFYLRACTGRGESPASSKPISINYRT
360	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYKIHLYTLTGRGESPASSKPISINYRT
361	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQVQAYTITGRGESPASSKPISINYRT
362	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYQFRVRIRTGRGESPASSKPISINYRT
363	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAQEVALVAYTGRGESPASSKPISINYRT
364	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYYFRIQARTGRGESPASSKPISINYRT
365	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVHYQFLLMARTGRGESPASSKPISINYRT
366	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYLIQVRCKTGRGESPASSKPISINYRT
367	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVAYNCSVTSFTGRGESPASSKPISINYRT
368	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVYYEVKVAAATGRGESPASSKPISINYRT
369	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYYFRVFSETGRGESPASSKPISINYRT
370	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYQVWLQAHTGRGESPASSKPISINYRT
371	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPFQIKVQAVTGRGESPASSKPISINYRT
372	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEFLVSFHTGRGESPASSKPISINYRT
373	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYVFKIQAKTGRGESPASSKPISINYRT
374	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYRVRLSVLTGRGESPASSKPISINYRT
375	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYNVNVSAITGRGESPASSKPISINYRT
376	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYFVCVRAKTGRGESPASSKPISINYRT
377	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYQFRVYAITGRGESPASSKPISINYRT
378	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYIIQIHCKTGRGESPASSKPISINYRT
379	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSVKVFASTGRGESPASSKPISINYRT
380	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVMAITGRGESPASSKPISINYRT
381	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYFFRVMAETGRGESPASSKPISINYRT
382	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIISVTAETGRGESPASSKPISINYRT
383	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEIRVLLTTGRGESPASSKPISINYRT
384	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAFEIKVQAVTGRGESPASSKPISINYRT
385	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYNFRICAITGRGESPASSKPISINYRT
386	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVYAVTGRGESPASSKPISINYRT
387	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVFRVAATTGRGESPASSKPISINYRT
388	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVPYVFRVAARTGRGESPASSKPISINYRT
389	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAFEIFIQAQTGRGESPASSKPISINYRT
390	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYVFKMRAITGRGESPASSKPISINYRT
391	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYVFRLAARTGRGESPASSKPISINYRT
392	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYTVEIFAQTGRGESPASSKPISINYRT
393	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYTVKILTQTGRGESPASSKPISINYRT
394	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKIRIYAMTGRGESPASSKPISINYRT
395	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEIRVAAVTGRGESPASSKPISINYRT
396	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVFRIMAQTGRGESPASSKPISINYRT
397	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYVMRVKAYTGRGESPASSKPISINYRT
398	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYNVNIISVTGRGESPASSKPISINYRT
399	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYLYRVKAVTGRGESPASSKPISINYRT
400	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYQLKLTVLTGRGESPASSKPISINYRT
401	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVHYTITIRGVTGRGESPASSKPISINYRT
402	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYSYVVKAKTGRGESPASSKPISINYRT
403	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYCFEIVPKTGRGESPASSKPISINYRT
404	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYGVQIKAVTGRGESPASSKPISINYRT
405	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVLRVRAVTGRGESPASSKPISINYRT
406	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEFRVVAITGRGESPASSKPISINYRT
407	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYEFRVYAETGRGESPASSKPISINYRT
408	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYEFRVKAVTGRGESPASSKPISINYRT
409	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVEYKVTITPITGRGESPASSKPISINYRT
410	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYGVGISAVTGRGESPASSKPISINYRT
411	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQLTVNVAAVTGRGESPASSKPISINYRT
412	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYQVDIFAETGRGESPASSKPISINYRT
413	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYSIQVAALTGRGESPASSKPISINYRT
414	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYAIRVQAVTGRGESPASSKPISINYRT
415	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYDFRVIAVTGRGESPASSKPISINYRT
416	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYEIKVRSITGRGESPASSKPISINYRT
417	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYGIGVTAVTGRGESPASSKPISINYRT
418	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYKFRVSAVTGRGESPASSKPISINYRT
419	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYSINITAKTGRGESPASSKPISINYRT
420	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYRISIYAATGRGESPASSKPISINYRT
421	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYRLWLEMYTGRGESPASSKPISINYRT
422	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYRLTLYTLTGRGESPASSKPISINYRT
423	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYEIKVRSITGRGESPASSKPISINYRT
424	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYQLDFITITGRGESPASSKPISINYRT
425	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKFRIKMITGRGESPASSKPISINYRT
426	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVHYRLLFTYKTGRGESPASSKPISINYRT
427	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYSFRVAAVTGRGESPASSKPISINYRT
428	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYQFRVYAVTGRGESPASSKPISINYRT
429	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYEFRIFAVTGRGESPASSKPISINYRT
430	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVTYQFFIRASTGRGESPASSKPISINYRT
431	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVYYQLDVRLYTGRGESPASSKPISINYRT
432	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYYTAVRAYTGRGESPASSKPISINYRT
433	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVFYSVKVRAATGRGESPASSKPISINYRT
434	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYKLNVRVVTGRGESPASSKPISINYRT
435	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQFRVVALTGRGESPASSKPISINYRT
436	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYEFQLKLITGRGESPASSKPISINYRT
437	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVVAVTGRGESPASSKPISINYRT
438	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYTVNVYEITGRGESPASSKPISINYRT
439	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYDISVITLTGRGESPASSKPISINYRT
440	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYSIGVSCMTGRGESPASSKPISINYRT
441	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTCSVRVAAVTGRGESPASSKPISINYRT
442	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKVEVISWTGRGESPASSKPISINYRT
443	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYRISVFATTGRGESPASSKPISINYRT
444	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYNVSLHIVTGRGESPASSKPISINYRT
445	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVFYNFVLTNRTGRGESPASSKPISINYRT
446	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYIFRLAAKTGRGESPASSKPISINYRT
447	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYTLTITAFTGRGESPASSKPISINYRT
448	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYFFRIAAETGRGESPASSKPISINYRT
449	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYSVRVTALTGRGESPASSKPISINYRT
450	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVHYTFRVTAITGRGESPASSKPISINYRT
451	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVLLLFRVRARTGRGESPASSKPISINYRT
452	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNWTFKISAITGRGESPASSKPISINYRT
453	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYKIKVTAVTGRGESPASSKPISINYRT
454	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYALTVSALTGRGESPASSKPISINYRT
455	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSLRFLAYTGRGESPASSKPISINYRT
456	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYVVQVRARTGRGESPASSKPISINYRT
457	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYRYRVFAVTGRGESPASSKPISINYRT
458	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYRISVQCVTGRGESPASSKPISINYRT
459	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVIAVTGRGESPASSKPISINYRT
460	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVFRVKSVTGRGESPASSKPISINYRT
461	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVIAVTGRGESPASSKPISINYRT
462	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYVVQVRARTGRGESPASSKPISINYRT
463	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYLVSIKVQTGRGESPASSKPISINYRT
464	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYVFRVCAETGRGESPASSKPISINYRT
465	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVIAVTGRGESPASSKPISINYRT
466	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYTFRIRCSTGRGESPASSKPISINYRT
467	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYLFRVKAVTGRGESPASSKPISINYRT
468	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFRVIAVTGRGESPASSKPISINYRT
469	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYDVLLAAETGRGESPASSKPISINYRT
470	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNVHVRLVLTTGRGESPASSKPISINYRT
471	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYIVTVTGNTGRGESPASSKPISINYRT
472	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVEYVFRIRCMTGRGESPASSKPISINYRT
473	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYNVSVYTVTGRGESPASSKPISINYRT
474	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYFFRVFAETGRGESPASSKPISINYRT
475	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYRVIVKAFTGRGESPASSKPISINYRT
476	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEVSLLAVTGRGESPASSKPISINYRT
477	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTLTVAPFTGRGESPASSKPISINYRT
478	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYAYYVKTYTGRGESPASSKPISINYRT
479	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYEVKVGVYTGRGESPASSKPISINYRT
480	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYHLTVWAETGRGESPASSKPISINYRT
481	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYWVKVTARTGRGESPASSKPISINYRT
482	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYMATVTSITGRGESPASSKPISINYRT
483	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYNIAITAVTGRGESPASSKPISINYRT
484	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIFRVRAVTGRGESPASSKPISINYRT
485	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYLIRVTAETGRGESPASSKPISINYRT
486	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVFYEFKIAAATGRGESPASSKPISINYRT
487	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYTFEVTALTGRGESPASSKPISINYRT
488	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYYFRVSGETGRGESPASSKPISINYRT
489	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTCTVRIAAVTGRGESPASSKPISINYRT
490	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVWYAYRVKTLTGRGESPASSKPISINYRT
491	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYNFTVWAETGRGESPASSKPISINYRT
492	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYSVTVTTKTGRGESPASSKPISINYRT
493	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVEYEVVLVSTTGRGESPASSKPISINYRT
494	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYVVTVTANTGRGESPASSKPISINYRT
495	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEIWVSAVTGRGESPASSKPISINYRT
496	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYKIWVLAGTGRGESPASSKPISINYRT
497	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVVSVYAQTGRGESPASSKPISINYRT
498	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYHIIVQLGTGRGESPASSKPISINYRT
499	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYVFKLRVSTGRGESPASSKPISINYRT
500	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYEGQLISVTGRGESPASSKPISINYRT
501	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYHLVLRAYTGRGESPASSKPISINYRT
502	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYTLQIRCITGRGESPASSKPISINYRT
503	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYKVNVFGMTGRGESPASSKPISINYRT
504	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYTVTITTRTGRGESPASSKPISINYRT
505	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYCVKARALTGRGESPASSKPISINYRT
506	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYTVCVMTNTGRGESPASSKPISINYRT
507	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYKMTVEAVTGRGESPASSKPISINYRT
508	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYQFRVIAVTGRGESPASSKPISINYRT
509	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTFQVMAVTGRGESPASSKPISINYRT
510	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYFVRVIAQTGRGESPASSKPISINYRT
511	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYEVSVYALTGRGESPASSKPISINYRT
512	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTVTVSAITGRGESPASSKPISINYRT
513	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYHLTVRAYTGRGESPASSKPISINYRT
514	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVLYLVNVTAETGRGESPASSKPISINYRT
515	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYAYTVRAFTGRGESPASSKPISINYRT
516	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYFVSVTATTGRGESPASSKPISINYRT
517	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYMFVIRAETGRGESPASSKPISINYRT
518	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSFRVVAYTGRGESPASSKPISINYRT
519	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTVQVVAYTGRGESPASSKPISINYRT
520	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFEVWAVTGRGESPASSKPISINYRT
521	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYYFRLAARTGRGESPASSKPISINYRT
522	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYFFRVKAVTGRGESPASSKPISINYRT
523	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYTVAMYATTGRGESPASSKPISINYRT
524	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYSFRIYALTGRGESPASSKPISINYRT
525	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYNVSVVSVTGRGESPASSKPISINYRT
526	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYAVSVIAETGRGESPASSKPISINYRT
527	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEYRVVAVTGRGESPASSKPISINYRT
528	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVGYEVMVSGFTGRGESPASSKPISINYRT
529	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYRYQVVSVTGRGESPASSKPISINYRT
530	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYYFRVSAVTGRGESPASSKPISINYRT
531	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYVLSLHAYTGRGESPASSKPISINYRT
532	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYALLVKTYTGRGESPASSKPISINYRT
533	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYSVRLNSFTGRGESPASSKPISINYRT
534	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYHLSVKAYTGRGESPASSKPISINYRT
535	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVPYAITVSSMTGRGESPASSKPISINYRT
536	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYSFWIEAVTGRGESPASSKPISINYRT
537	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYYFRVTAVTGRGESPASSKPISINYRT
538	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYQFQVAAMTGRGESPASSKPISINYRT
539	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVLEFRVFAKTGRGESPASSKPISINYRT
540	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYLVQVQALTGRGESPASSKPISINYRT
541	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYIFHIRVRTGRGESPASSKPISINYRT
542	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYTFTVRAVTGRGESPASSKPISINYRT
543	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYIFRVKAVTGRGESPASSKPISINYRT
544	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYQFSLYPITGRGESPASSKPISINYRT
545	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYVVQVRCRTGRGESPASSKPISINYRT
546	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYEVTVVSVTGRGESPASSKPISINYRT
547	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTIRLRLLLSTGRGESPASSKPISINYRT
548	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYVFRVIARTGRGESPASSKPISINYRT
549	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYVFLVRAETGRGESPASSKPISINYRT
550	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYHIHLMAATGRGESPASSKPISINYRT
551	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYILEVQAVTGRGESPASSKPISINYRT
552	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYNVTLYGVTGRGESPASSKPISINYRT
553	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYNVTIFPETGRGESPASSKPISINYRT
554	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKIQILTVTGRGESPASSKPISINYRT
555	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYQIMIASVTGRGESPASSKPISINYRT
556	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVTYIAAVVARTGRGESPASSKPISINYRT
557	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYTVRVQAYTGRGESPASSKPISINYRT
558	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYTIQVQSVTGRGESPASSKPISINYRT
559	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEFHLTVLAYTGRGESPASSKPISINYRT
560	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTVTVYAMTGRGESPASSKPISINYRT
561	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVWYQFRVAAVTGRGESPASSKPISINYRT
562	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYAFRVLAVTGRGESPASSKPISINYRT
563	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQVTVIALTGRGESPASSKPISINYRT
564	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYVFRVAAETGRGESPASSKPISINYRT
565	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYVFSITPGTGRGESPASSKPISINYRT
566	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYTFLVFAKTGRGESPASSKPISINYRT
567	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYSYTVTAVTGRGESPASSKPISINYRT
568	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYEFRVVSVTGRGESPASSKPISINYRT
569	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYSVTVAAYTGRGESPASSKPISINYRT
570	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYNITIYAVTGRGESPASSKPISINYRT
571	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFTVVVRTGRGESPASSKPISINYRT
572	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYLIAVQALTGRGESPASSKPISINYRT
573	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYIIRVATATGRGESPASSKPISINYRT
574	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYFLRIVGKTGRGESPASSKPISINYRT
575	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFTVVVRTGRGESPASSKPISINYRT
576	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNLTVCVAAYTGRGESPASSKPISINYRT
577	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVNYTLALSMITGRGESPASSKPISINYRT
578	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYAVYVKAVTGRGESPASSKPISINYRT
579	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYRLSLTVLTGRGESPASSKPISINYRT
580	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYLFRVIAQTGRGESPASSKPISINYRT
581	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYNITVHPFTGRGESPASSKPISINYRT
582	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVCYTVSAWAWTGRGESPASSKPISINYRT
583	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIFRVSAVTGRGESPASSKPISINYRT
584	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYQVVIYLRTGRGESPASSKPISINYRT
585	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYEFFLTPFTGRGESPASSKPISINYRT
586	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYIFRVRAVTGRGESPASSKPISINYRT
587	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYVFRVRAITGRGESPASSKPISINYRT
588	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYFFRVFAETGRGESPASSKPISINYRT
589	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYIFRVRAVTGRGESPASSKPISINYRT
590	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYSFRVIAVTGRGESPASSKPISINYRT
591	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVVDLRVRVVTGRGESPASSKPISINYRT
592	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVCYEIRVYALTGRGESPASSKPISINYRT
593	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTIVASILAHTGRGESPASSKPISINYRT
594	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYNIIVEAVTGRGESPASSKPISINYRT
595	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEVVVKTITGRGESPASSKPISINYRT
596	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYLVSVSSVTGRGESPASSKPISINYRT
597	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYTITVYAVTGRGESPASSKPISINYRT
598	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVTYKFRVLAVTGRGESPASSKPISINYRT
599	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYQLRVLVFTGRGESPASSKPISINYRT
600	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYQFQVCAYTGRGESPASSKPISINYRT
601	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFRVIAVTGRGESPASSKPISINYRT
602	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKFLLRVSAVTGRGESPASSKPISINYRT
603	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYSITVAAYTGRGESPASSKPISINYRT
604	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTVSLFAYTGRGESPASSKPISINYRT
605	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYKLLLYGLTGRGESPASSKPISINYRT
606	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYHMTLSAVTGRGESPASSKPISINYRT
607	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLFNLSVTPYTGRGESPASSKPISINYRT
608	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYKFEVTAVTGRGESPASSKPISINYRT
609	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPFFFRVLAETGRGESPASSKPISINYRT
610	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTVALTMITGRGESPASSKPISINYRT
611	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTVSLVALTGRGESPASSKPISINYRT
612	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKIFVRVKAVTGRGESPASSKPISINYRT
613	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYDITVLPITGRGESPASSKPISINYRT
614	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYVFKVRAVTGRGESPASSKPISINYRT
615	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYRFRVIAITGRGESPASSKPISINYRT
616	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYYFRVSAVTGRGESPASSKPISINYRT
617	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVVSIASYTGRGESPASSKPISINYRT
618	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYEVLMVARTGRGESPASSKPISINYRT
619	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVNYALQIIAFTGRGESPASSKPISINYRT
620	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTINIVAITGRGESPASSKPISINYRT
621	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYEMRVYAVTGRGESPASSKPISINYRT
622	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYTFRVIAFTGRGESPASSKPISINYRT
623	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYAVFVRAITGRGESPASSKPISINYRT
624	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVYYNFRVTMVTGRGESPASSKPISINYRT
625	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYVFKLAVKTGRGESPASSKPISINYRT
626	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYQIAVFAITGRGESPASSKPISINYRT
627	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQFSVYTVTGRGESPASSKPISINYRT
628	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRFPFRVVAVTGRGESPASSKPISINYRT
629	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYSLRVLAFTGRGESPASSKPISINYRT
630	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYNISVYAVTGRGESPASSKPISINYRT
631	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYAVSLVAVTGRGESPASSKPISINYRT
632	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYSFRVFAYTGRGESPASSKPISINYRT
633	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYEIKIFVTTGRGESPASSKPISINYRT
634	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYTCSVWVETGRGESPASSKPISINYRT
635	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQVAVSVLTGRGESPASSKPISINYRT
636	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYVFAVSAYTGRGESPASSKPISINYRT
637	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYRITVAPLTGRGESPASSKPISINYRT
638	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYTFVIVGRTGRGESPASSKPISINYRT
639	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYHFKVFAVTGRGESPASSKPISINYRT
640	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVIYFVRVNVRTGRGESPASSKPISINYRT
641	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYVYSVTAVTGRGESPASSKPISINYRT
642	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYEVSVTAITGRGESPASSKPISINYRT
643	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTVAIFSLTGRGESPASSKPISINYRT
644	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYIVRVRMLTGRGESPASSKPISINYRT
645	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKFLLYGLTGRGESPASSKPISINYRT
646	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYVLEVITWTGRGESPASSKPISINYRT
647	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYMVRMAAYTGRGESPASSKPISINYRT
648	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYTFKISVFTGRGESPASSKPISINYRT
649	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYTVRVAAVTGRGESPASSKPISINYRT
650	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYFASVRAYTGRGESPASSKPISINYRT
651	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYKMVIVTHTGRGESPASSKPISINYRT
652	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYLIRMAVVTGRGESPASSKPISINYRT
653	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYEFSVMVTTGRGESPASSKPISINYRT
654	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVFYEMRVFAVTGRGESPASSKPISINYRT
655	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYTVSIYGVTGRGESPASSKPISINYRT
656	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVYYEVSLLAYTGRGESPASSKPISINYRT
657	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVGYEIGLYYITGRGESPASSKPISINYRT
658	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIFRVMAVTGRGESPASSKPISINYRT
659	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYQVTVIAVTGRGESPASSKPISINYRT
660	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNVSVKLILMTGRGESPASSKPISINYRT
661	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVAYSINVIALTGRGESPASSKPISINYRT
662	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYRFLLYGLTGRGESPASSKPISINYRT
663	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYIFTVKAITGRGESPASSKPISINYRT
664	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYEFYVIAVTGRGESPASSKPISINYRT
665	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVNYEFFVIPYTGRGESPASSKPISINYRT
666	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYRICLVPLTGRGESPASSKPISINYRT
667	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYVLWVVTHTGRGESPASSKPISINYRT
668	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYIMTATALTGRGESPASSKPISINYRT
669	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYIVRIAAVTGRGESPASSKPISINYRT
670	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYTFIVRAFTGRGESPASSKPISINYRT
671	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYIFRVMAQTGRGESPASSKPISINYRT
672	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYNVRVVVVTGRGESPASSKPISINYRT
673	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYLFRVQAVTGRGESPASSKPISINYRT
674	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYKITVVAVTGRGESPASSKPISINYRT
675	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIVNLVAVTGRGESPASSKPISINYRT
676	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYKFVLYGFTGRGESPASSKPISINYRT
677	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIVNVVALTGRGESPASSKPISINYRT
678	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYAVVVQAITGRGESPASSKPISINYRT
679	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIVNVFSVTGRGESPASSKPISINYRT
680	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRYLVTLYGFTGRGESPASSKPISINYRT
681	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYTYSVVAVTGRGESPASSKPISINYRT
682	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVKYTIGVIAFTGRGESPASSKPISINYRT
683	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYHVAVVCYTGRGESPASSKPISINYRT
684	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYIITVNATTGRGESPASSKPISINYRT
685	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYYIWLAARTGRGESPASSKPISINYRT
686	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYFVKIAAVTGRGESPASSKPISINYRT
687	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYTVTVVPVTGRGESPASSKPISINYRT
688	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYTYTIVALTGRGESPASSKPISINYRT
689	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYNVKVVAVTGRGESPASSKPISINYRT
690	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVYYYFRVAAYTGRGESPASSKPISINYRT
691	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYIFSLTPVTGRGESPASSKPISINYRT
692	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVMYTFRVVAYTGRGESPASSKPISINYRT
693	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYEFSVYVTTGRGESPASSKPISINYRT
694	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVFYNITVCPVTGRGESPASSKPISINYRT
695	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVRLLFRVISVTGRGESPASSKPISINYRT
696	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYTYVVKLYTGRGESPASSKPISINYRT
697	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVQYKILVLTITGRGESPASSKPISINYRT
698	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVYTISVLTGRGESPASSKPISINYRT
699	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYSLTLYGLTGRGESPASSKPISINYRT
700	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYYVYVTAKTGRGESPASSKPISINYRT
701	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDYVISVVPFTGRGESPASSKPISINYRT
702	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYTISVLAFTGRGESPASSKPISINYRT
703	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVTYVLWMTALTGRGESPASSKPISINYRT
704	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYTFSIISVTGRGESPASSKPISINYRT
705	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYKVKITLITGRGESPASSKPISINYRT
706	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYFIRVFAITGRGESPASSKPISINYRT
707	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYVFQIRAFTGRGESPASSKPISINYRT
708	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVAYYFAITAVTGRGESPASSKPISINYRT
709	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYLVFITGYTGRGESPASSKPISINYRT
710	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYTIYVIALTGRGESPASSKPISINYRT
711	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYEIIVTPLTGRGESPASSKPISINYRT
712	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIVSFQVRVFTGRGESPASSKPISINYRT
713	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYLIKVVAITGRGESPASSKPISINYRT
714	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYGFRILAVTGRGESPASSKPISINYRT
715	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVTYIVQVYAVTGRGESPASSKPISINYRT
716	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYTVKVYAFTGRGESPASSKPISINYRT
717	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYLIYIYGTTGRGESPASSKPISINYRT
718	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVPYAFCVLAVTGRGESPASSKPISINYRT
719	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYLFLVRAATGRGESPASSKPISINYRT
720	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVYYTVRIVAITGRGESPASSKPISINYRT
721	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYVVSIIAVTGRGESPASSKPISINYRT
722	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVKYIVTLIAYTGRGESPASSKPISINYRT
723	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVSYVLQIVAITGRGESPASSKPISINYRT
724	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE

Seq. #[1]	Amino Acid Sequence[2]
	FTVPKNVYTATISGLKPGVAYYFSVYSFTGRGESPASSKPISINYRT
725	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVEYIFRVVWITGRGESPASSKPISINYRT
726	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYYISIAAYTGRGESPASSKPISINYRT
727	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVFYEFFITPFTGRGESPASSKPISINYRT
728	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYSVVVTTVTGRGESPASSKPISINYRT
729	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVDFIVYLYGITGRGESPASSKPISINYRT
730	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYYYKVVAVTGRGESPASSKPISINYRT
731	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYFIRIKVVTGRGESPASSKPISINYRT
732	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYMFTLWGVTGRGESPASSKPISINYRT
733	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVVYKLWVFPITGRGESPASSKPISINYRT
734	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYTVFLWAYTGRGESPASSKPISINYRT
735	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVIYKFRVIVITGRGESPASSKPISINYRT
736	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYYVILTVSTGRGESPASSKPISINYRT
737	VSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQE
	FTVPKNVYTATISGLKPGVLYIATVMVFTGRGESPASSKPISINYRT

[1]Arbitrary order
[2]Single letter amino acid code; β-strand F highlighted

B.2 Full Amino Acid Sequences of Scaffold Mutants

Table B.2: Full amino acid sequences (as encoded by the DNA constructs cloned into pET-9d expression vectors) of the scaffold mutants in Table 5.1.

Identifier[1]	**Amino Acid Sequence**[2]
Parent	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYAVTLLRDYGPISINYRTEIDKPSQHHHHHH
T71E	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTI**E**VYAVTLLRDYGPISINYRTEIDKPSQHHHHHH
T71R	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTI**R**VYAVTLLRDYGPISINYRTEIDKPSQHHHHHH
Y73E	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITV**E**AVTLLRDYGPISINYRTEIDKPSQHHHHHH
Y73R	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITV**R**AVTLLRDYGPISINYRTEIDKPSQHHHHHH
V75E	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYA**E**TLLRDYGPISINYRTEIDKPSQHHHHHH
V75R	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTITVYA**R**TLLRDYGPISINYRTEIDKPSQHHHHHH
T71R/V75E	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTI**R**VYA**E**TLLRDYGPISINYRTEIDKPSQHHHHHH
T71R/V75R	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVP KNVYTATISGLKPGVDYTI**R**VYA**R**TLLRDYGPISINYRTEIDKPSQHHHHHH

[1]Identifiers indicate mutations relative to the Parent Adnectin
[2]Single letter amino acid code; sites of mutation highlighted

B.3 DNA Sequences of Scaffold Mutants

Table B.3: DNA sequences of Adnectin scaffold mutants.

Identifier[1]	DNA Sequence[2]
T71E	ATGGGAGTTTCTGATGTGCCGCGCGACCTGGAAGTGGTTGCTGCCACCCCCACCAGC CTGCTGATCAGCTGGTCTGCGCGTCTGAAAGTTGCGCGATATTACCGCATCACTTAC GGCGAAACAGGAGGCAATAGCCCTGTCCAGGAGTTCACTGTGCCTAAAAACGTTTAC ACAGCTACCATCAGCGGCCTTAAACCTGGCGTTGATTATACCATCGAGGTGTATGCT GTCACTTTGTTGAGGGACTACGGGCCAATTTCCATTAATTACCGCACAGAAATTGAC AAACCATCCCAGCACCATCACCACCACCACTGA
T71R	ATGGGAGTTTCTGATGTGCCGCGCGACCTGGAAGTGGTTGCTGCCACCCCCACCAGC CTGCTGATCAGCTGGTCTGCGCGTCTGAAAGTTGCGCGATATTACCGCATCACTTAC GGCGAAACAGGAGGCAATAGCCCTGTCCAGGAGTTCACTGTGCCTAAAAACGTTTAC ACAGCTACCATCAGCGGCCTTAAACCTGGCGTTGATTATACCATCCGCGTGTATGCT GTCACTTTGTTGAGGGACTACGGGCCAATTTCCATTAATTACCGCACAGAAATTGAC AAACCATCCCAGCACCATCACCACCACCACTGA
Y73E	ATGGGAGTTTCTGATGTGCCGCGCGACCTGGAAGTGGTTGCTGCCACCCCCACCAGC CTGCTGATCAGCTGGTCTGCGCGTCTGAAAGTTGCGCGATATTACCGCATCACTTAC GGCGAAACAGGAGGCAATAGCCCTGTCCAGGAGTTCACTGTGCCTAAAAACGTTTAC ACAGCTACCATCAGCGGCCTTAAACCTGGCGTTGATTATACCATCACTGTGGAGGCT GTCACTTTGTTGAGGGACTACGGGCCAATTTCCATTAATTACCGCACAGAAATTGAC AAACCATCCCAGCACCATCACCACCACCACTGA
Y73R	ATGGGAGTTTCTGATGTGCCGCGCGACCTGGAAGTGGTTGCTGCCACCCCCACCAGC CTGCTGATCAGCTGGTCTGCGCGTCTGAAAGTTGCGCGATATTACCGCATCACTTAC GGCGAAACAGGAGGCAATAGCCCTGTCCAGGAGTTCACTGTGCCTAAAAACGTTTAC ACAGCTACCATCAGCGGCCTTAAACCTGGCGTTGATTATACCATCACTGTGCGCGCT GTCACTTTGTTGAGGGACTACGGGCCAATTTCCATTAATTACCGCACAGAAATTGAC AAACCATCCCAGCACCATCACCACCACCACTGA
V75E	ATGGGAGTTTCTGATGTGCCGCGCGACCTGGAAGTGGTTGCTGCCACCCCCACCAGC CTGCTGATCAGCTGGTCTGCGCGTCTGAAAGTTGCGCGATATTACCGCATCACTTAC GGCGAAACAGGAGGCAATAGCCCTGTCCAGGAGTTCACTGTGCCTAAAAACGTTTAC ACAGCTACCATCAGCGGCCTTAAACCTGGCGTTGATTATACCATCACTGTGTATGCT GAGACTTTGTTGAGGGACTACGGGCCAATTTCCATTAATTACCGCACAGAAATTGAC AAACCATCCCAGCACCATCACCACCACCACTGA
V75R	ATGGGAGTTTCTGATGTGCCGCGCGACCTGGAAGTGGTTGCTGCCACCCCCACCAGC CTGCTGATCAGCTGGTCTGCGCGTCTGAAAGTTGCGCGATATTACCGCATCACTTAC GGCGAAACAGGAGGCAATAGCCCTGTCCAGGAGTTCACTGTGCCTAAAAACGTTTAC ACAGCTACCATCAGCGGCCTTAAACCTGGCGTTGATTATACCATCACTGTGTATGCT CGCACTTTGTTGAGGGACTACGGGCCAATTTCCATTAATTACCGCACAGAAATTGAC AAACCATCCCAGCACCATCACCACCACCACTGA

[1]Identifiers indicate mutations relative to the amino acid sequence of the Parent Adnectin
[2]Single letter nucleotide code; sites of mutation relative to the Parent sequence are highlighted

Appendix C

Amide H/D Exchange: Supplemental Information

C.1 ^{1}H-^{15}N HSQC Spectra

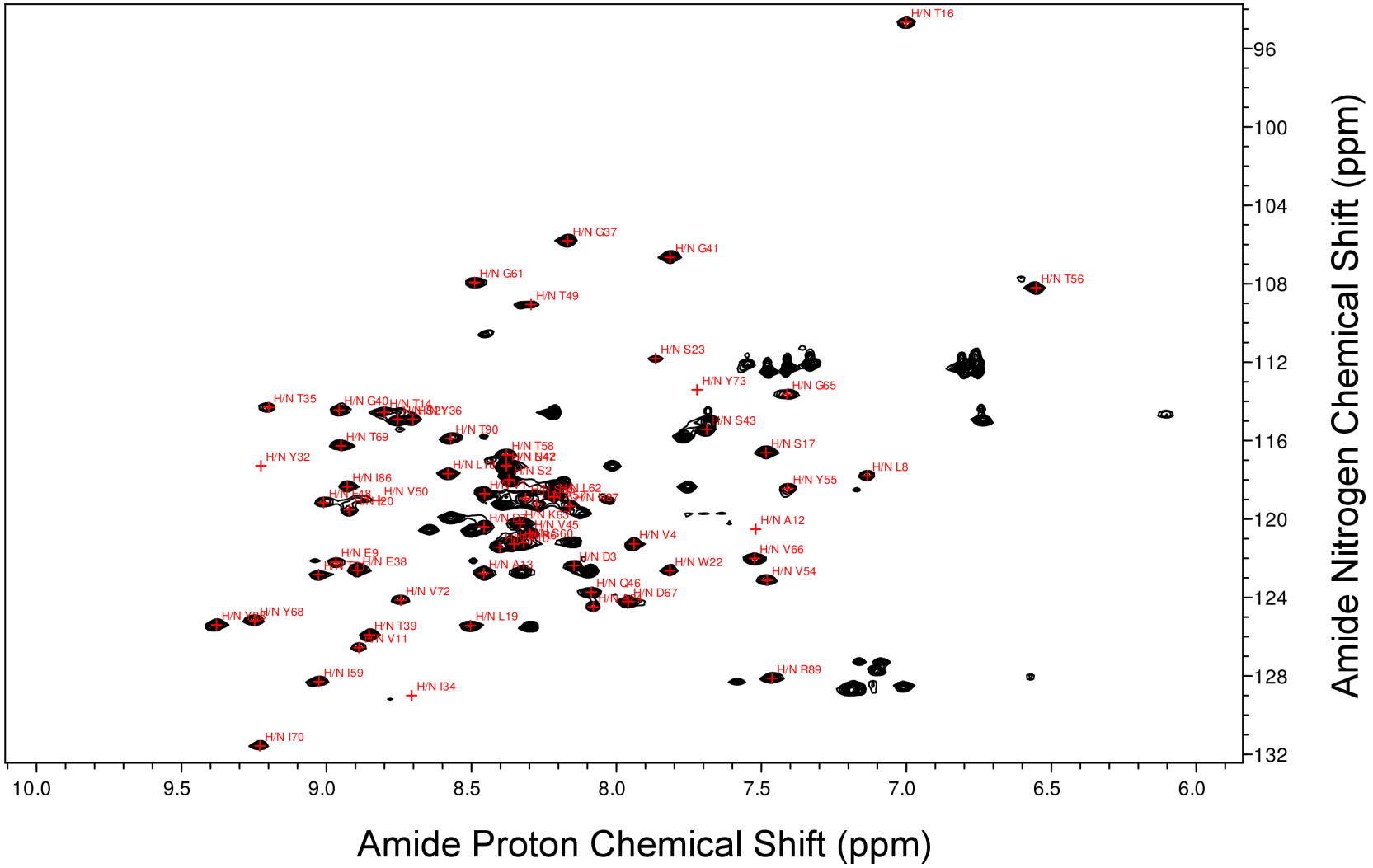


Figure C.1: A Parent Adnectin ^{1}H-^{15}N HSQC spectrum. Consistent with its greater propensity for self-association, the peaks visible at this contour level are fewer and less uniform in shape and intensity relative to V75R (Fig. C.2).

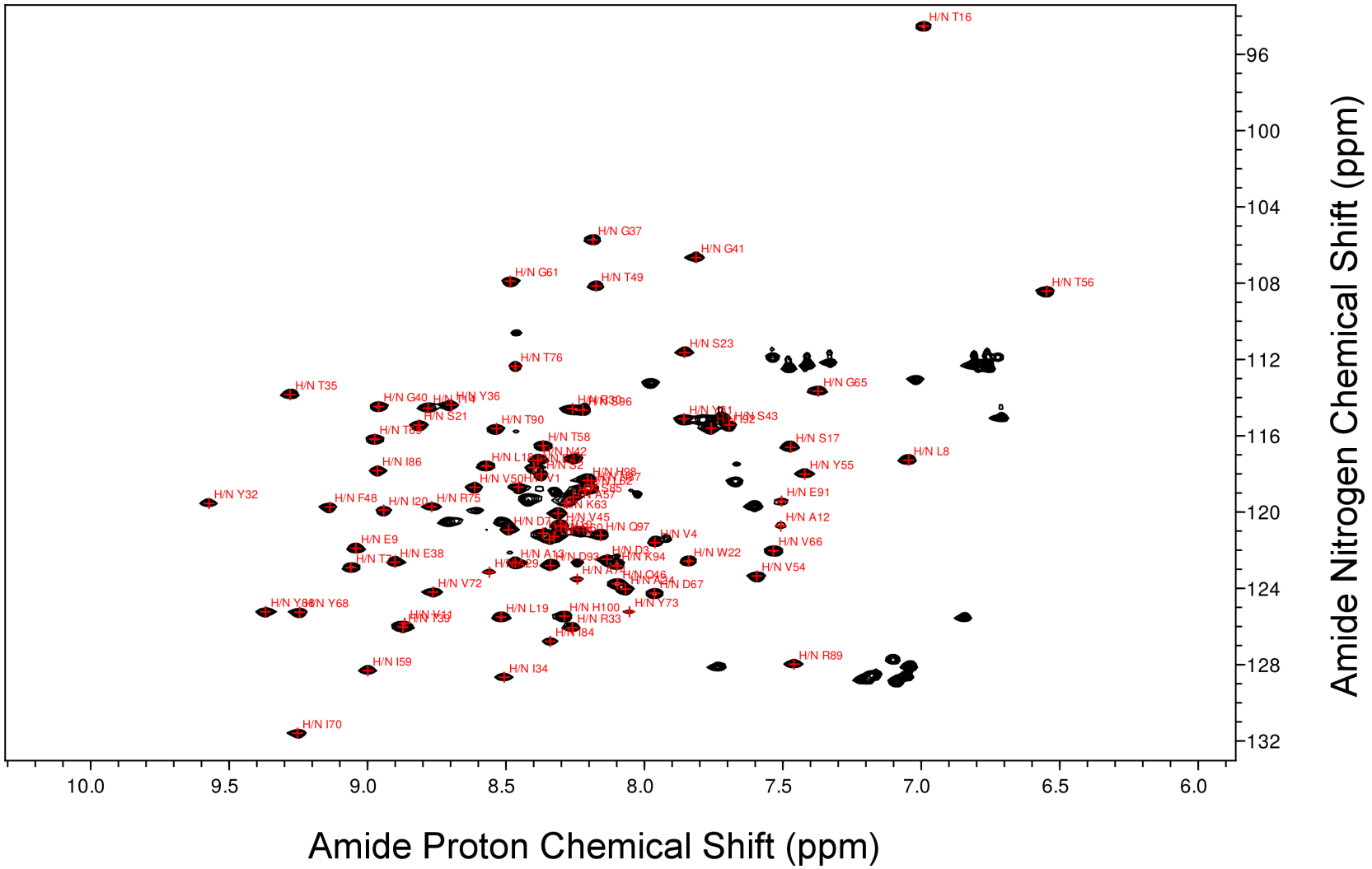


Figure C.2: A V75R Adnectin ^{1}H-^{15}N HSQC spectrum. Consistent with its lower propensity for self-association, the peaks visible at this contour level are greater in number and more uniform in shape and intensity relative to the Parent (Fig. C.1).

C.2 Exchange Rate Constants

Table C.1: Adnectin amide H/D exchange rate constants (standard error)[1]

		k_{ex} (1/s)			
		pD 4.6			pD 7.0
Res.[2]	**β H-bond[3]**	**pWT**	**Parent**	**V75R**	**V75R**
4		$2.0 \cdot 10^{-4}$ ($3.9 \cdot 10^{-7}$)	Unprotected	Unprotected	Unprotected
8		$8.4 \cdot 10^{-5}$ ($1.6 \cdot 10^{-7}$)	$6.1 \cdot 10^{-4}$ ($8.5 \cdot 10^{-5}$)	$1.0 \cdot 10^{-3}$ ($5.7 \cdot 10^{-6}$)	$1.1 \cdot 10^{-3}$ ($2.2 \cdot 10^{-5}$)
9	$\beta_A - \beta_B$	Protected	$4.6 \cdot 10^{-6}$ ($8.8 \cdot 10^{-8}$)	Protected	N/D
10		N/D	$9.7 \cdot 10^{-5}$ ($5.4 \cdot 10^{-7}$)	N/D	Unprotected
11	$\beta_A - \beta_B$	N/D	$9.2 \cdot 10^{-7}$ ($1.2 \cdot 10^{-7}$)	Protected	$2.2 \cdot 10^{-5}$ ($5.2 \cdot 10^{-7}$)
12	$\beta_A - \beta_B$	$3.1 \cdot 10^{-5}$ ($9.1 \cdot 10^{-8}$)	$6.3 \cdot 10^{-5}$ ($5.7 \cdot 10^{-7}$)	$4.7 \cdot 10^{-5}$ ($1.6 \cdot 10^{-7}$)	Unprotected
13		Protected	N/D	Unprotected	Unprotected
14	$\beta_A - \beta_B$	N/D	Protected	$4.9 \cdot 10^{-6}$ ($4.4 \cdot 10^{-8}$)	N/D
16		$3.8 \cdot 10^{-4}$ ($1.8 \cdot 10^{-6}$)	$3.1 \cdot 10^{-4}$ ($3.1 \cdot 10^{-6}$)	$2.7 \cdot 10^{-4}$ ($9.6 \cdot 10^{-7}$)	Unprotected
17	$\beta_B - \beta_A$	$2.8 \cdot 10^{-6}$ ($4.9 \cdot 10^{-8}$)	$3.6 \cdot 10^{-6}$ ($9.3 \cdot 10^{-8}$)	$2.4 \cdot 10^{-6}$ ($3.9 \cdot 10^{-8}$)	$4.5 \cdot 10^{-4}$ ($1.7 \cdot 10^{-6}$)
18	$\beta_B - \beta_E$	Protected	Protected	Protected	$3.8 \cdot 10^{-6}$ ($7.1 \cdot 10^{-8}$)
19	$\beta_B - \beta_A$	Protected	Protected	Protected	$1.1 \cdot 10^{-6}$ ($1.4 \cdot 10^{-7}$)
20	$\beta_B - \beta_E$	Protected	N/D	Protected	Protected
21	$\beta_B - \beta_A$	Protected	N/D	Protected	$2.2 \cdot 10^{-6}$ ($1.8 \cdot 10^{-7}$)
22	$\beta_B - \beta_E$	Protected	Protected	Protected	$3.3 \cdot 10^{-6}$ ($7.6 \cdot 10^{-8}$)
23	$\beta_B - \beta_A$	$6.9 \cdot 10^{-4}$ ($4.7 \cdot 10^{-6}$)	$2.4 \cdot 10^{-4}$ ($1.8 \cdot 10^{-5}$)	$2.9 \cdot 10^{-5}$ ($4.7 \cdot 10^{-8}$)	$3.7 \cdot 10^{-4}$ ($2.0 \cdot 10^{-6}$)
24		$2.2 \cdot 10^{-5}$ ($2.4 \cdot 10^{-7}$)		N/D	Unprotected
26		$2.6 \cdot 10^{-3}$ ($4.7 \cdot 10^{-5}$)	N/A	N/A	N/A
29		$7.0 \cdot 10^{-4}$ ($4.1 \cdot 10^{-6}$)	N/A	N/D	Unprotected
30		$3.9 \cdot 10^{-4}$ ($7.0 \cdot 10^{-6}$)	N/A	N/D	Unprotected
31	$\beta_C - \beta_F$	Protected	N/A	Unprotected	Unprotected
32	$\beta_C - \beta_F$	Protected	Protected	Protected	$4.6 \cdot 10^{-5}$ ($2.2 \cdot 10^{-7}$)
33	$\beta_C - \beta_F$	N/D	N/A	Protected	$5.7 \cdot 10^{-5}$ ($1.9 \cdot 10^{-7}$)
34	$\beta_C - \beta_F$	Protected	Protected	Protected	$1.1 \cdot 10^{-6}$ ($2.6 \cdot 10^{-7}$)
35	$\beta_C - \beta_F$	Protected	Protected	Protected	$2.0 \cdot 10^{-6}$ ($1.7 \cdot 10^{-7}$)
36	$\beta_C - \beta_F$	Protected	Protected	Protected	$8.9 \cdot 10^{-3}$ ($5.9 \cdot 10^{-3}$)
37	$\beta_C - \beta_F$	Protected	$1.4 \cdot 10^{-6}$ ($1.1 \cdot 10^{-7}$)	Protected	$1.1 \cdot 10^{-5}$ ($1.4 \cdot 10^{-7}$)
39		N/D	$2.4 \cdot 10^{-4}$ ($1.8 \cdot 10^{-6}$)	Protected	Unprotected
43		$3.1 \cdot 10^{-4}$ ($6.0 \cdot 10^{-7}$)	$3.1 \cdot 10^{-4}$ ($1.9 \cdot 10^{-6}$)	$2.3 \cdot 10^{-4}$ ($1.9 \cdot 10^{-6}$)	Unprotected
48	$\beta_D - \beta_C$	Protected	N/D	Protected	$4.4 \cdot 10^{-6}$ ($5.5 \cdot 10^{-8}$)
49		N/D	N/D	$5.9 \cdot 10^{-5}$ ($1.5 \cdot 10^{-7}$)	Unprotected
50	$\beta_D - \beta_C$	Protected	N/D	Protected	$8.6 \cdot 10^{-6}$ ($9.1 \cdot 10^{-8}$)
52		$1.9 \cdot 10^{-5}$ ($6.9 \cdot 10^{-8}$)	N/A	N/A	N/A
54		$2.8 \cdot 10^{-5}$ ($1.3 \cdot 10^{-7}$)	N/D	Protected	$4.8 \cdot 10^{-5}$ ($1.4 \cdot 10^{-7}$)
55		$2.5 \cdot 10^{-3}$ ($4.9 \cdot 10^{-5}$)	$7.6 \cdot 10^{-5}$ ($5.9 \cdot 10^{-7}$)	$1.9 \cdot 10^{-5}$ ($6.8 \cdot 10^{-8}$)	$4.4 \cdot 10^{-4}$ ($3.8 \cdot 10^{-6}$)
56		$1.3 \cdot 10^{-6}$ ($4.0 \cdot 10^{-8}$)	$1.0 \cdot 10^{-6}$ ($1.2 \cdot 10^{-7}$)	Protected	$2.2 \cdot 10^{-5}$ ($3.6 \cdot 10^{-7}$)
58		$5.6 \cdot 10^{-4}$ ($2.9 \cdot 10^{-6}$)	N/D	$3.3 \cdot 10^{-4}$ ($1.3 \cdot 10^{-6}$)	$2.5 \cdot 10^{-4}$ ($1.2 \cdot 10^{-6}$)
59		Protected	Protected	Protected	$1.9 \cdot 10^{-6}$ ($1.9 \cdot 10^{-7}$)
61		$1.5 \cdot 10^{-3}$ ($2.6 \cdot 10^{-5}$)	$2.1 \cdot 10^{-3}$ ($2.7 \cdot 10^{-4}$)	$1.5 \cdot 10^{-3}$ ($4.6 \cdot 10^{-5}$)	Unprotected
62		N/A	$1.7 \cdot 10^{-6}$ ($7.2 \cdot 10^{-8}$)	N/D	$1.7 \cdot 10^{-4}$ ($6.0 \cdot 10^{-7}$)
65		$1.9 \cdot 10^{-5}$ ($5.0 \cdot 10^{-8}$)	$2.2 \cdot 10^{-5}$ ($1.0 \cdot 10^{-7}$)	$1.6 \cdot 10^{-5}$ ($1.4 \cdot 10^{-7}$)	$1.2 \cdot 10^{-3}$ ($2.8 \cdot 10^{-5}$)
66		Protected	Protected	Protected	$2.1 \cdot 10^{-6}$ ($1.1 \cdot 10^{-7}$)
68	$\beta_F - \beta_G$	Protected	Protected	Protected	$1.8 \cdot 10^{-6}$ ($2.3 \cdot 10^{-7}$)
69	$\beta_F - \beta_C$	Protected	Protected	Protected	$2.0 \cdot 10^{-6}$ ($1.8 \cdot 10^{-7}$)
70	$\beta_F - \beta_G$	Protected	Protected	Protected	$1.5 \cdot 10^{-6}$ ($2.3 \cdot 10^{-7}$)
71	$\beta_F - \beta_C$	Protected	Protected	Protected	N/D
72	$\beta_F - \beta_G$	Protected	Protected	Protected	$7.4 \cdot 10^{-7}$ ($1.5 \cdot 10^{-7}$)
73	$\beta_F - \beta_C$	Protected	N/D	Protected	$1.3 \cdot 10^{-5}$ ($8.1 \cdot 10^{-7}$)
74		$3.0 \cdot 10^{-5}$ ($1.1 \cdot 10^{-7}$)	N/A	$3.8 \cdot 10^{-5}$ ($6.9 \cdot 10^{-8}$)	Unprotected
75	$\beta_F - \beta_C$	Protected	N/A	$1.9 \cdot 10^{-4}$ ($1.1 \cdot 10^{-6}$)	Unprotected
76		$5.7 \cdot 10^{-5}$ ($4.2 \cdot 10^{-7}$)	N/A	$2.3 \cdot 10^{-4}$ ($6.0 \cdot 10^{-7}$)	Unprotected
77		$5.9 \cdot 10^{-5}$ ($2.5 \cdot 10^{-7}$)	N/A	N/A	N/A

Table C.1 (continued): Adnectin amide H/D exchange rate constants (standard error)[1]

		k_{ex} (1/s)			
		pD 4.6			pD 7.0
Res.[2]	**β H-bond[3]**	**pWT**	**Parent**	**V75R**	**V75R**
88/84	$\beta_G - \beta_F$	$2.5 \cdot 10^{-6}$ ($3.6 \cdot 10^{-8}$)	N/A	$1.8 \cdot 10^{-6}$ ($3.8 \cdot 10^{-8}$)	$2.1 \cdot 10^{-4}$ ($1.7 \cdot 10^{-6}$)
89/85		$5.9 \cdot 10^{-5}$ ($1.5 \cdot 10^{-7}$)	N/D	N/D	N/D
90/86	$\beta_G - \beta_F$	$5.8 \cdot 10^{-7}$ ($8.0 \cdot 10^{-8}$)	$1.2 \cdot 10^{-6}$ ($9.6 \cdot 10^{-8}$)	Protected	$1.2 \cdot 10^{-5}$ ($1.7 \cdot 10^{-7}$)
91/87		$7.1 \cdot 10^{-5}$ ($1.8 \cdot 10^{-7}$)	N/D	N/D	N/D
92/88	$\beta_G - \beta_F$	$2.0 \cdot 10^{-6}$ ($5.2 \cdot 10^{-8}$)	$2.8 \cdot 10^{-6}$ ($8.5 \cdot 10^{-8}$)	Protected	$5.0 \cdot 10^{-6}$ ($9.2 \cdot 10^{-8}$)
93/89		$4.7 \cdot 10^{-5}$ ($1.8 \cdot 10^{-7}$)	$5.2 \cdot 10^{-5}$ ($2.5 \cdot 10^{-7}$)	$4.8 \cdot 10^{-5}$ ($8.7 \cdot 10^{-8}$)	Unprotected
94/90		N/A	$1.7 \cdot 10^{-5}$ ($1.3 \cdot 10^{-7}$)	$1.2 \cdot 10^{-5}$ ($1.4 \cdot 10^{-7}$)	$2.5 \cdot 10^{-4}$ ($1.2 \cdot 10^{-6}$)

[1]'Protected': less than one exchange half-life is observed; 'Unprotected': exchange is too fast to measure; 'N/A': missing assignment; 'N/D': no data (severely overlapped or otherwise ambiguous)
[2][pWT residue #]/[other Adnectins residue #], where necessary
[3]Intramolecular hydrogen bonds inferred between β-strands

Table C.2: Intrinsic exchange rate constants[1]

	k_{int} (1/s)			
	pD 4.0			pD 7.0
Res.[2]	**pWT**	**Parent**	**V75R**	**V75R**
2	$7.67 \cdot 10^{-1}$	$7.67 \cdot 10^{-1}$	$7.67 \cdot 10^{-1}$	4.62
3	$3.05 \cdot 10^{-2}$	$3.05 \cdot 10^{-2}$	$3.05 \cdot 10^{-2}$	2.74
4	$3.94 \cdot 10^{-3}$	$3.94 \cdot 10^{-3}$	$3.94 \cdot 10^{-3}$	0.36
5				
6	$8.27 \cdot 10^{-3}$	$8.27 \cdot 10^{-3}$	$8.27 \cdot 10^{-3}$	1.88
7	$2.55 \cdot 10^{-2}$	$2.55 \cdot 10^{-2}$	$2.55 \cdot 10^{-2}$	2.28
8	$4.95 \cdot 10^{-3}$	$4.95 \cdot 10^{-3}$	$4.95 \cdot 10^{-3}$	0.48
9	$5.13 \cdot 10^{-3}$	$5.13 \cdot 10^{-3}$	$5.13 \cdot 10^{-3}$	0.52
10	$3.43 \cdot 10^{-3}$	$3.43 \cdot 10^{-3}$	$3.43 \cdot 10^{-3}$	0.39
11	$2.34 \cdot 10^{-3}$	$2.34 \cdot 10^{-3}$	$2.34 \cdot 10^{-3}$	0.39
12	$8.63 \cdot 10^{-3}$	$8.63 \cdot 10^{-3}$	$8.63 \cdot 10^{-3}$	1.97
13	$1.16 \cdot 10^{-2}$	$1.16 \cdot 10^{-2}$	$1.16 \cdot 10^{-2}$	2.72
14	$1.00 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$	2.32
15				
16	$6.08 \cdot 10^{-3}$	$6.08 \cdot 10^{-3}$	$6.08 \cdot 10^{-3}$	1.33
17	$4.10 \cdot 10^{-2}$	$4.10 \cdot 10^{-2}$	$4.10 \cdot 10^{-2}$	$1.01 \cdot 10^{1}$
18	$6.46 \cdot 10^{-3}$	$6.46 \cdot 10^{-3}$	$6.46 \cdot 10^{-3}$	1.43
19	$2.53 \cdot 10^{-3}$	$2.53 \cdot 10^{-3}$	$2.53 \cdot 10^{-3}$	0.44
20	$2.01 \cdot 10^{-3}$	$2.01 \cdot 10^{-3}$	$2.01 \cdot 10^{-3}$	0.31
21	$1.57 \cdot 10^{-2}$	$1.57 \cdot 10^{-2}$	$1.57 \cdot 10^{-2}$	3.76
22	$9.18 \cdot 10^{-3}$	$9.18 \cdot 10^{-3}$	$9.18 \cdot 10^{-3}$	2.11
23	$1.23 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	4.96
24	$1.67 \cdot 10^{-2}$	$2.24 \cdot 10^{-2}$	$2.24 \cdot 10^{-2}$	5.43
25		$1.38 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$	3.27
26	$7.02 \cdot 10^{-3}$	$5.50 \cdot 10^{-3}$	$5.50 \cdot 10^{-3}$	1.19
27	$2.93 \cdot 10^{-3}$	$6.87 \cdot 10^{-3}$	$6.87 \cdot 10^{-3}$	1.53
28	$7.45 \cdot 10^{-3}$	$3.62 \cdot 10^{-3}$	$3.62 \cdot 10^{-3}$	0.72
29	$4.20 \cdot 10^{-3}$	$8.63 \cdot 10^{-3}$	$8.63 \cdot 10^{-3}$	1.97
30	$1.02 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$	3.27
31	$1.04 \cdot 10^{-2}$	$1.04 \cdot 10^{-2}$	$1.04 \cdot 10^{-2}$	2.43
32	$7.30 \cdot 10^{-3}$	$7.30 \cdot 10^{-3}$	$7.30 \cdot 10^{-3}$	1.64
33	$1.54 \cdot 10^{-2}$	$1.54 \cdot 10^{-2}$	$1.54 \cdot 10^{-2}$	3.67
34	$4.12 \cdot 10^{-3}$	$4.12 \cdot 10^{-3}$	$4.12 \cdot 10^{-3}$	0.84
35	$6.20 \cdot 10^{-3}$	$6.20 \cdot 10^{-3}$	$6.20 \cdot 10^{-3}$	1.37
36	$1.00 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$	2.32

Table C.2 (continued): Intrinsic exchange rate constants[1]

	k_{int} (1/s)			
	pD 4.0			pD 7.0
Res.[2]	**pWT**	**Parent**	**V75R**	**V75R**
37	$2.34 \cdot 10^{-2}$	$2.34 \cdot 10^{-2}$	$2.34 \cdot 10^{-2}$	5.69
38	$1.12 \cdot 10^{-2}$	$1.12 \cdot 10^{-2}$	$1.12 \cdot 10^{-2}$	1.25
39	$1.21 \cdot 10^{-2}$	$1.21 \cdot 10^{-2}$	$1.21 \cdot 10^{-2}$	1.65
40	$3.28 \cdot 10^{-2}$	$3.28 \cdot 10^{-2}$	$3.28 \cdot 10^{-2}$	8.04
41	$3.06 \cdot 10^{-2}$	$3.06 \cdot 10^{-2}$	$3.06 \cdot 10^{-2}$	7.50
42	$5.03 \cdot 10^{-2}$	$5.03 \cdot 10^{-2}$	$5.03 \cdot 10^{-2}$	$1.24 \cdot 10^{1}$
43	$5.39 \cdot 10^{-2}$	$5.39 \cdot 10^{-2}$	$5.39 \cdot 10^{-2}$	$1.33 \cdot 10^{1}$
44				
45	$2.01 \cdot 10^{-3}$	$2.01 \cdot 10^{-3}$	$2.01 \cdot 10^{-3}$	0.31
46	$9.79 \cdot 10^{-3}$	$9.79 \cdot 10^{-3}$	$9.79 \cdot 10^{-3}$	2.27
47	$1.20 \cdot 10^{-2}$	$1.20 \cdot 10^{-2}$	$1.20 \cdot 10^{-2}$	1.34
48	$8.43 \cdot 10^{-3}$	$8.43 \cdot 10^{-3}$	$8.43 \cdot 10^{-3}$	1.11
49	$1.14 \cdot 10^{-2}$	$1.14 \cdot 10^{-2}$	$1.14 \cdot 10^{-2}$	2.66
50	$4.20 \cdot 10^{-3}$	$4.20 \cdot 10^{-3}$	$4.20 \cdot 10^{-3}$	0.86
51				
52	$1.24 \cdot 10^{-2}$	$6.46 \cdot 10^{-3}$	$6.46 \cdot 10^{-3}$	1.43
53	$3.84 \cdot 10^{-2}$	$4.49 \cdot 10^{-2}$	$4.49 \cdot 10^{-2}$	$1.11 \cdot 10^{1}$
54	$2.05 \cdot 10^{-2}$	$5.29 \cdot 10^{-3}$	$5.29 \cdot 10^{-3}$	1.14
55	$3.43 \cdot 10^{-2}$	$4.99 \cdot 10^{-3}$	$4.99 \cdot 10^{-3}$	1.06
56	$1.92 \cdot 10^{-2}$	$1.11 \cdot 10^{-2}$	$1.11 \cdot 10^{-2}$	2.60
57	$1.80 \cdot 10^{-2}$	$1.80 \cdot 10^{-2}$	$1.80 \cdot 10^{-2}$	4.32
58	$1.00 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$	2.32
59	$3.97 \cdot 10^{-3}$	$3.97 \cdot 10^{-3}$	$3.97 \cdot 10^{-3}$	0.80
60	$1.57 \cdot 10^{-2}$	$1.57 \cdot 10^{-2}$	$1.57 \cdot 10^{-2}$	3.76
61	$4.10 \cdot 10^{-2}$	$4.10 \cdot 10^{-2}$	$4.10 \cdot 10^{-2}$	$1.01 \cdot 10^{1}$
62	$5.00 \cdot 10^{-3}$	$5.00 \cdot 10^{-3}$	$5.00 \cdot 10^{-3}$	1.06
63	$6.87 \cdot 10^{-3}$	$6.87 \cdot 10^{-3}$	$6.87 \cdot 10^{-3}$	1.53
64				
65	$1.24 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	2.92
66	$3.97 \cdot 10^{-3}$	$3.97 \cdot 10^{-3}$	$3.97 \cdot 10^{-3}$	0.80
67	$1.16 \cdot 10^{-2}$	$1.16 \cdot 10^{-2}$	$1.16 \cdot 10^{-2}$	1.00
68	$9.30 \cdot 10^{-3}$	$9.30 \cdot 10^{-3}$	$9.30 \cdot 10^{-3}$	0.97
69	$1.11 \cdot 10^{-2}$	$1.11 \cdot 10^{-2}$	$1.11 \cdot 10^{-2}$	2.60
70	$3.97 \cdot 10^{-3}$	$3.97 \cdot 10^{-3}$	$3.97 \cdot 10^{-3}$	0.80
71	$6.20 \cdot 10^{-3}$	$6.20 \cdot 10^{-3}$	$6.20 \cdot 10^{-3}$	1.37
72	$4.20 \cdot 10^{-3}$	$4.20 \cdot 10^{-3}$	$4.20 \cdot 10^{-3}$	0.86

Table C.2 (continued): Intrinsic exchange rate constants[1]

	k_{int} (1/s)			
	pD 4.0			pD 7.0
Res.[2]	**pWT**	**Parent**	**V75R**	**V75R**
73	$4.99 \cdot 10^{-3}$	$4.99 \cdot 10^{-3}$	$4.99 \cdot 10^{-3}$	1.06
74	$1.29 \cdot 10^{-2}$	$1.29 \cdot 10^{-2}$	$1.29 \cdot 10^{-2}$	3.06
75	$2.93 \cdot 10^{-3}$	$2.93 \cdot 10^{-3}$	$1.38 \cdot 10^{-2}$	0.54
76	$7.45 \cdot 10^{-3}$	$7.45 \cdot 10^{-3}$	$1.61 \cdot 10^{-2}$	1.68
77	$3.28 \cdot 10^{-2}$	$5.29 \cdot 10^{-3}$	$5.29 \cdot 10^{-3}$	1.14
78	$2.01 \cdot 10^{-2}$	$2.53 \cdot 10^{-3}$	$2.53 \cdot 10^{-3}$	0.44
79	$3.43 \cdot 10^{-2}$	$8.81 \cdot 10^{-3}$	$8.81 \cdot 10^{-3}$	2.02
80	$1.12 \cdot 10^{-2}$	$2.55 \cdot 10^{-2}$	$2.55 \cdot 10^{-2}$	2.28
81	$3.20 \cdot 10^{-2}$	$9.30 \cdot 10^{-3}$	$9.30 \cdot 10^{-3}$	0.97
82		$2.34 \cdot 10^{-2}$	$2.34 \cdot 10^{-2}$	5.69
83	$7.02 \cdot 10^{-3}$			
84	$2.62 \cdot 10^{-2}$			
85	$5.15 \cdot 10^{-2}$			
86	$2.05 \cdot 10^{-2}$			
87/83				
88/84	$1.93 \cdot 10^{-3}$	$1.93 \cdot 10^{-3}$	$1.93 \cdot 10^{-3}$	0.29
89/85	$1.57 \cdot 10^{-2}$	$1.57 \cdot 10^{-2}$	$1.57 \cdot 10^{-2}$	3.76
90/86	$4.79 \cdot 10^{-3}$	$4.79 \cdot 10^{-3}$	$4.79 \cdot 10^{-3}$	1.02
91/87	$2.05 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	4.96
92/88	$1.29 \cdot 10^{-2}$	$1.29 \cdot 10^{-2}$	$1.29 \cdot 10^{-2}$	3.06
93/89	$1.54 \cdot 10^{-2}$	$1.54 \cdot 10^{-2}$	$1.54 \cdot 10^{-2}$	3.67
94/90	$1.61 \cdot 10^{-2}$	$1.61 \cdot 10^{-2}$	$1.61 \cdot 10^{-2}$	3.85

[1]Determined from peptide data as described in Chapter 6

[2][pWT residue #]/[other Adnectins residue #], where necessary

Table C.3: Apparent free energies of exchange[1]

		ΔG_{ex}^{app} (kcal/mol)[2]			
		pD 4.6			pD 7.0
Res. #[3]	**β H-bond[4]**	**pWT**	**Parent**	**V75R**	**V75R**
4		1.8	U/P	U/P	U/P
8		2.4	1.2	1.0	3.6
9	$\beta_A - \beta_B$	High	4.2	High	N/D
10		N/D	2.1	N/D	U/P
11	$\beta_A - \beta_B$	N/D	4.6	High	5.8
12	$\beta_A - \beta_B$	3.3	2.9	3.1	U/P
13		High	N/D	U/P	U/P
14	$\beta_A - \beta_B$	N/D	High	4.5	N/D
16		1.6	1.8	1.8	U/P
17	$\beta_B - \beta_A$	5.7	5.5	5.8	5.9
18	$\beta_B - \beta_E$	High	High	High	7.6
19	$\beta_B - \beta_A$	High	High	High	7.6
20	$\beta_B - \beta_E$	High	N/D	High	High
21	$\beta_B - \beta_A$	High	N/D	High	8.5
22	$\beta_B - \beta_E$	High	High	High	7.9
23	$\beta_B - \beta_A$	1.7	2.6	3.9	5.6
24		3.9	N/D	N/D	U/P
26		0.6	N/A	N/A	N/A
29		1.1	N/A	N/D	U/P
30		1.9	N/A	N/D	U/P
31	$\beta_C - \beta_F$	High	N/A	U/P	U/P
32	$\beta_C - \beta_F$	High	High	High	6.2
33	$\beta_C - \beta_F$	N/D	N/A	High	6.6
34	$\beta_C - \beta_F$	High	High	High	8.0
35	$\beta_C - \beta_F$	High	High	High	8.0
36	$\beta_C - \beta_F$	High	High	High	3.3
37	$\beta_C - \beta_F$	High	5.8	High	7.8
39		N/D	2.3	High	U/P
43		3.1	3.1	3.2	U/P
48	$\beta_D - \beta_C$	High	N/D	High	7.4
49		N/D	N/D	3.1	U/P
50	$\beta_D - \beta_C$	High	N/D	High	6.8
52		3.8	N/A	N/A	N/A
54		3.9	N/D	High	6.0
55		1.6	2.5	3.3	4.6

Table C.3 (continued): Apparent free energies of exchange[1]

		ΔG_{ex}^{app} (kcal/mol)[2]			
		pD 4.6			pD 7.0
Res. #[3]	**β H-bond[4]**	**pWT**	**Parent**	**V75R**	**V75R**
56		5.7	5.5	High	6.9
58		1.7	N/D	2.0	5.4
59		High	High	High	7.7
61		2.0	1.8	2.0	U/P
62		N/A	4.7	N/D	5.2
65		3.8	3.8	3.9	4.6
66		High	High	High	7.6
68	$\beta_F - \beta_G$	High	High	High	7.8
69	$\beta_F - \beta_C$	High	High	High	8.3
70	$\beta_F - \beta_G$	High	High	High	7.8
71	$\beta_F - \beta_C$	High	High	High	N/D
72	$\beta_F - \beta_G$	High	High	High	8.3
73	$\beta_F - \beta_C$	High	N/D	High	6.7
74		3.6	N/A	3.5	U/P
75	$\beta_F - \beta_C$	High	N/A	2.5	U/P
76		2.9	N/A	2.5	U/P
77		3.7	N/A	N/A	N/A
88/84	$\beta_G - \beta_F$	3.9	N/A	4.1	4.3
89/85		3.3	N/D	N/D	N/D
90/86	$\beta_G - \beta_F$	5.3	4.9	High	6.7
91/87		3.4	N/D	N/D	N/D
92/88	$\beta_G - \beta_F$	5.2	5.0	High	7.9
93/89		3.4	3.4	3.4	U/P
94/90		N/A	4.1	4.3	5.7

[1]Calculated as described in Chapter 6

[2]'High': less than one exchange half-life is observed; 'U/P': exchange is too fast to measure; 'N/A': missing assignment; 'N/D': no data (severely overlapped or otherwise ambiguous)

[3][pWT residue #]/[other Adnectins residue #], where necessary

[4]Intramolecular hydrogen bonds inferred between β-strands

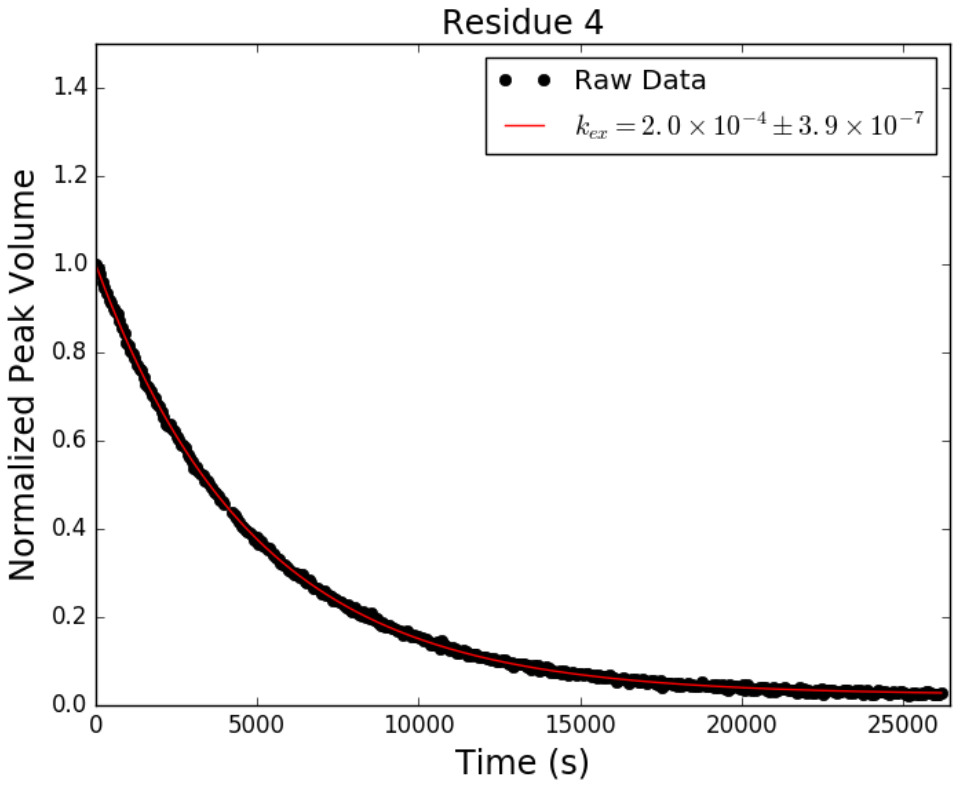


Figure C.3: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 4 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

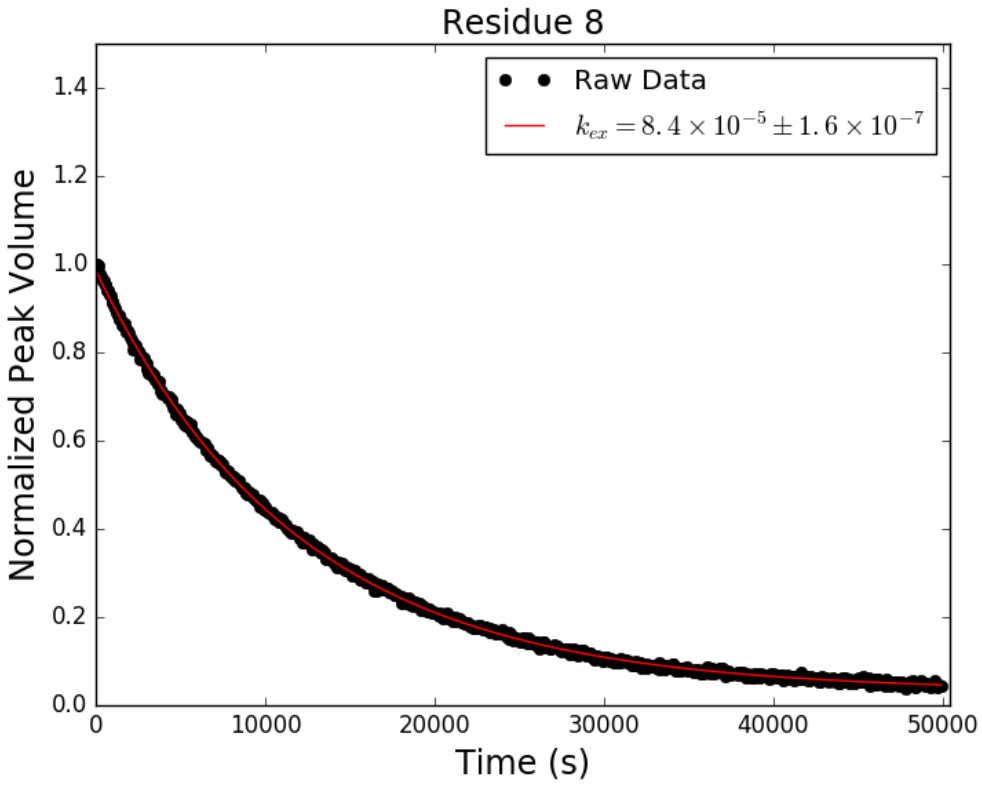


Figure C.4: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 8 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

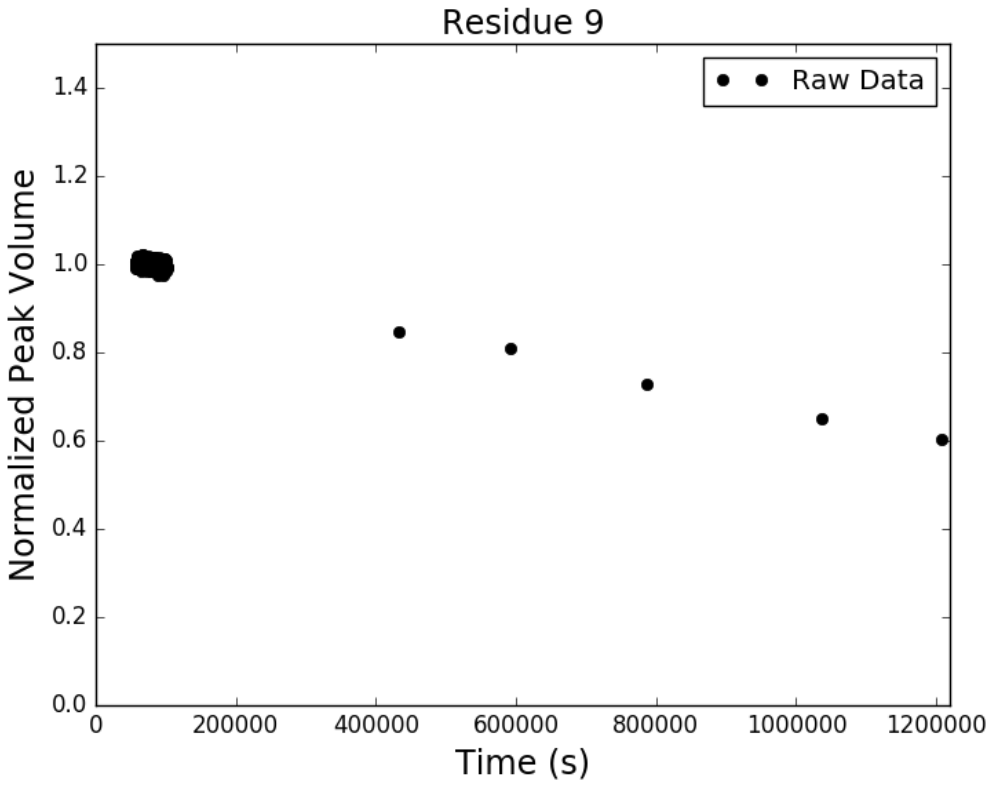


Figure C.5: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 9 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

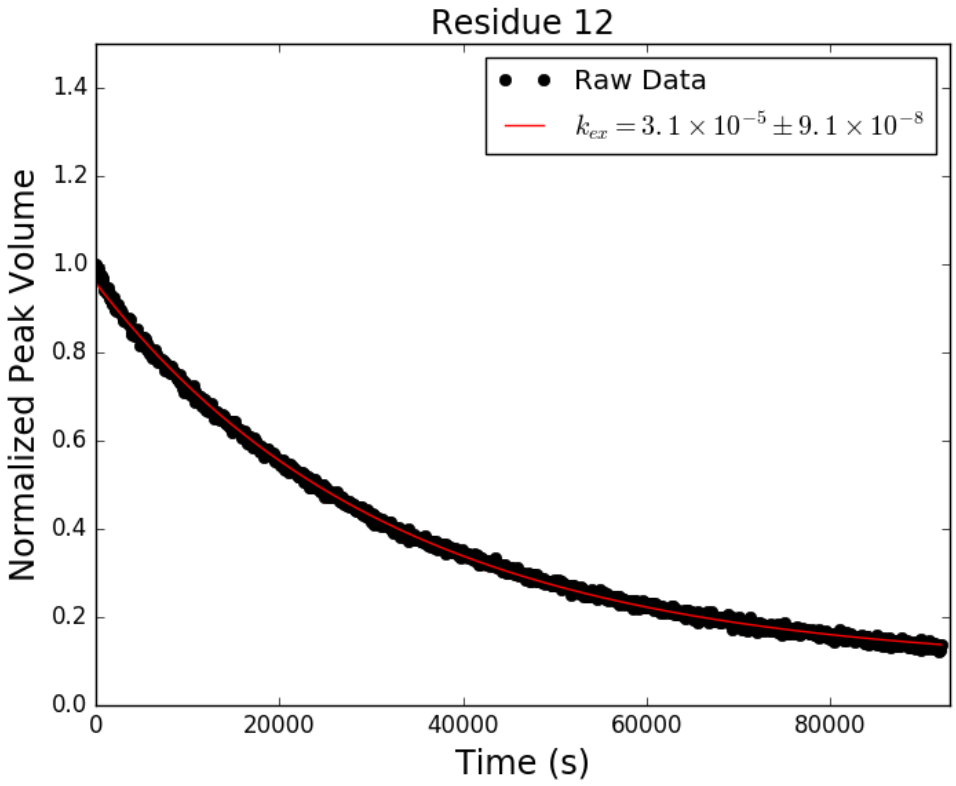


Figure C.6: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 12 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

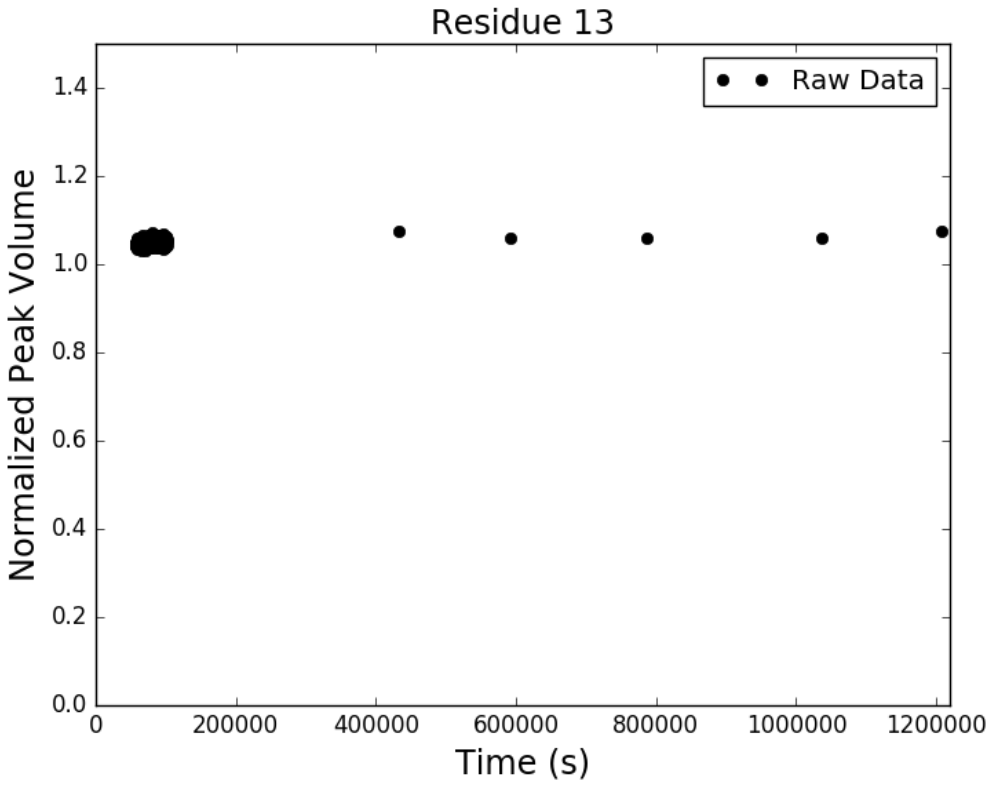


Figure C.7: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 13 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

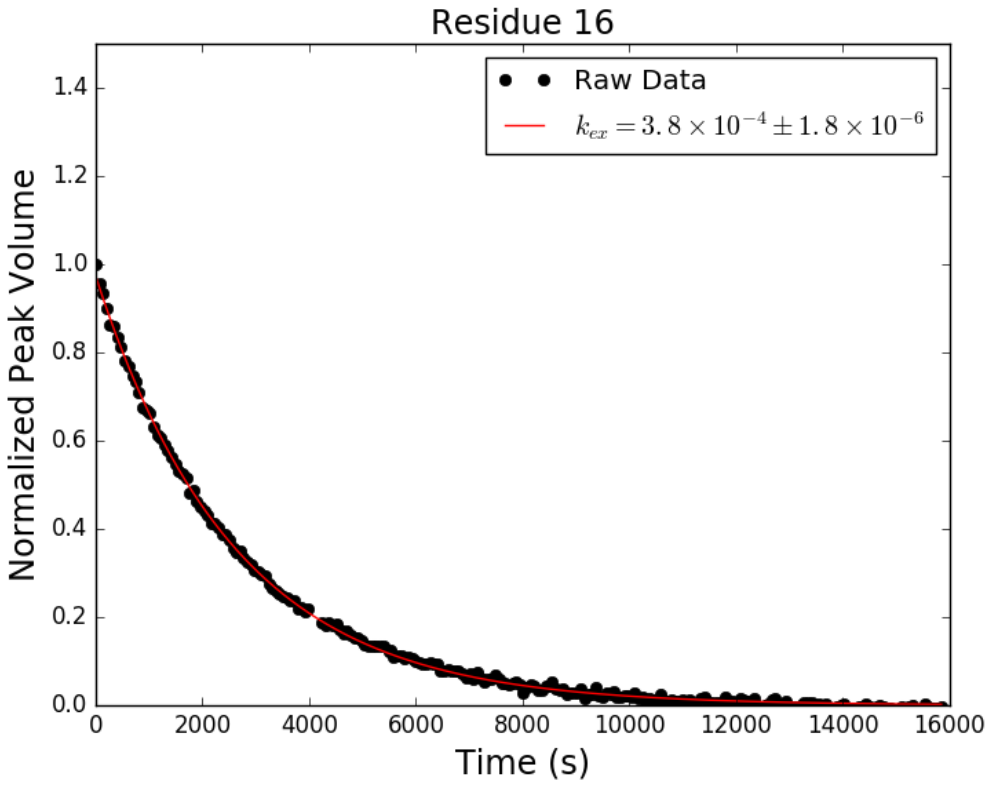


Figure C.8: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 16 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

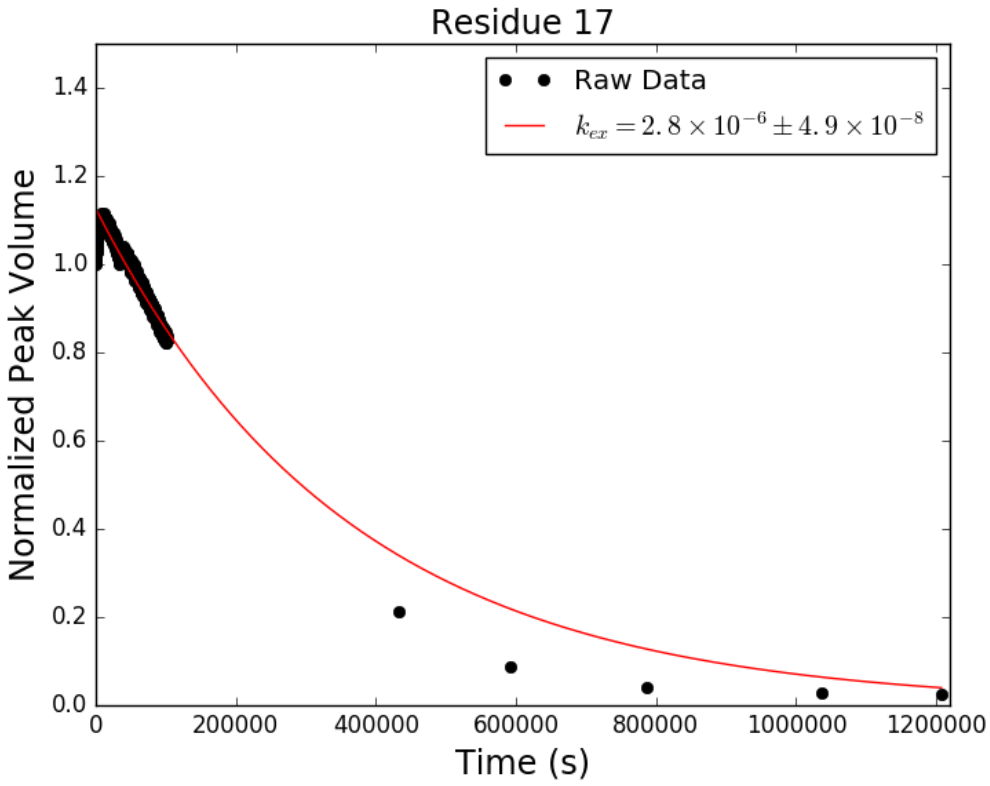


Figure C.9: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 17 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

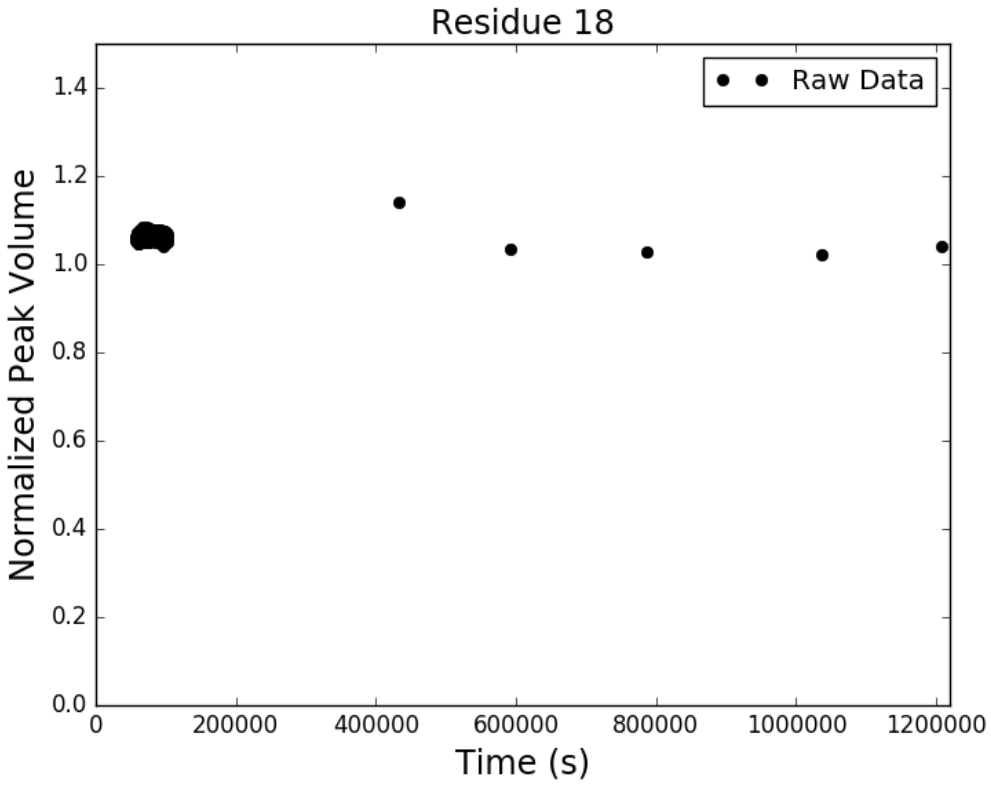


Figure C.10: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 18 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

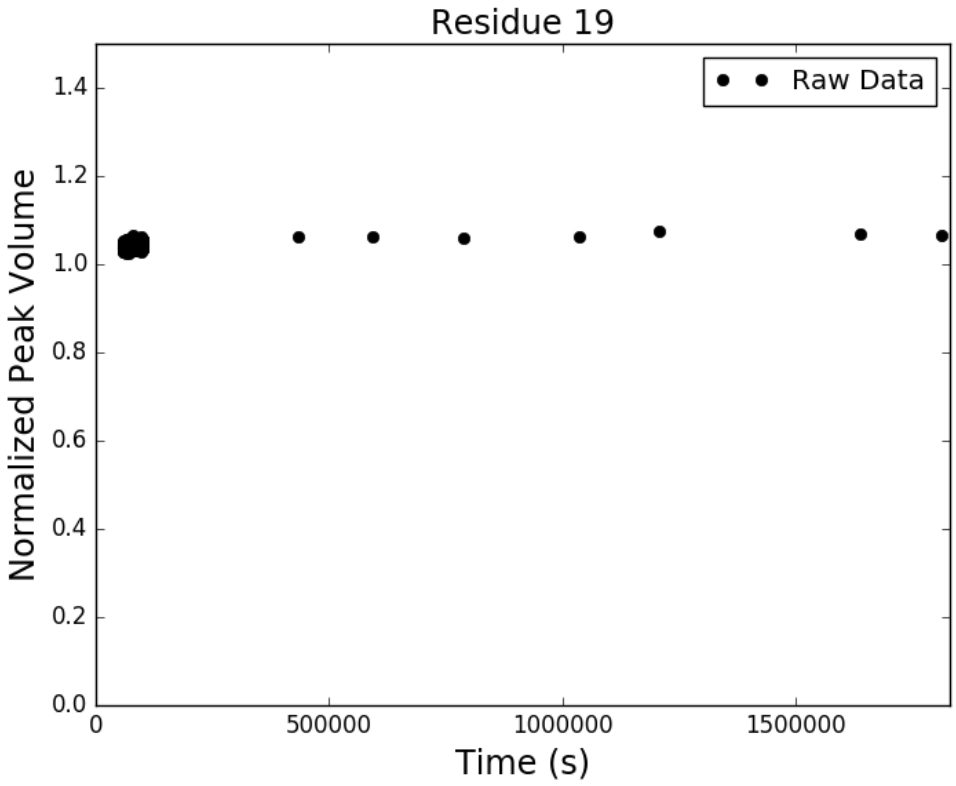


Figure C.11: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 19 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

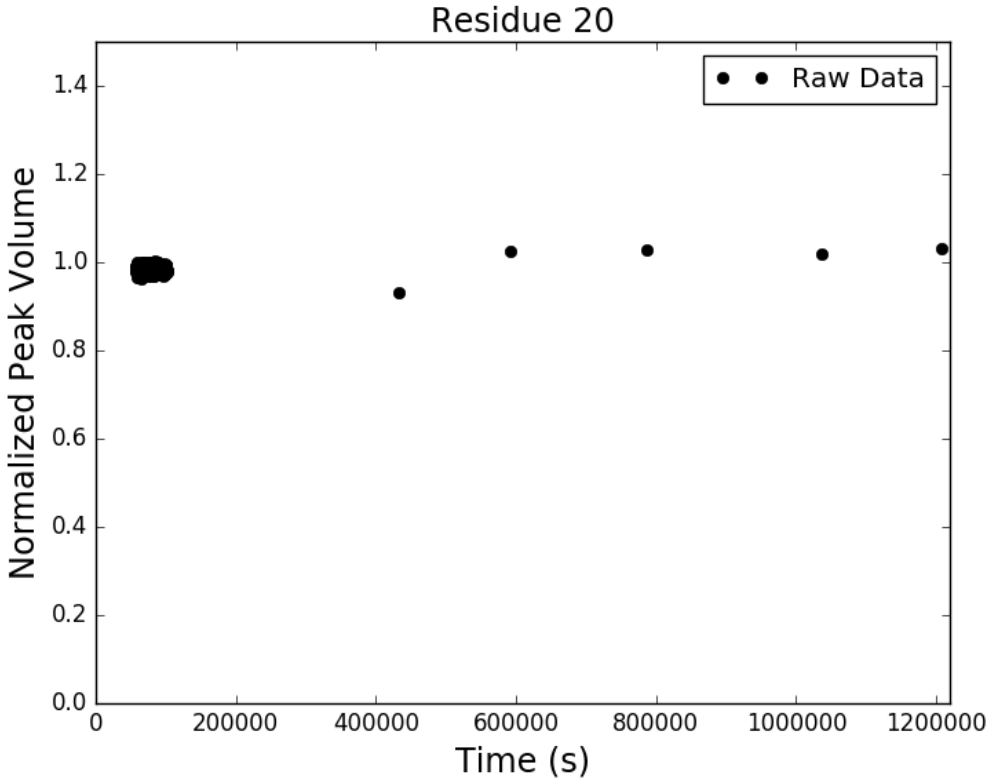


Figure C.12: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 20 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

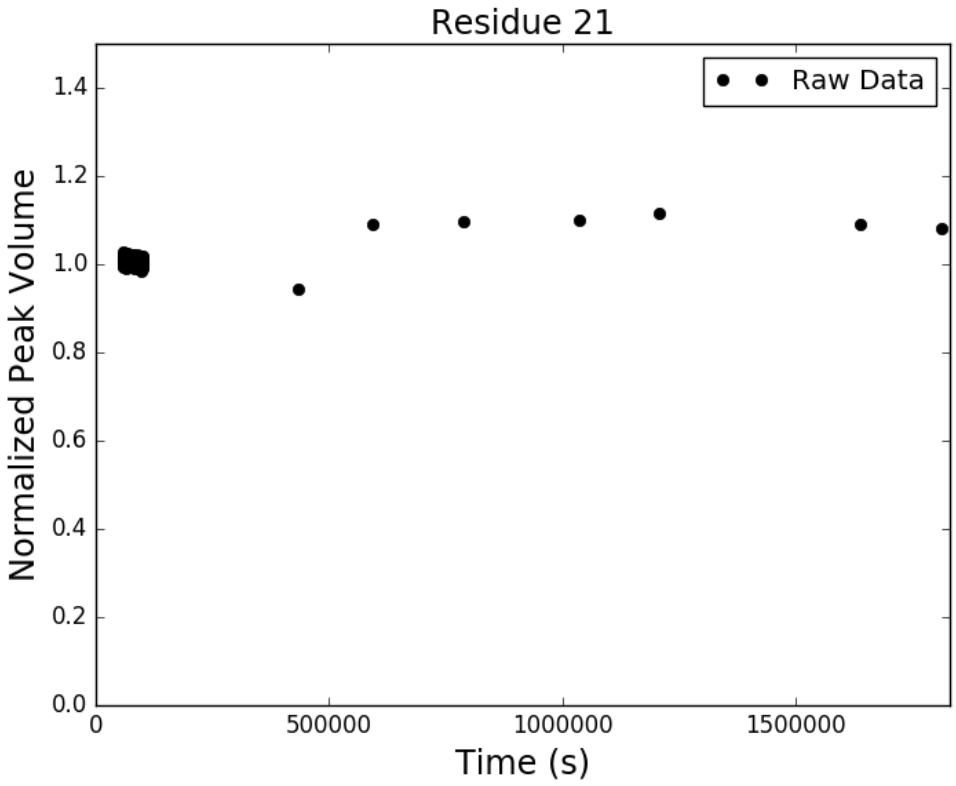


Figure C.13: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 21 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

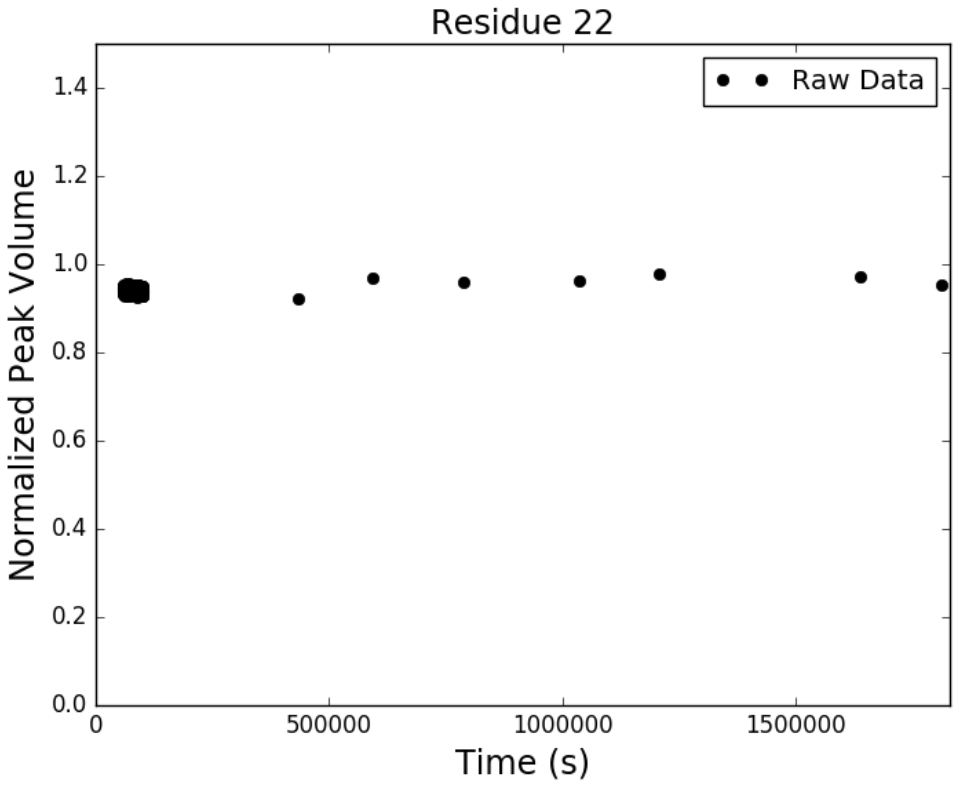


Figure C.14: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 22 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

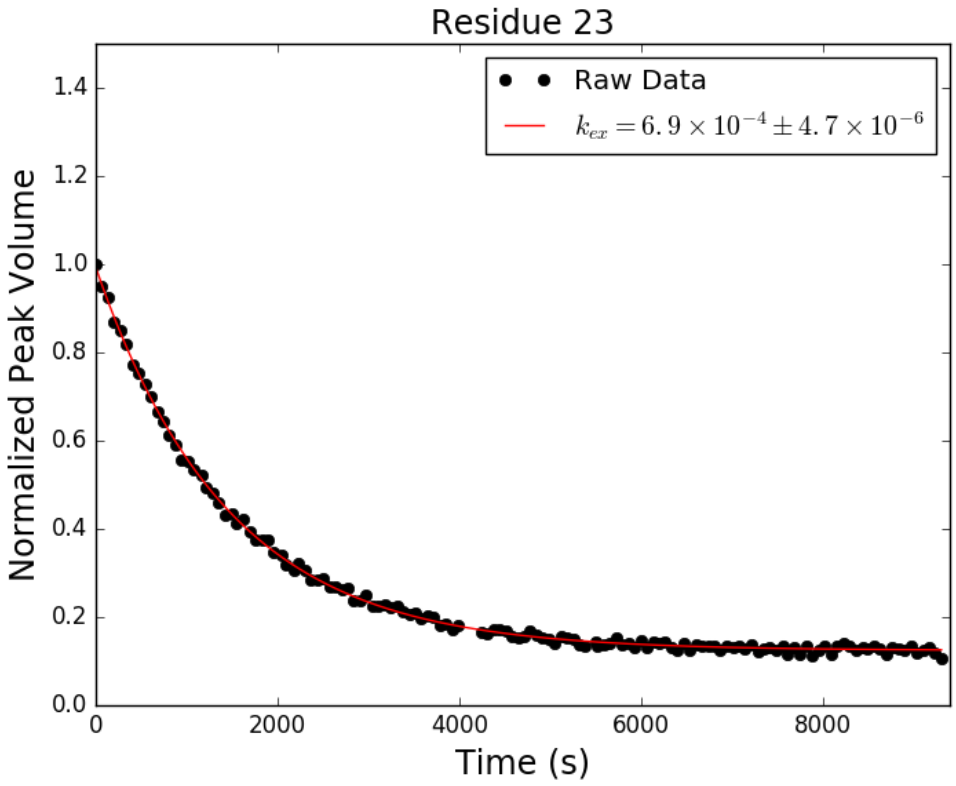


Figure C.15: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 23 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

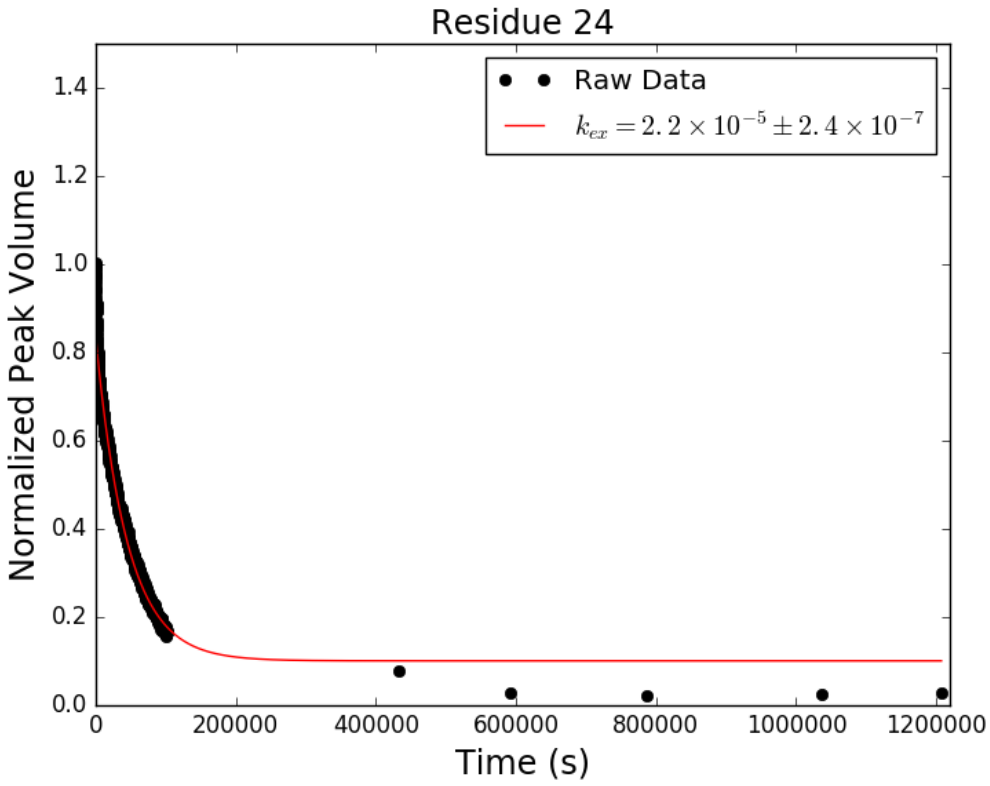


Figure C.16: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 24 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

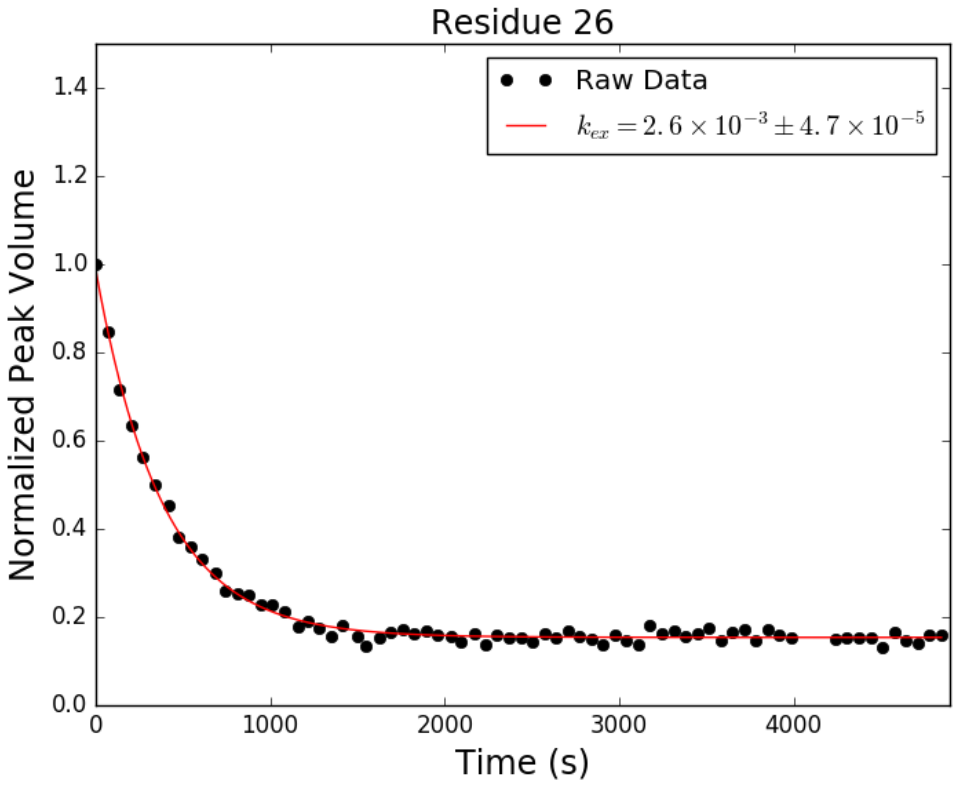


Figure C.17: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 26 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

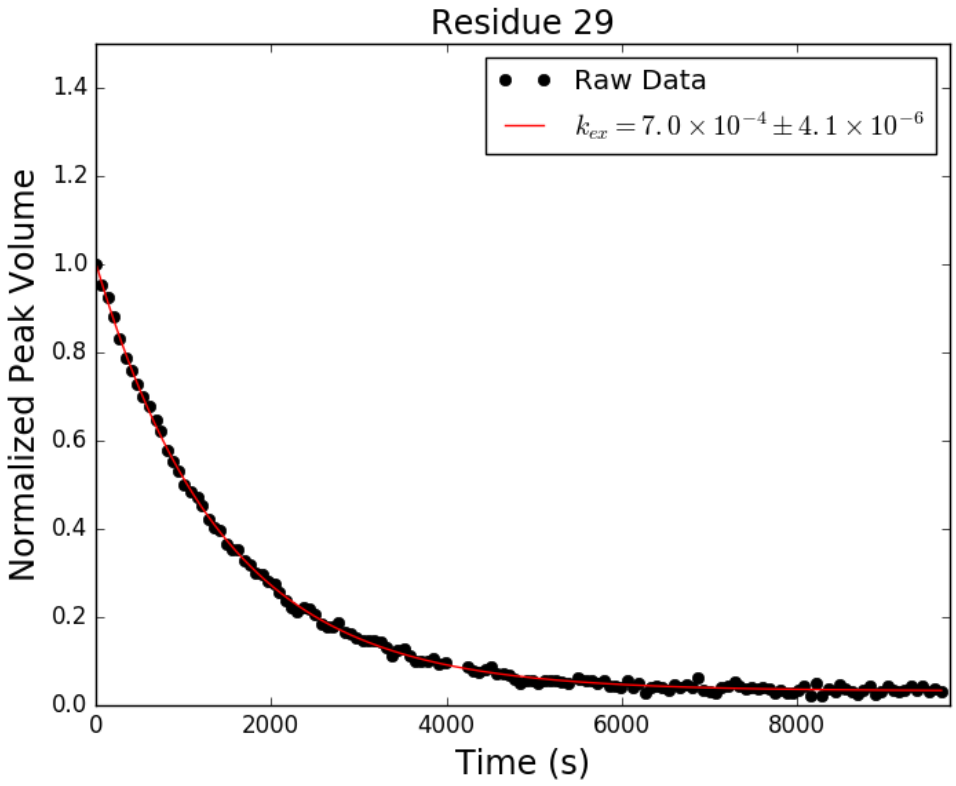


Figure C.18: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 29 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

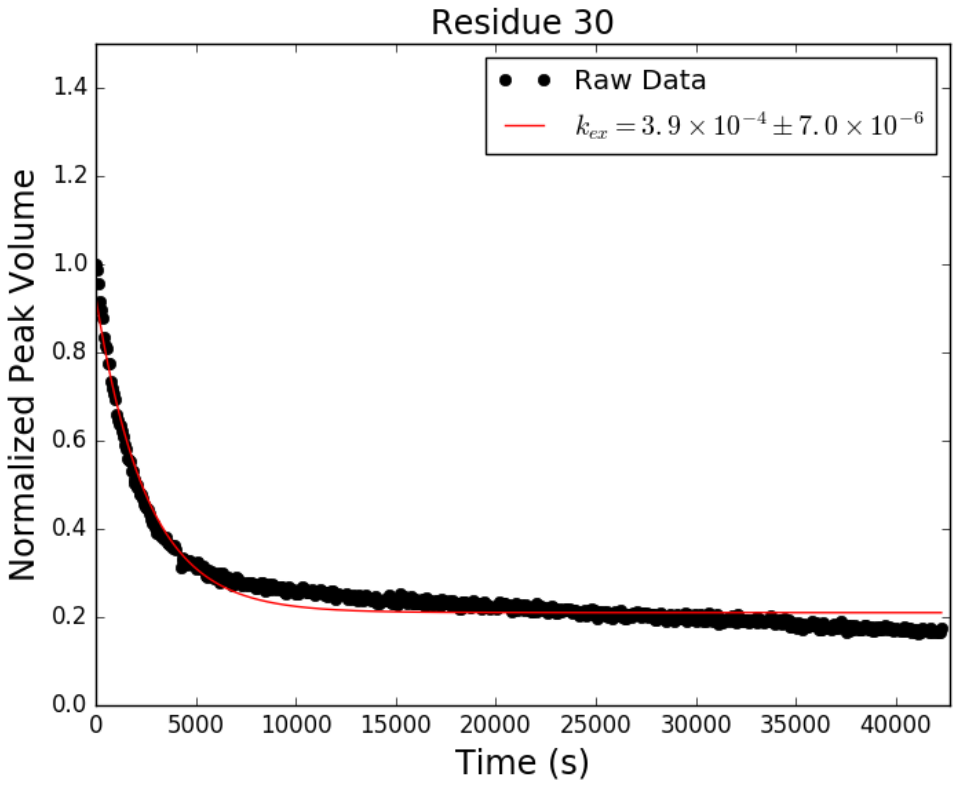


Figure C.19: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 30 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

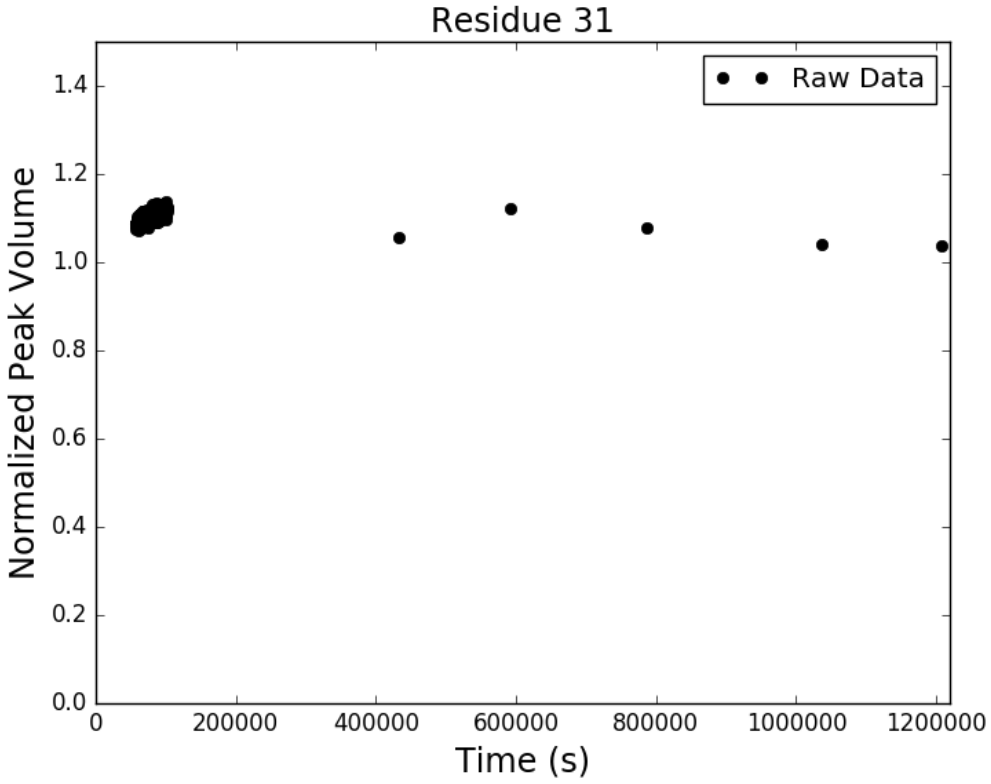


Figure C.20: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 31 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

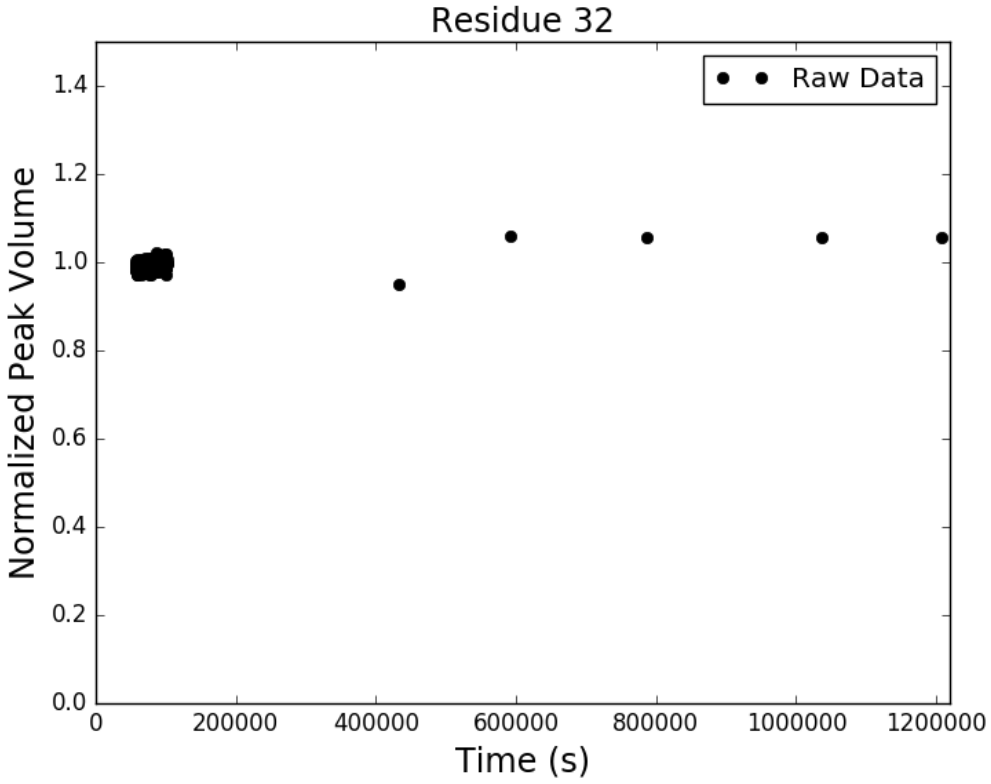


Figure C.21: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 32 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

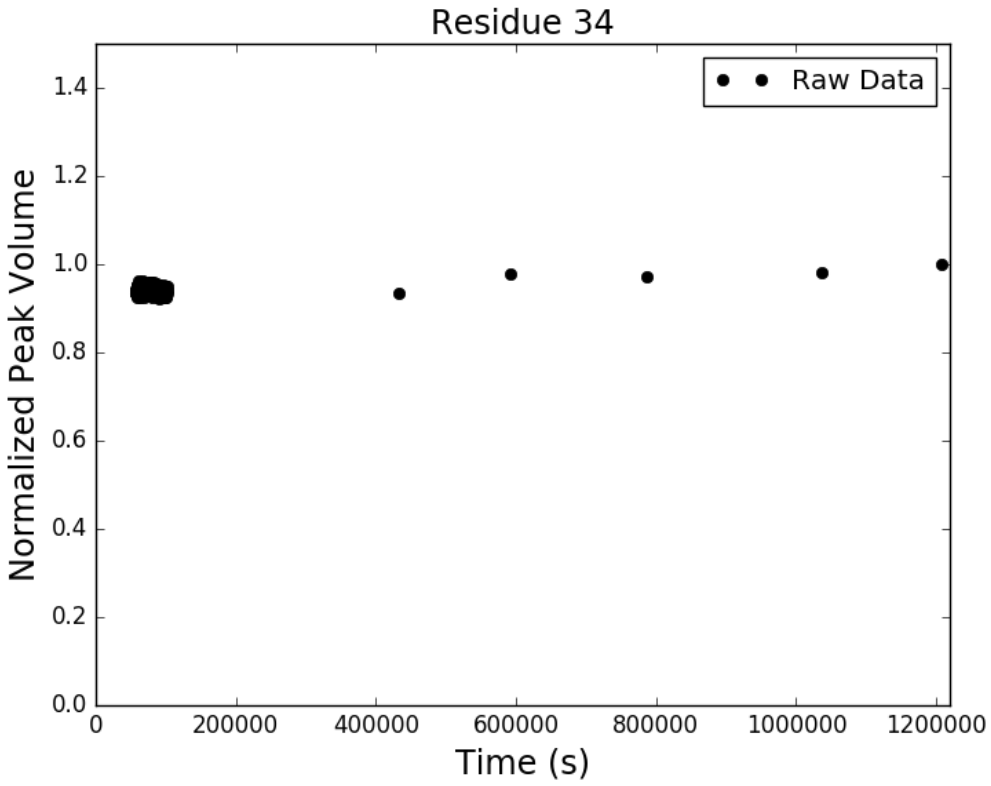


Figure C.22: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 34 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

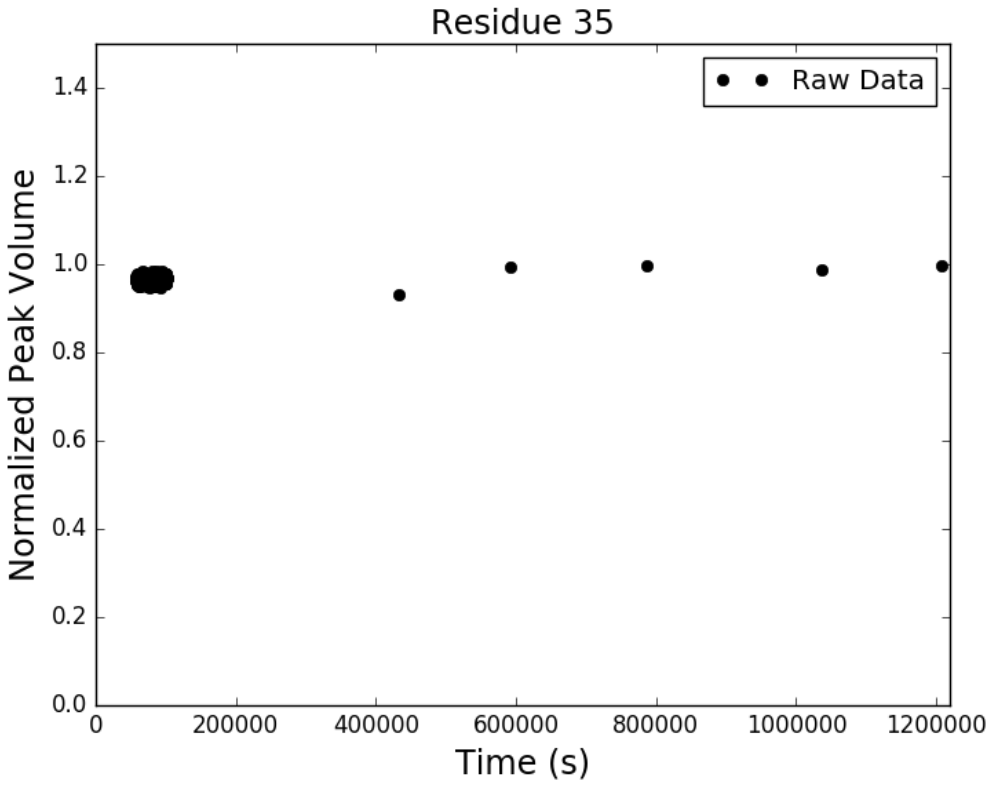


Figure C.23: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 35 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

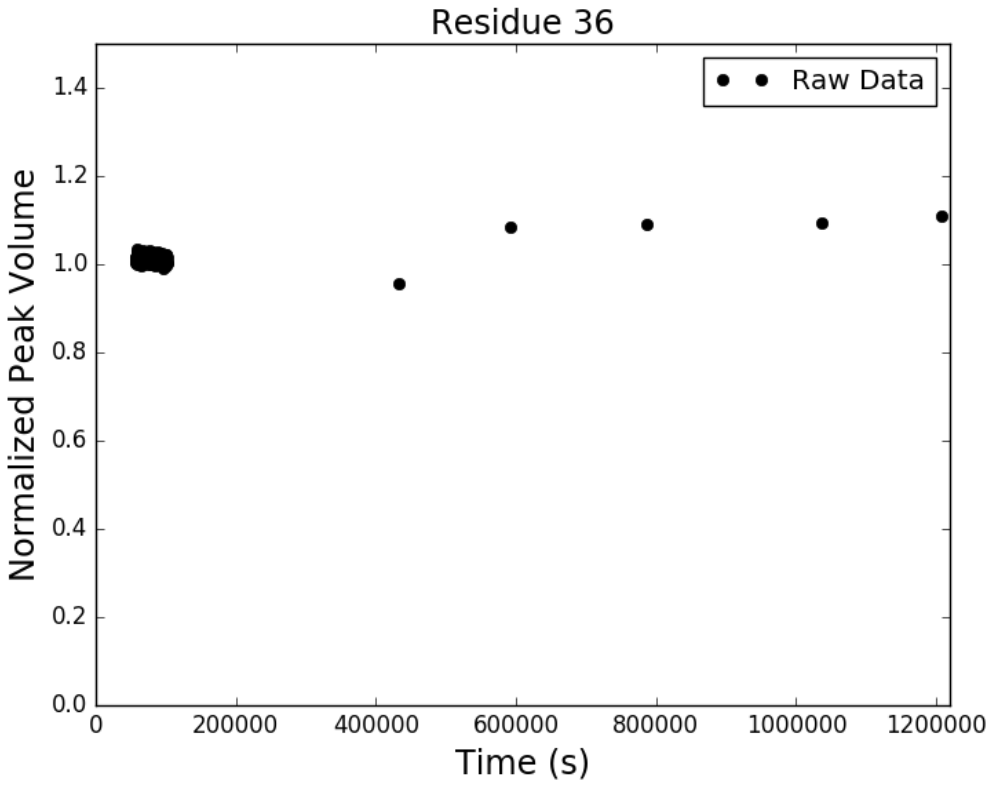


Figure C.24: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 36 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

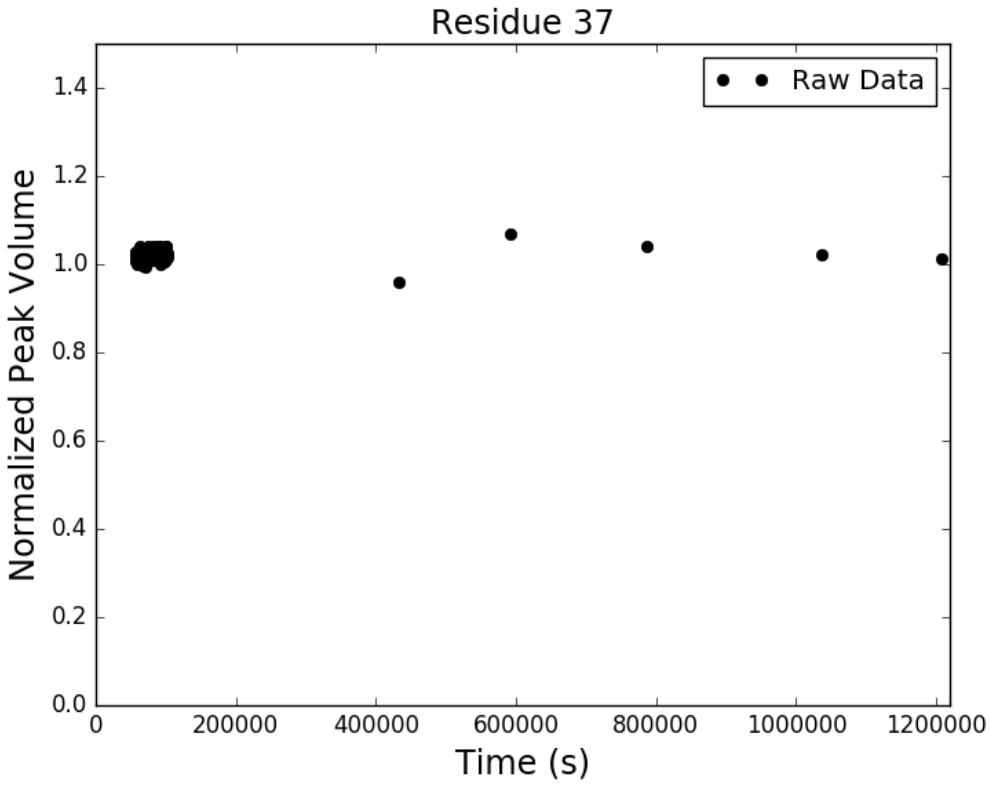


Figure C.25: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 37 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

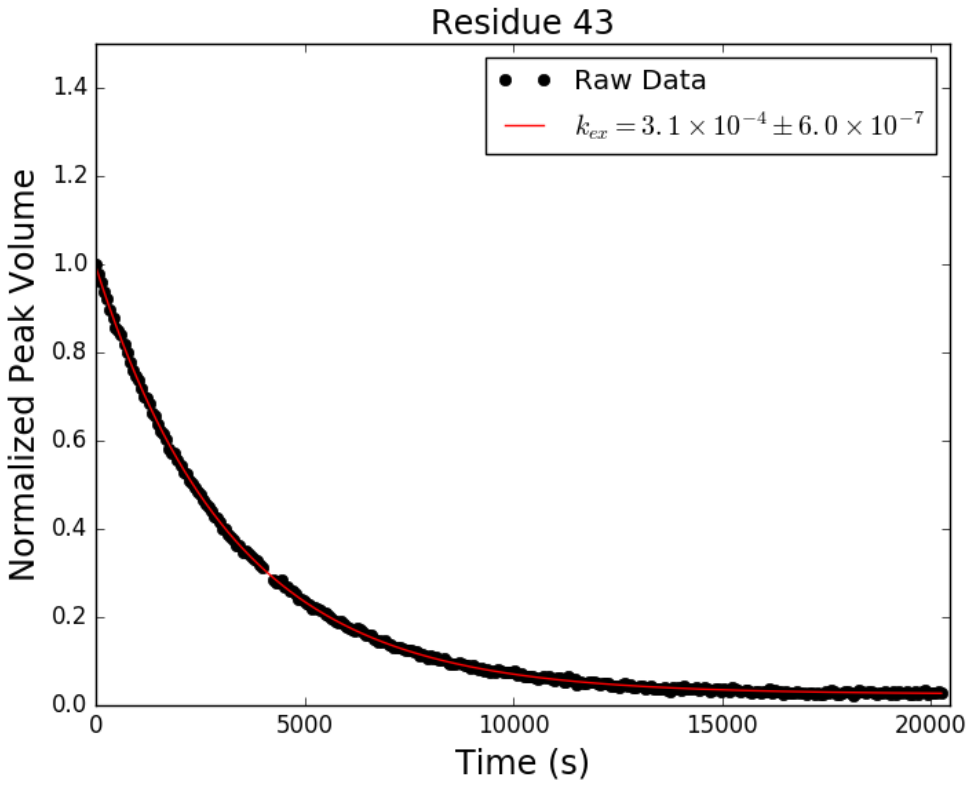


Figure C.26: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 43 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

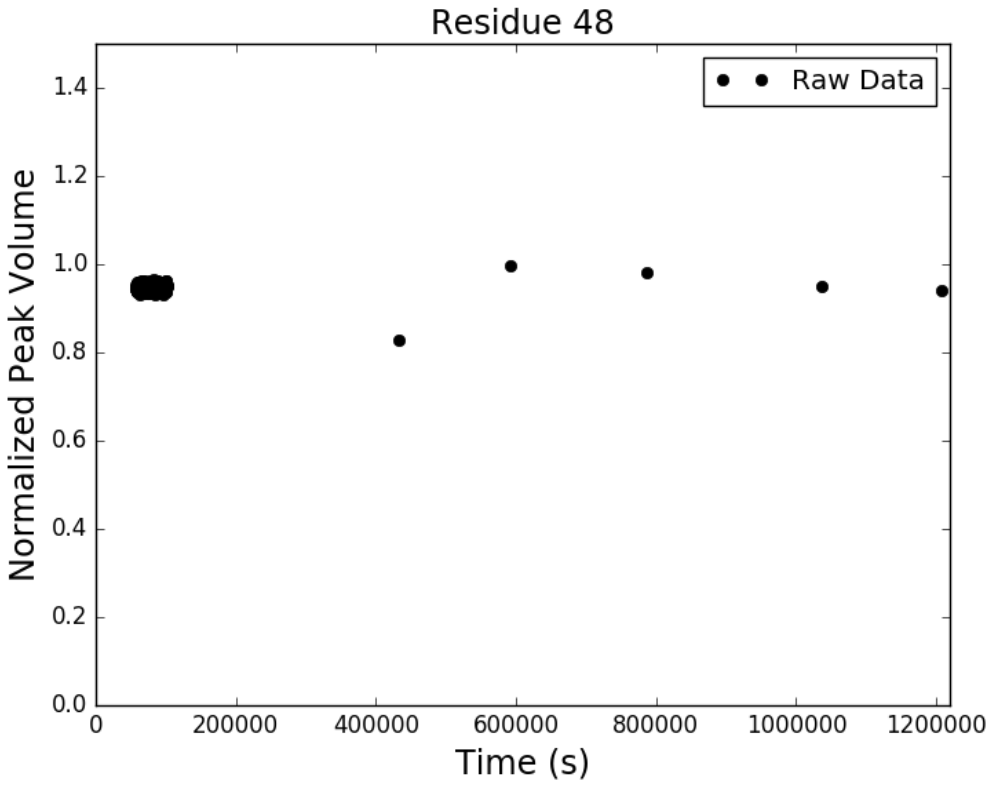


Figure C.27: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 48 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

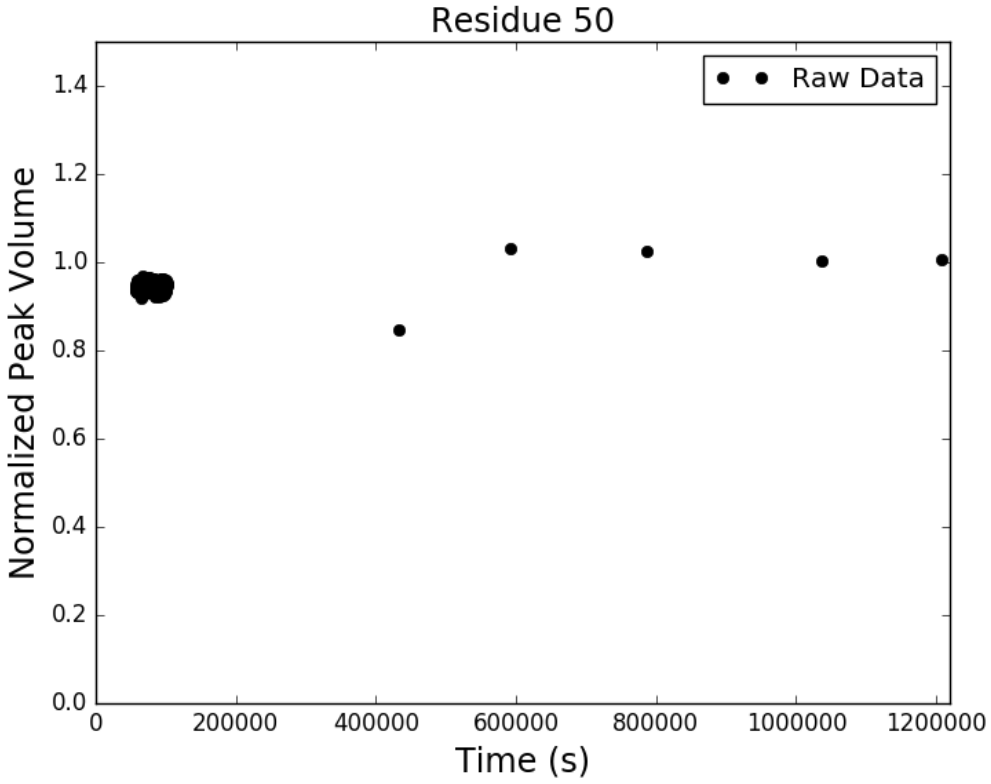


Figure C.28: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 50 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

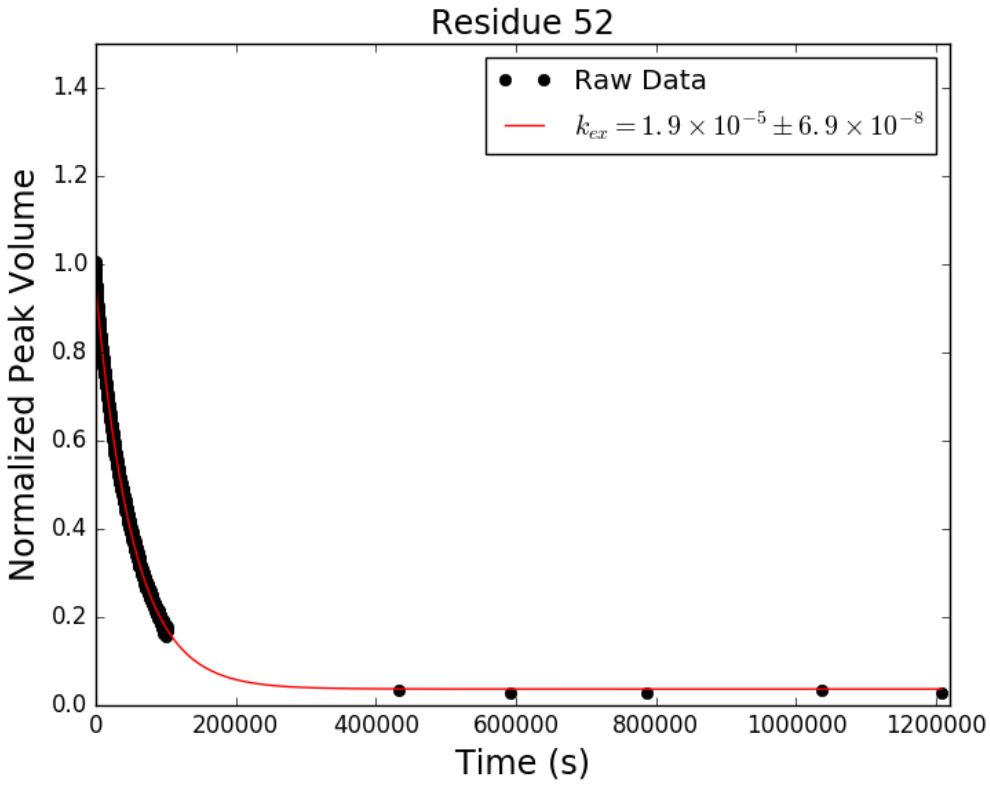


Figure C.29: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 52 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

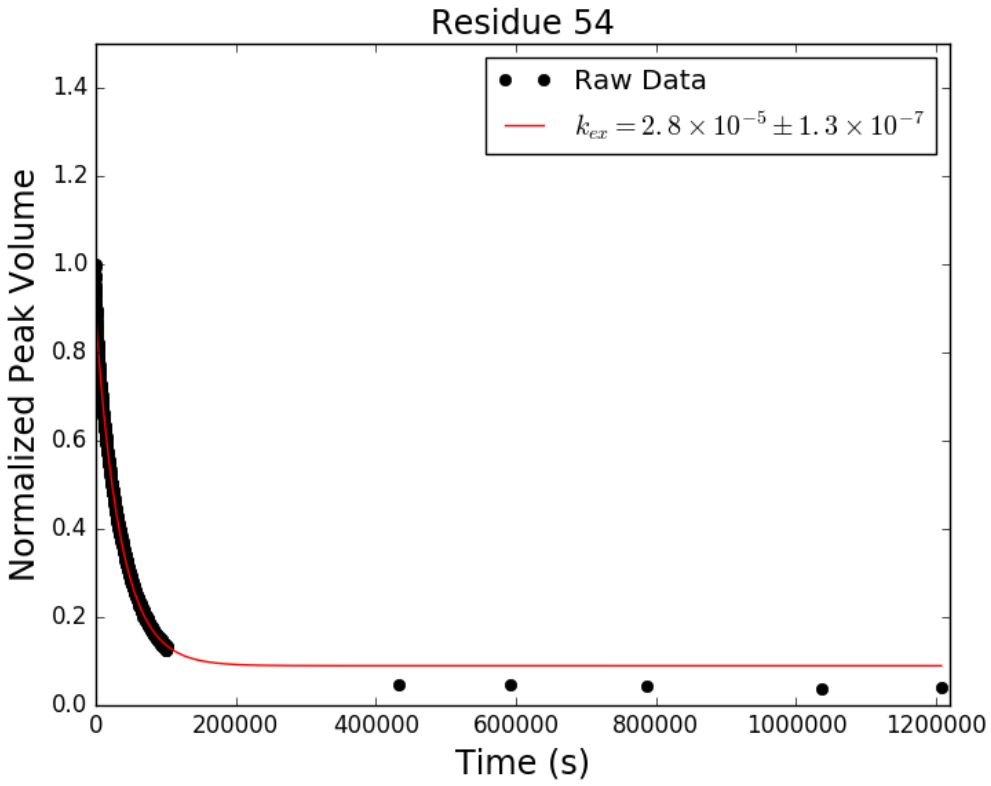


Figure C.30: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 54 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

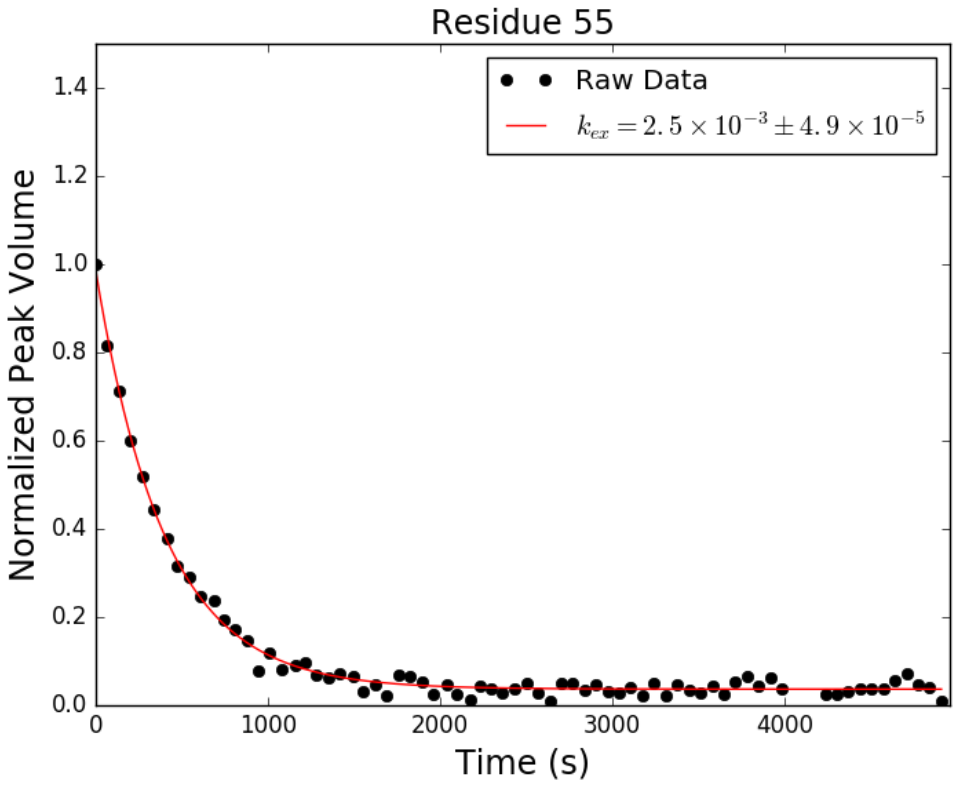


Figure C.31: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 55 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

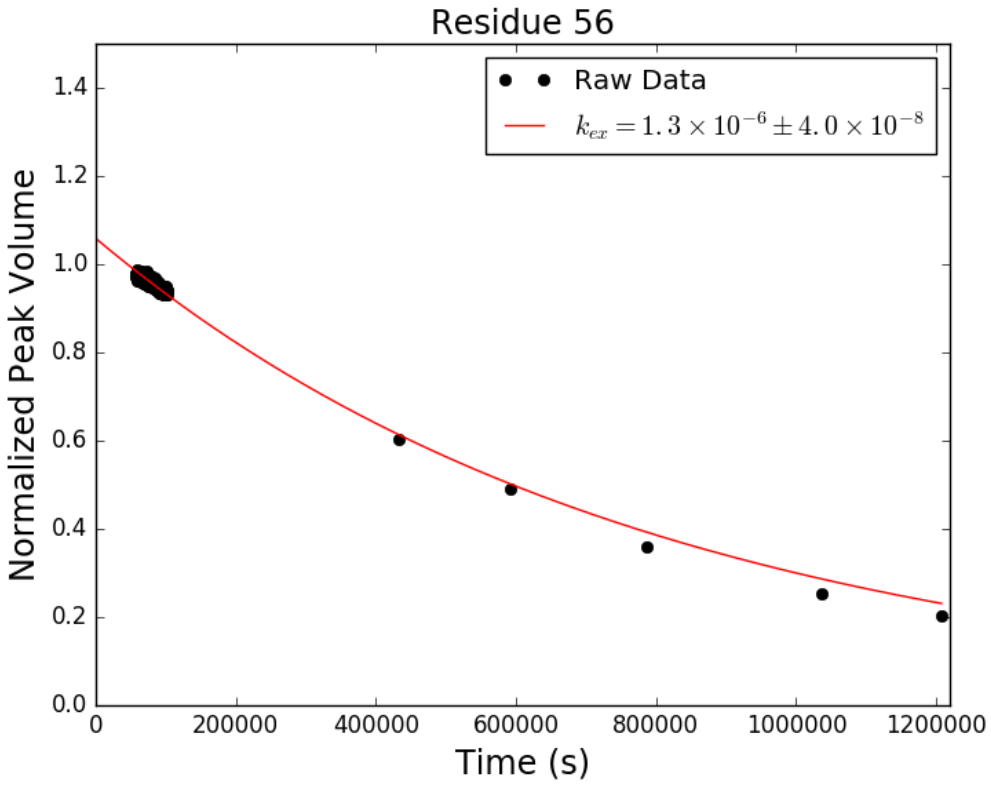


Figure C.32: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 56 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

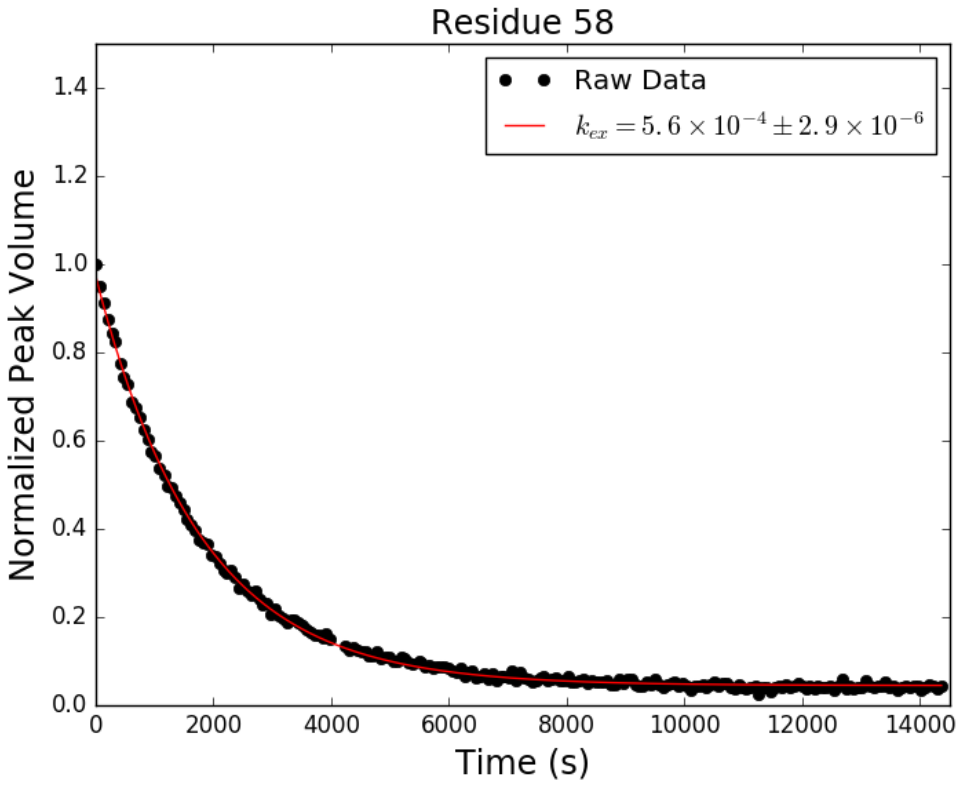


Figure C.33: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 58 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

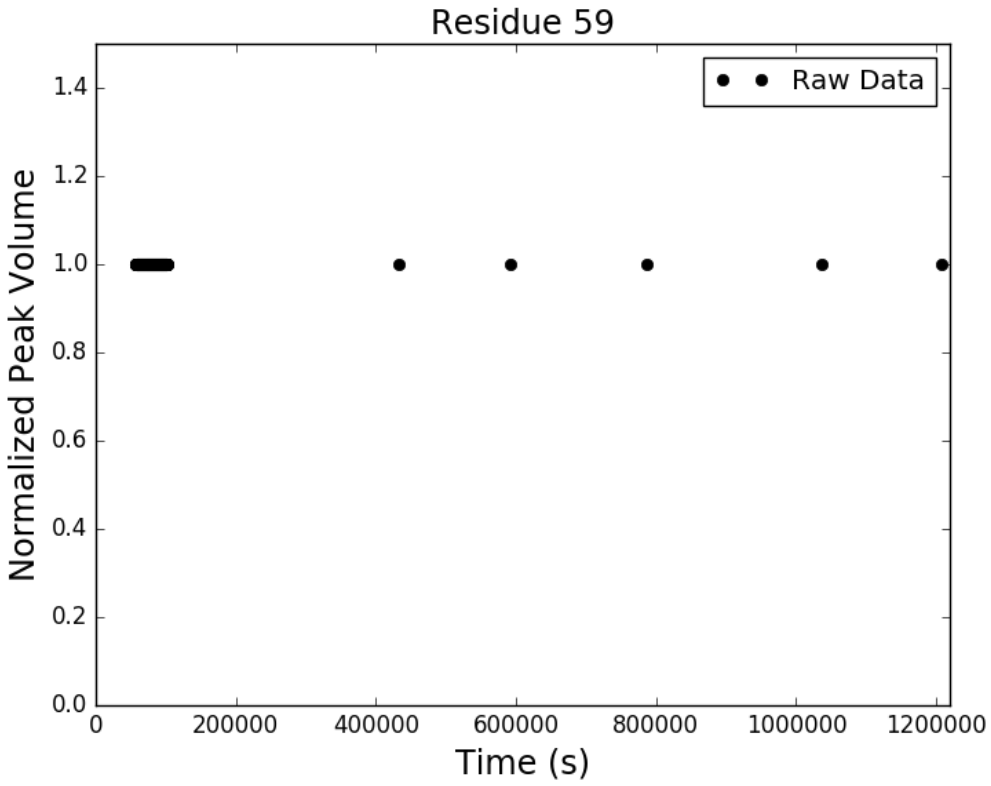


Figure C.34: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 59 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

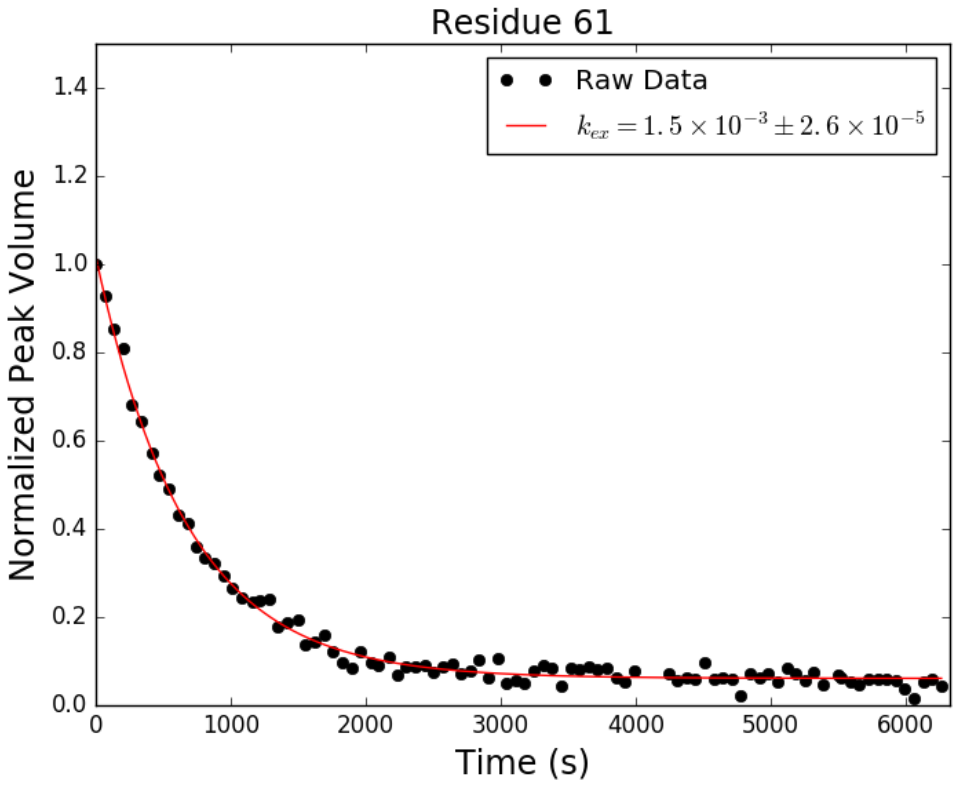


Figure C.35: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 61 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

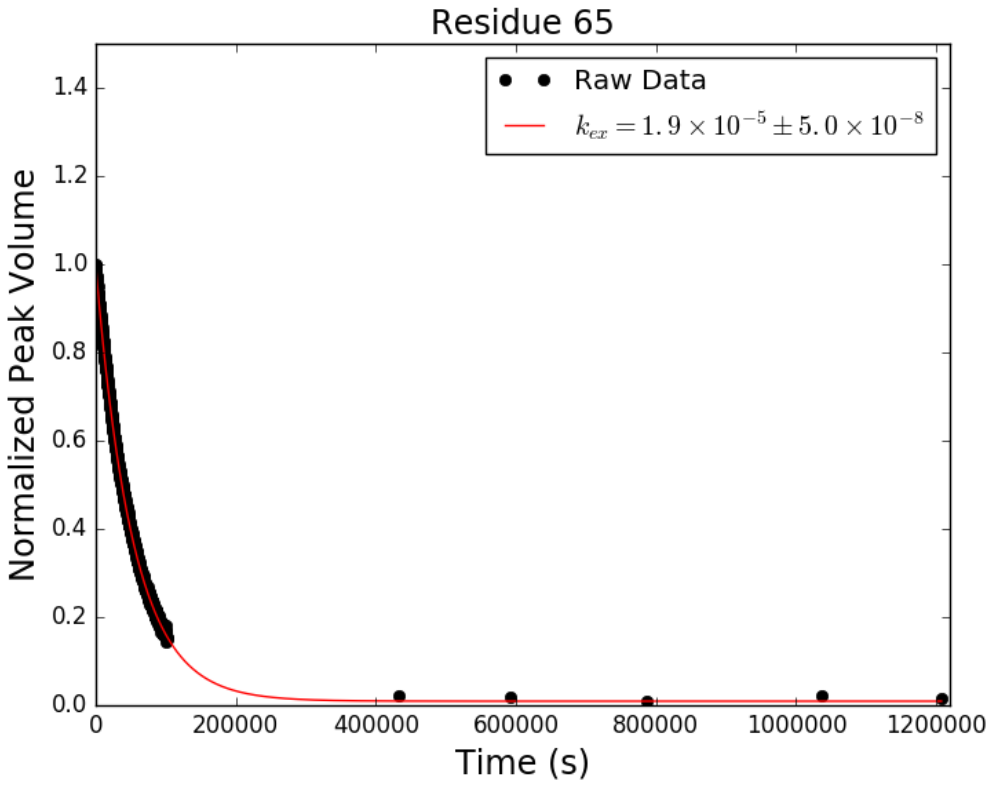


Figure C.36: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 65 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

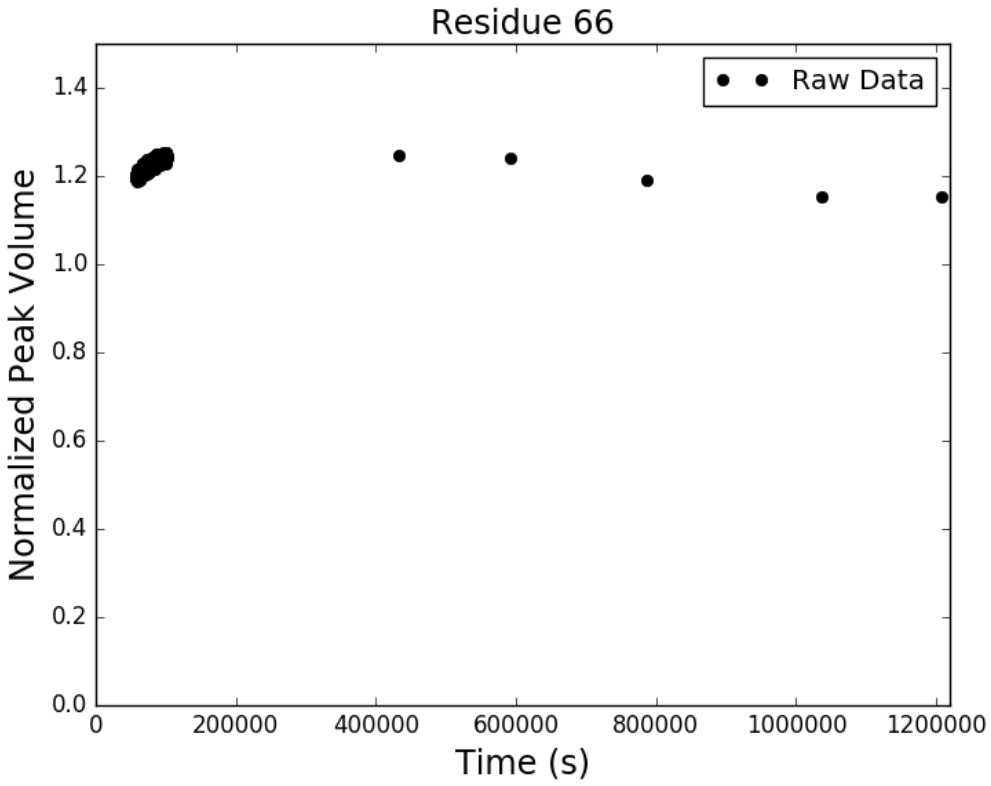


Figure C.37: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 66 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

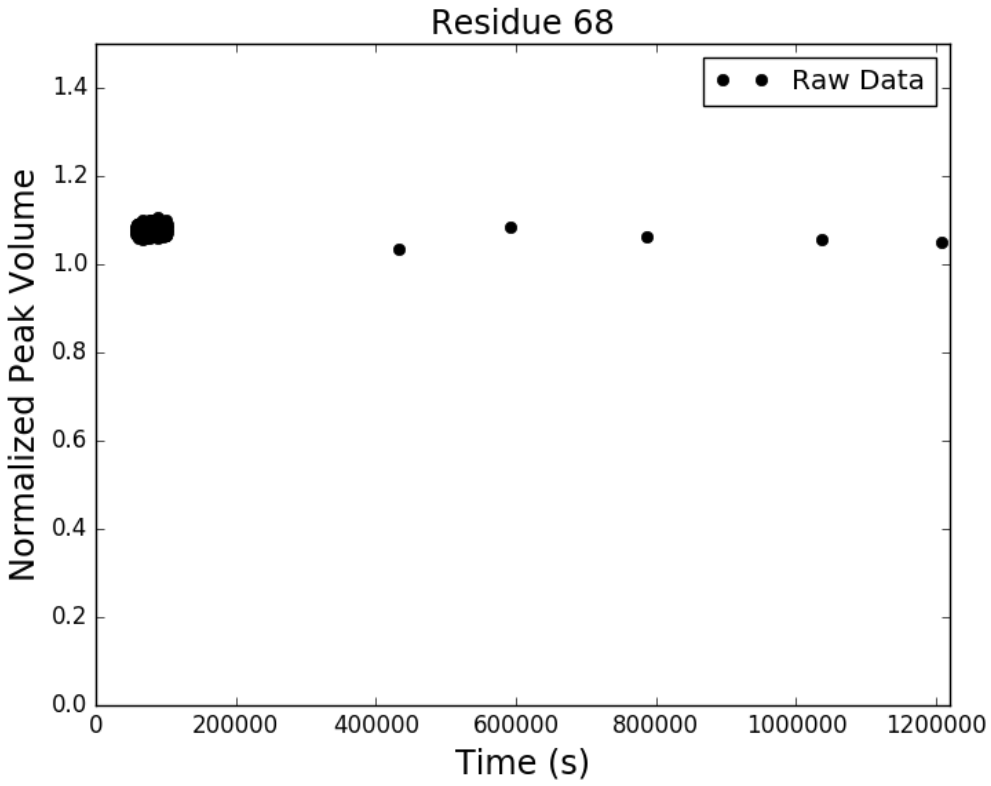


Figure C.38: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 68 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

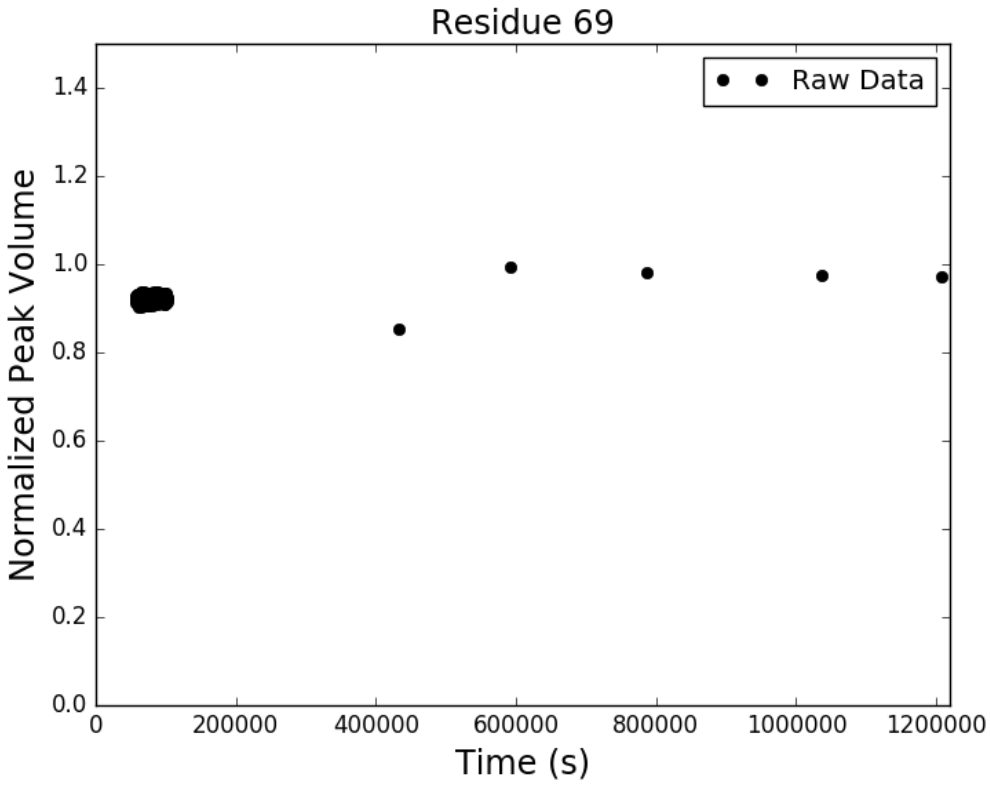


Figure C.39: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 69 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

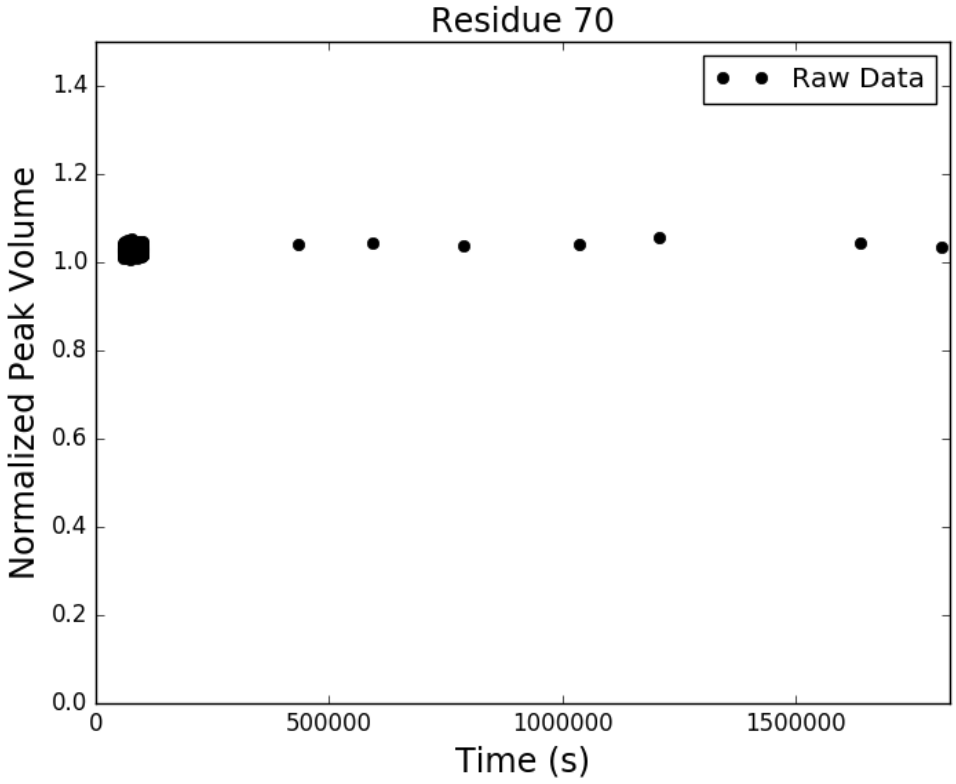


Figure C.40: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 70 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

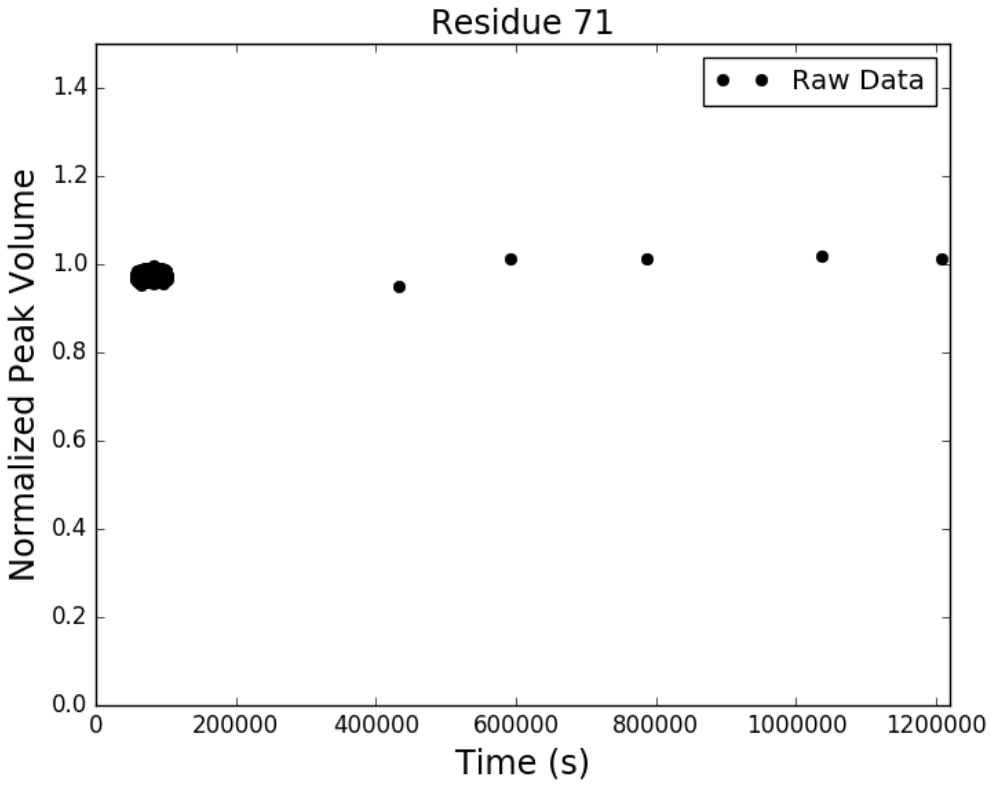


Figure C.41: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 71 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

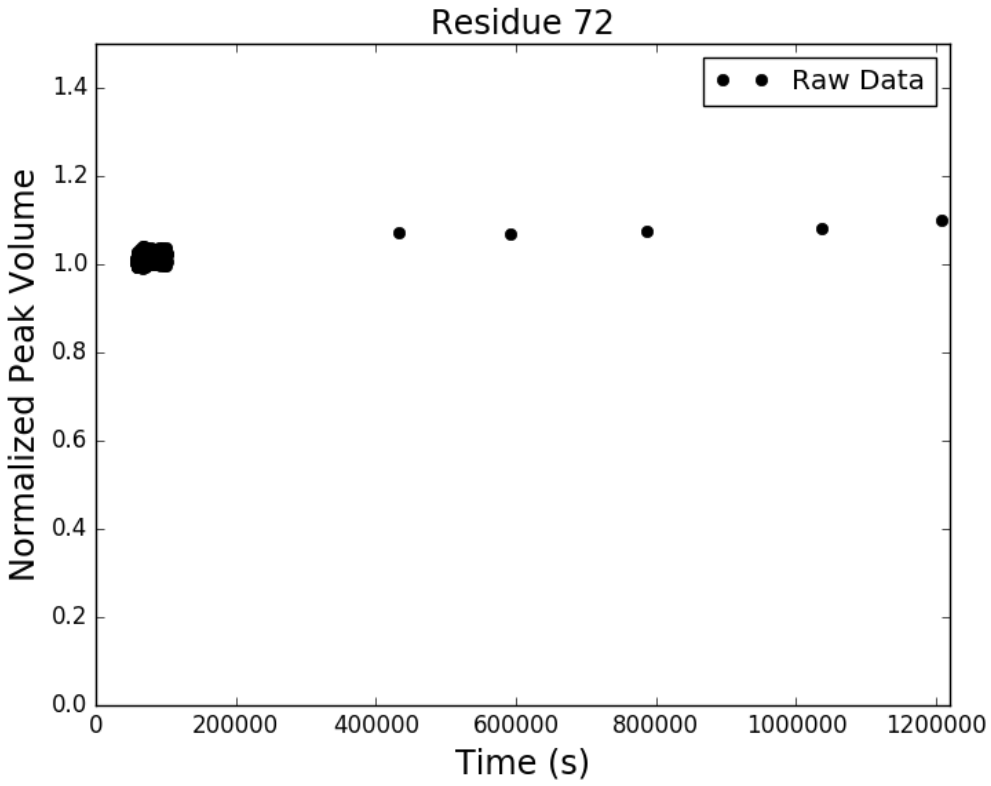


Figure C.42: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 72 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

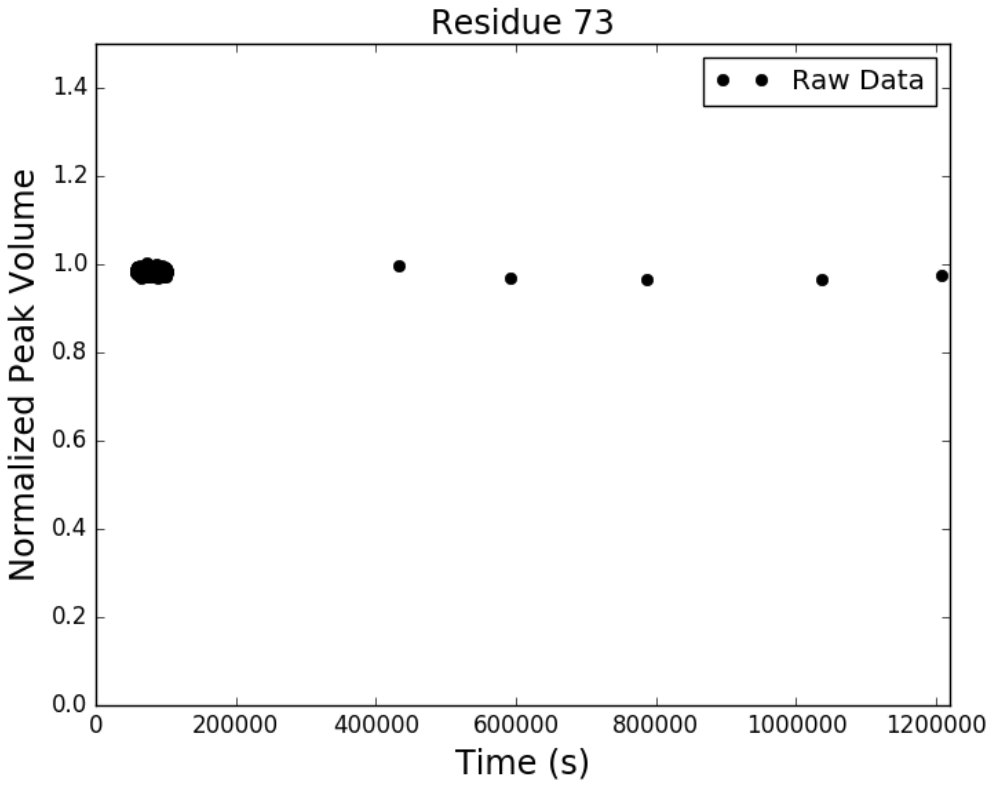


Figure C.43: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 73 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

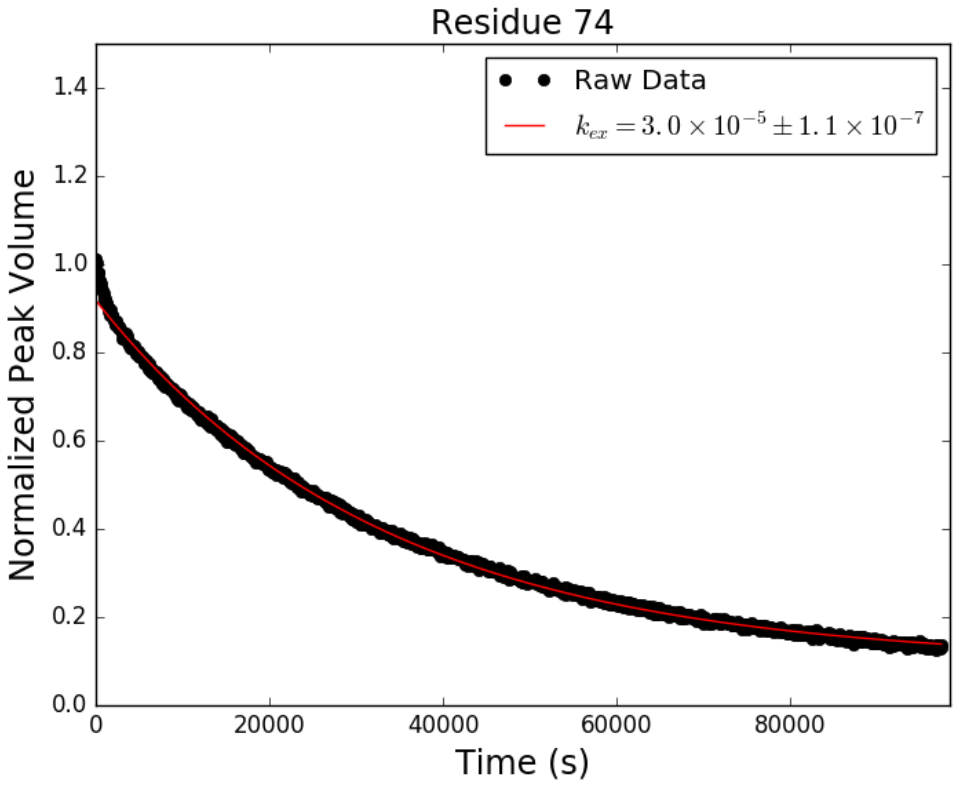


Figure C.44: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 74 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

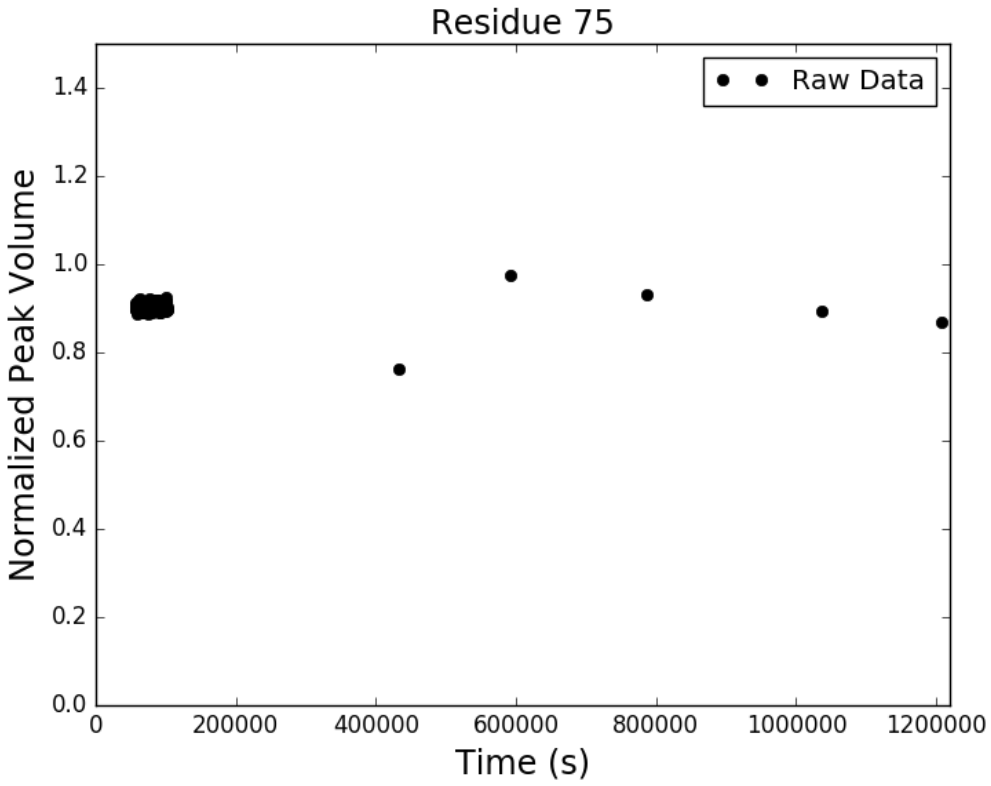


Figure C.45: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 75 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

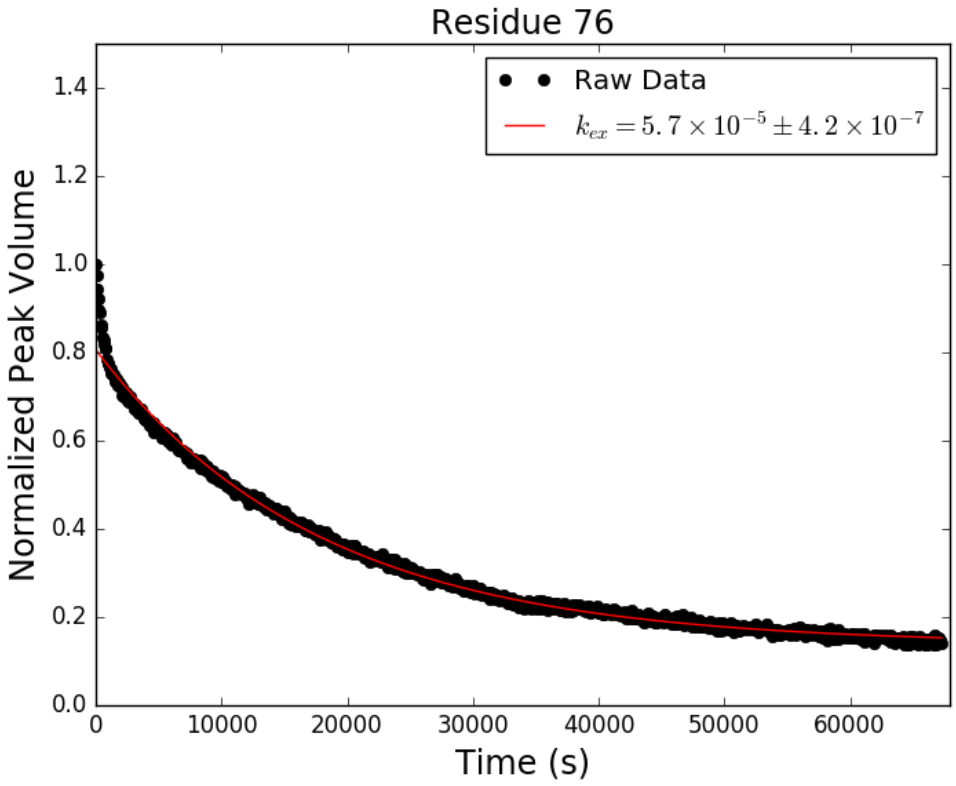


Figure C.46: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 76 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

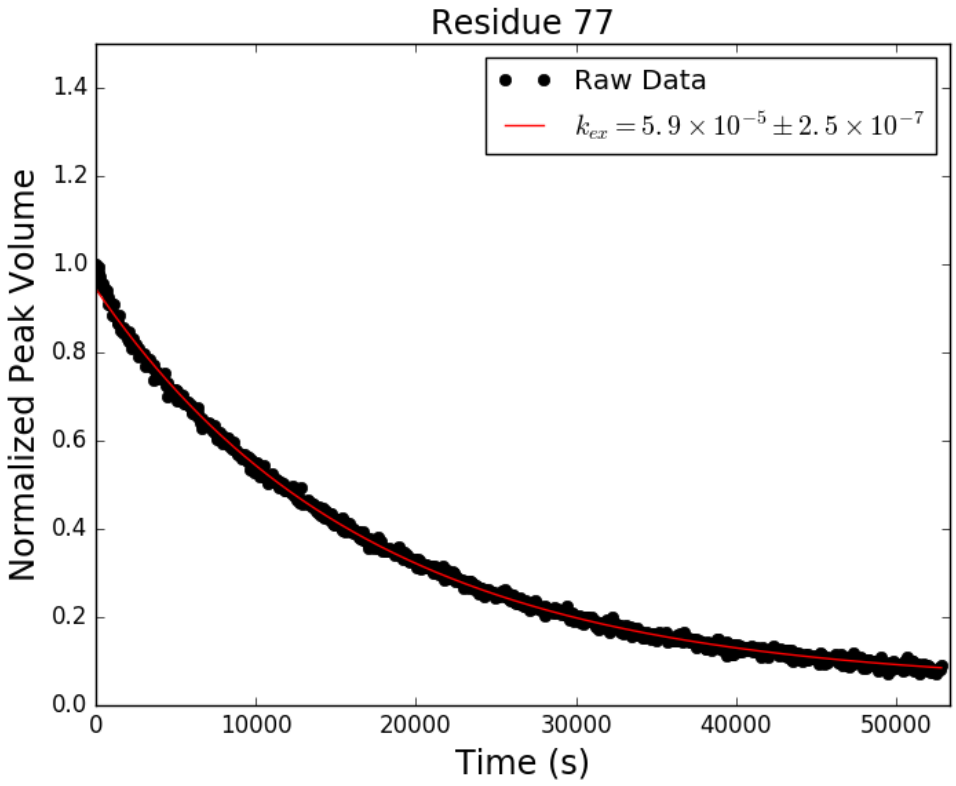


Figure C.47: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 77 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

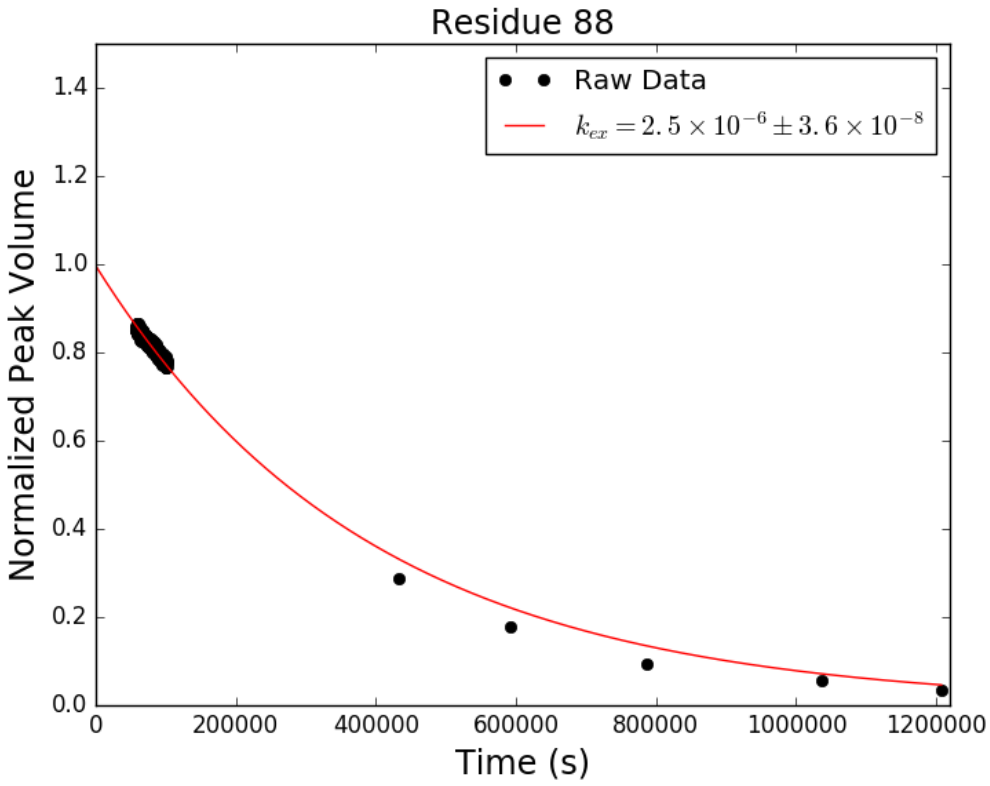


Figure C.48: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 88 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

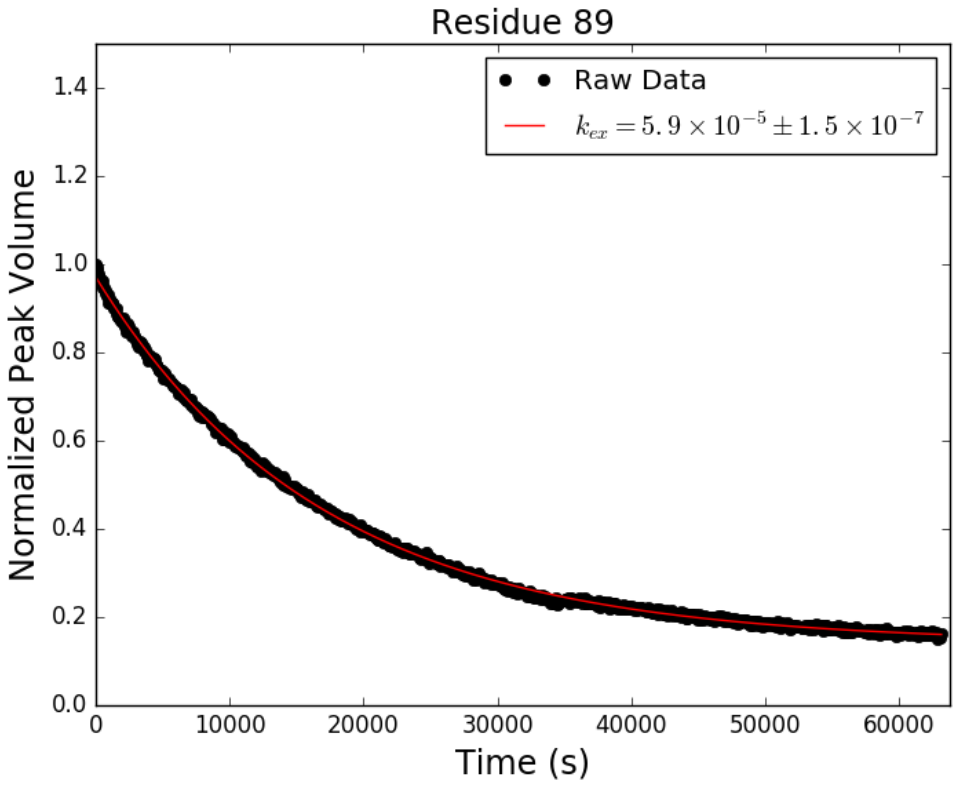


Figure C.49: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 89 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

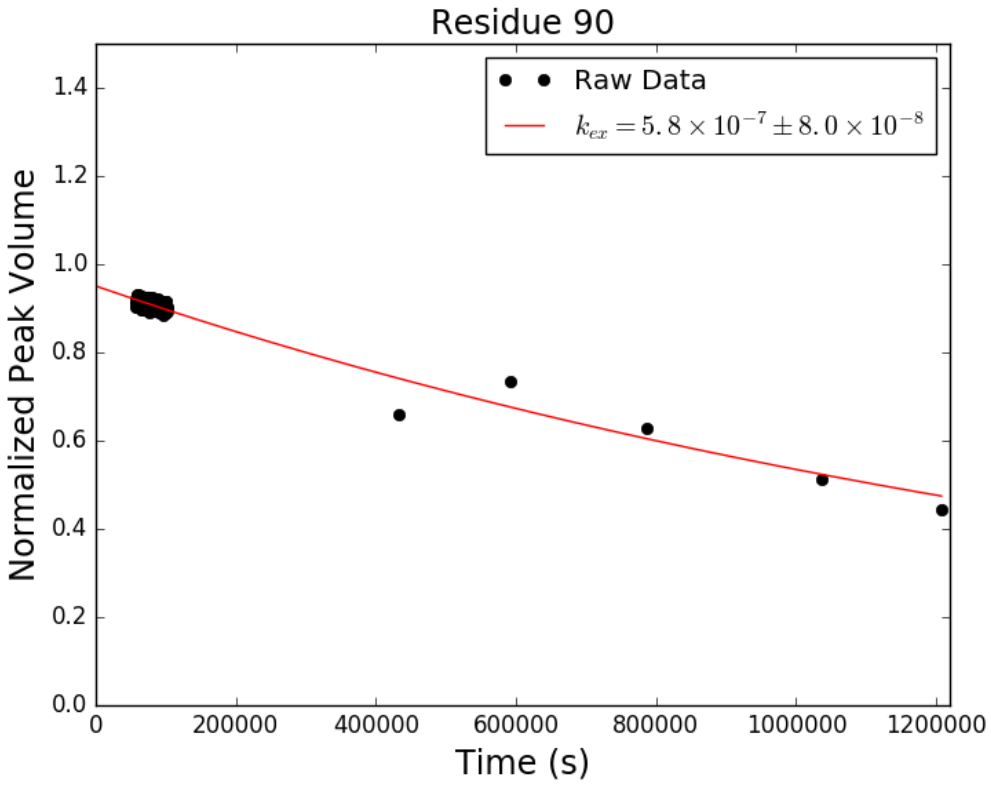


Figure C.50: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 90 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

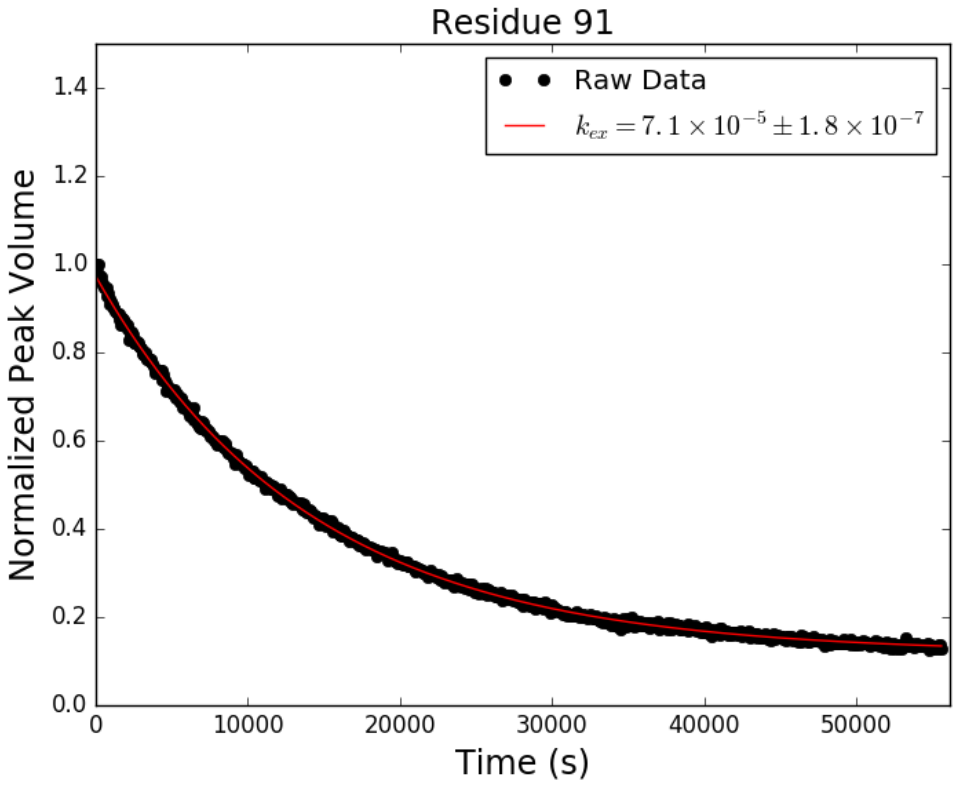


Figure C.51: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 91 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

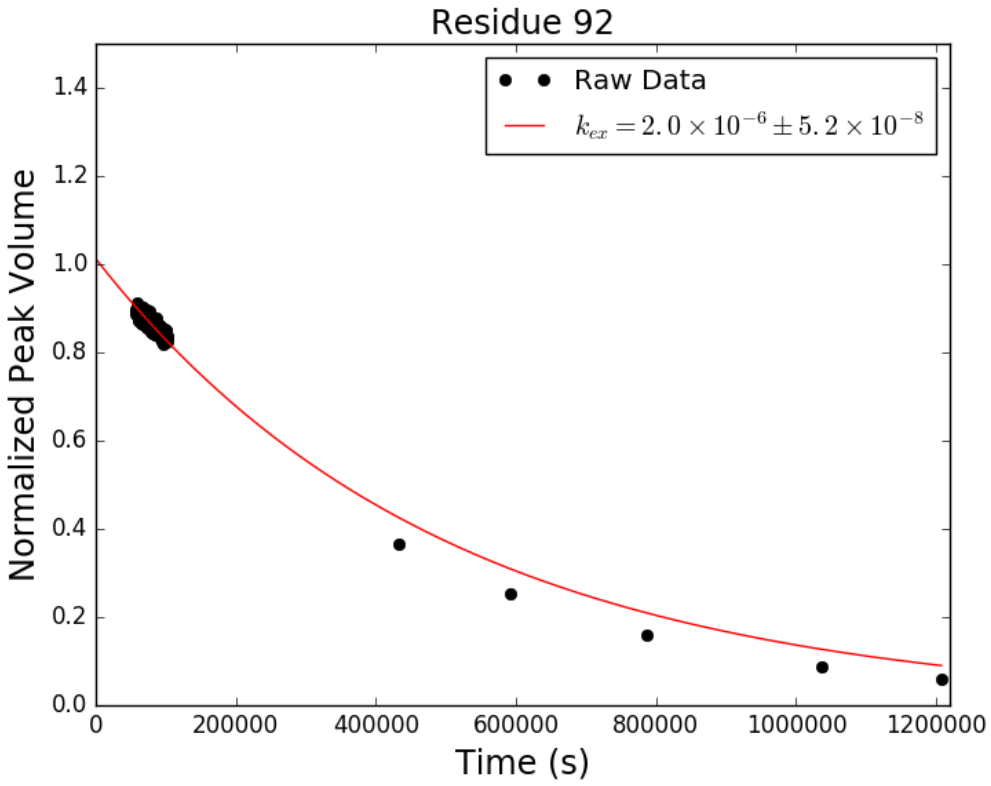


Figure C.52: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 92 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

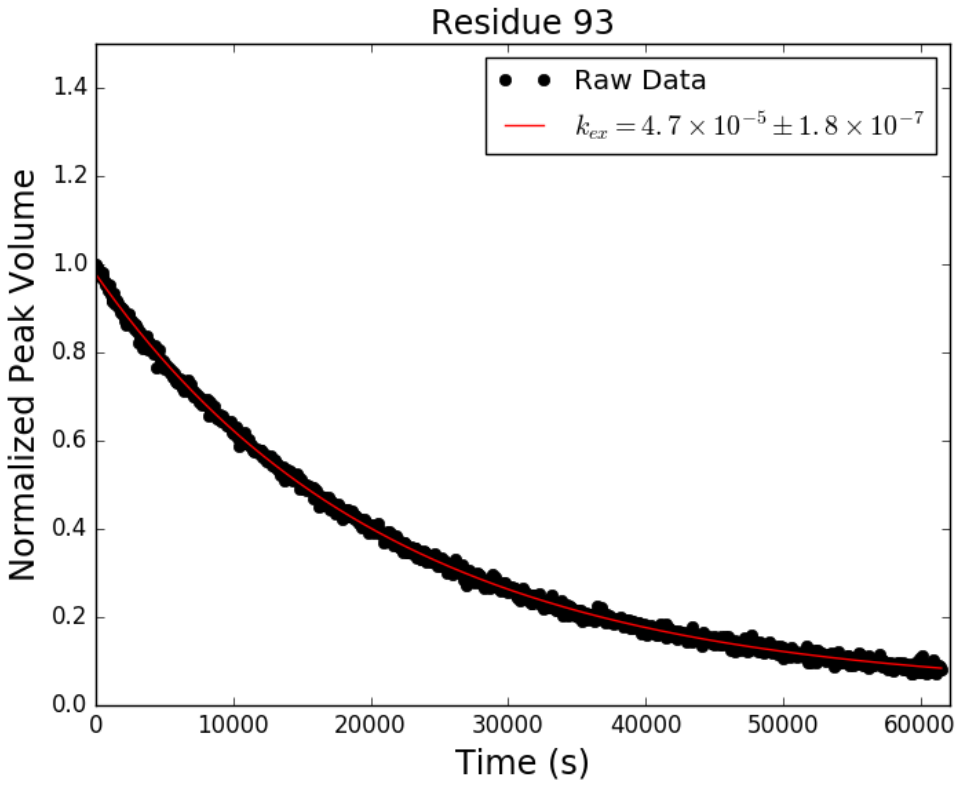


Figure C.53: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 93 of pWT. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

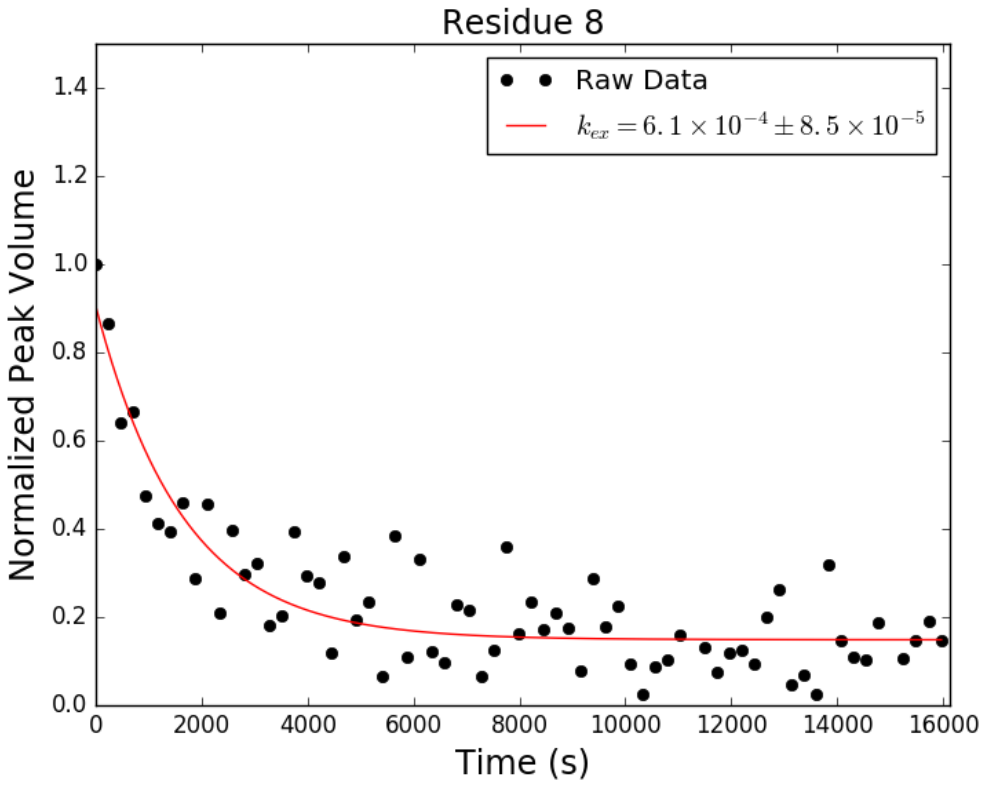


Figure C.54: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 8 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

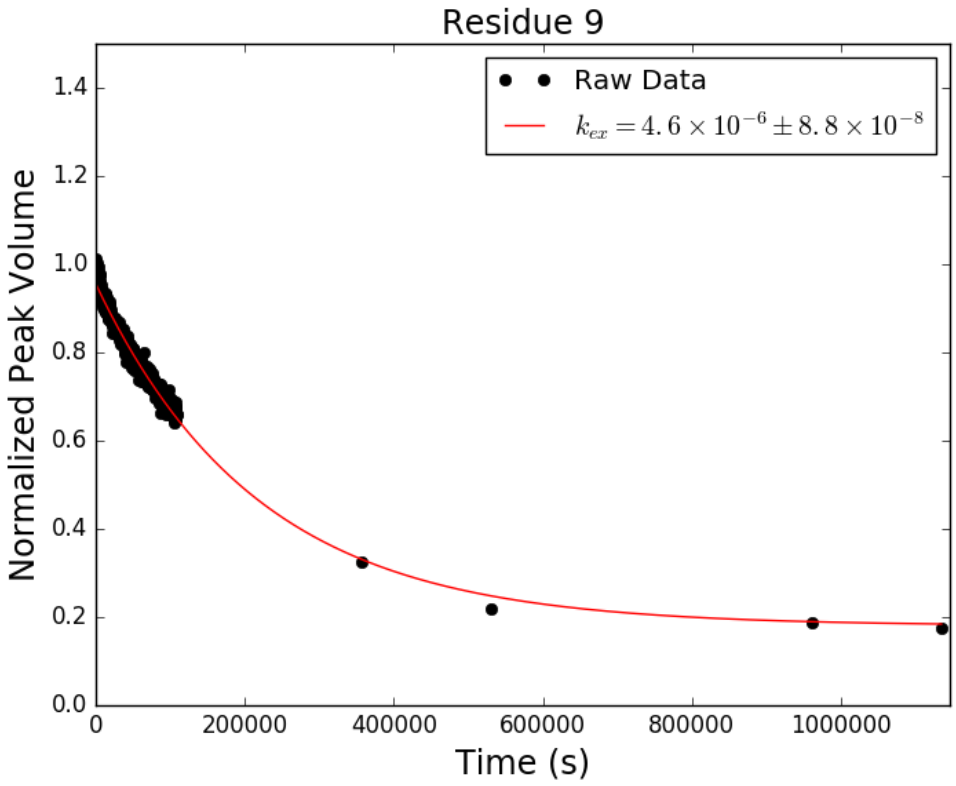


Figure C.55: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 9 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

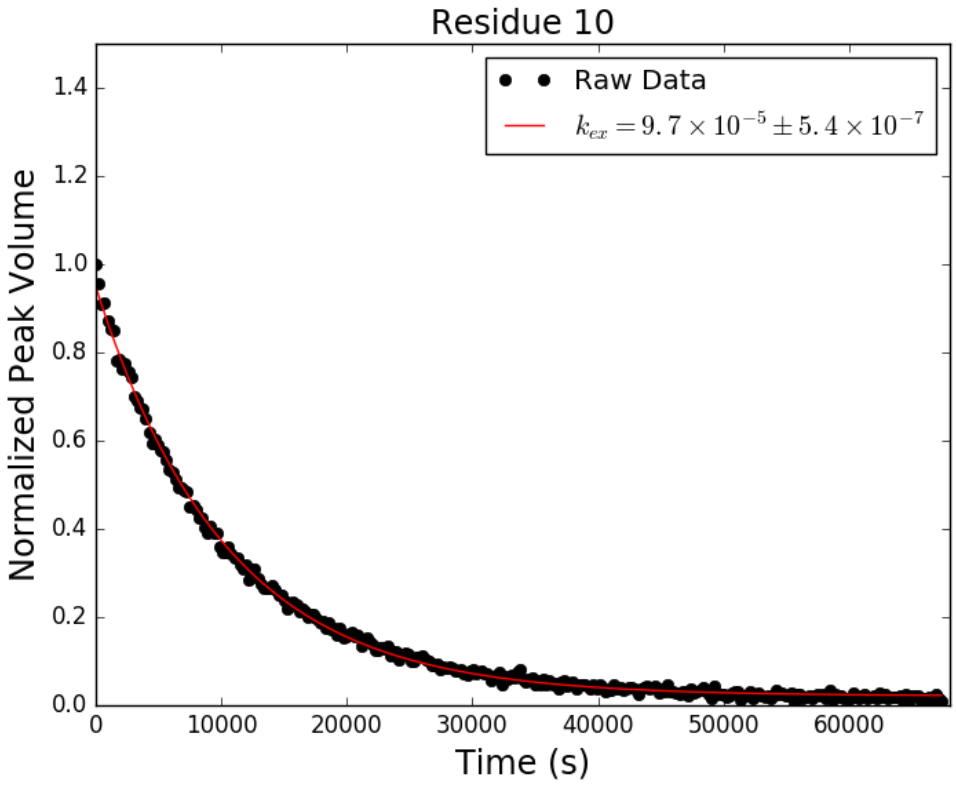


Figure C.56: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 10 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

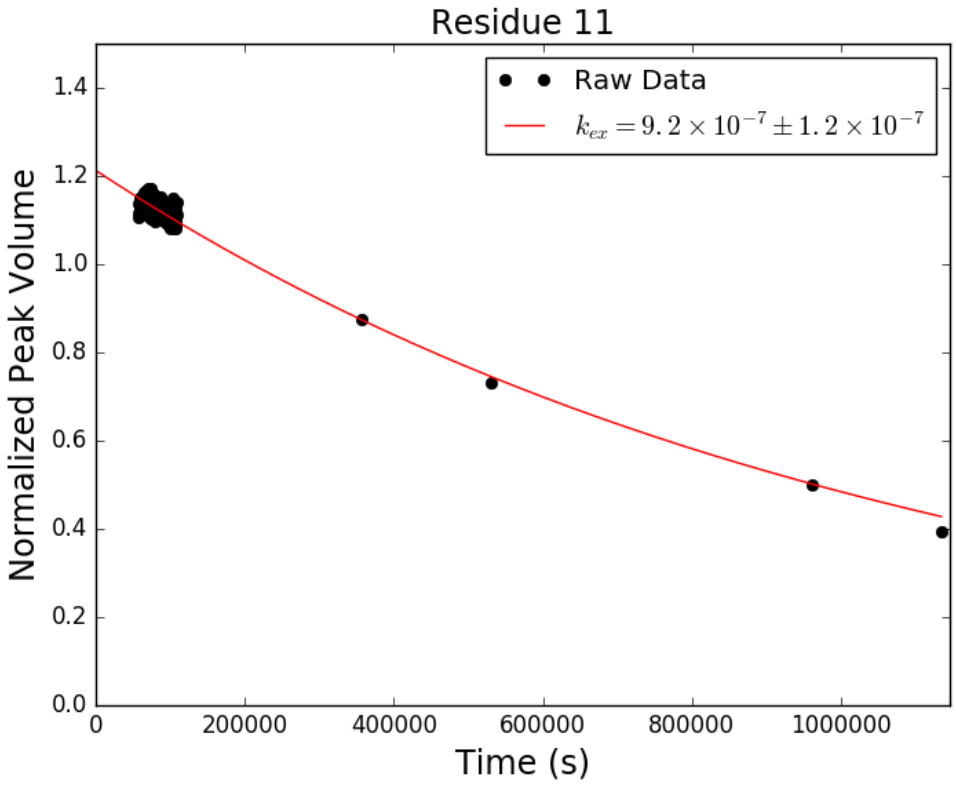


Figure C.57: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 11 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

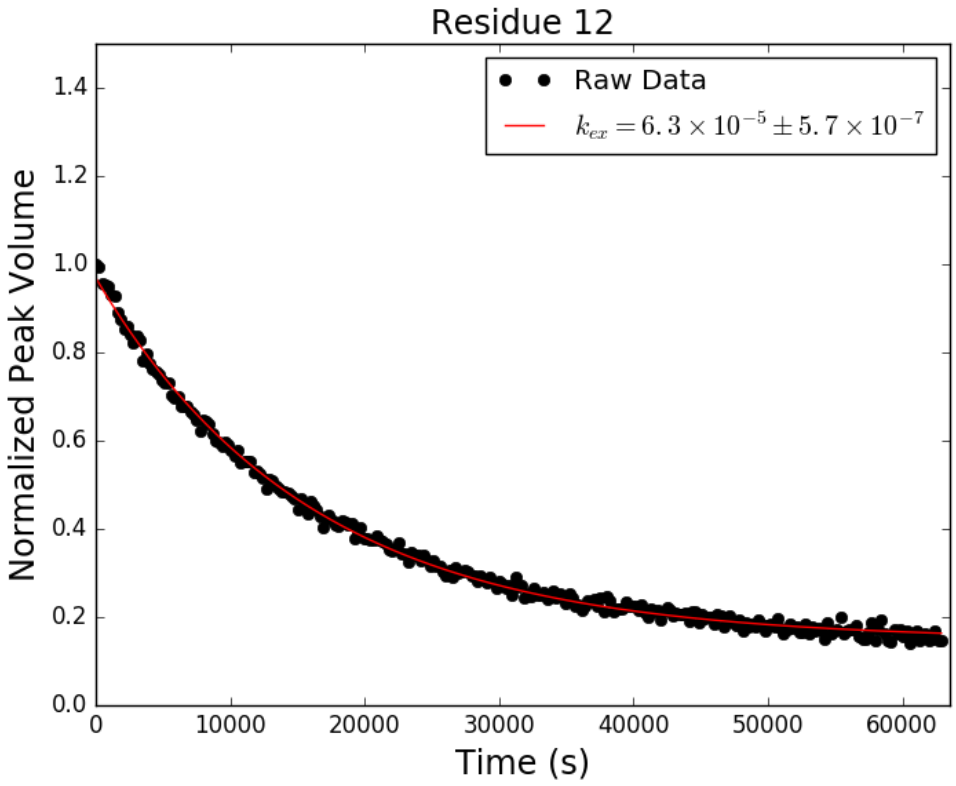


Figure C.58: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 12 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

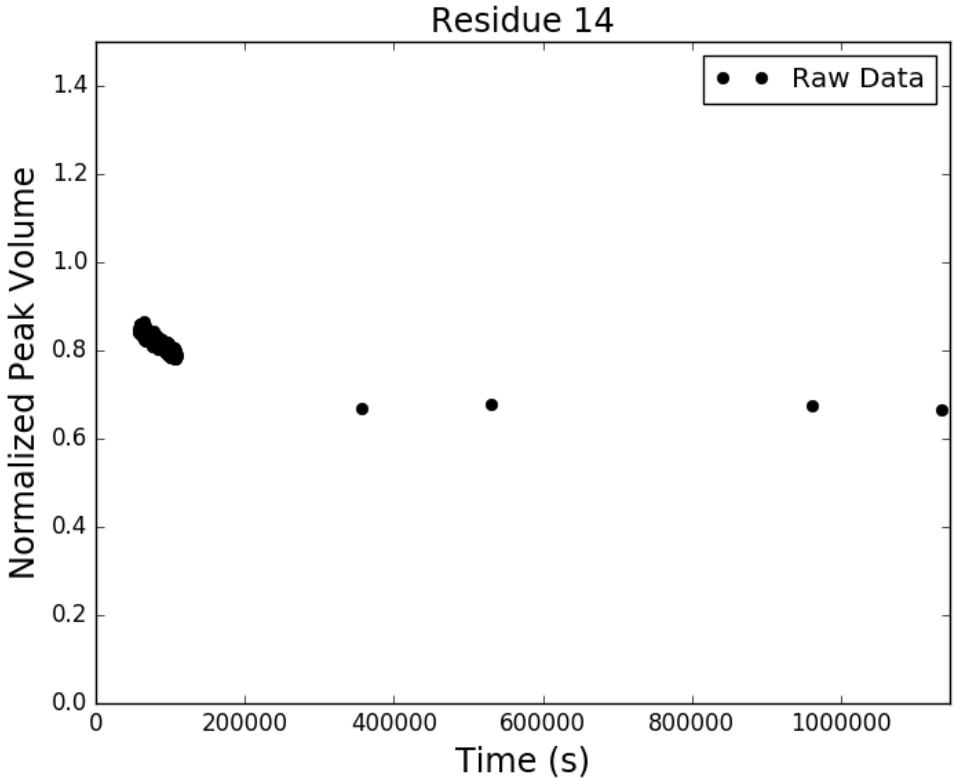


Figure C.59: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 14 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

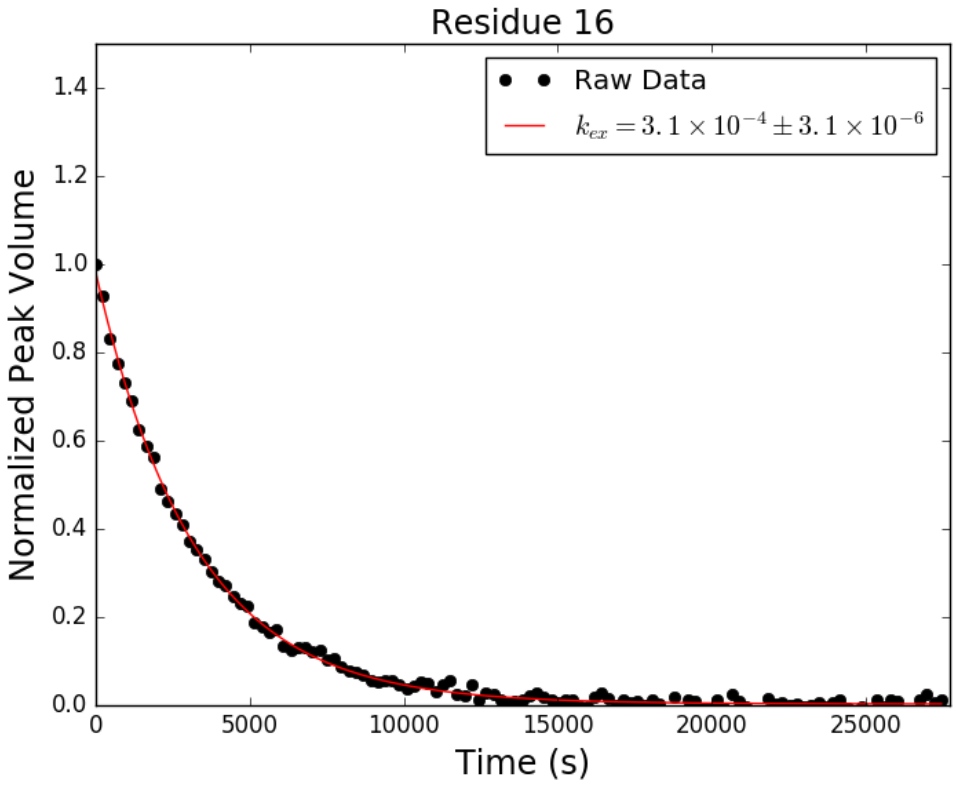


Figure C.60: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 16 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

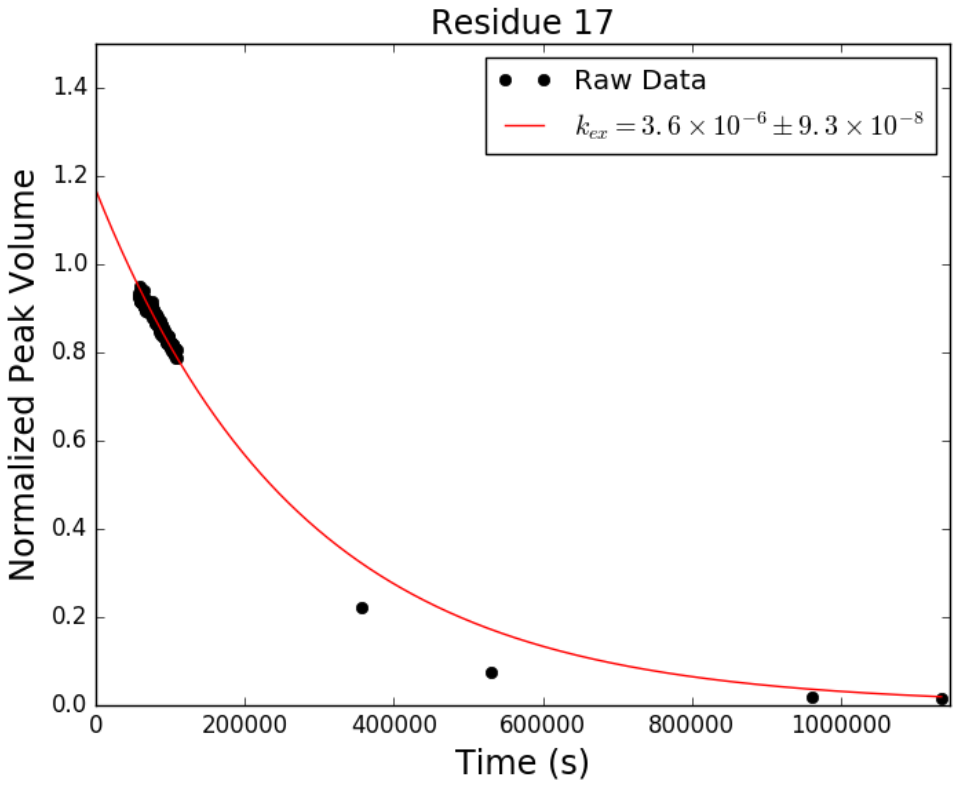


Figure C.61: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 17 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

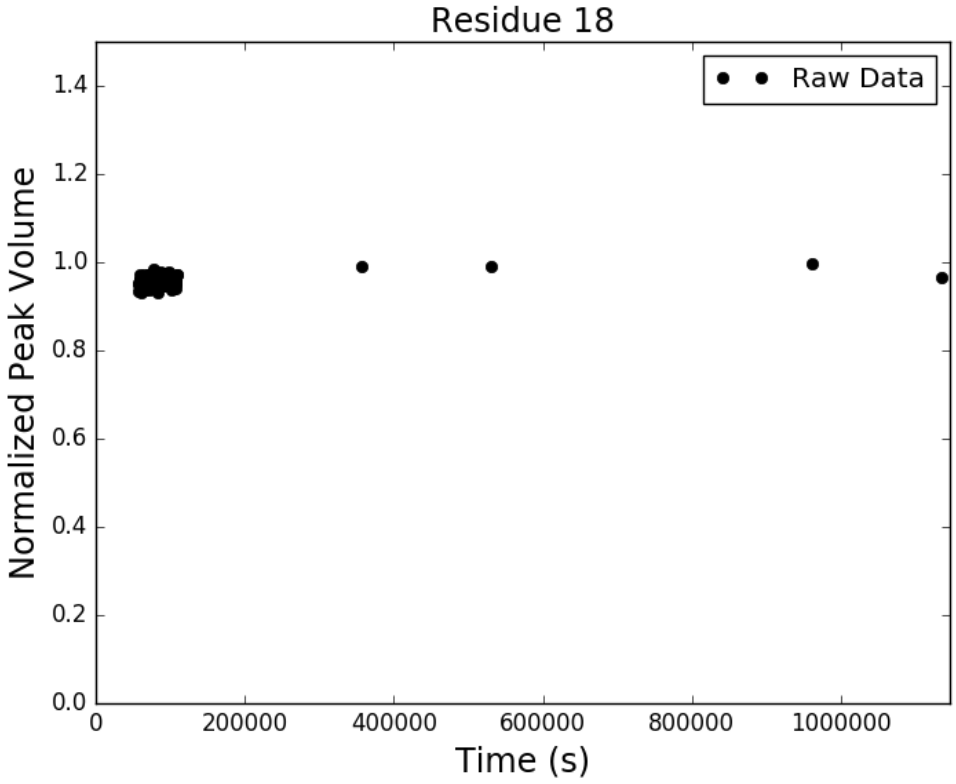


Figure C.62: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 18 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

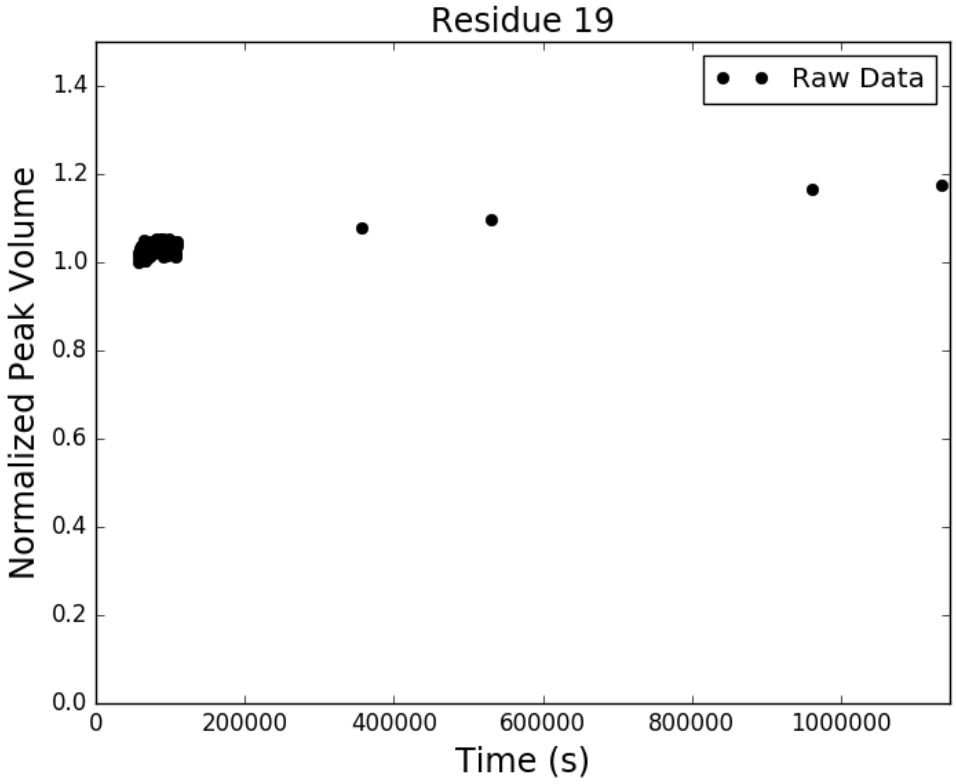


Figure C.63: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 19 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

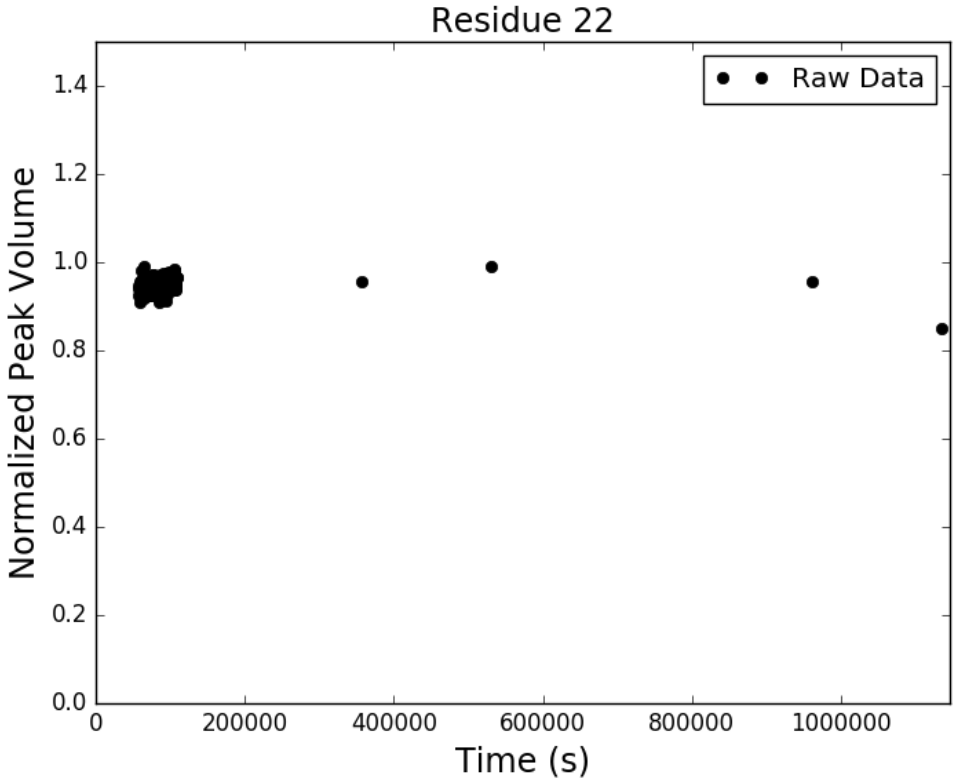


Figure C.64: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 22 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

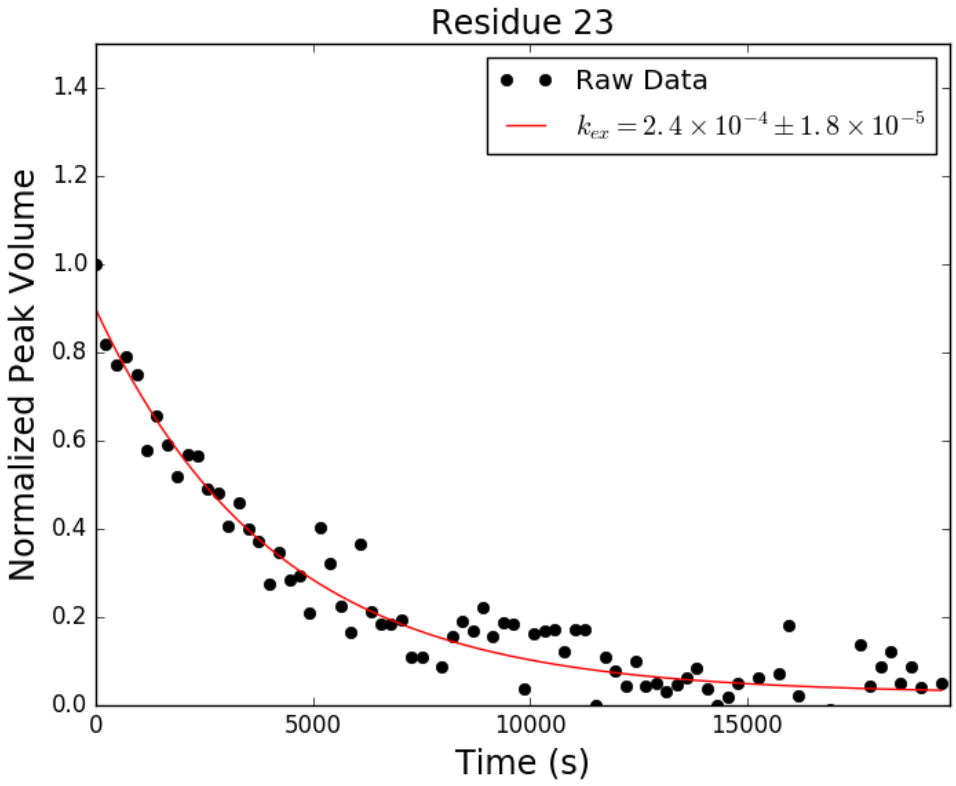


Figure C.65: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 23 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

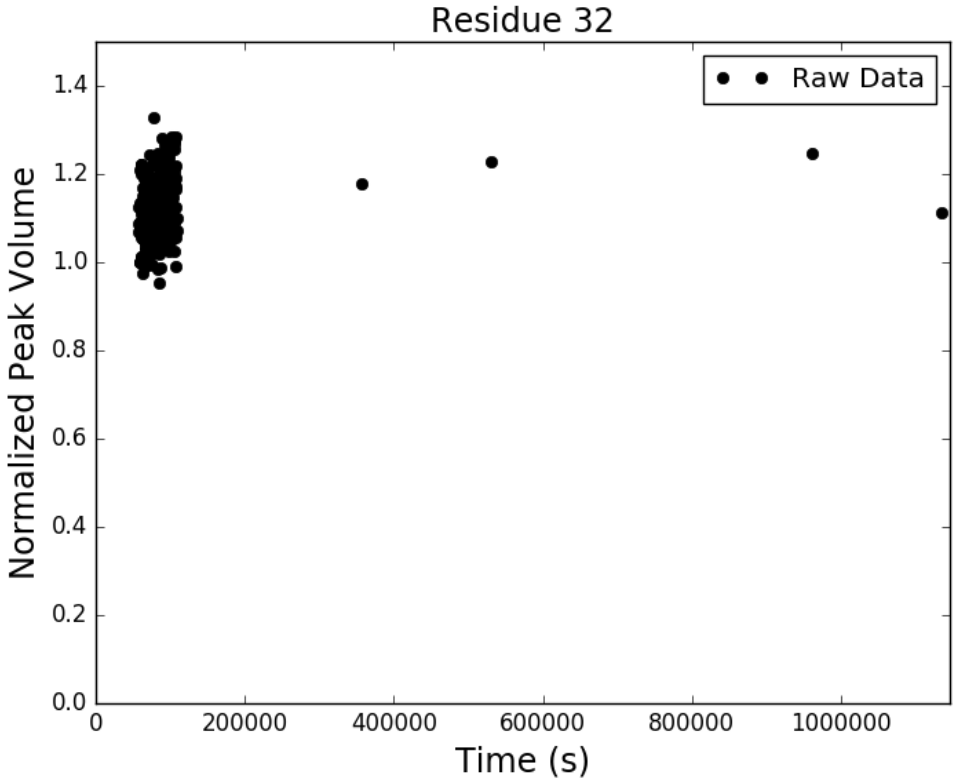


Figure C.66: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 32 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

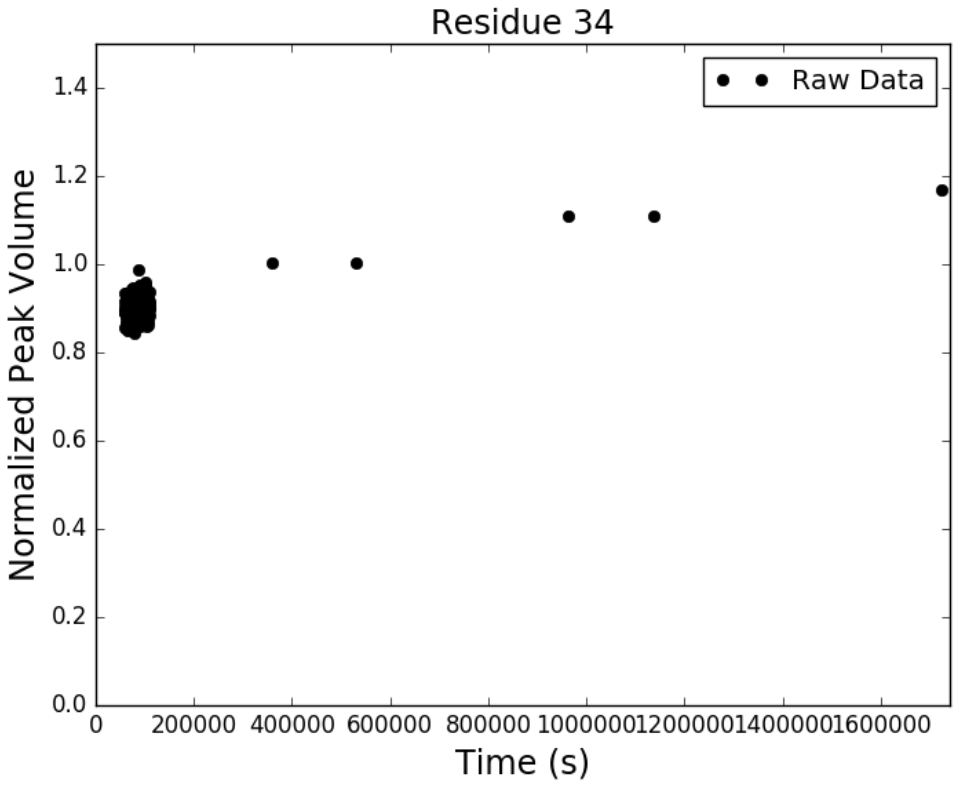


Figure C.67: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 34 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

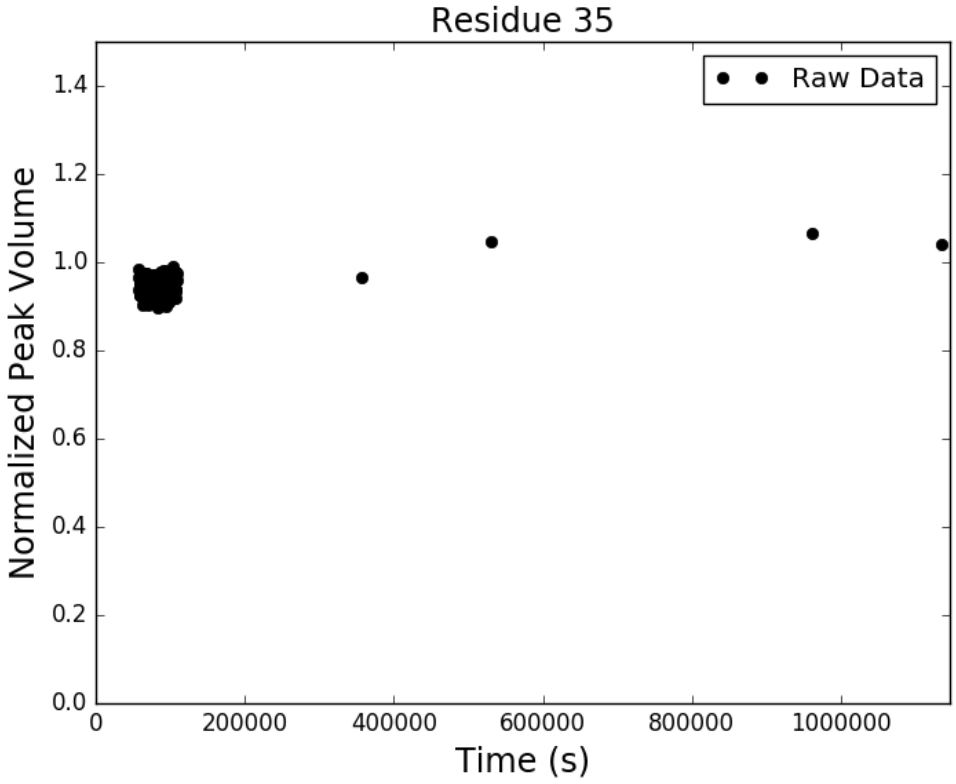


Figure C.68: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 35 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

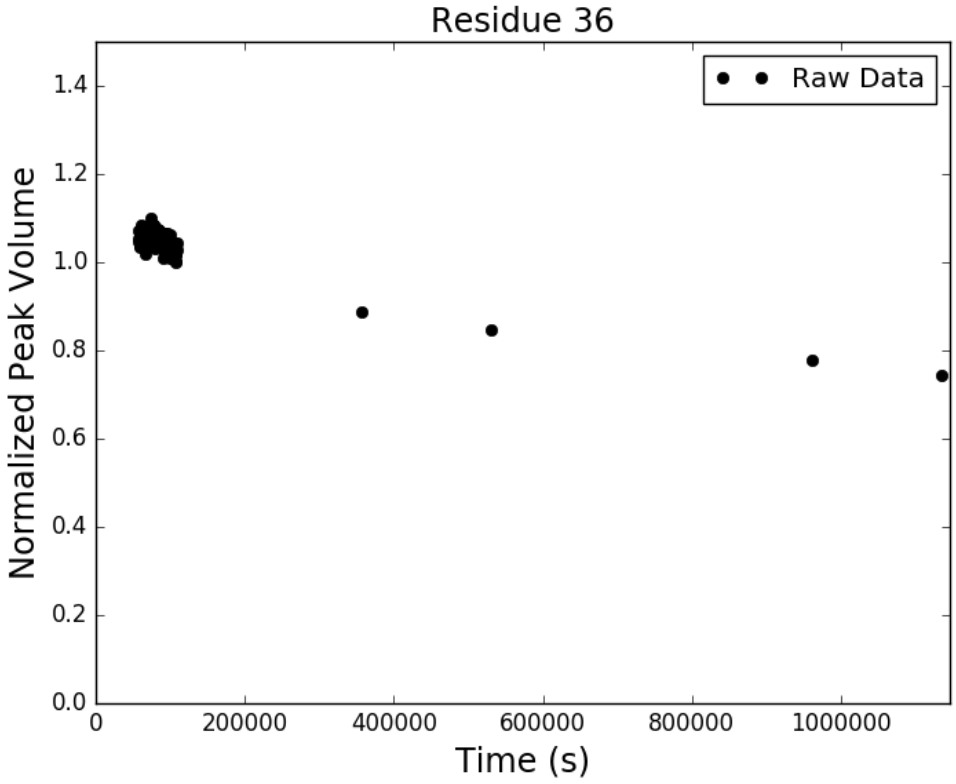


Figure C.69: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 36 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

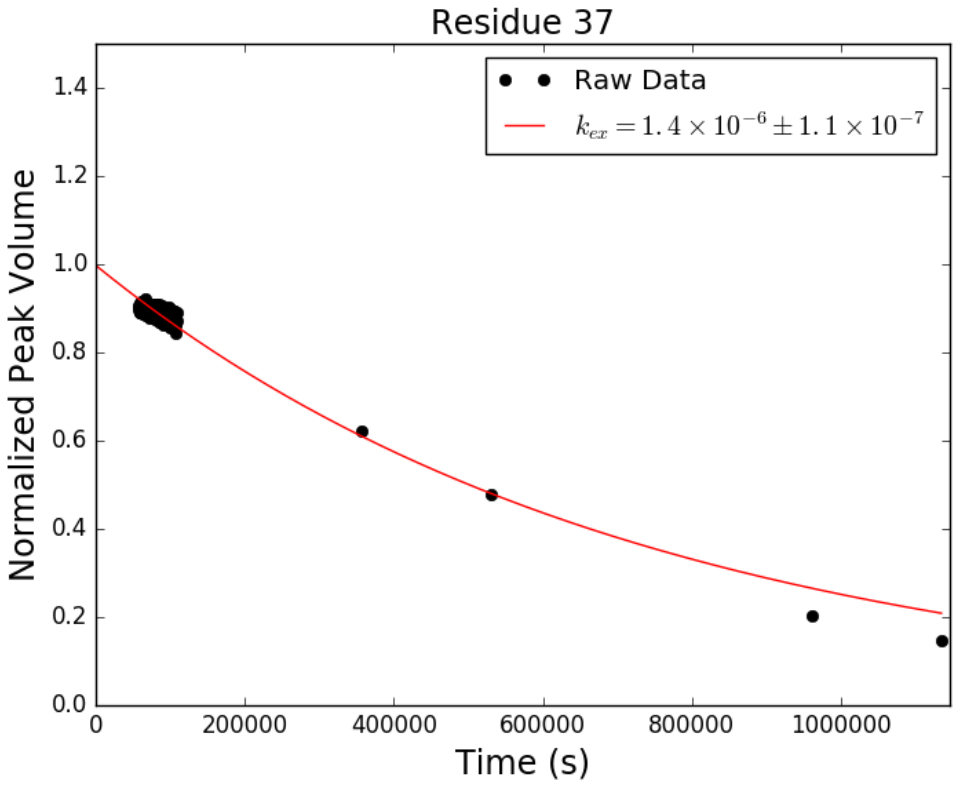


Figure C.70: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 37 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

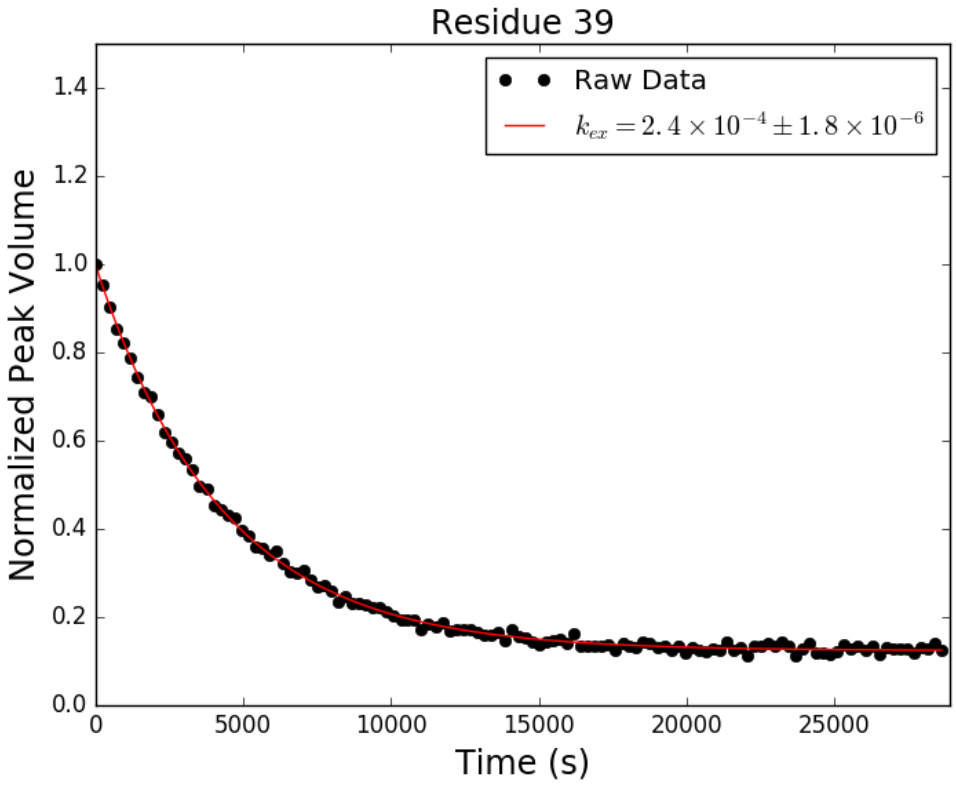


Figure C.71: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 39 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

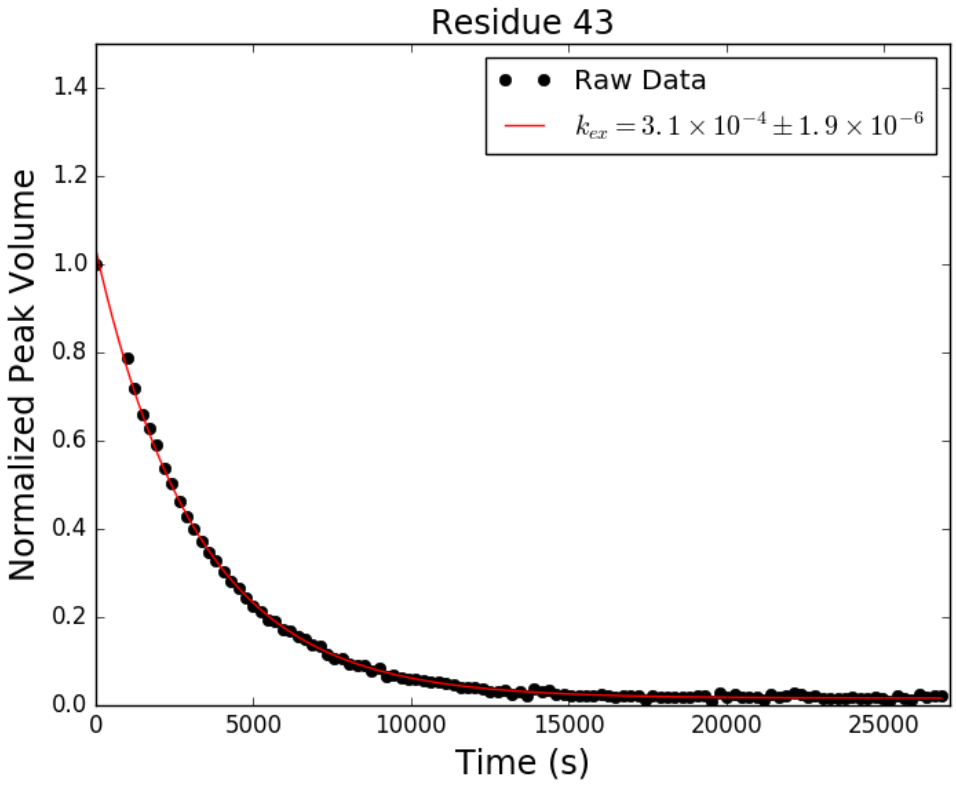


Figure C.72: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 43 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

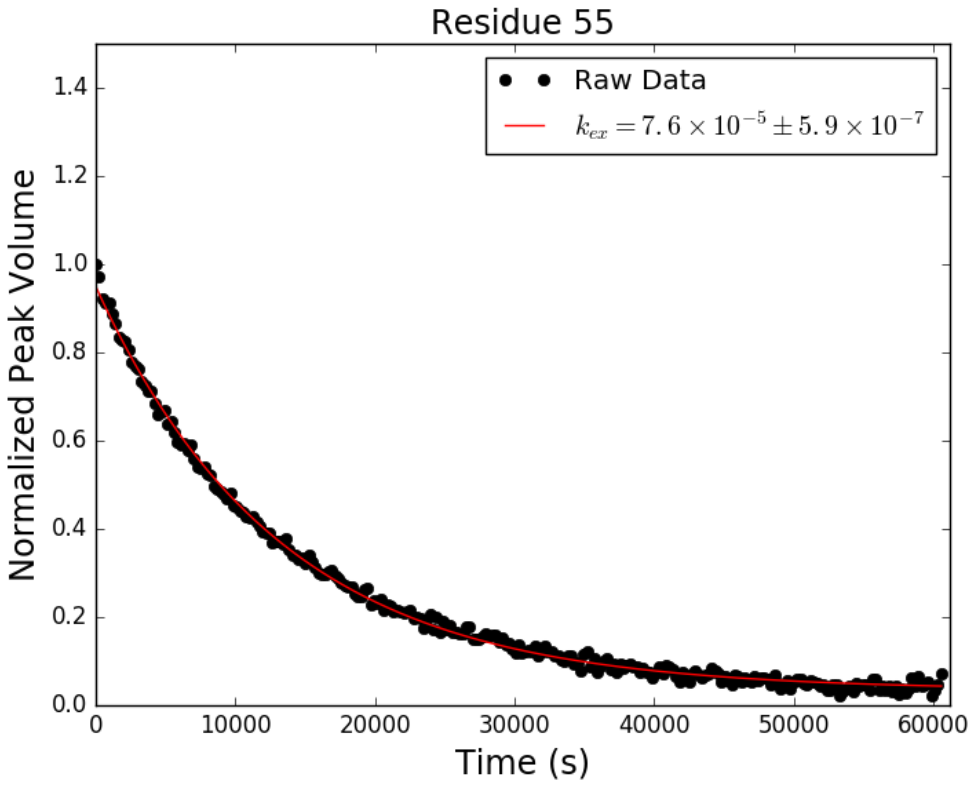


Figure C.73: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 55 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

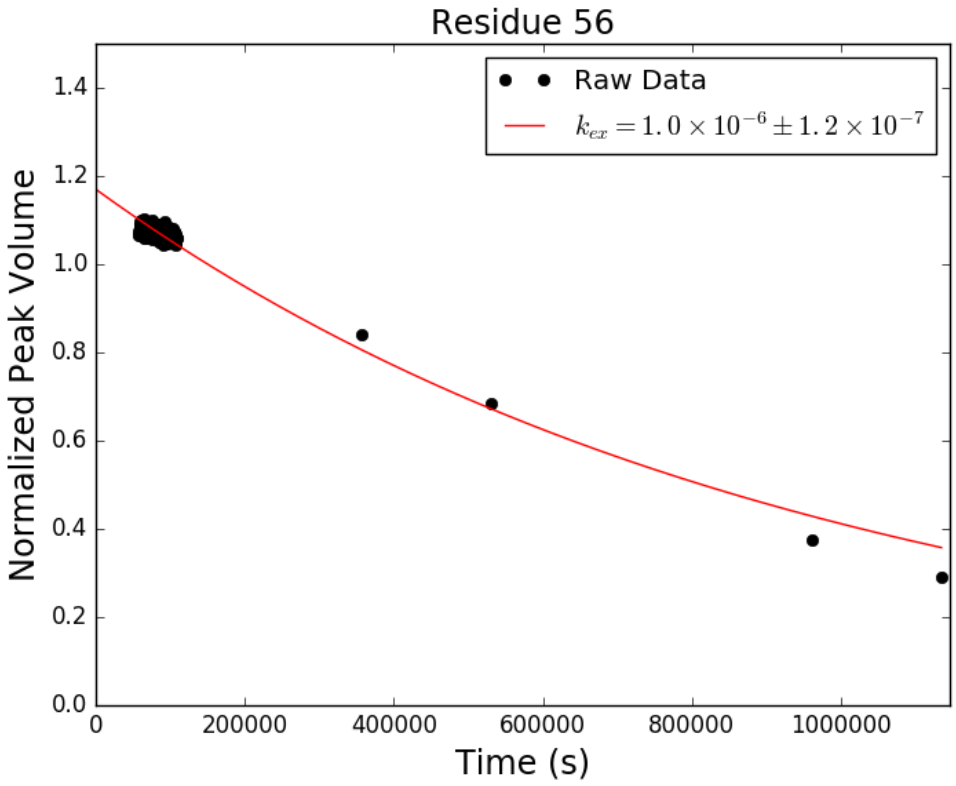


Figure C.74: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 56 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

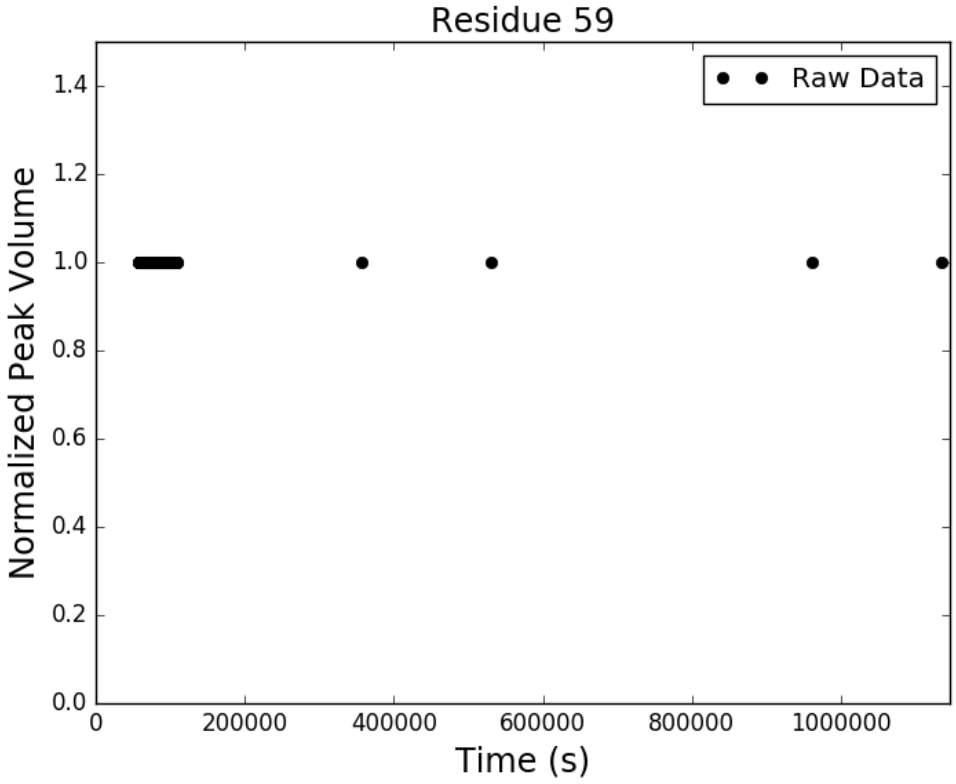


Figure C.75: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 59 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

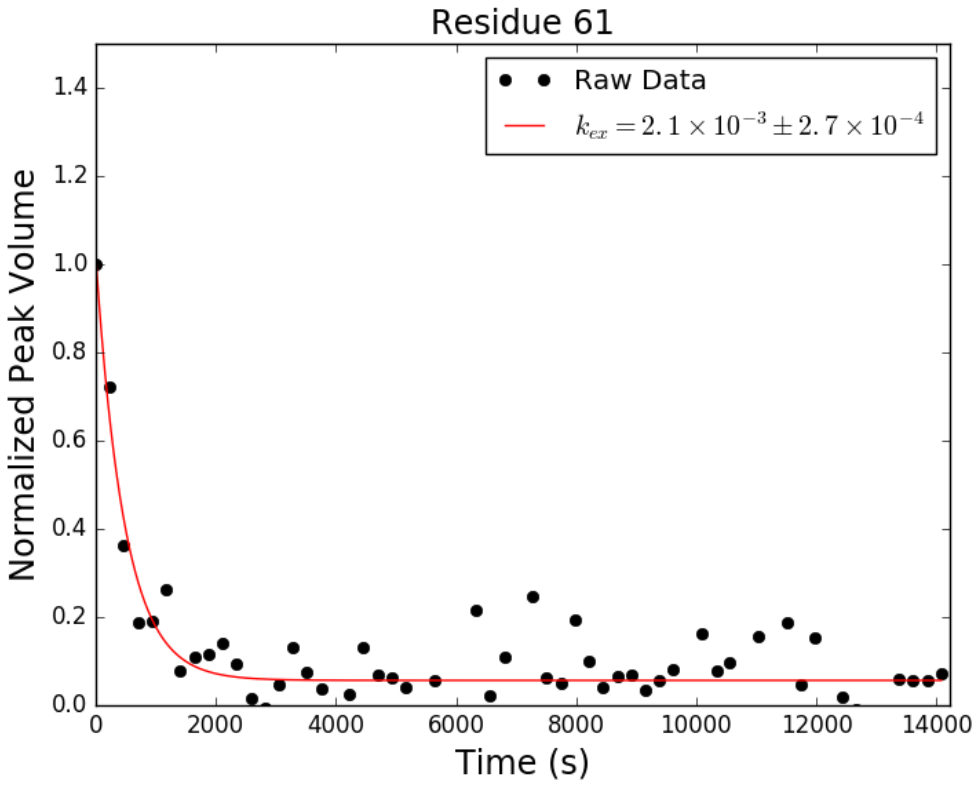


Figure C.76: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 61 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

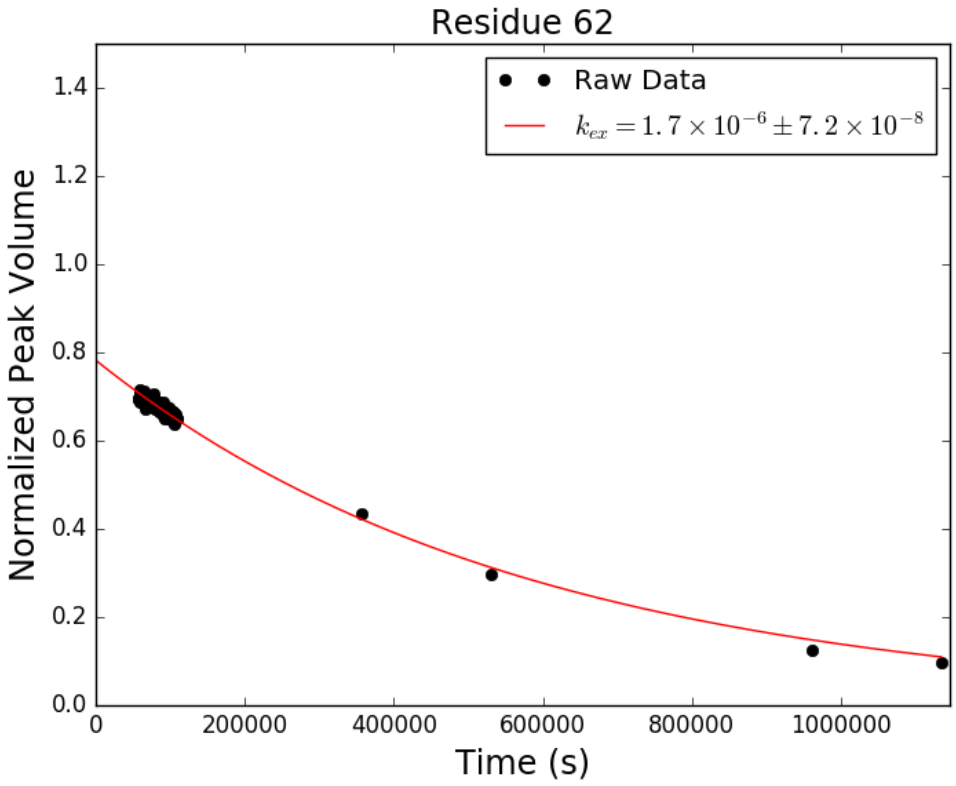


Figure C.77: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 62 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

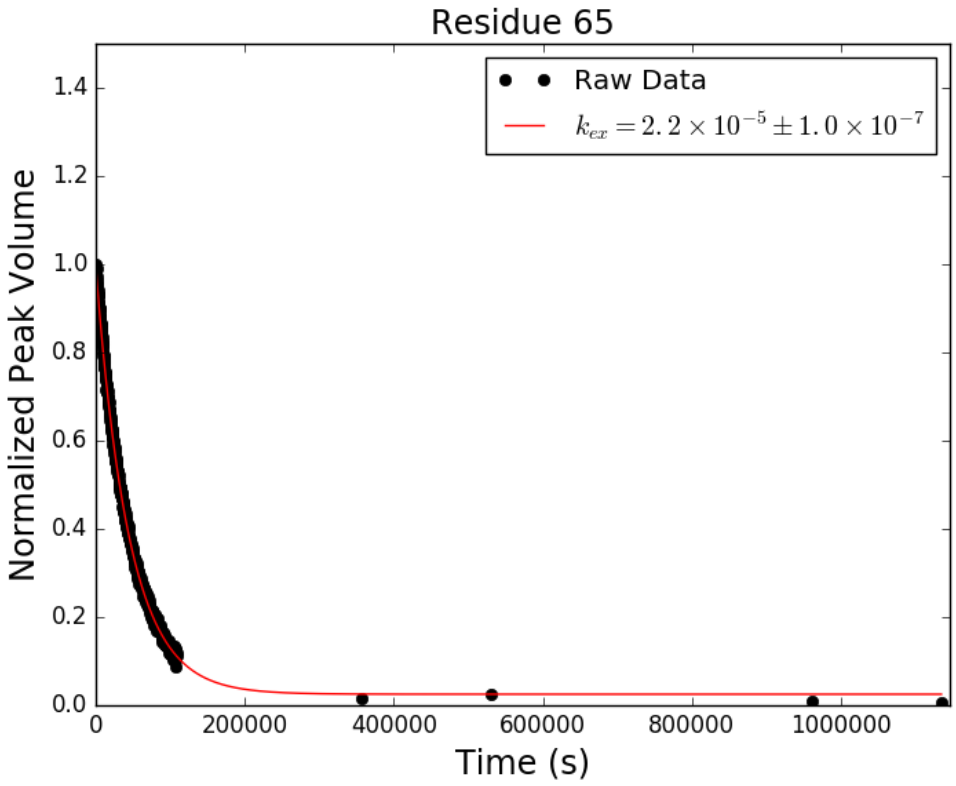


Figure C.78: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 65 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

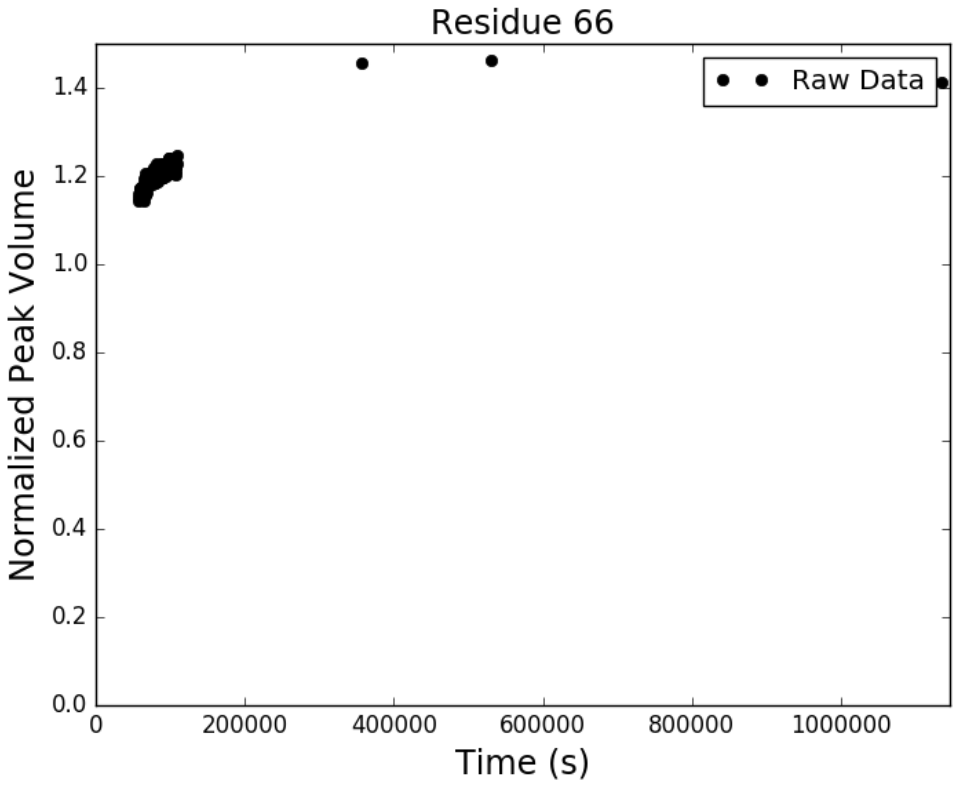


Figure C.79: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 66 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

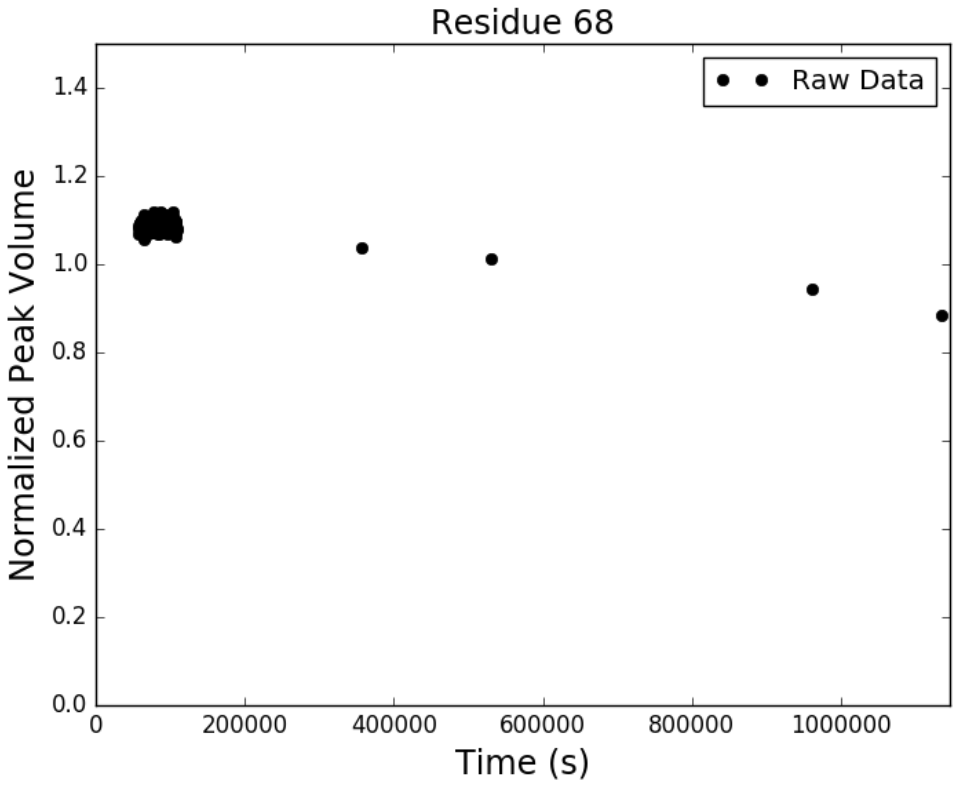


Figure C.80: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 68 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

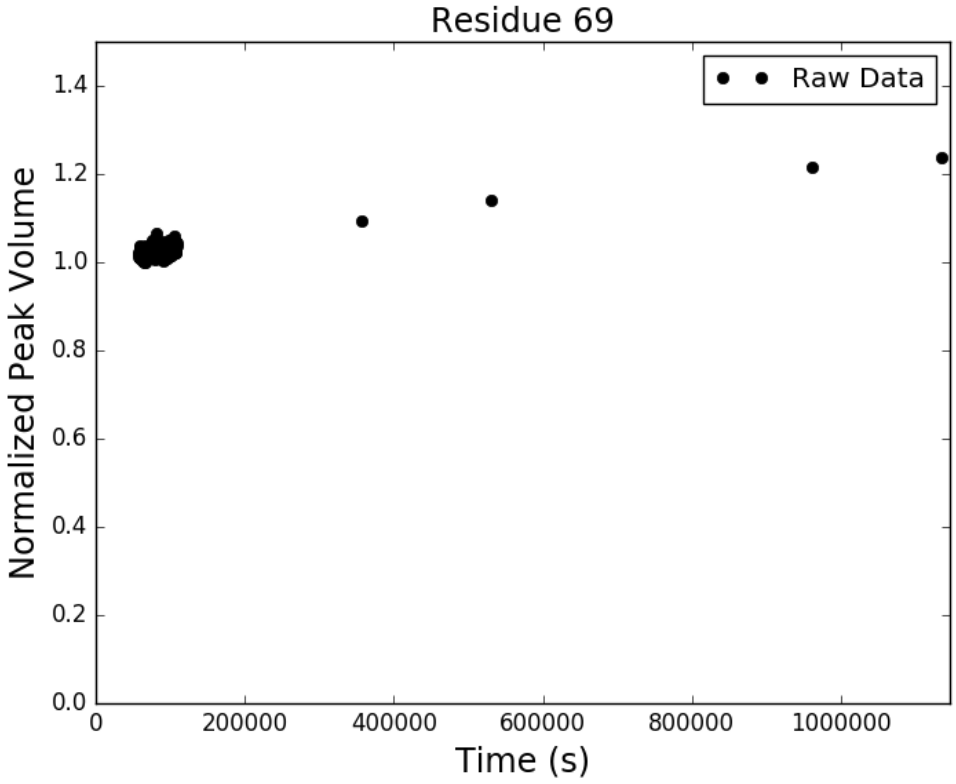


Figure C.81: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 69 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

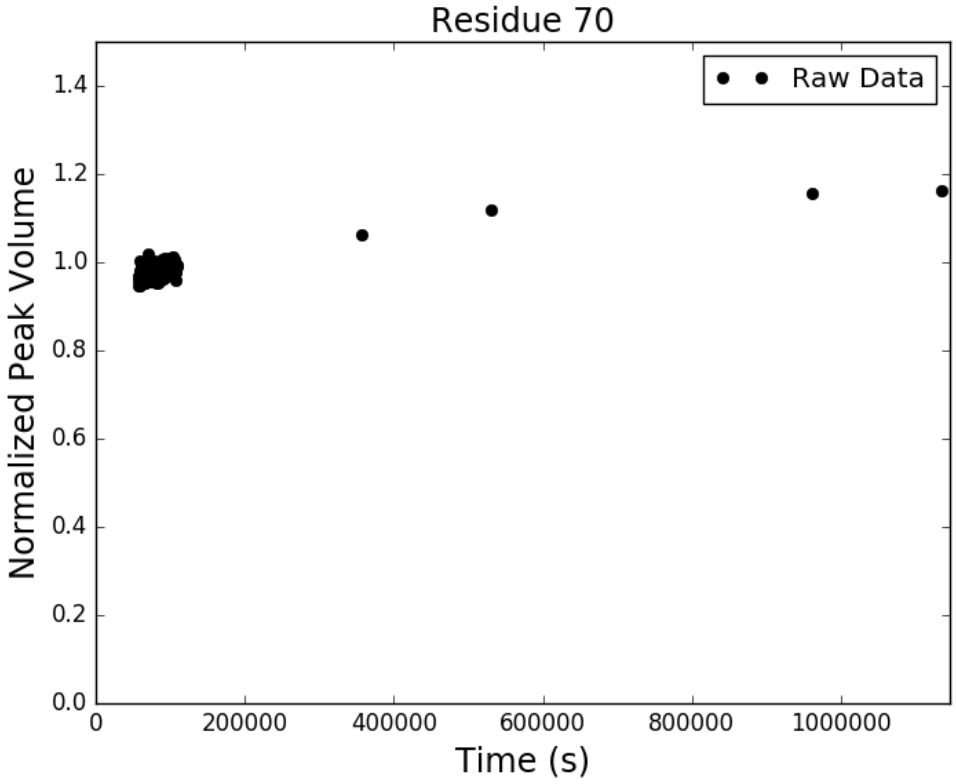


Figure C.82: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 70 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

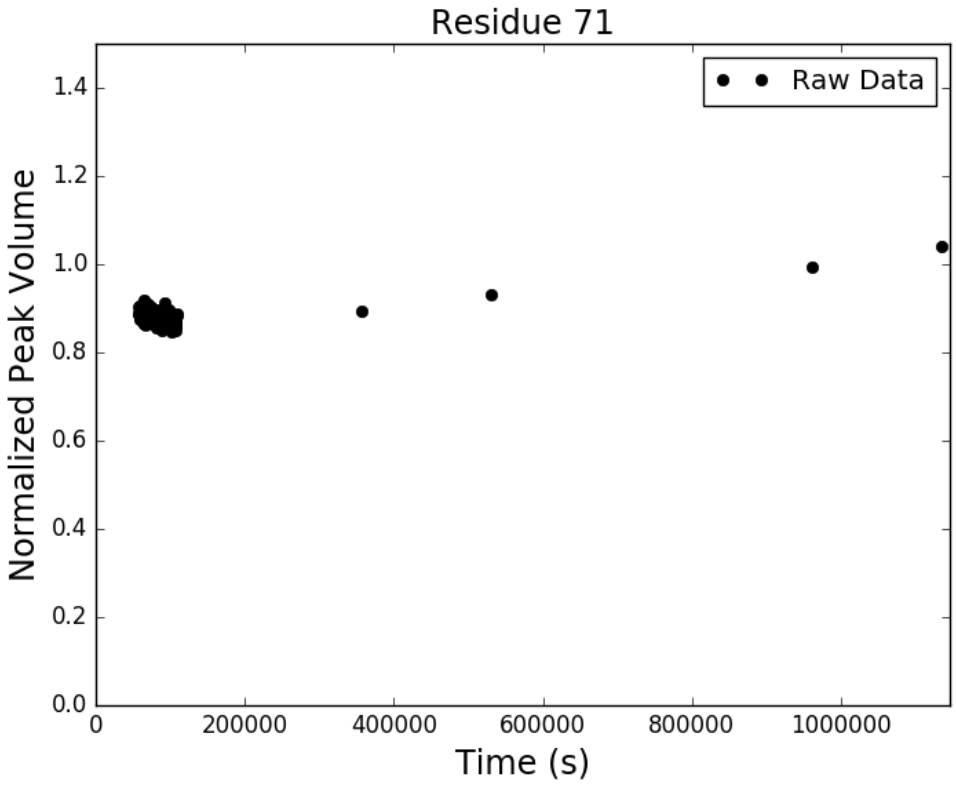


Figure C.83: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 71 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

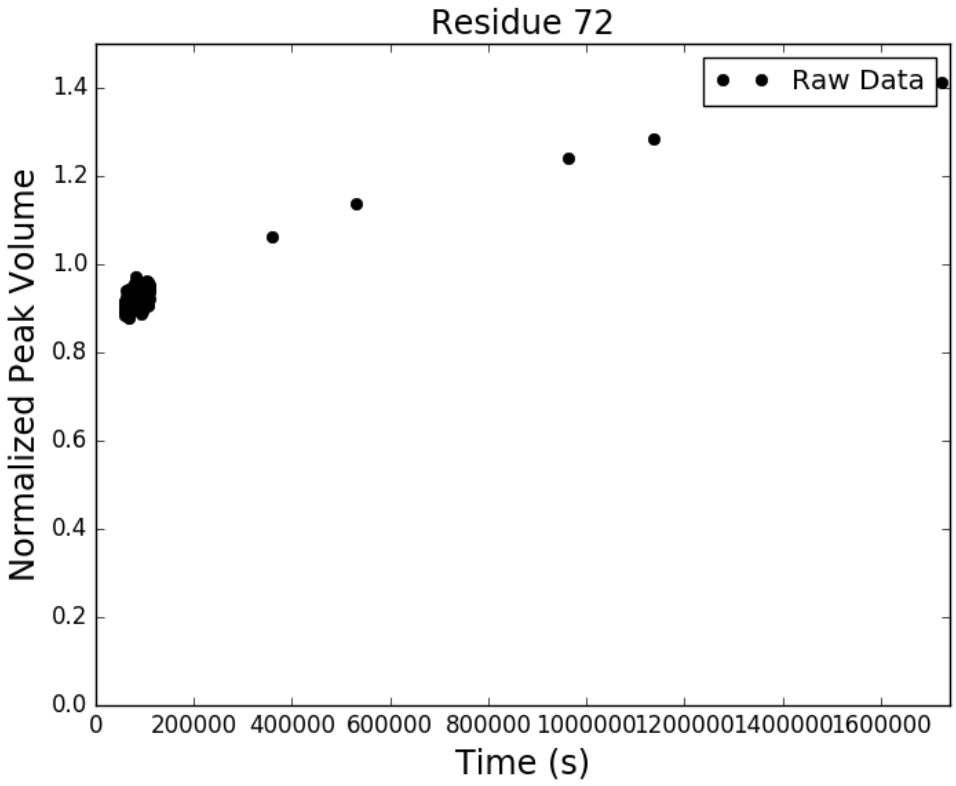


Figure C.84: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 72 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

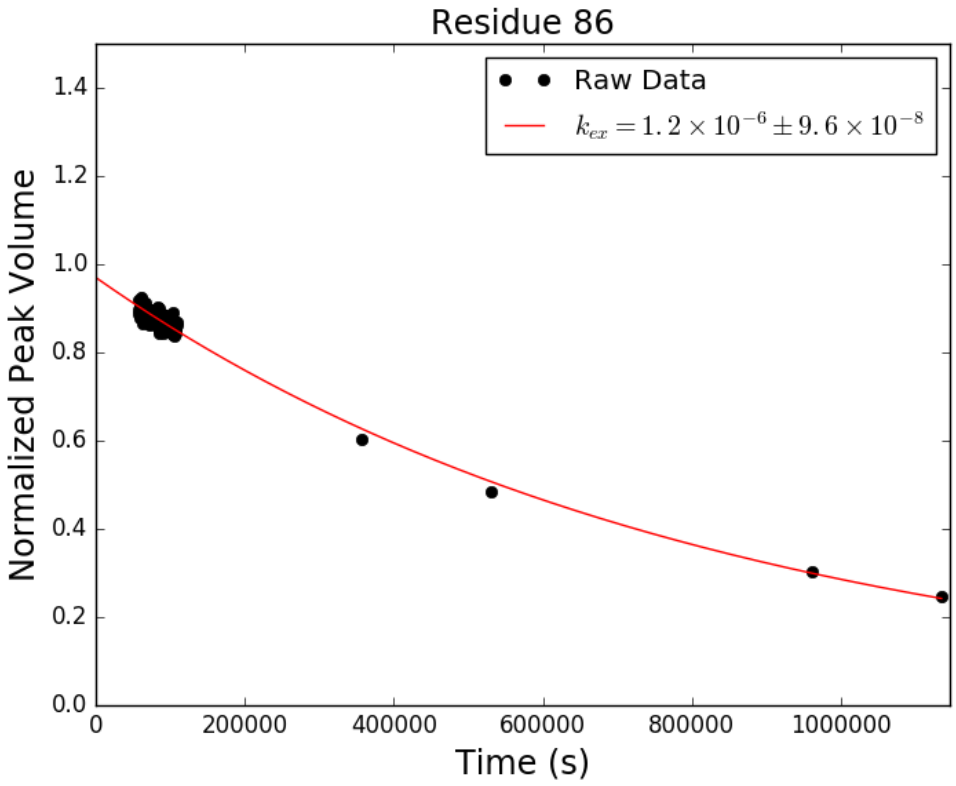


Figure C.85: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 86 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

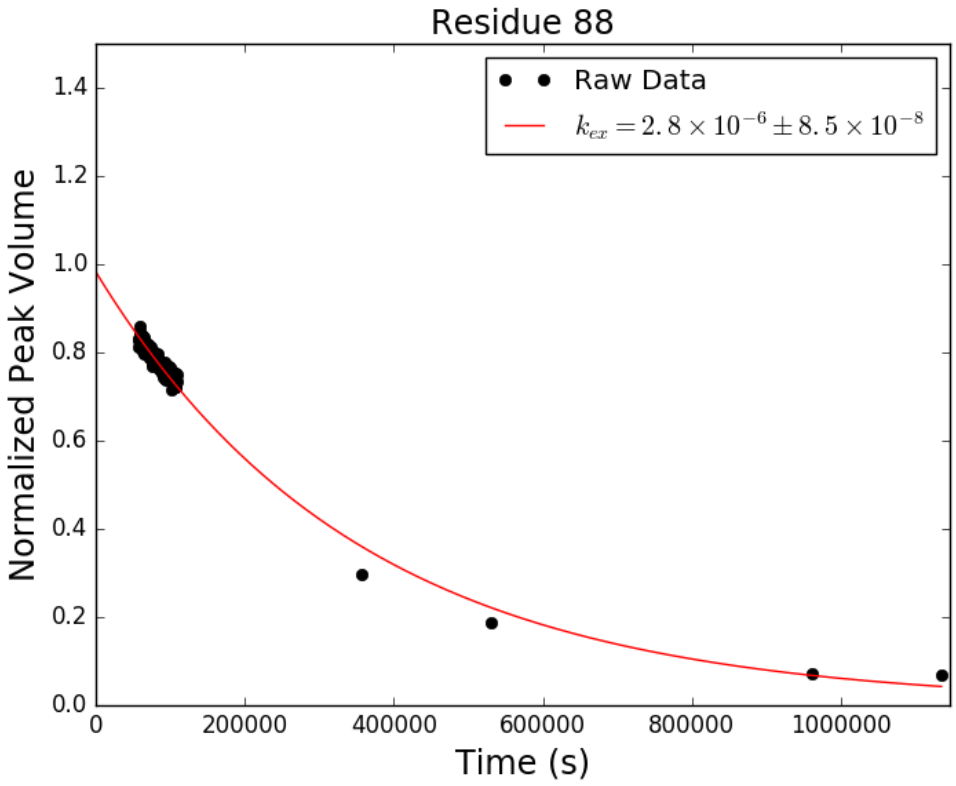


Figure C.86: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 88 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

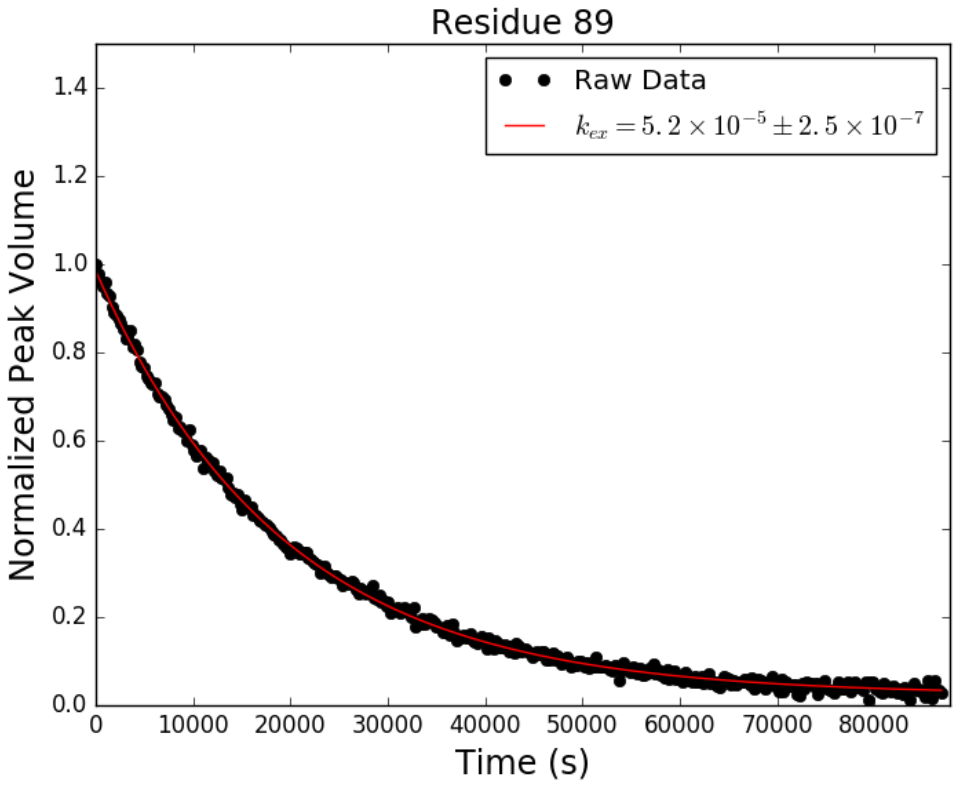


Figure C.87: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 89 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

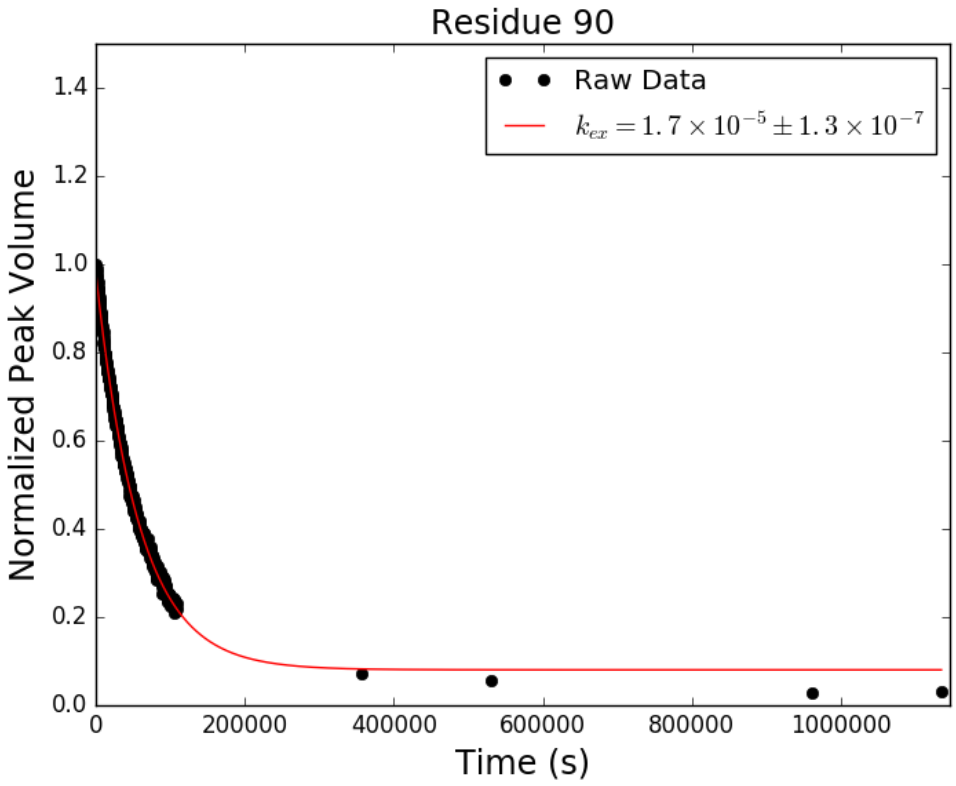


Figure C.88: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 90 of Parent. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

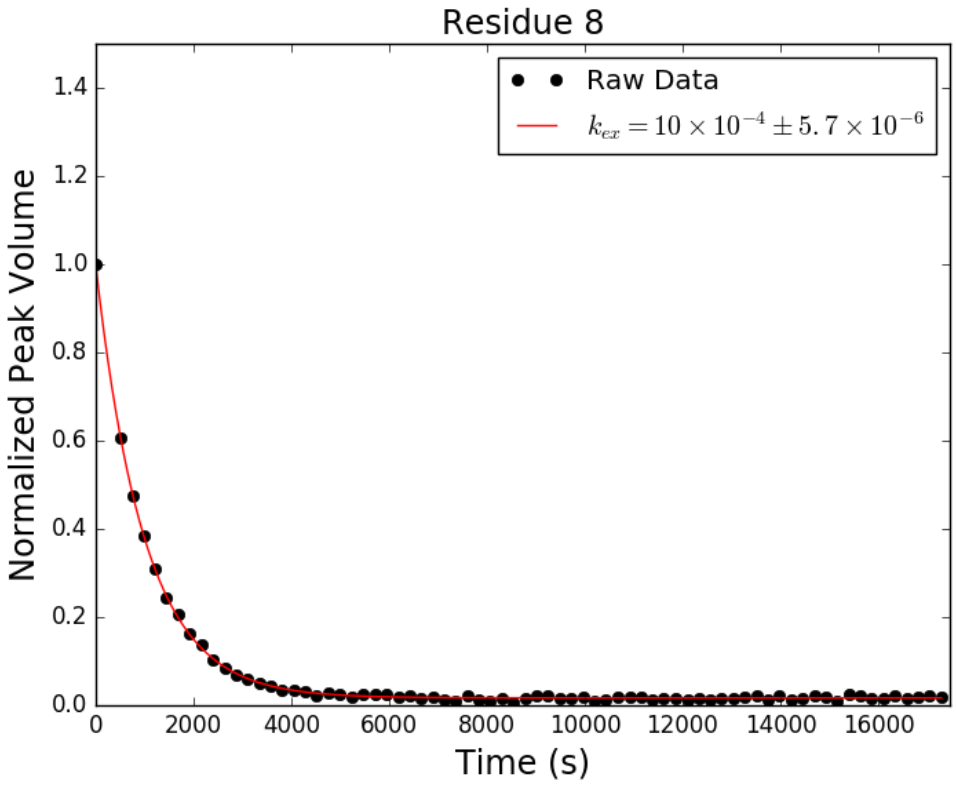


Figure C.89: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 8 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

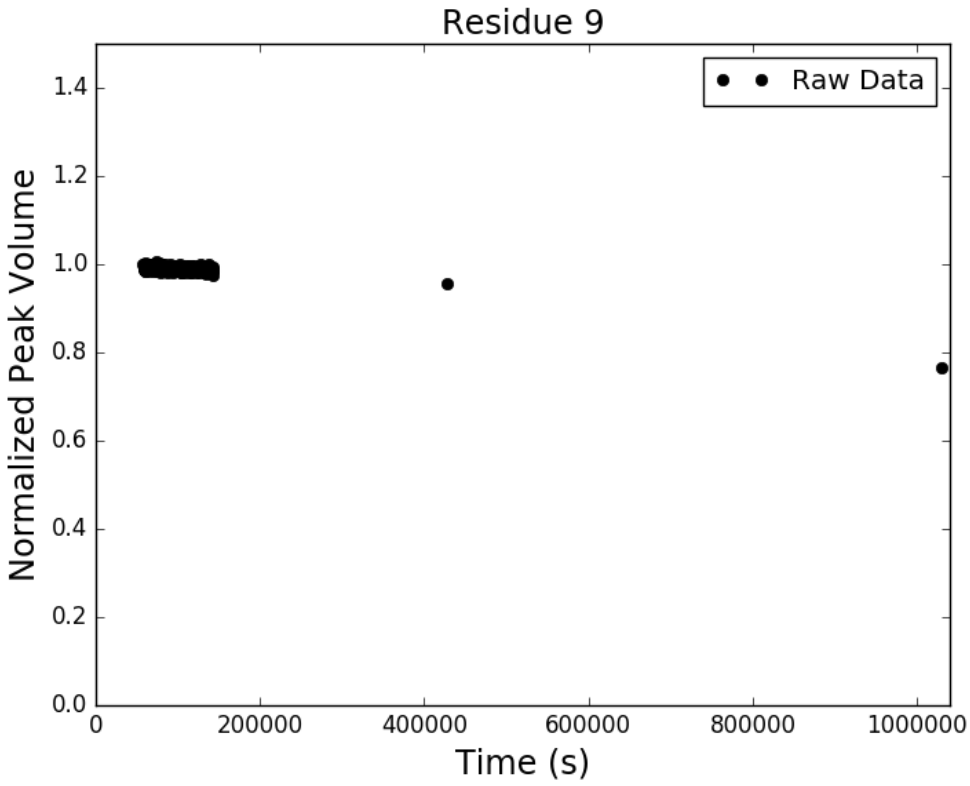


Figure C.90: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 9 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

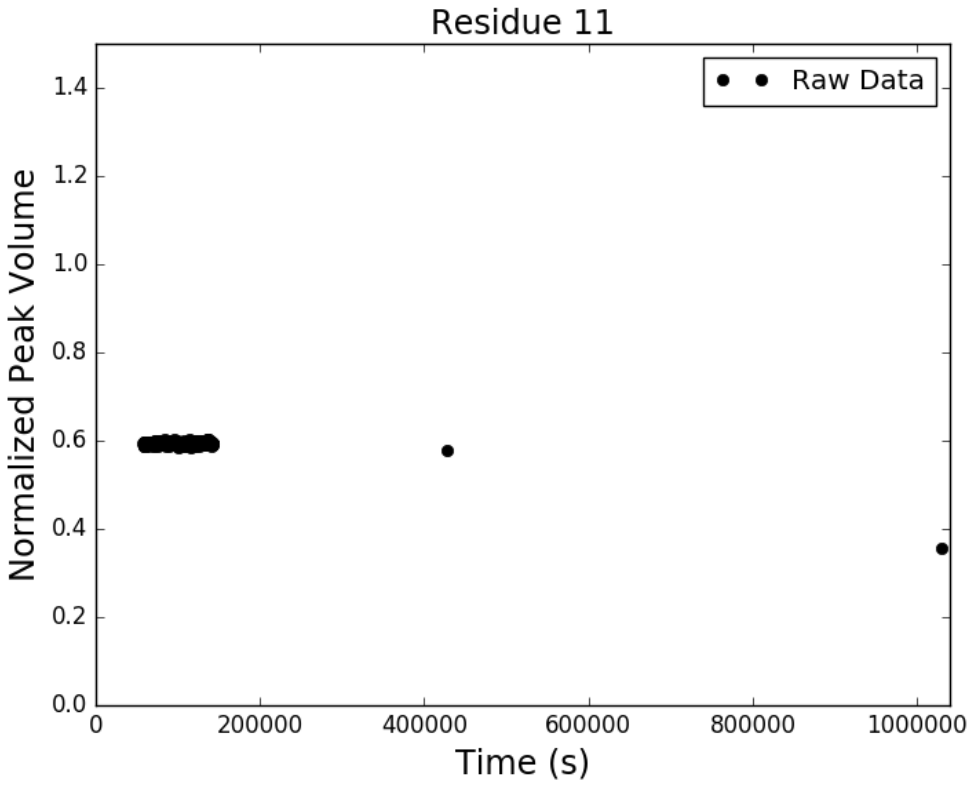


Figure C.91: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 11 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

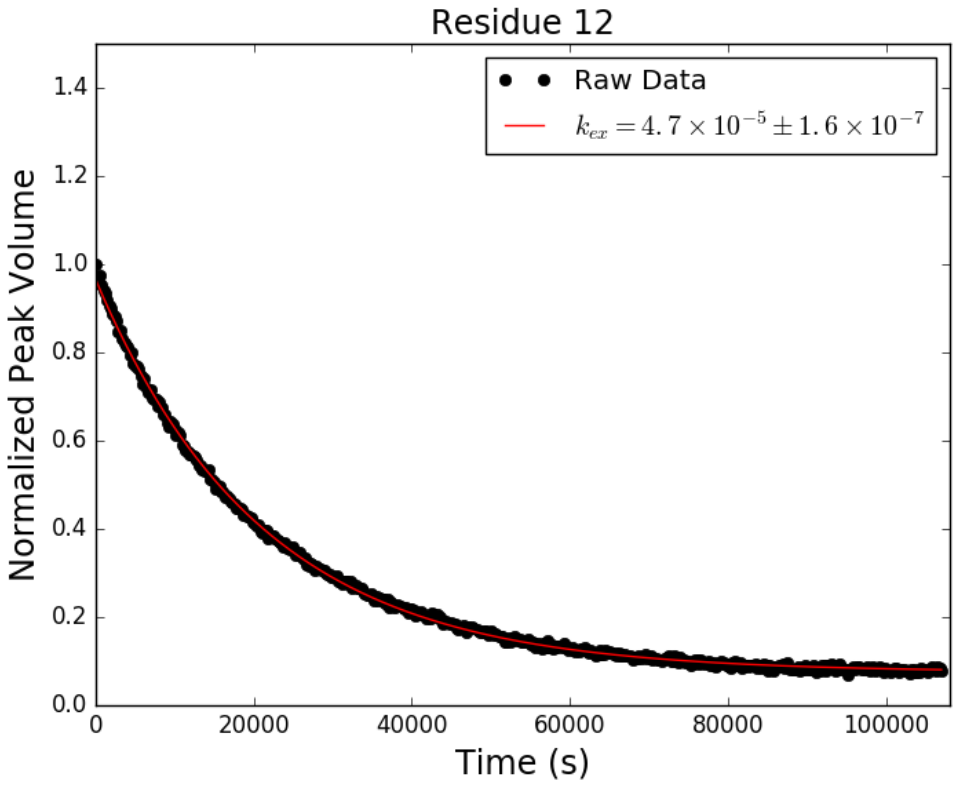


Figure C.92: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 12 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

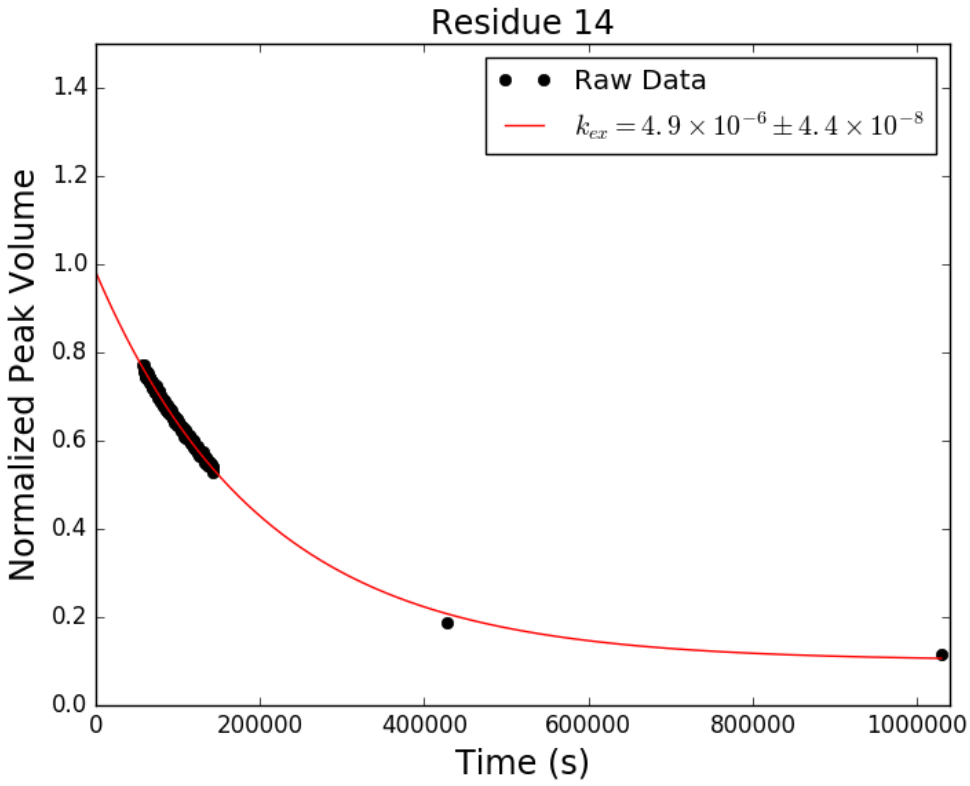


Figure C.93: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 14 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

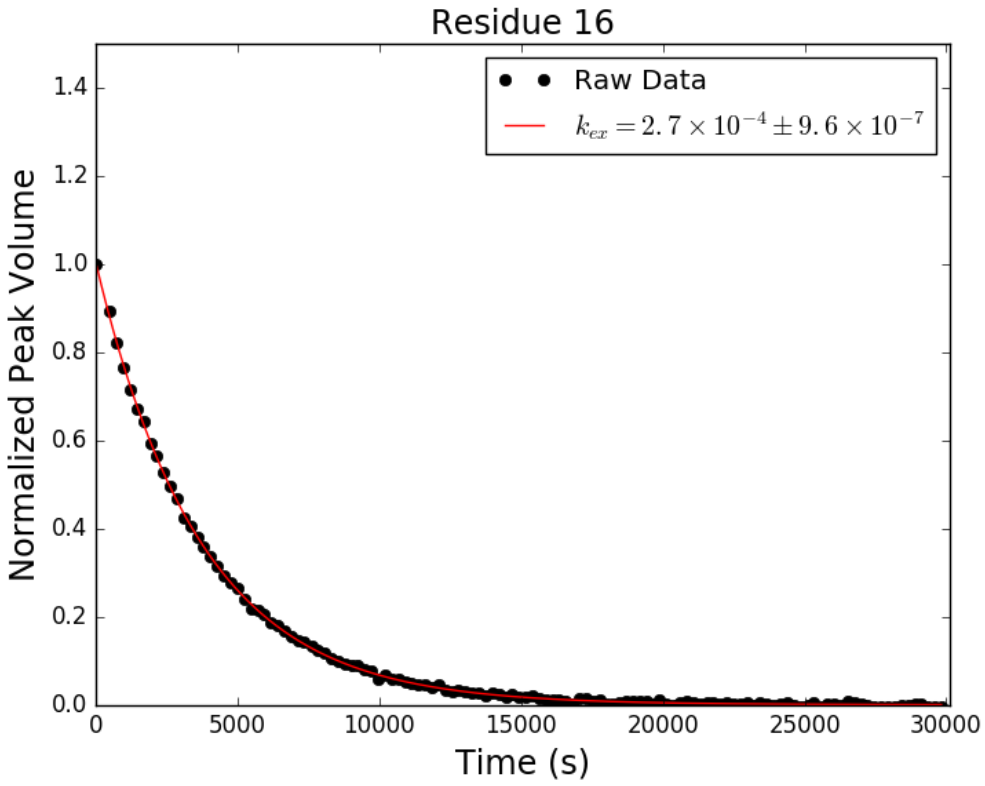


Figure C.94: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 16 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

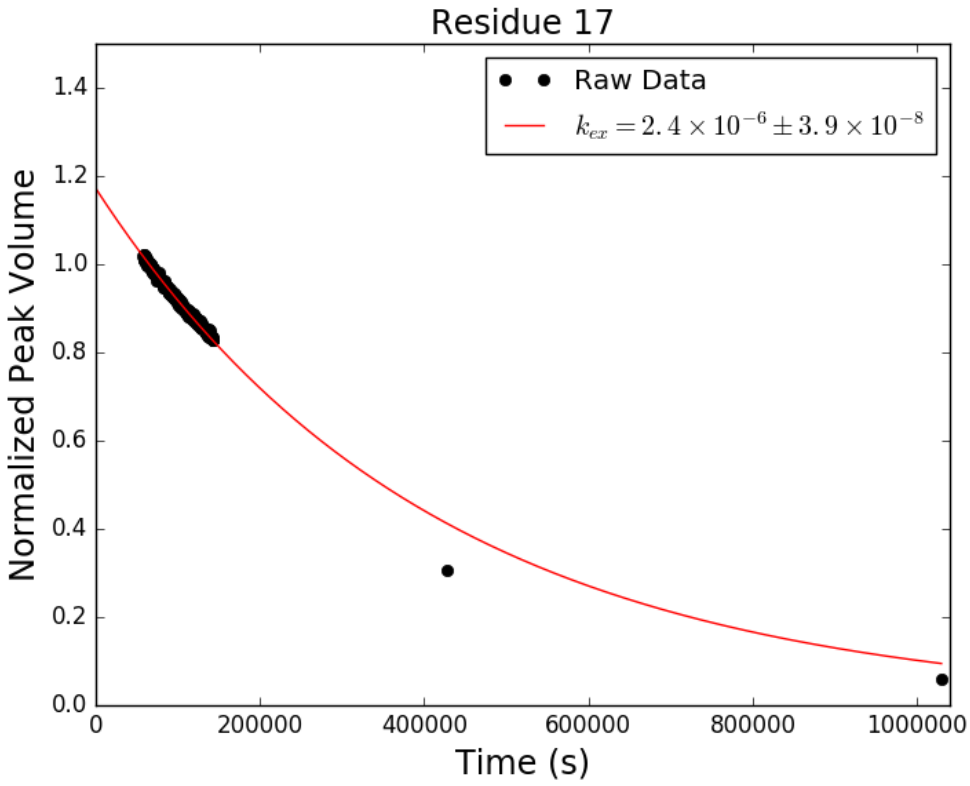


Figure C.95: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 17 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

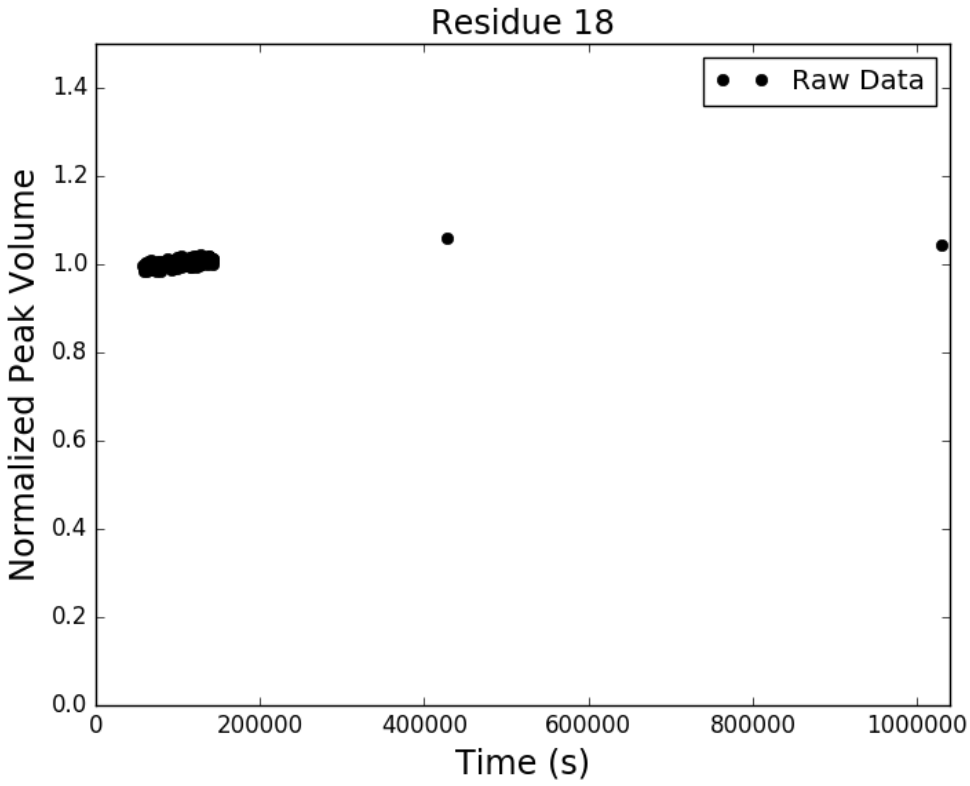


Figure C.96: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 18 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

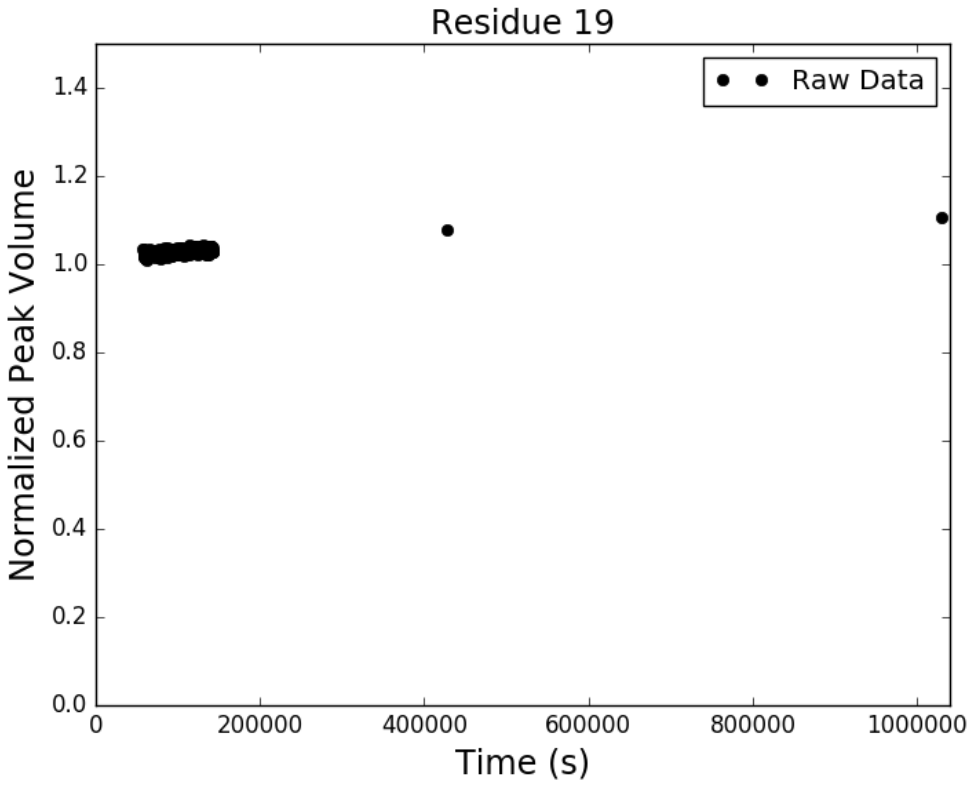


Figure C.97: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 19 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

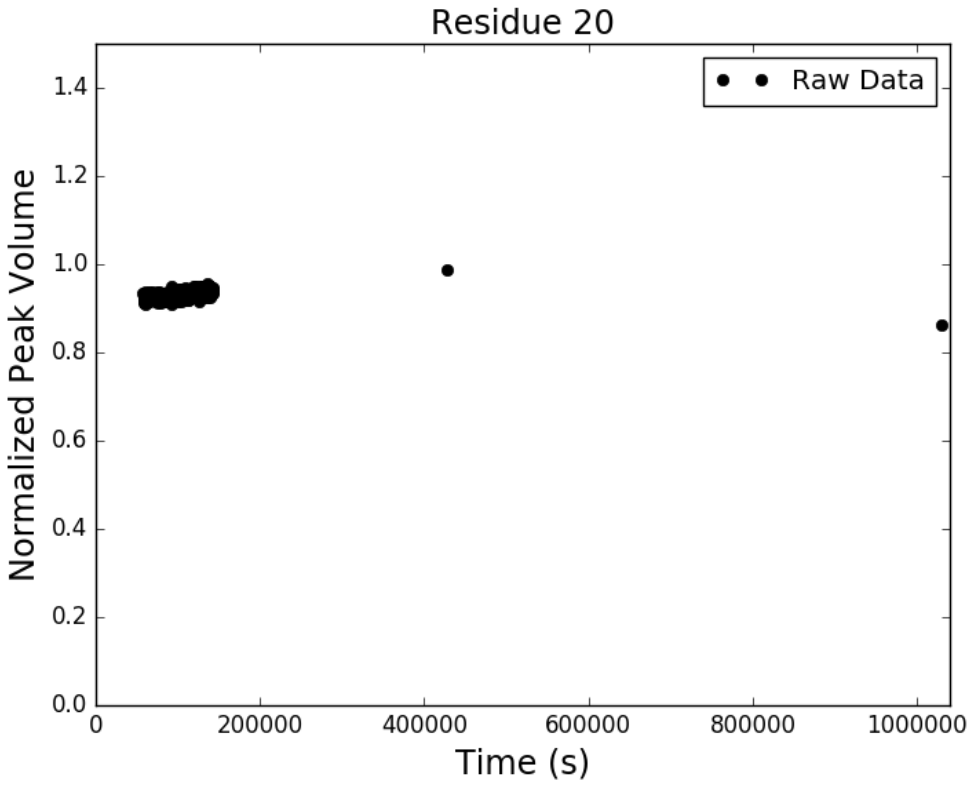


Figure C.98: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 20 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

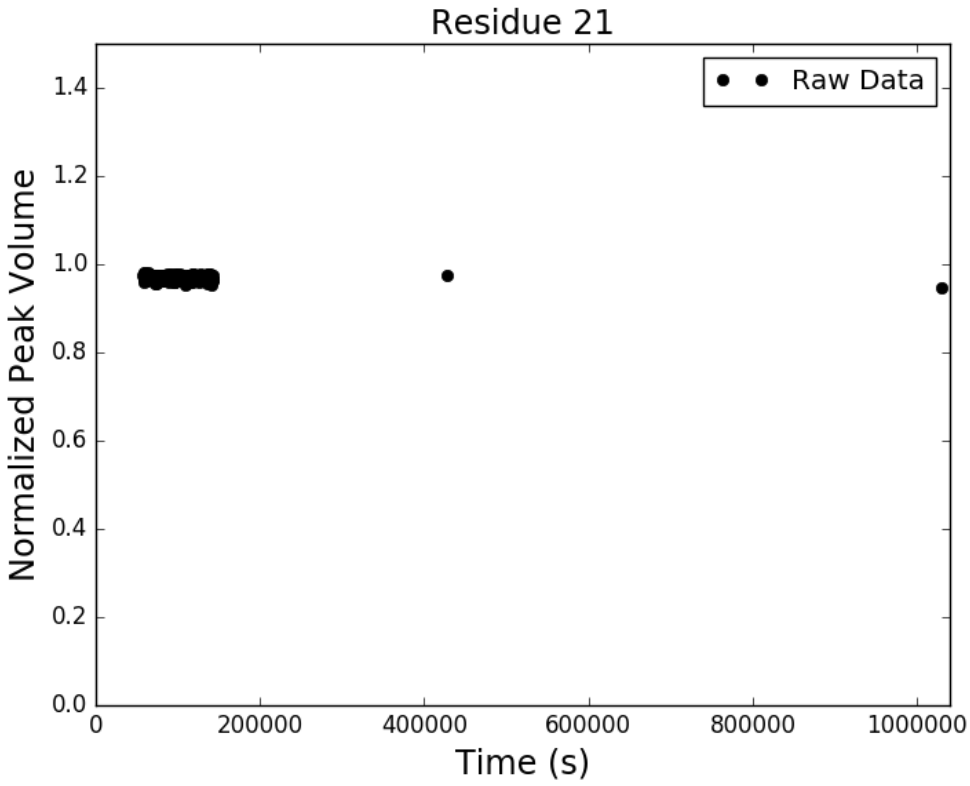


Figure C.99: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 21 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

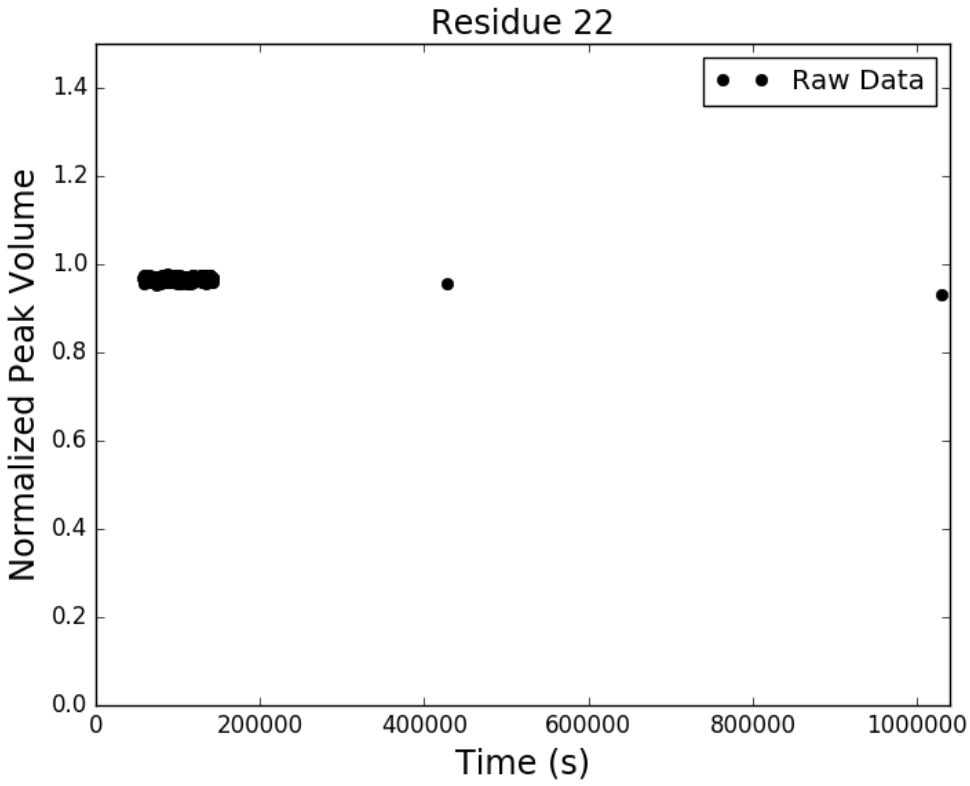


Figure C.100: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 22 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

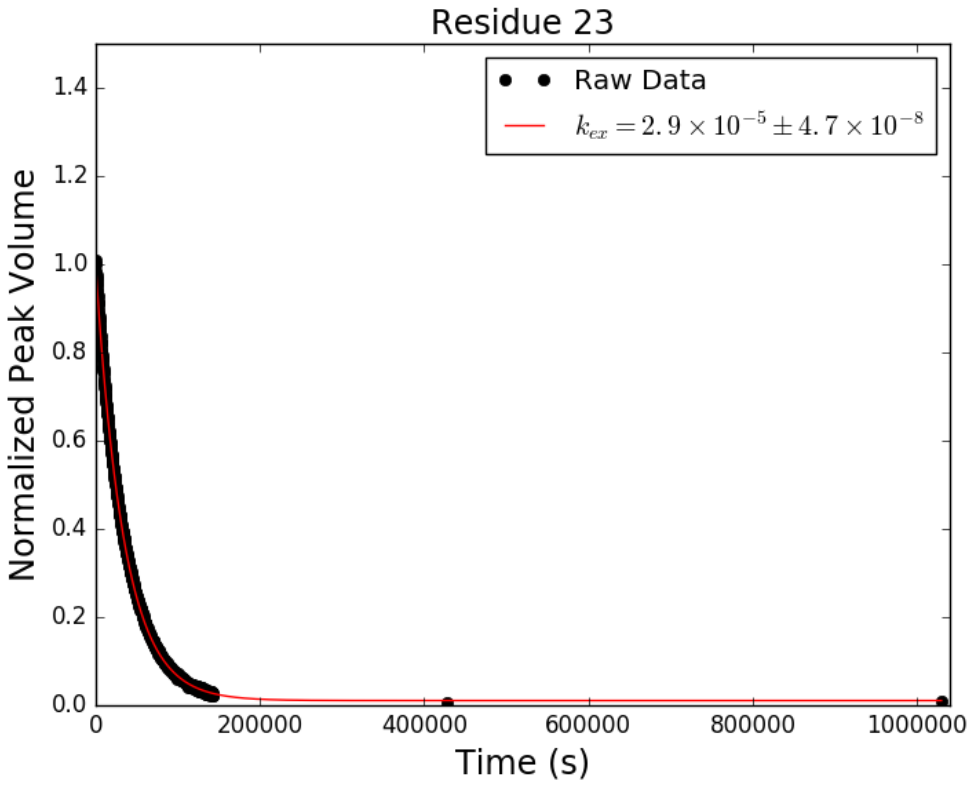


Figure C.101: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 23 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

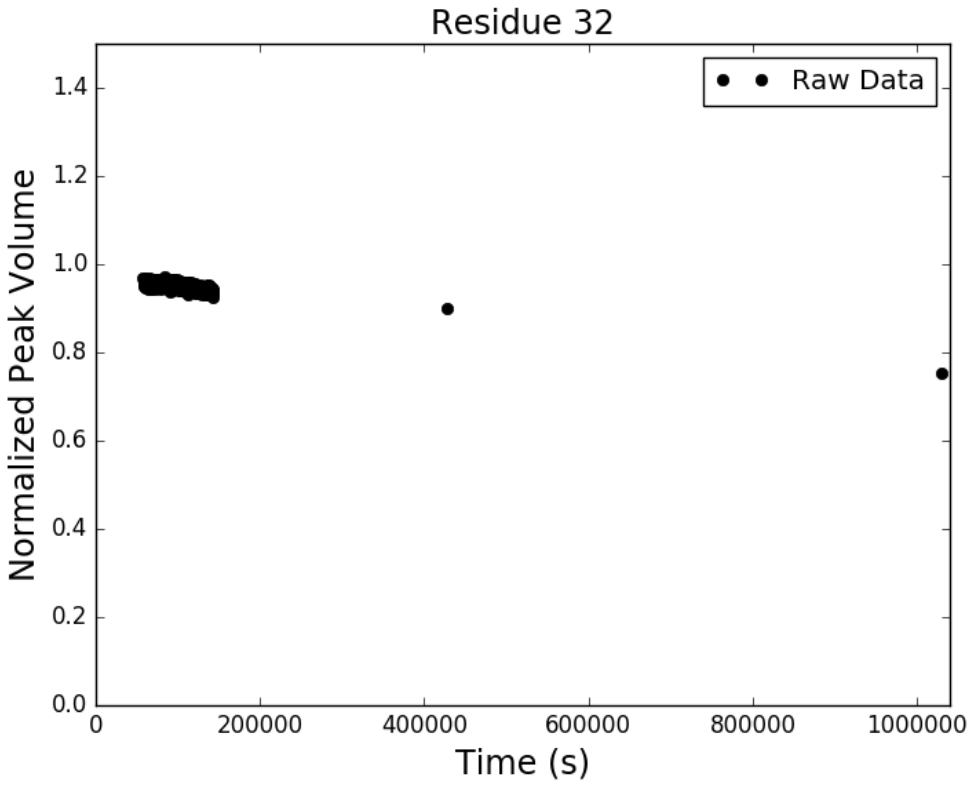


Figure C.102: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 32 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

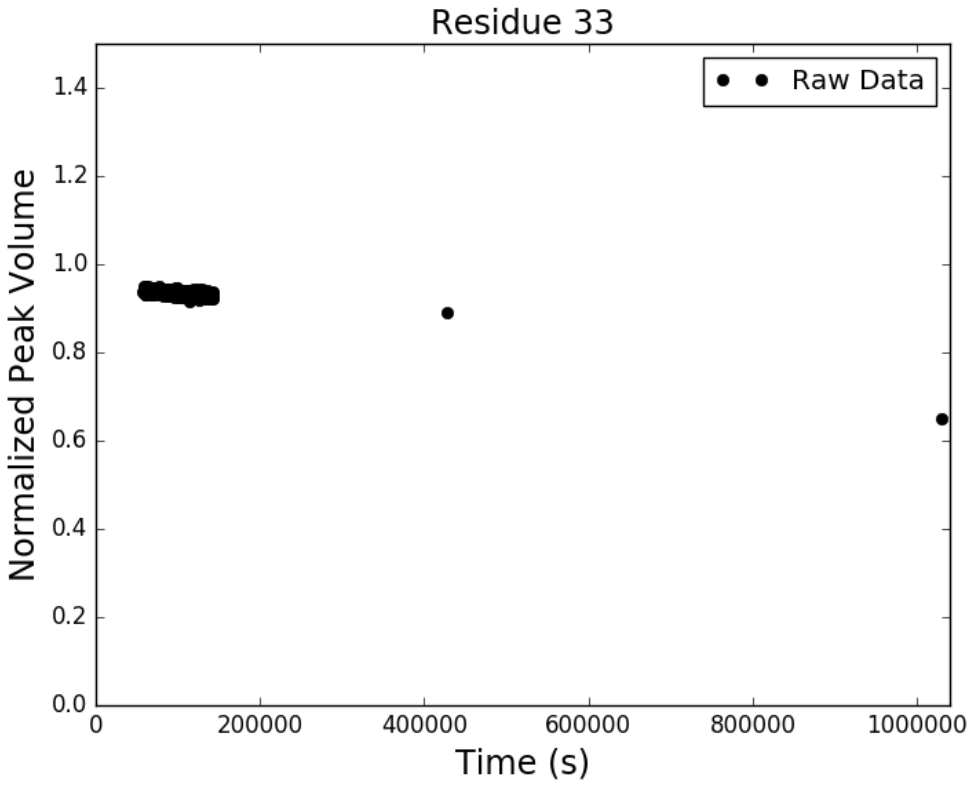


Figure C.103: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 33 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

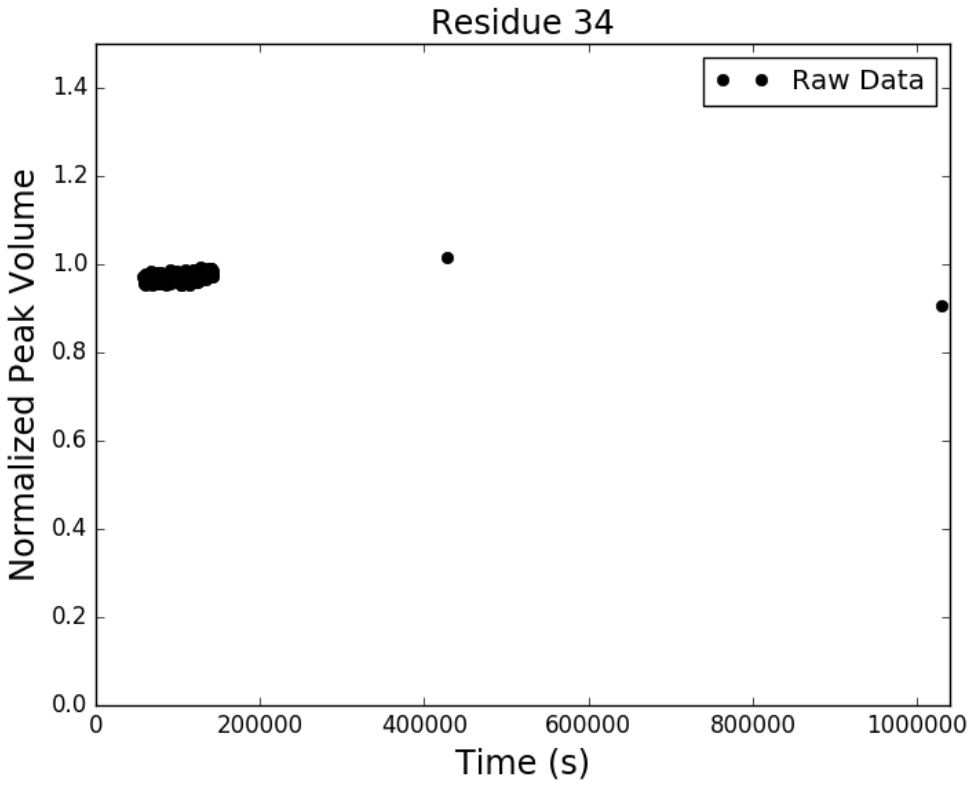


Figure C.104: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 34 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

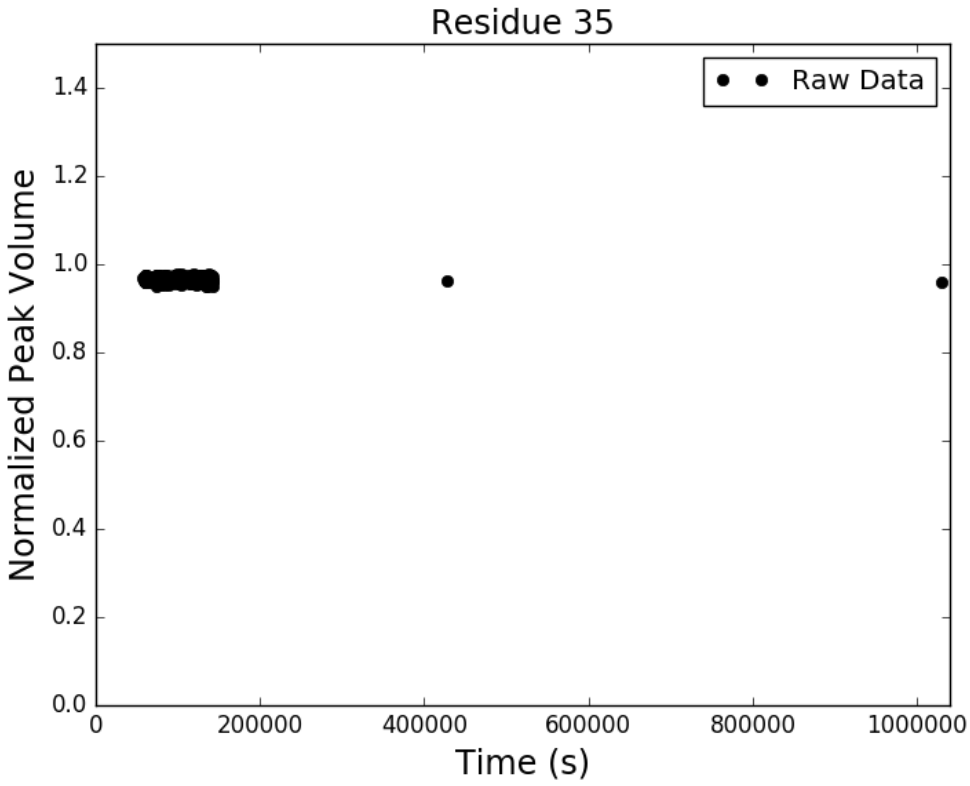


Figure C.105: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 35 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

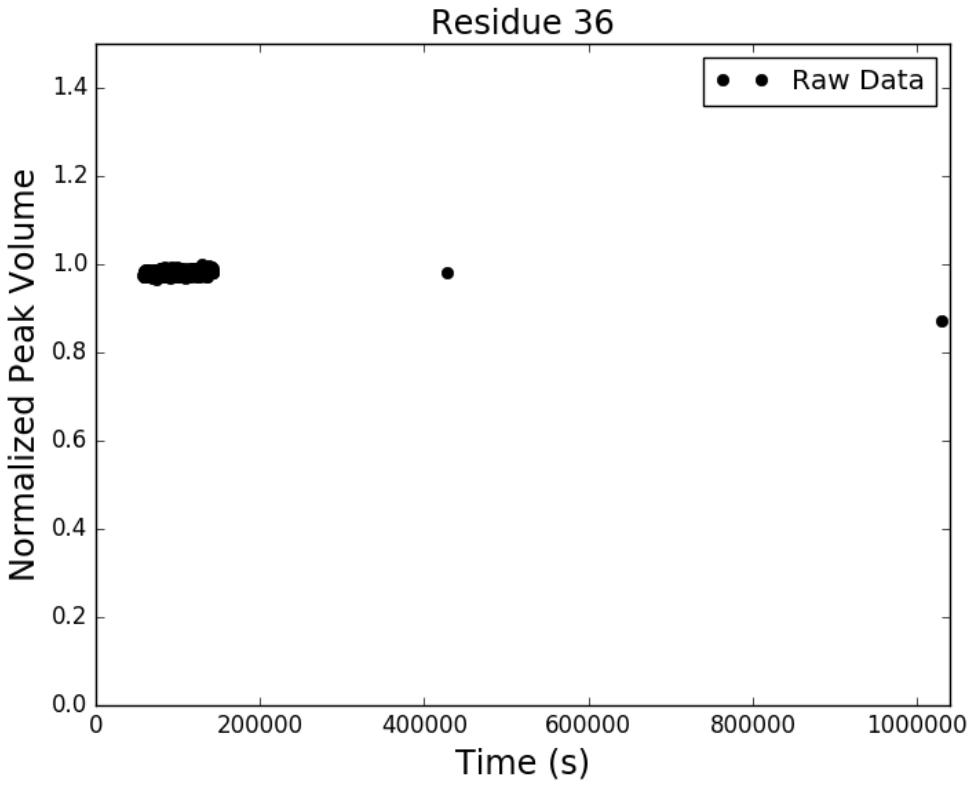


Figure C.106: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 36 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

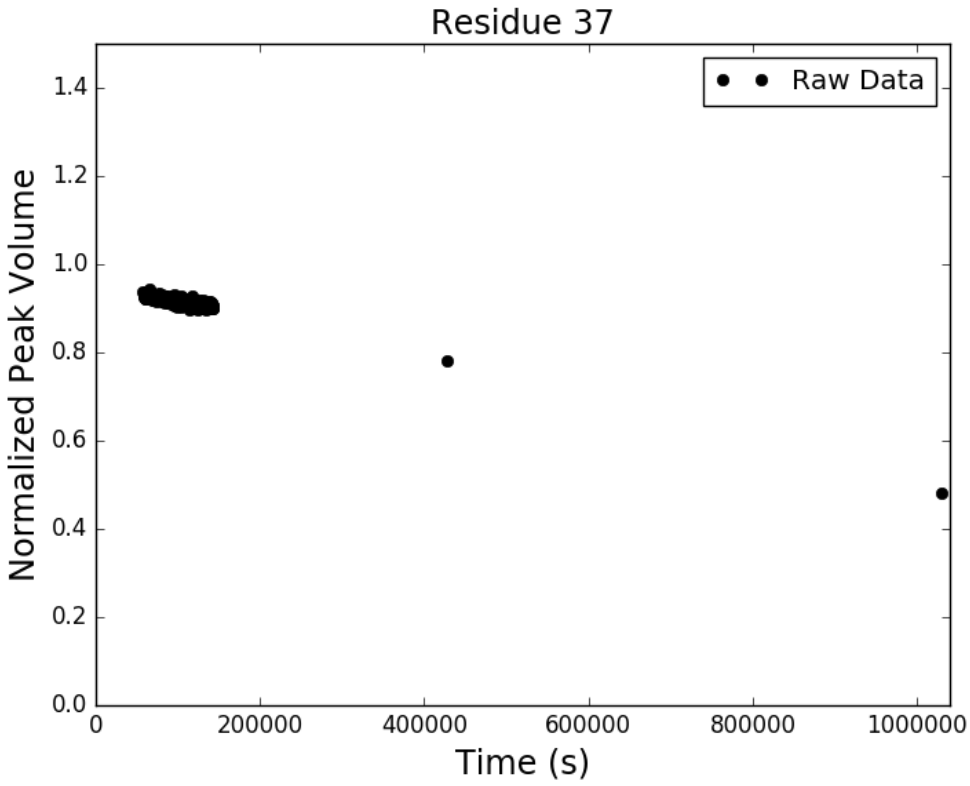


Figure C.107: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 37 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

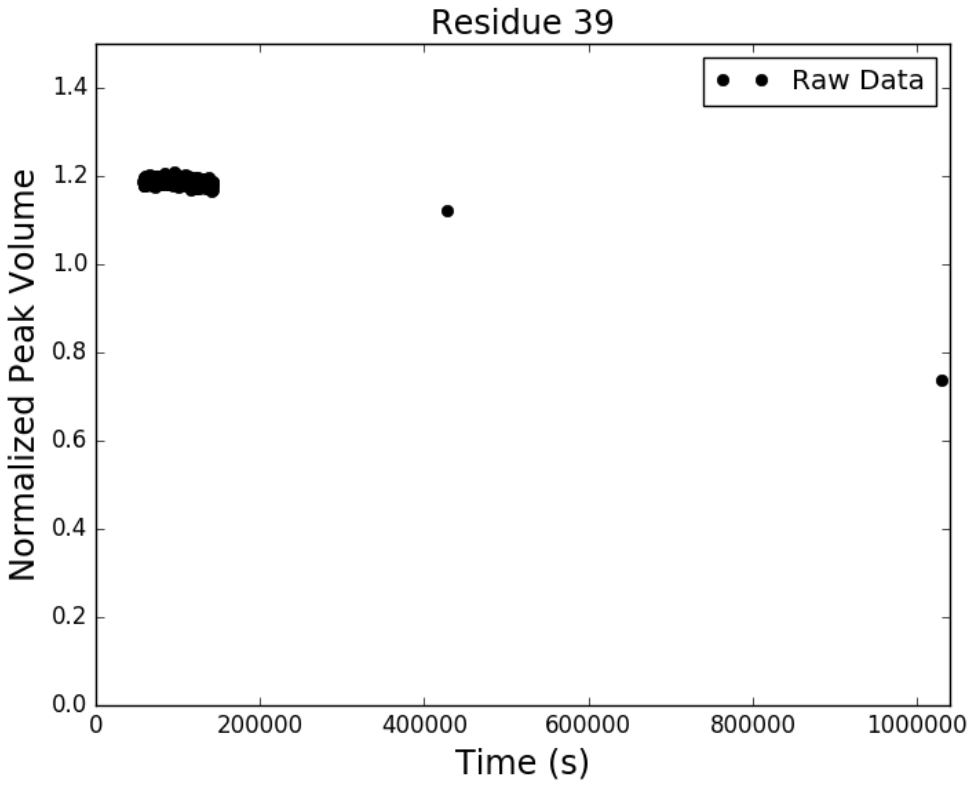


Figure C.108: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 39 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

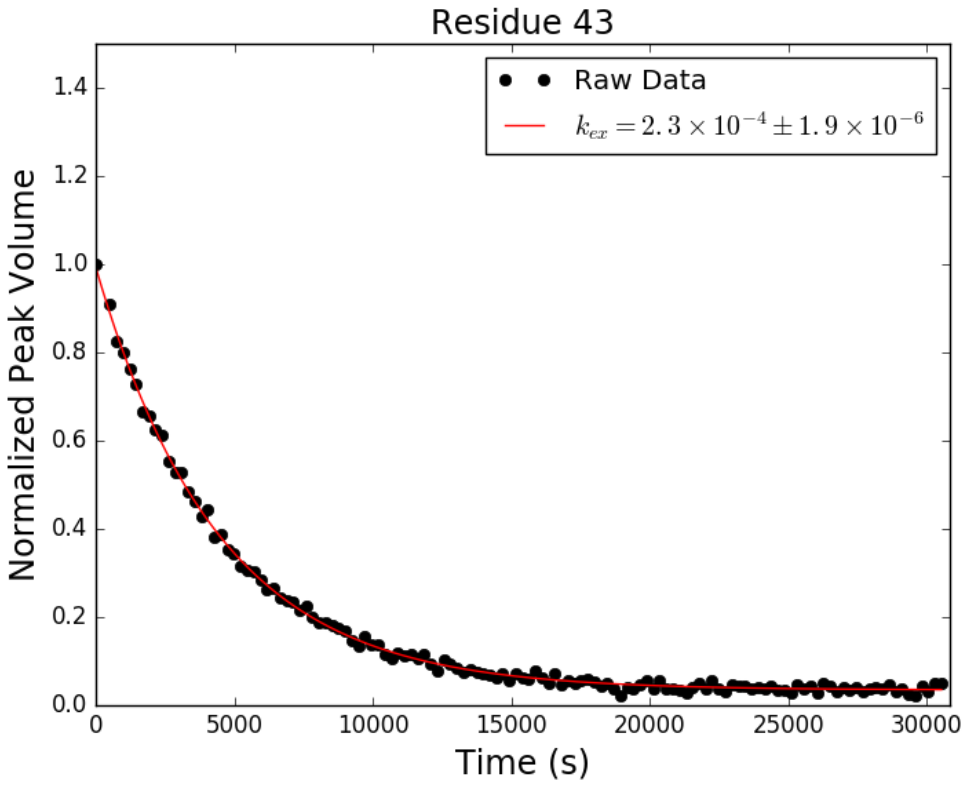


Figure C.109: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 43 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

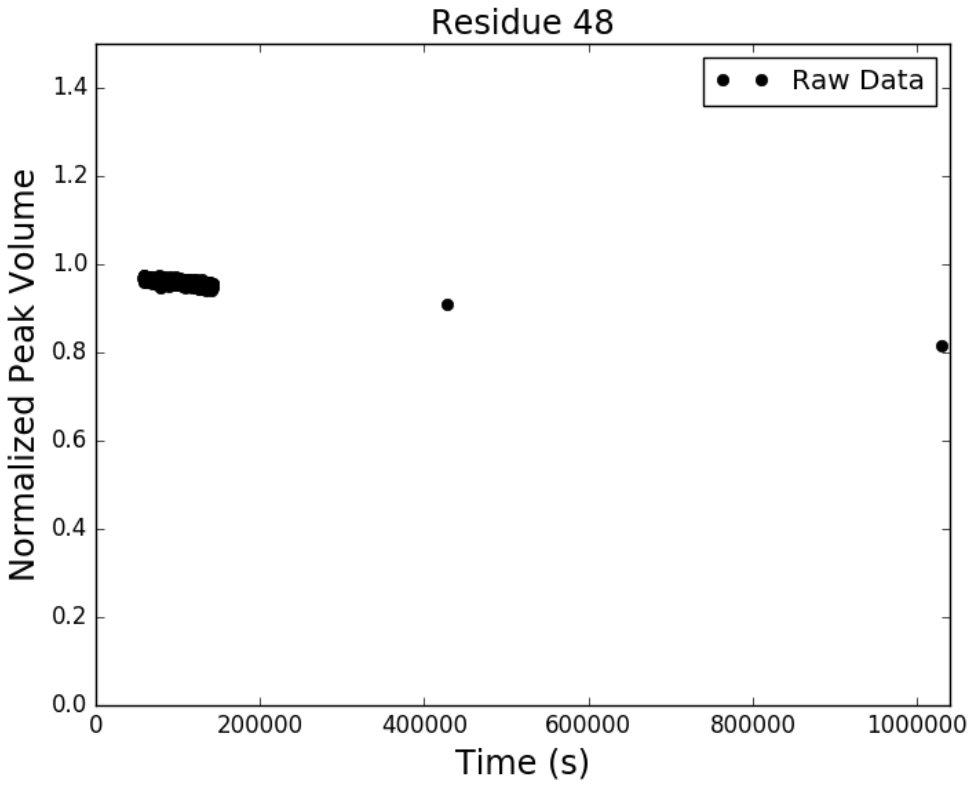


Figure C.110: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 48 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

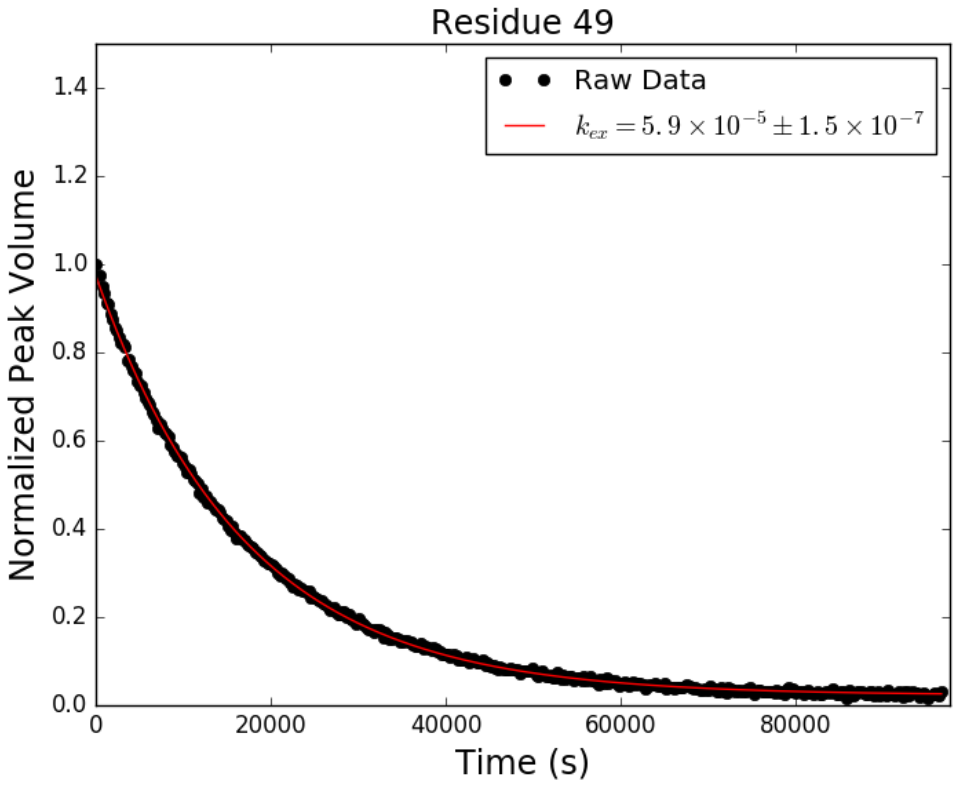


Figure C.111: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 49 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

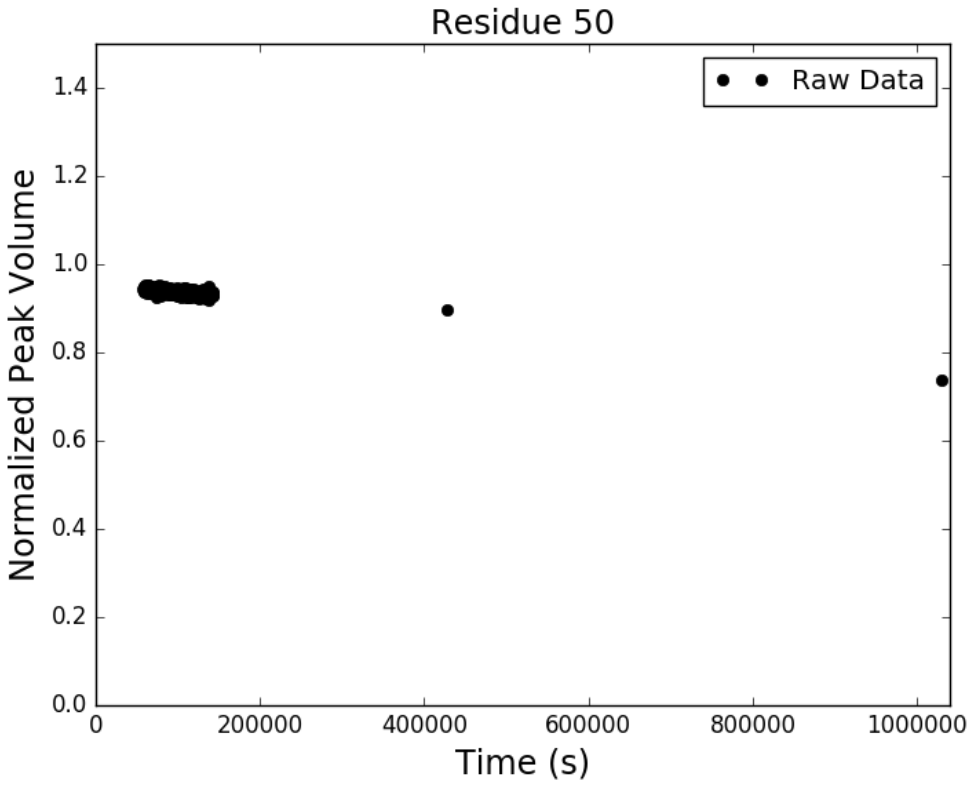


Figure C.112: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 50 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

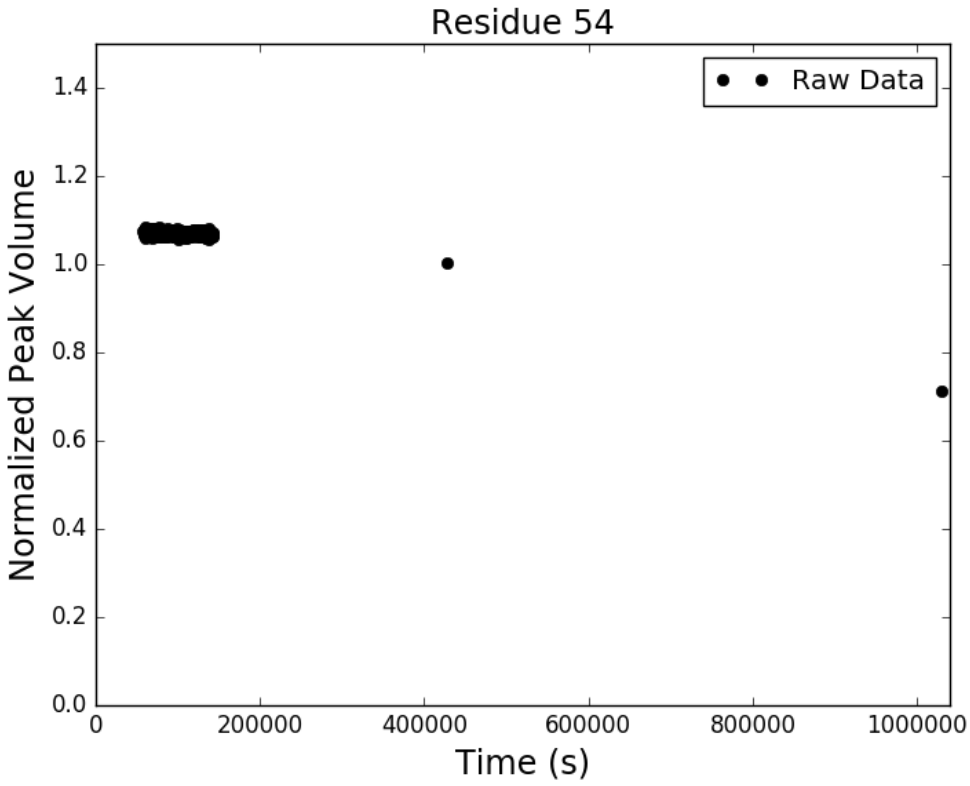


Figure C.113: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 54 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

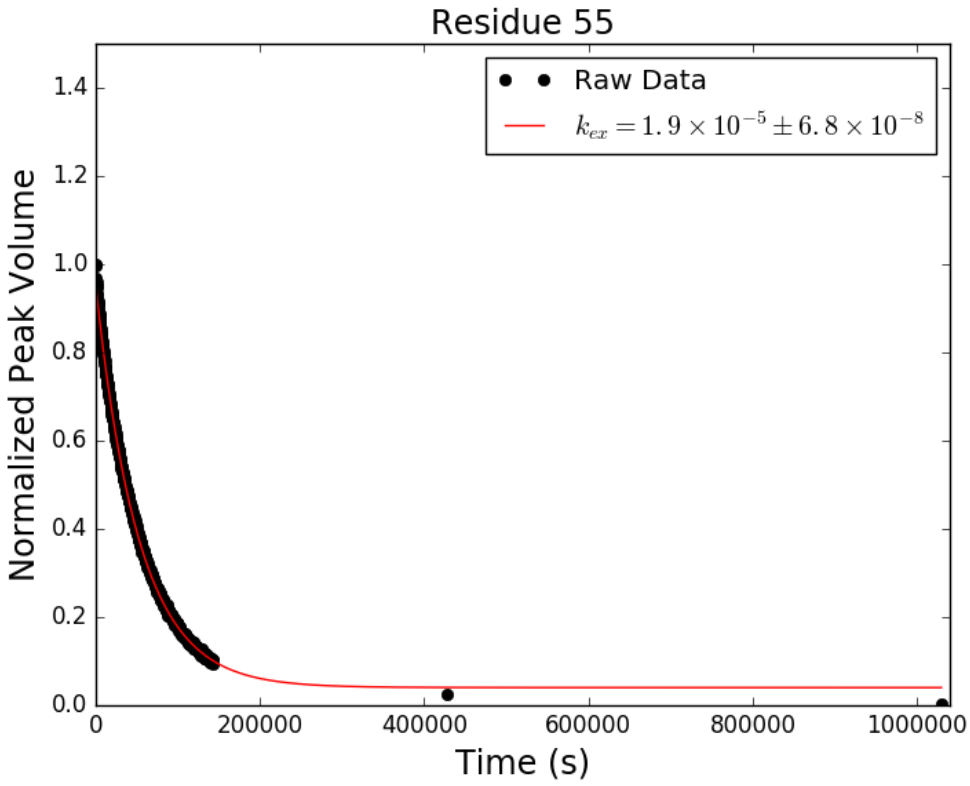


Figure C.114: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 55 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

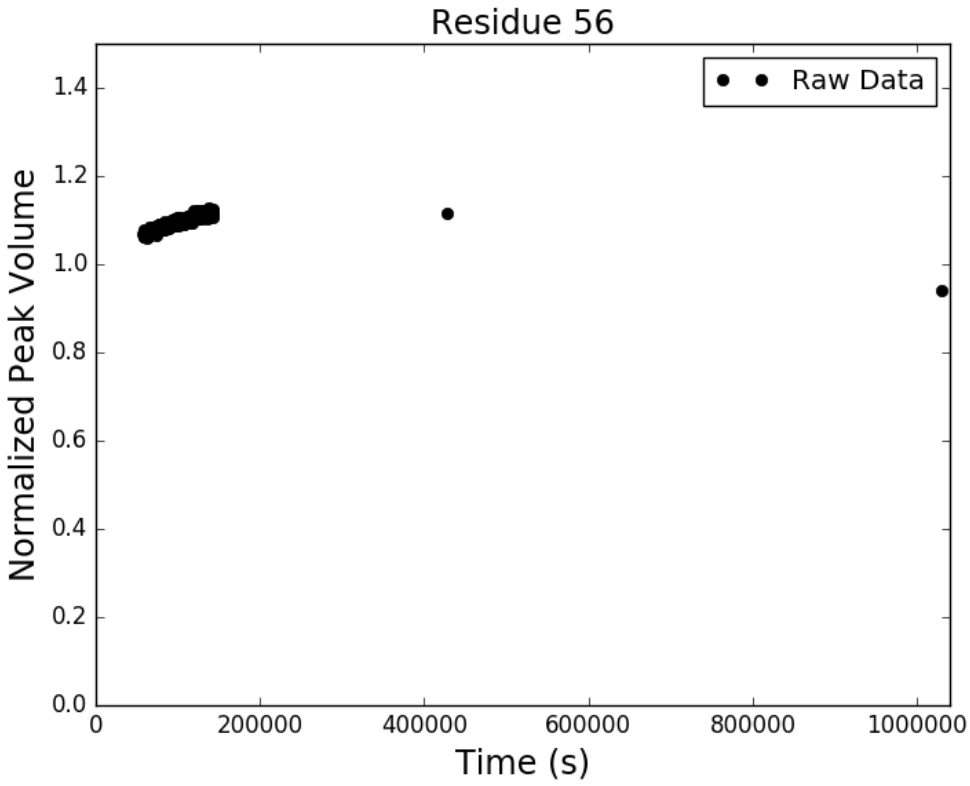


Figure C.115: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 56 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

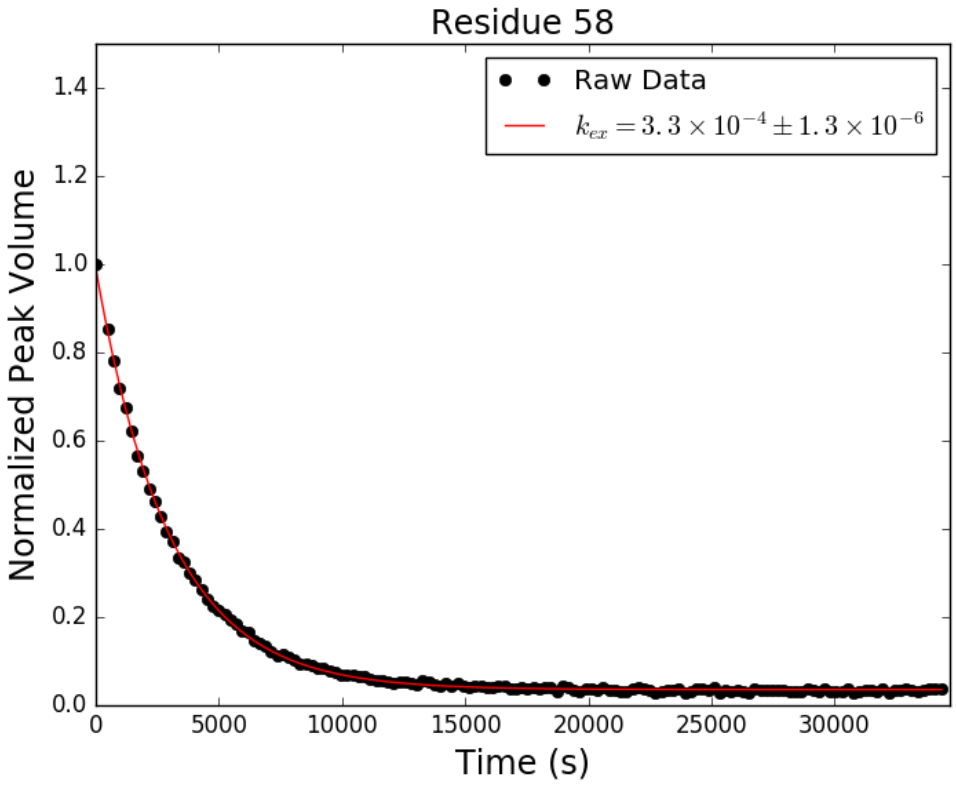


Figure C.116: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 58 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

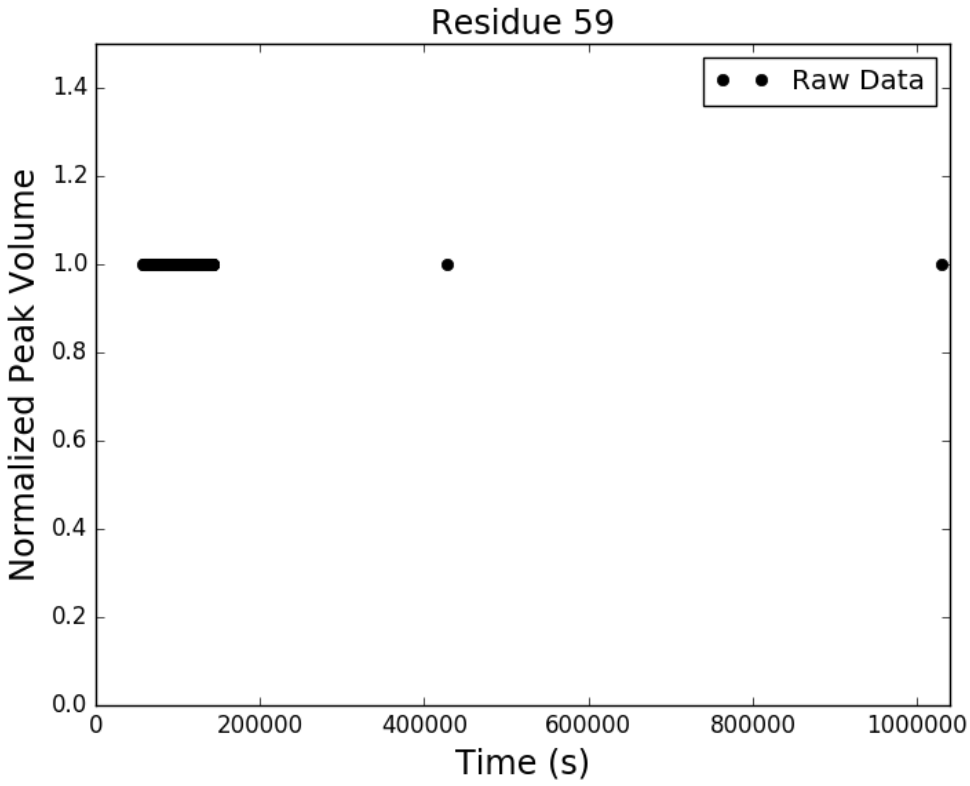


Figure C.117: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 59 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

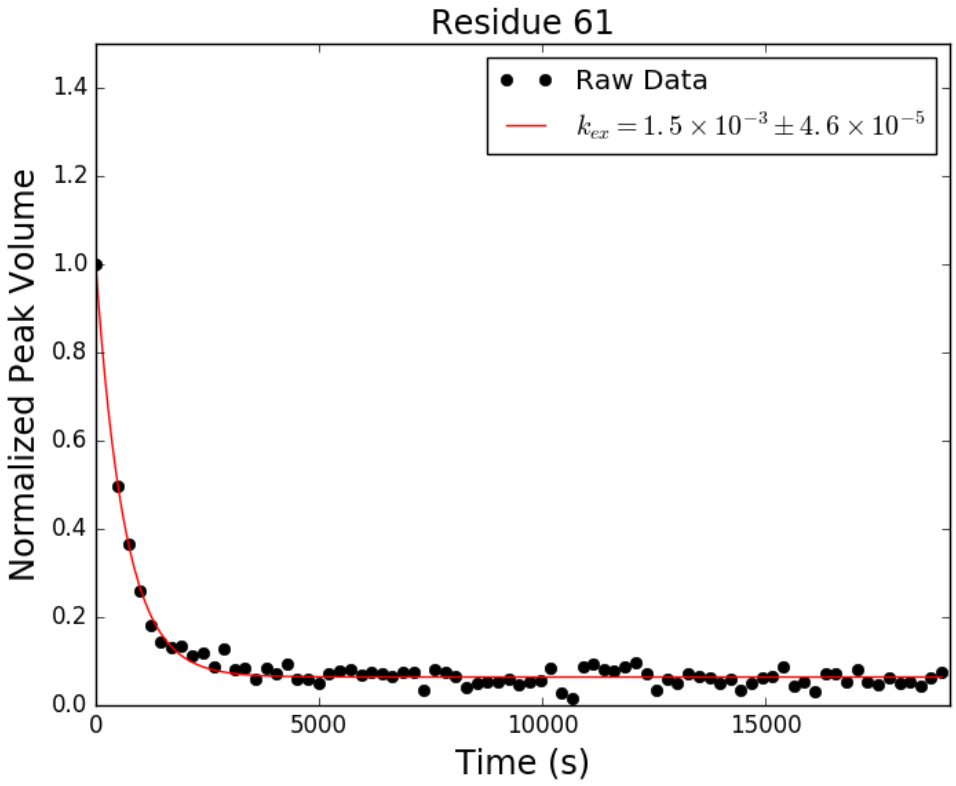


Figure C.118: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 61 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

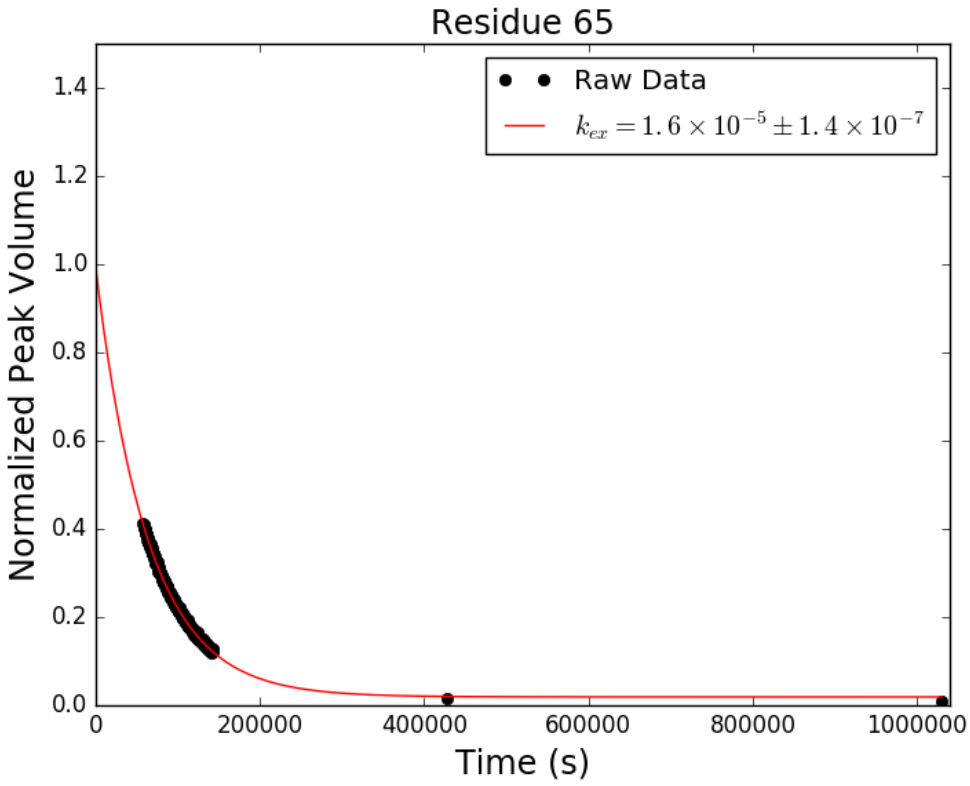


Figure C.119: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 65 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

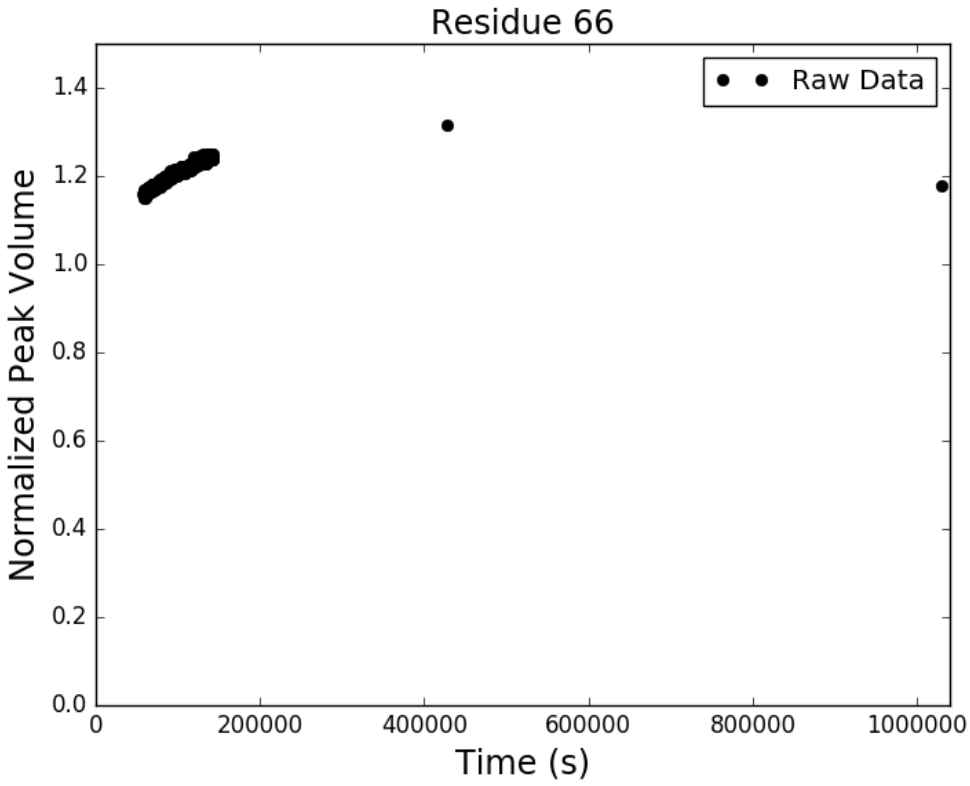


Figure C.120: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 66 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

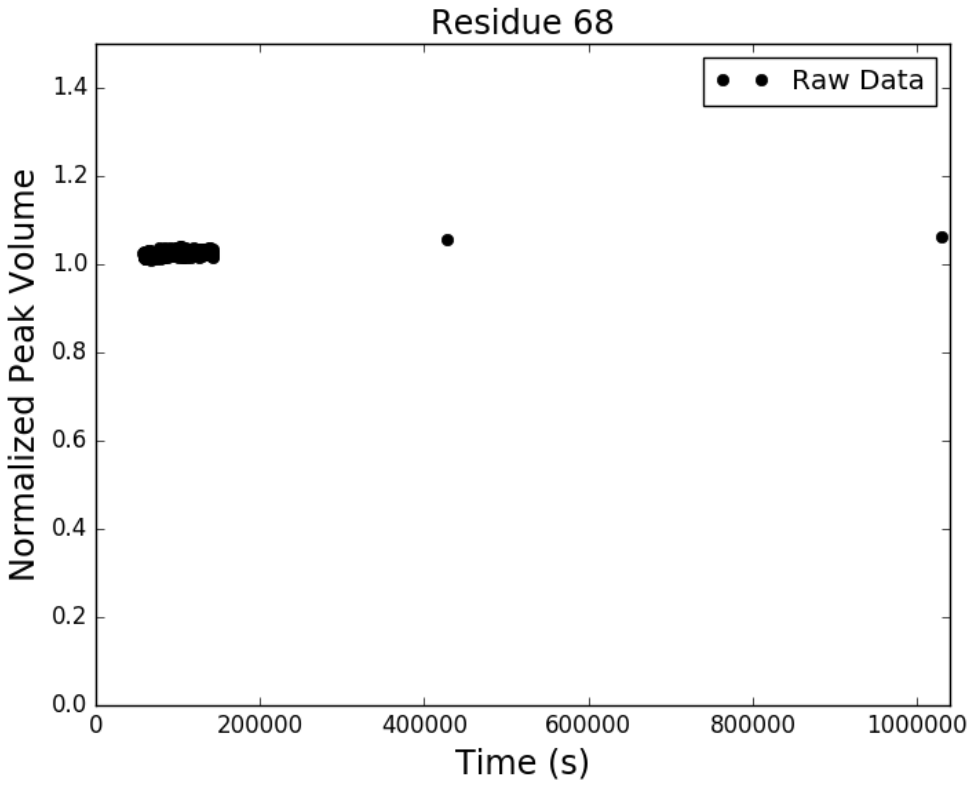


Figure C.121: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 68 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

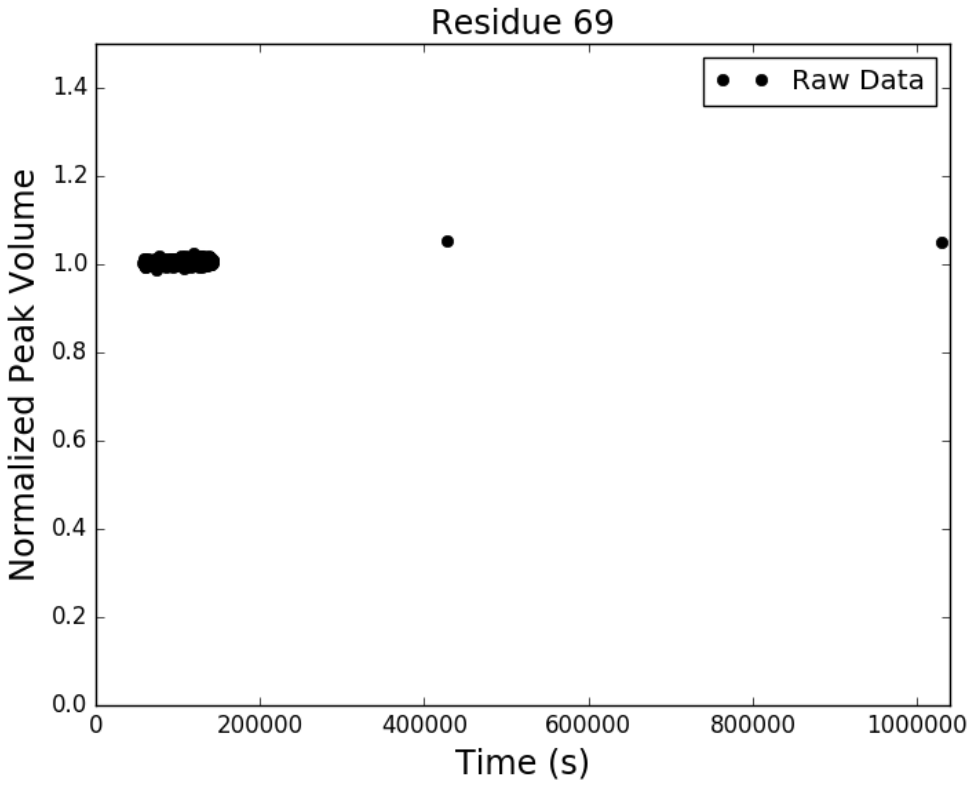


Figure C.122: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 69 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

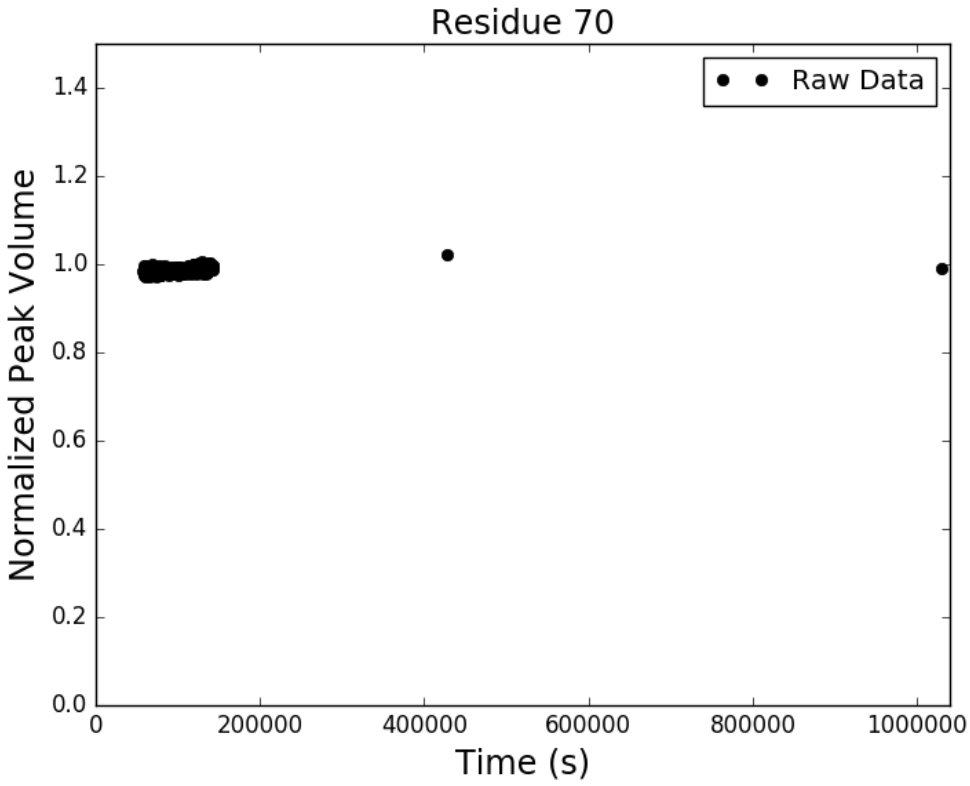


Figure C.123: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 70 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

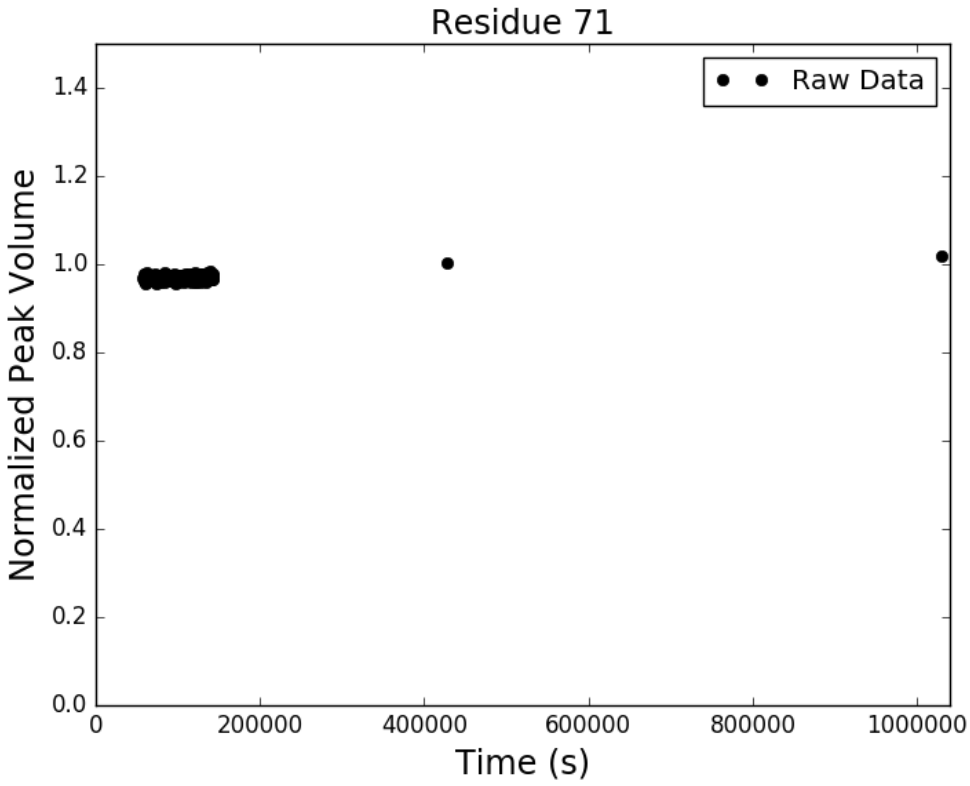


Figure C.124: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 71 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

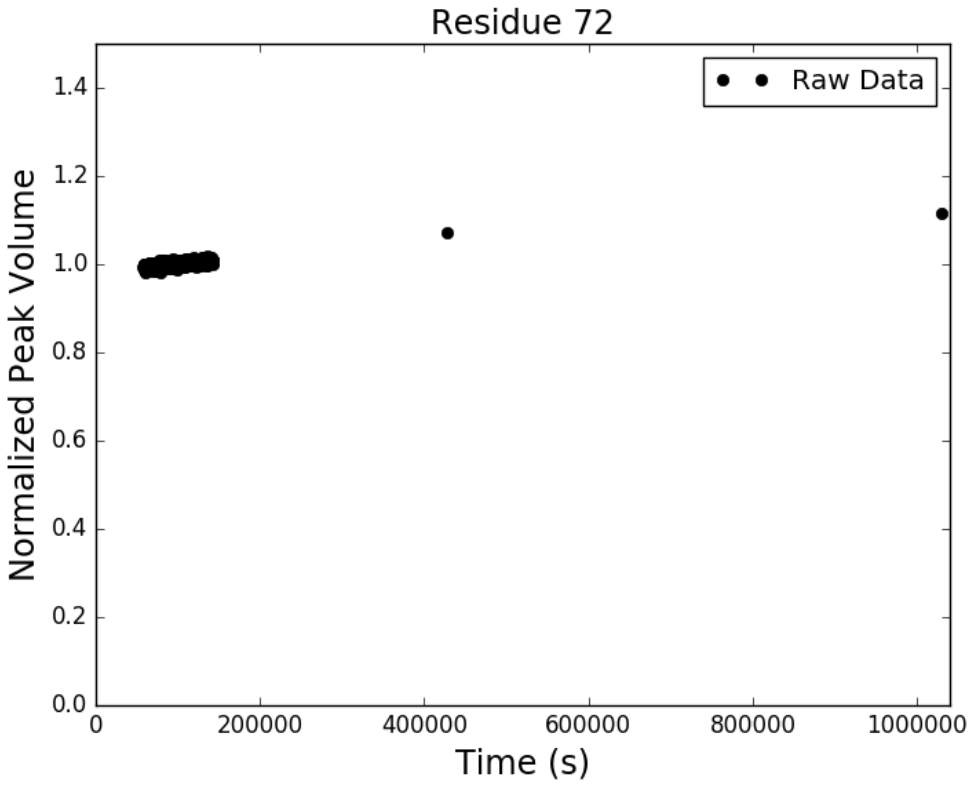


Figure C.125: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 72 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

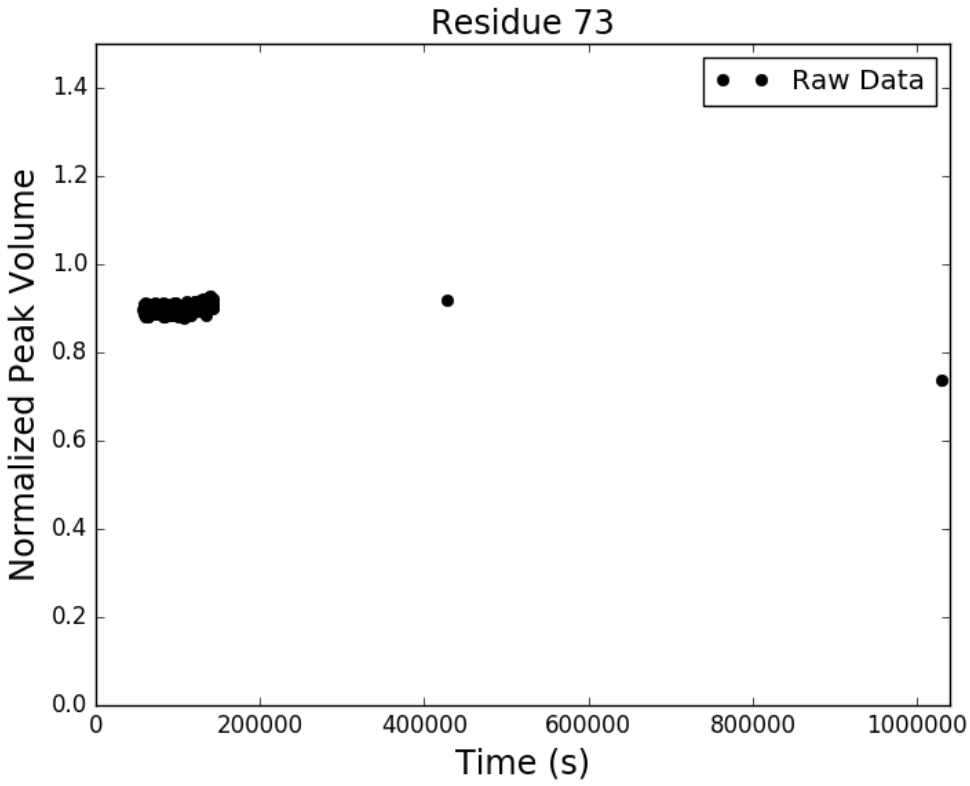


Figure C.126: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 73 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

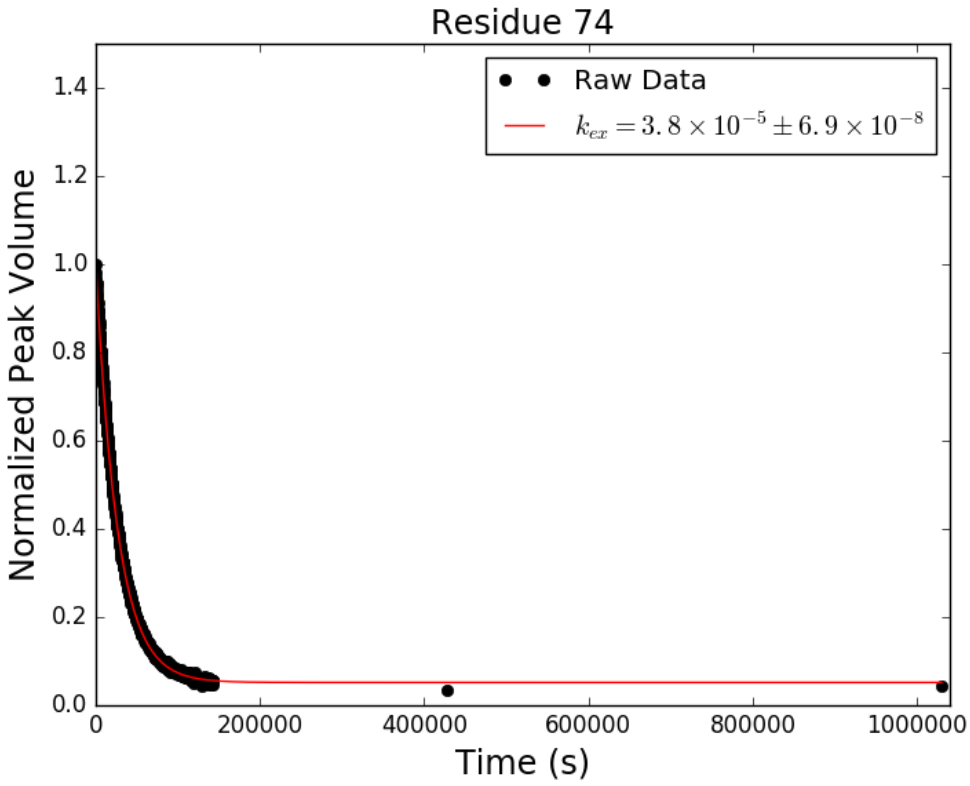


Figure C.127: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 74 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

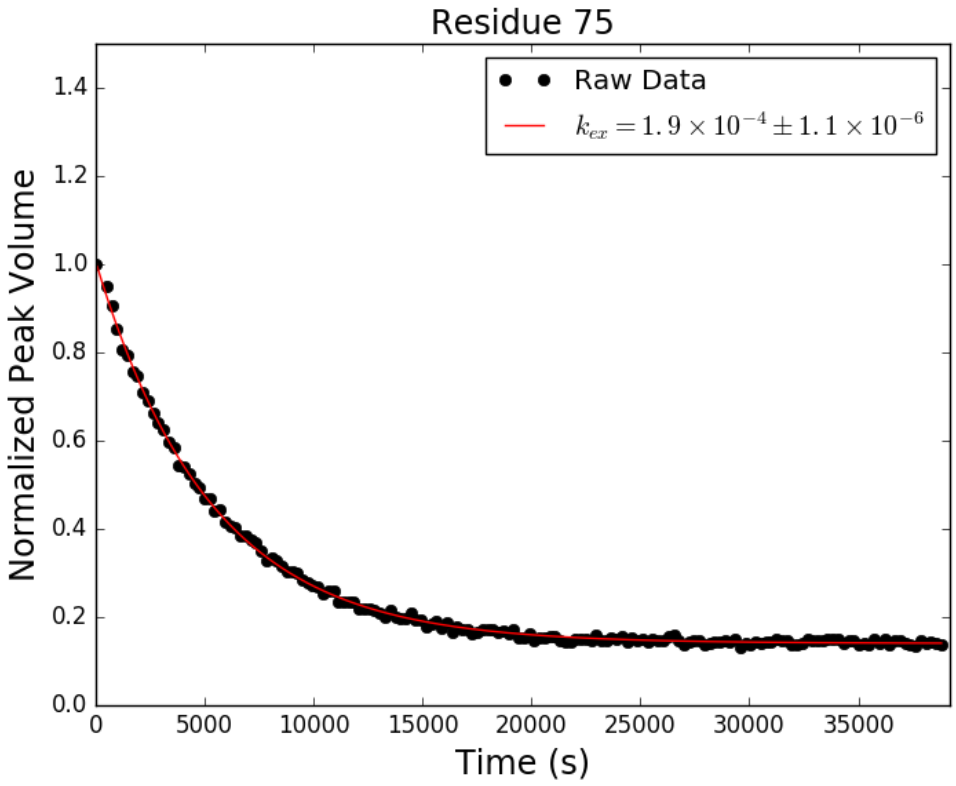


Figure C.128: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 75 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

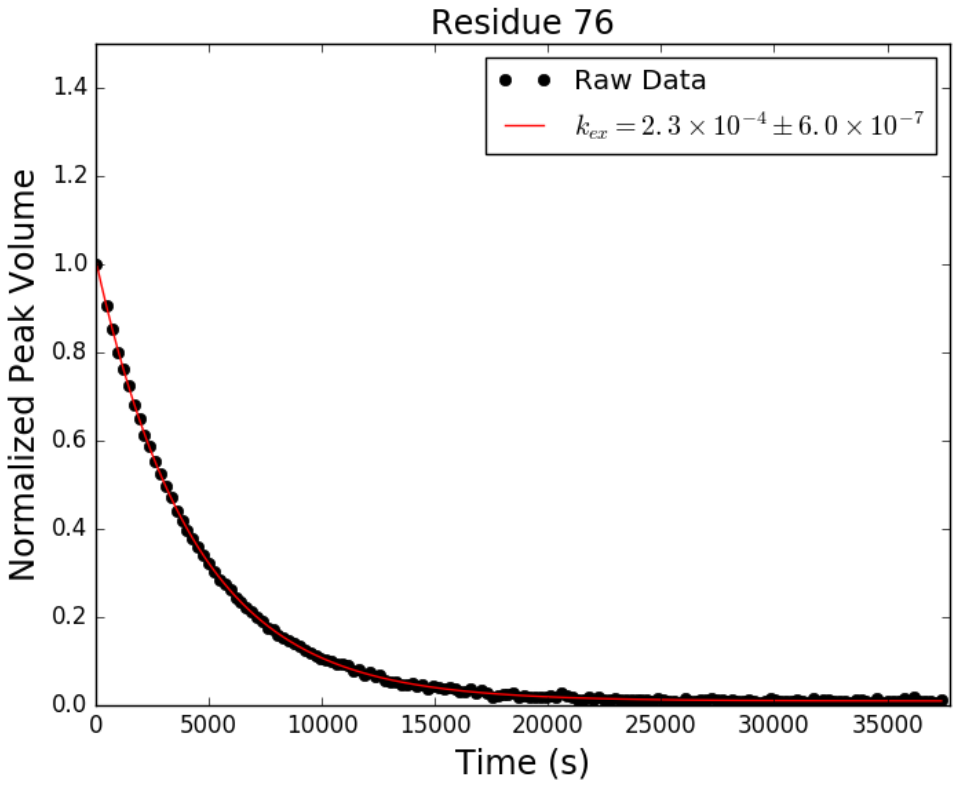


Figure C.129: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 76 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

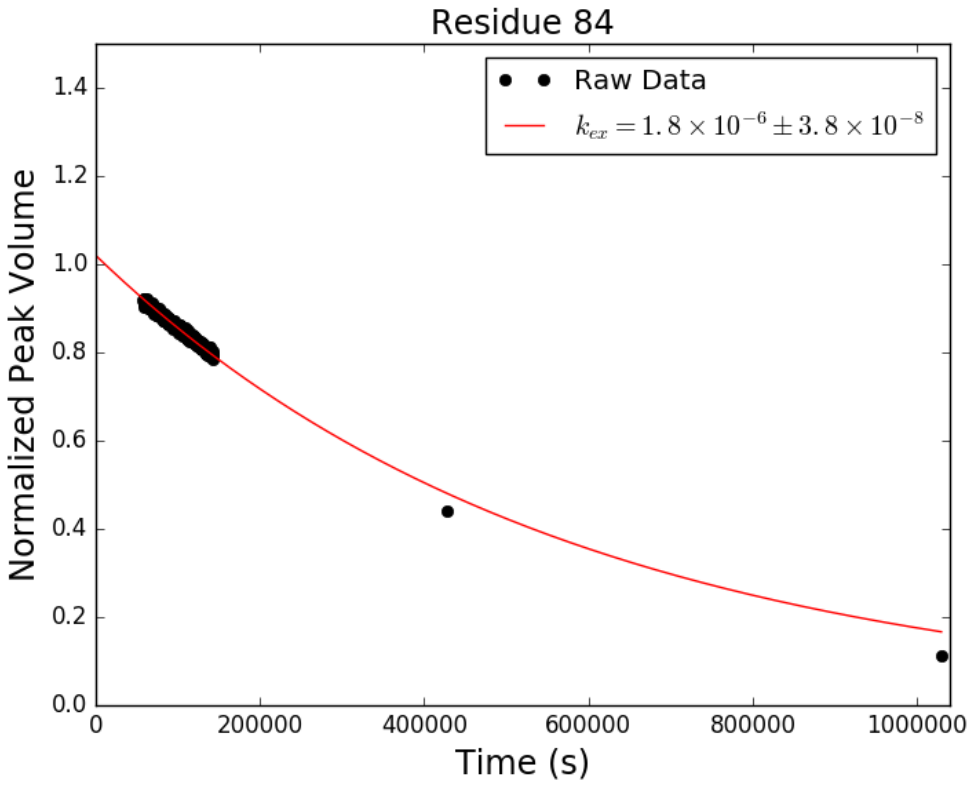


Figure C.130: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 84 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

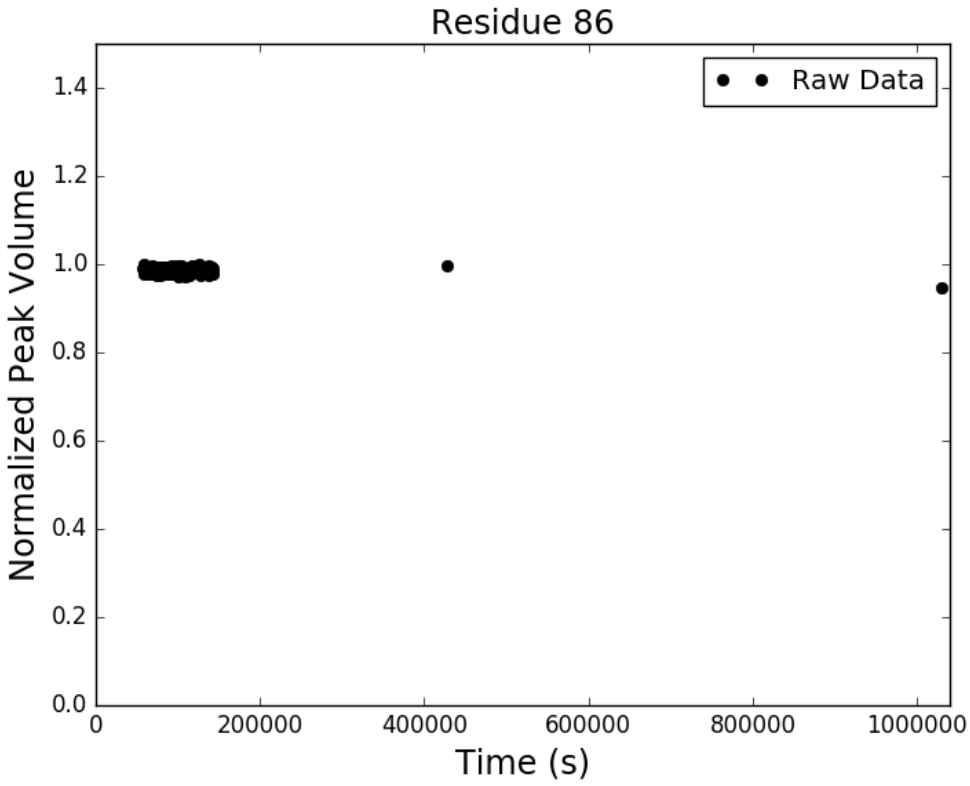


Figure C.131: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 86 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

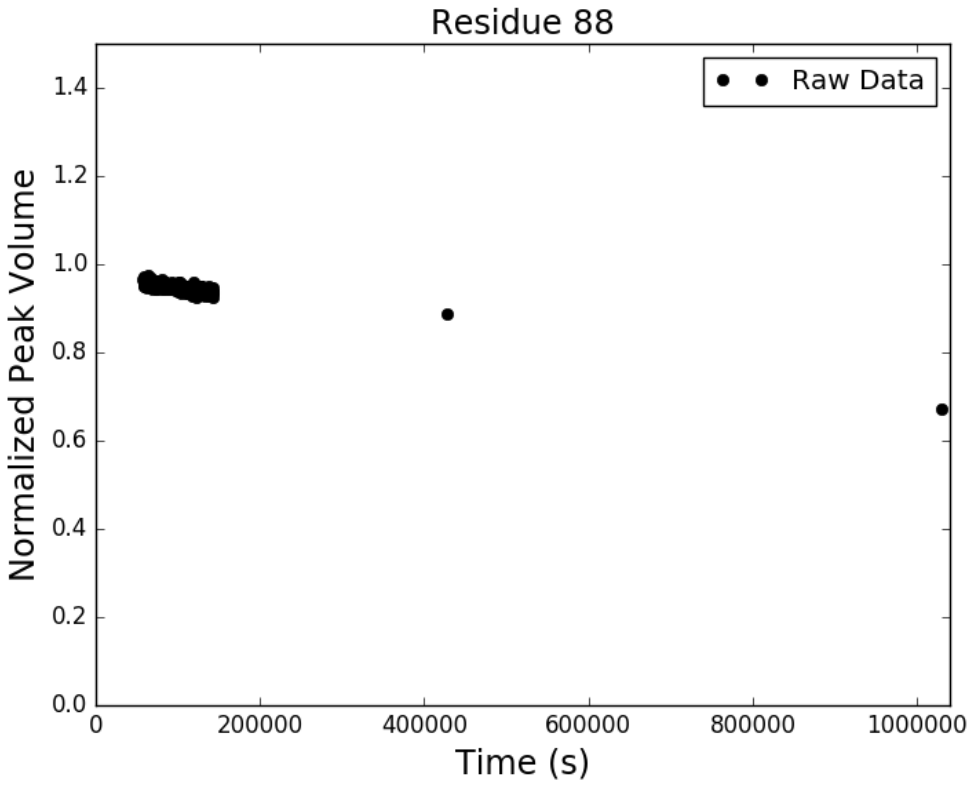


Figure C.132: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 88 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

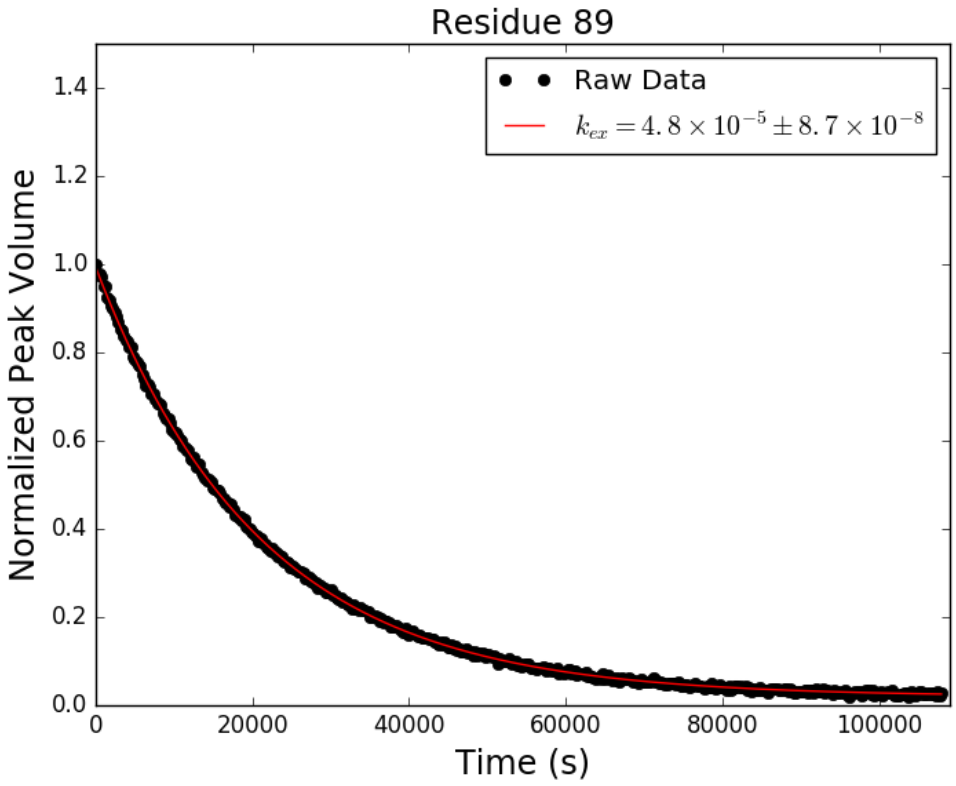


Figure C.133: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 89 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

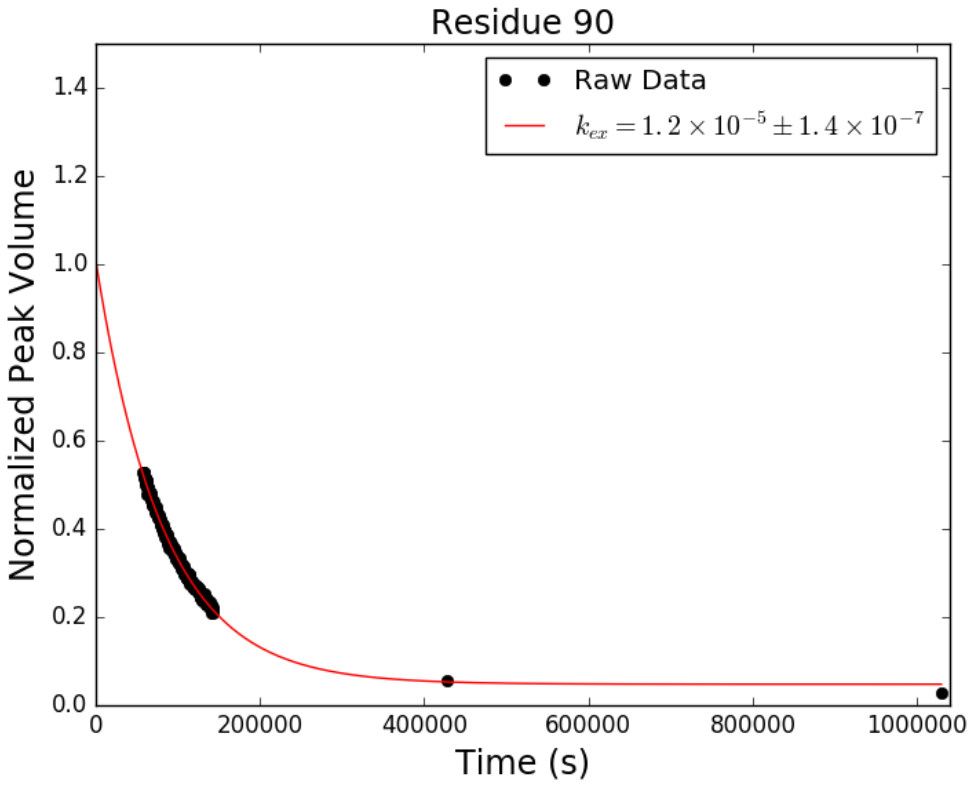


Figure C.134: Adnectin amide H-D exchange monitored by ^{1}H-^{15}N SOFAST-HMQC: residue 90 of V75R. Black dots show integrated peak volumes as a fraction of the initial peak volume for this residue and normalized using an internal standard (the peak volume of residue I59, which does not measurably exchange on the timescale shown). If at least one H-D exchange half-life is captured in the experimental data, the result of fitting said data to a single exponential decay of the form $A \cdot e^{-k_{ex}t} + C$ is shown in red.

Appendix D

Full Amino Acid Sequences of Adnectins Studied by NMR

Table D.1: Full amino acid sequences (as encoded by the DNA constructs cloned into pET-9d expression vectors) of Adnectins studied by NMR.

Identifier[1]	**Amino Acid Sequence**[2]
pWT	MGVSDVPRDLEVVAATPTSLLISWDAPAVTVRYYRITYGETGGNSPVQEFTVPGS KSTATISGLKPGVDYTITVYAVTGRGESPASSKPISINYRTEIDKPSQHHHHHH
Parent	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVPKN VYTATISGLKPGVDYTITVYAVTLLRDYGPISINYRTEIDKPSQHHHHHH
L78I	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVPKN VYTATISGLKPGVDYTITVYAVTLIRDYGPISINYRTEIDKPSQHHHHHH
V75R	MGVSDVPRDLEVVAATPTSLLISWSARLKVARYYRITYGETGGNSPVQEFTVPKN VYTATISGLKPGVDYTITVYARTLLRDYGPISINYRTEIDKPSQHHHHHH
L18V/Y88F	MGVSDVPRDLEVVAATPTSVLISWSARLKVARYYRITYGETGGNSPVQEFTVPKN VYTATISGLKPGVDYTITVYAVTLLRDYGPISINFRTEIDKPSQHHHHHH

[1]Except for pWT (pseudo wild-type ^{10}Fn3), identifiers indicate mutations relative to the Parent Adnectin

[2]Single letter amino acid code; sites of mutation relative to the Parent Adnectin highlighted

Appendix E

NMR Pulse Programs

E.1 SOFAST HMQC (sfhmqcf3gpph)

```
;sfhmqcf3gpph
;avance-version (09/11/18)
;SOFAST HMQC
;2D H-1/X correlation via heteronuclear zero and double quantum
;   coherence
;phase sensitive
;with decoupling during acquisition
;
;P.Schanda and B. Brutscher, J. Am. Chem. Soc. 127, 8014 (2005)
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


prosol relations=<triple>


#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"d11=30m"
"d12=20u"
"d21=1s/(cnst4*2)"


"in0=inf1"

"d0=in0/2-p21*4/3.1415"


"DELTA1=d21-p16-d16-p39*cnst39"
"DELTA2=p39*cnst39-de-4u"


"spoff23=bf1*(cnst19/1000000)-o1"
```

```
"spoff24=bf1*(cnst19/1000000)-o1"


1 ze
  d11 pl26:f3
2 d1 do:f3
3 d12 pl3:f3
  50u UNBLKGRAD

  p16:gp2
  d16

  (p39:sp23 ph1):f1
  p16:gp1
  d16

#   ifdef LABEL_CN
  (center (p40:sp24 ph2):f1 (p8:sp13 ph1):f2 (DELTA1 p21 ph3 d0 p21 ph4 DELTA1):f3 )
#   else
  (center (p40:sp24 ph2):f1 (DELTA1 p21 ph3 d0 p21 ph4 DELTA1):f3 )
#   endif /*LABEL_CN*/


  DELTA2
  p16:gp1
  d16 pl26:f3
  4u BLKGRAD
  go=2 ph31 cpd3:f3
  d1 do:f3 mc #0 to 2
     F1PH(ip3, id0)
exit


ph1=0
ph2=0
ph3=0 2
ph4=0 0 2 2
ph31=0 2 2 0


;pl3 : f3 channel - power level for pulse (default)
;pl26: f3 channel - power level for CPD/BB decoupling (low power)
;sp13: f2 channel - shaped pulse 180 degree (adiabatic)
;sp23: f1 channel - shaped pulse 120 degree
;                   (Pc9_4_120.1000 or Q5.1000)
;sp24: f1 channel - shaped pulse 180 degree (Rsnob.1000)
;p8 : f2 channel - 180 degree shaped pulse for inversion (adiabatic)
;p16: homospoil/gradient pulse                       [1 msec]
;p21: f3 channel -  90 degree high power pulse
;p39: f1 channel - 120 degree shaped pulse for excitation
;                    Pc9_4_120.1000 (120o)    (3.0ms at 600.13 MHz)
;                 (or Q5.1000 (90o)            (2.0ms at 600.13 MHz) )
;p40: f1 channel - 180 degree shaped pulse for refocussing
;                    Rsnob.1000               (1.0ms at 600.13 MHz)
;d0 : incremented delay (2D) = in0/2-p21*4/3.1415
;d1 : relaxation delay
;d11: delay for disk I/O                             [30 msec]
;d12: delay for power switching                      [20 usec]
;d16: delay for homospoil/gradient recovery
;d21 : 1/(2J)NH
;cnst4: = J(NH)
;cnst19: H(N) chemical shift (offset, in ppm)
;cnst39: compensation of chemical shift evolution during p39
;          Pc9_4_120.1000: 0.529
;          Q5.1000: -0.07
```

```
;inf1: 1/SW(N) = 2 * DW(N)
;in0: 1/ SW(N) = 2 * DW(N)
;nd0: 1
;NS: 2 * n
;DS: 16
;aq: <= 50 msec
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEC
;cpd3: decoupling according to sequence defined by cpdprg3: garp4.p62
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence
;          use pulse of >= 350 usec


;use gradient ratio:    gp 1 : gp 2
;                         11 :    7


;for z-only gradients:
;gpz1: 11%
;gpz2:  7%

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100


                                          ;preprocessor-flags-start
;LABEL_CN: for C-13 and N-15 labeled samples start experiment with
;             option -DLABEL_CN (eda: ZGOPTNS)
                                          ;preprocessor-flags-end


;Processing

;PHC0(F1): 90
;PHC1(F1): -180
;FCOR(F1): 1



;$Id: sfhmqcf3gpph,v 1.1.2.8 2009/11/18 11:19:58 ber Exp $
```

E.2 HSQC (hsqcetfpf3gpsi)

```
;hsqcetfpf3gpsi
;avance-version (02/07/15)
;HSQC
;2D H-1/X correlation via double inept transfer
;   using sensitivity improvement
;phase sensitive using Echo/Antiecho-TPPI gradient selection
;with decoupling during acquisition
;using f3 - channel
;using flip-back pulse
;
;A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn.
;   Reson. 93, 151-170 (1991)
;L.E. Kay, P. Keifer & T. Saarinen, J. Am. Chem. Soc. 114,
;   10663-5 (1992)
;J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky,
;   S.J. Glaser, O.W. Sorensen & C. Griesinger, J. Biomol. NMR 4,
;   301-306 (1994)
;S. Grzesiek & A. Bax, J. Am. Chem. Soc. 115, 12593-12594 (1993)
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


prosol relations=<triple>


#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"p2=p1*2"
"p22=p21*2"
"d0=3u"
"d11=30m"
"d26=1s/(cnst4*4)"


"DELTA1=p16+d16+8u"

#   ifdef LABEL_CN
"DELTA=p16+d16+larger(p2,p8)+d0*2"
#   else
"DELTA=p16+d16+p2+d0*2"
#   endif /*LABEL_CN*/


1 ze
  d11 pl16:f3
2 d1 do:f3
3 (p1 ph1)
  d26 pl3:f3
  (center (p2 ph2) (p22 ph6):f3 )
  d26 UNBLKGRAD
  (p1 ph2)
  4u pl0:f1
  (p11:sp1 ph1:r):f1
  4u
  p16:gp1
```

```
  d16 pl1:f1
  (p21 ph3):f3
  d0

#   ifdef LABEL_CN
  (center (p2 ph7) (p8:sp13 ph1):f2 )
#   else
  (p2 ph7)
#   endif /*LABEL_CN*/

  d0
  p16:gp2*EA
  d16
  (p22 ph4):f3
  DELTA
  (center (p1 ph1) (p21 ph4):f3 )
  d24
  (center (p2 ph1) (p22 ph1):f3 )
  d24
  (center (p1 ph2) (p21 ph5):f3 )
  d26
  (center (p2 ph1) (p22 ph1):f3 )
  d26
  (p1 ph1)
  DELTA1
  (p2 ph1)
  4u
  p16:gp3
  d16 pl16:f3
  4u BLKGRAD
  go=2 ph31 cpd3:f3
  d1 do:f3 mc #0 to 2
     F1EA(igrad EA & ip5*2, id0 & ip3*2 & ip6*2 & ip31*2)
exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
ph5=1 1 3 3
ph6=0
ph7=0 0 2 2
ph31=2 0 0 2


;pl0 : 120dB
;pl1 : f1 channel - power level for pulse (default)
;pl3 : f3 channel - power level for pulse (default)
;pl16: f3 channel - power level for CPD/BB decoupling
;sp1: f1 channel - shaped pulse  90 degree
;sp13: f2 channel - shaped pulse 180 degree (adiabatic)
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p8 : f2 channel - 180 degree shaped pulse for inversion (adiabatic)
;p11: f1 channel -  90 degree shaped pulse
;p16: homospoil/gradient pulse                          [1 msec]
;p21: f3 channel -  90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;d0 : incremented delay (2D)                            [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O                                [30 msec]
;d16: delay for homospoil/gradient recovery
;d24: 1/(4J)YH for YH
;     1/(8J)YH for all multiplicities
```

```
;d26: 1/(4J(YH))
;cnst4: = J(YH)
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 1 * n
;DS: >= 16
;td1: number of experiments
;FnMODE: echo-antiecho
;cpd3: decoupling according to sequence defined by cpdprg3
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence


;use gradient ratio:    gp 1 : gp 2 : gp 3
;                         50 :   80 : 20.1       for C-13
;                         50 :   80 :  8.1       for N-15

;for z-only gradients:
;gpz1: 50%
;gpz2: 80%
;gpz3: 20.1% for C-13, 8.1% for N-15

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100
;gpnam3: SINE.100


                                          ;preprocessor-flags-start
;LABEL_CN: for C-13 and N-15 labeled samples start experiment with
;             option -DLABEL_CN (eda: ZGOPTNS)
                                          ;preprocessor-flags-end



;$Id: hsqcetfpf3gpsi,v 1.2.8.1 2005/11/10 13:18:57 ber Exp $
```

E.3 NOESY-HSQC (noesyhsqcetf3gp3d)

```
;noesyhsqcetf3gp3d
;avance-version (05/10/20)
;NOESY-HSQC
;3D sequence with
;   homonuclear correlation via dipolar coupling
;   dipolar coupling may be due to noe or chemical exchange.
;   H-1/X correlation via double inept transfer
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho-TPPI gradient selection (t2)
;using trim pulses in inept transfer
;with decoupling during acquisition
;(use parameterset NOESYHSQCETF3GP3D)
;
;A.L. Davis, J. Keeler, E.D. Laue & D. Moskau, J. Magn. Reson. 98,
;   207-216 (1992)
;
;$CLASS=HighRes
;$DIM=3D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"p2=p1*2"
"p22=p21*2"
"d0=3u"
"d10=3u"
"d11=30m"
"d12=20u"
"d13=4u"
"d26=1s/(cnst4*4)"


"DELTA2=d26-p16-d13-4u"

#   ifdef LABEL_CN
"DELTA=larger(p14,p22)+d0*2"
"DELTA1=p16+d16+larger(p2,p14)+d10*2"
#   else
"DELTA=p22+d0*2"
"DELTA1=p16+d16+p2+d10*2"
#   endif /*LABEL_CN*/


aqseq 321


1 ze
  d11 pl16:f3
2 d1 do:f3
3 d12 pl3:f3
  (p1 ph7)
  DELTA
  (p2 ph8)
  d0

#   ifdef LABEL_CN
```

```
  (center (p14:sp3 ph1):f2 (p22 ph1):f3 )
#   else
  (p22 ph1):f3
#   endif /*LABEL_CN*/

  d0
  (p1 ph1):f1
  d8
  (p1 ph1):f1
  d26
  (center (p2 ph1) (p22 ph6):f3 )
  d26 UNBLKGRAD
  p28 ph1
  d13
  (p1 ph2) (p21 ph3):f3
  d10

#   ifdef LABEL_CN
  (center (p2 ph5) (p14:sp3 ph1):f2 )
#   else
  (p2 ph5)
#   endif /*LABEL_CN*/

  d10
  p16:gp1*EA
  d16
  (p22 ph4):f3
  DELTA1
  (ralign (p1 ph1) (p21 ph4):f3 )
  d26
  (center (p2 ph1) (p22 ph1):f3 )
  d13
  p16:gp2
  DELTA2 pl16:f3
  4u BLKGRAD
  go=2 ph31 cpd3:f3
  d1 do:f3 mc #0 to 2
     F1PH(rd10 & rp3 & rp6 & rp31 & ip7 & ip8, id0)
     F2EA(igrad EA, id10 & ip3*2 & ip6*2 & ip31*2)
exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph5=0 0 2 2
ph6=0
ph7=0 0 0 0 2 2 2 2
ph8=1 1 1 1 3 3 3 3
ph31=0 2 0 2 2 0 2 0 2 0 2 0 0 2 0 2


;pl1 : f1 channel - power level for pulse (default)
;pl3 : f3 channel - power level for pulse (default)
;pl16: f3 channel - power level for CPD/BB decoupling
;sp3: f2 channel - shaped pulse 180 degree (adiabatic)
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p14: f2 channel - 180 degree shaped pulse for inversion (adiabatic)
;p16: homospoil/gradient pulse
;p21: f3 channel -  90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;p28: f1 channel - trim pulse
;d0 : incremented delay (F1 in 3D)                      [3 usec]
```

```
;d1 : relaxation delay; 1-5 * T1
;d8 : mixing time
;d10: incremented delay (F2 in 3D)                      [3 usec]
;d11: delay for disk I/O                               [30 msec]
;d12: delay for power switching                        [20 usec]
;d13: short delay                                      [4 usec]
;d16: delay for homospoil/gradient recovery
;d26: 1/(4J)YH
;cnst4: = J(YH)
;in0: 1/(2 * SW(H)) = DW(H)
;nd0: 2
;in10: 1/(2 * SW(X)) = DW(X)
;nd10: 2
;NS: 8 * n
;DS: >= 16
;td1: number of experiments in F1
;td2: number of experiments in F2
;FnMODE: States-TPPI (or TPPI) in F1
;FnMODE: echo-antiecho in F2
;cpd3: decoupling according to sequence defined by cpdprg3
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence


;use gradient ratio:    gp 1 : gp 2
;                         80 : 20.1    for C-13
;                         80 :  8.1    for N-15

;for z-only gradients:
;gpz1: 80%
;gpz2: 20.1% for C-13, 8.1% for N-15

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100


                                        ;preprocessor-flags-start
;LABEL_CN: for C-13 and N-15 labeled samples start experiment with
;             option -DLABEL_CN (eda: ZGOPTNS)
                                        ;preprocessor-flags-end


;for older datasets use AQORDER : 3 - 2 - 1



;$Id: noesyhsqcetf3gp3d,v 1.4.2.1 2005/11/10 13:18:59 ber Exp $
```

E.4 TOCSY-HSQC (dipsihsqcf3gpsi3d)

```
;dipsihsqcf3gpsi3d
;avance-version (05/10/20)
;TOCSY-HSQC
;3D sequence with
;   homonuclear Hartman-Hahn transfer using DIPSI2 sequence
;      for mixing
;   H-1/X correlation via double inept transfer
;      using sensitivity improvement
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho-TPPI gradient selection (t2)
;using trim pulses in inept transfer
;using f3 - channel
;(use parameterset DIPSIHSQCF3GPSI3D)
;
;A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn.
;   Reson. 93, 151-170 (1991)
;L.E. Kay, P. Keifer & T. Saarinen, J. Am. Chem. Soc. 114,
;   10663-5 (1992)
;J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky,
;   S.J. Glaser, O.W. Sorensen & C. Griesinger, J. Biomol. NMR 4,
;   301-306 (1994)
;
;$CLASS=HighRes
;$DIM=3D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"p2=p1*2"
"p22=p21*2"
"d0=3u"
"d10=3u"
"d11=30m"
"d12=20u"
"d13=4u"
"d26=1s/(cnst4*4)"


"DELTA1=d13+p16+d16+4u"

#   ifdef LABEL_CN
"DELTA=p16+d16+larger(p2,p14)+d10*2"
"DELTA2=larger(p14,p22)+d0*2"
#   else
"DELTA=p16+d16+p2+d10*2"
"DELTA2=p22+d0*2"
#   endif /*LABEL_CN*/


"FACTOR1=(d9/(p6*115.112))/2+0.5"
"l1=FACTOR1*2"


aqseq 321
```

```
1 ze
  d11 pl16:f3
2 d11 do:f3
3 d12 pl9:f1
  d1 cw:f1 ph29
  4u do:f1
  d12 pl1:f1

  (p1 ph8)

  DELTA2 pl3:f3
  (p2 ph9)
  d0

#   ifdef LABEL_CN
  (center (p14:sp3 ph1):f2 (p22 ph1):f3 )
#   else
  (p22 ph1):f3
#   endif /*LABEL_CN*/

  d0

  (p1 ph10)
  d20 pl10:f1

                                                    ;begin DIPSI2
7 p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23

  p6*3.556 ph25
  p6*4.556 ph23
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
  p6*2.722 ph23
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25

  p6*3.556 ph25
  p6*4.556 ph23
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
  p6*2.722 ph23
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25

  p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23
```

```
  lo to 7 times l1
                                                    ;end DIPSI2

  d21 pl1:f1
  (p1 ph11)

  d26
  (center (p2 ph1) (p22 ph6):f3 )
  d26 UNBLKGRAD
  p28 ph1
  d13
  (p1 ph2)
  3u
  p16:gp1
  d16
  (p21 ph3):f3
  d10

#   ifdef LABEL_CN
  (center (p2 ph7) (p14:sp3 ph1):f2 )
#   else
  p2 ph7
#   endif /*LABEL_CN*/

  d10
  p16:gp2*EA
  d16
  (p22 ph4):f3
  DELTA
  (center (p1 ph1) (p21 ph4):f3 )
  d24
  (center (p2 ph1) (p22 ph1):f3 )
  d24
  (center (p1 ph2) (p21 ph5):f3 )
  d26
  (center (p2 ph1) (p22 ph1):f3 )
  d26
  (p1 ph1)
  DELTA1
  (p2 ph1)
  d13
  p16:gp3
  d16 pl16:f3
  4u BLKGRAD
  go=2 ph31 cpd3:f3
  d11 do:f3 mc #0 to 2
     F1PH(rd10 & rp3 & rp6 & rp31 & ip8 & ip9 & ip29, id0)
     F2EA(igrad EA & ip5*2, id10 & ip3*2 & ip6*2 & ip31*2)
exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
ph5=1 1 3 3
ph6=0
ph7=0 0 2 2
ph8=0 0 0 0 2 2 2 2
ph9=1 1 1 1 3 3 3 3
ph10=2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
ph11=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph23=1
ph25=3
```

```
ph29=0
ph31=0 2 2 0 2 0 0 2 2 0 0 2 0 2 2 0
     2 0 0 2 0 2 2 0 0 2 2 0 2 0 0 2


;pl0 : 120dB
;pl1 : f1 channel - power level for pulse (default)
;pl3 : f3 channel - power level for pulse (default)
;pl9 : f1 channel - power level for presaturation
;pl10: f1 channel - power level for TOCSY-spinlock
;pl16: f3 channel - power level for CPD/BB decoupling
;sp3 : f2 channel - shaped pulse 180 degree (adiabatic)
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p6 : f1 channel -  90 degree low power pulse
;p14: f2 channel - 180 degree shaped pulse for inversion (adiabatic)
;p16: homospoil/gradient pulse                          [1 msec]
;p21: f3 channel -  90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;p28: f1 channel - trim pulse                           [1 msec]
;d0 : incremented delay (F1 in 3D)                      [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d9 : TOCSY mixing time
;d10: incremented delay (F2 in 3D)                      [3 usec]
;d11: delay for disk I/O                                [30 msec]
;d12: delay for power switching                         [20 usec]
;d13: short delay                                       [4 usec]
;d16: delay for homospoil/gradient recovery
;d20: first z-filter delay                             [10 usec]
;d21: second z-filter delay                            [10 usec]
;d24: 1/(4J)YH for YH
;     1/(8J)YH for all multiplicities
;d26: 1/(4J(YH))
;cnst4: = J(YH)
;l1: loop for DIPSI cycle: ((p6*115.112) * l1) = mixing time
;in0: 1/(2 * SW(H)) = DW(H)
;nd0: 2
;in10: 1/(2 * SW(X)) = DW(X)
;nd10: 2
;NS: 8 * n
;DS: >= 16
;td1: number of experiments
;td2: number of experiments in F2
;FnMODE: States-TPPI (or TPPI) in F1
;FnMODE: echo-antiecho in F2
;cpd3: decoupling according to sequence defined by cpdprg3
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence


;use gradient ratio:    gp 1 : gp 2 : gp 3
;                         50 :   80 :  8.1      for N-15
;                         50 :   80 : 20.1      for C-13

;for z-only gradients:
;gpz1: 50%
;gpz2: 80%
;gpz3: 8.1% for N-15, 20.1% for C-13

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100
;gpnam3: SINE.100


;set pl9 to 120dB when presaturation is not required
```

```
;   use 70 - 80dB to reduce radiation damping


                                          ;preprocessor-flags-start
;LABEL_CN: for C-13 and N-15 labeled samples start experiment with
;             option -DLABEL_CN (eda: ZGOPTNS)
                                          ;preprocessor-flags-end


;$Id: dipsihsqcf3gpsi3d,v 1.2.8.1 2005/11/10 13:18:56 ber Exp $
```

Appendix F

Shift-T Source Code

F.1 ShiftTrack

```
#!/usr/bin/python3

# ShiftTrack.py
#
# This is a program for the automatic calculation of temperature coefficients
# from 1H-15N correlation spectra. It operates on the assumption that the
# peaks corresponding to a particular amide resonance at different temperatures
# fall (approximately) on a line in the 1H-15N plane, and that there is an upper
# limit on the distance between consecutive points. A weak constraint on the
# standard deviation of the distances between the points has also been found to
# be useful.
#
# Kyle Trainor, March 2018

# ------------
# Imports
# ------------

# useful stuff for parsing CSV
import pandas as pd

# useful stuff for linear regression
import numpy as np

# to get command-line arguments
import sys

# to work with files and directories
import os

# to manipulate iterators
from itertools import chain

# plotting
import matplotlib
matplotlib.use('AGG')
import matplotlib.pyplot as plt

# scipy
import scipy, scipy.stats as stats
```

```
41
42  from celery import Celery
43  import time
44
45  # manipulate zip files
46  import zipfile
47
48  # get pathnames matching pattern
49  import glob
50
51  # for emailing results
52  import smtplib
53  from email import encoders
54  from email.mime.base import MIMEBase
55  from email.mime.text import MIMEText
56  from email.mime.multipart import MIMEMultipart
57  from email.mime.application import MIMEApplication
58
59  app = Celery('shifttrack', backend='rpc://', broker='pyamqp://guest@localhost//')
60
61  # ------------
62  # Constants
63  # ------------
64
65  # relevant column headers in peak data exported from TopSpin
66  f1 = "ν(F1) [ppm]"
67  f2 = "ν(F2) [ppm]"
68
69  # alternate column headers specified in FAQ
70  f1_alt = "15N"
71  f2_alt = "1H"
72
73  # ------------
74  # Empirically tuned parameters
75  # ------------
76
77  # RSS cut-off for linearity; for best results keep this
78  # relaxed (i.e. not too small).
79  rss_cutoff = 0.5
80
81  # std dev cut-offs for point spacing in each dimension
82  stdev_h_cutoff = 0.015
83  stdev_n_cutoff = 0.100
84
85  # largest spacing outlier allowed (in std devs; 5 is very generous)
86  outlier_h_cutoff = 5
87  outlier_n_cutoff = 5
88
89
90  # ------------
91  # The heavy lifting
92  # ------------
93
94  def increasing(L):
95      return all(x<=y for x, y in zip(L, L[1:]))
96
97  def decreasing(L):
98      return all(x>=y for x, y in zip(L, L[1:]))
99
100 def isFloat(string):
101     try:
102         float(string)
103         return True
104     except ValueError:
105         return False
```

```
106
107 # 'main_program' called from section below
108 def main_program():
109   sol_count = 0
110   ass_count = 0
111
112   # initialize empty dictionary (will contain temperature data)
113   temps = {}
114
115   # load variable temperature data into 'temps' dictionary
116   files = [f.strip() for f in peaklists.split(",")]
117   for i,fname in enumerate(files):
118     full_fname = directory + "/" + fname
119     if not(os.path.isfile(full_fname)):
120       log.write("File Not Found: "+full_fname+'\n')
121       return 1
122
123     df = pd.read_csv(full_fname)
124
125     if (f1 in df.axes[1]) and (f2 in df.axes[1]):
126       f1ppm = df[f1]
127       f2ppm = df[f2]
128     elif ('15N' in df.axes[1]) and ('1H' in df.axes[1]):
129       f1ppm = df['15N']
130       f2ppm = df['1H']
131     else:
132       log.write("Error: "+full_fname+" peak list format error\n")
133       return 1
134
135     temps[fname] = list(zip(f1ppm,f2ppm))
136
137   if not(os.path.isfile(assignments.strip())):
138     log.write("File Not Found: "+assignments+'\n')
139     return 1
140
141   peaks = []
142   infile = open(assignments.strip())
143   for line in infile:
144     splitline = line.split(",")
145     # import peak list; name of assignment stored in last position of tuple
146     peaks.append([splitline[1].strip(),splitline[2].strip(),splitline[0].strip()])
147
148   # make sure that number of reference peaks matches number of temperatures
149   if len(references.split(",")) != len(peaklists.split(",")):
150     log.write("Error: "+ \
151       "The number of reference peaks must match the number of peak lists\n")
152     return 1
153
154   # make sure that number of reference peaks matches number of temperatures
155   if len(nom_temperatures.split(",")) != len(peaklists.split(",")):
156     log.write("Error: "+ \
157       "The number of temperatures must match the number of peak lists\n")
158     return 1
159
160   # make sure that the reference peaks are all numbers
161   for ref in references.split(","):
162     ref = ref.strip()
163     if not(isFloat(ref)):
164       log.write("Error: "+"One of the reference peaks is not a number\n")
165       return 1
166
167   # make sure that the nominal temperatures are all numbers
168   # and calculate temperatures based upon DSS peaks
169   templist = []
170   refs = references.split(',')
```

```
171   for idx, temp in enumerate(nom_temperatures.split(",")):
172     temp = temp.strip()
173     if not(isFloat(temp)):
174       log.write("Error: "+"One of the temperatures is not a number\n")
175       return 1
176     else:
177       if len(templist) == 0 or calc_temp != 1:
178         last_temp = float(temp)
179         templist.append(format(last_temp, '.1f'))
180       else:
181         delta_ref = float(refs[idx])-float(refs[idx-1])
182         deltaT = delta_ref/0.0119 # based upon known 11.9 ppb/deg temp. dependence of water
183         last_temp = last_temp + deltaT
184         templist.append(format(last_temp, '.1f'))
185
186     temperatures = ', '.join(templist)
187
188   # make sure that starting temperature index is a valid number
189   if not(start_index.isnumeric()):
190     log.write("Error: "+"Starting temperature index must be an integer\n")
191     return 1
192   if (int(start_index) < 1 or int(start_index) > len(temperatures.split(","))):
193     log.write("Error: "+"Starting temperature index out of range\n")
194     return 1
195
196   # open output file handle
197   outhandle = open(csvfile, "w")
198
199   # write headers
200   outlist = []
201   outlist.append("Project: "+projname)
202   outhandle.write(",".join(outlist)+"\n")
203   outlist = []
204   outlist.append("Reference Peaks (ppm):")
205   outlist.append(references)
206   outhandle.write(",".join(outlist)+"\n")
207   outlist = []
208   outlist.append("Nominal Temperatures (K):")
209   outlist.append(nom_temperatures)
210   outhandle.write(",".join(outlist)+"\n")
211   outlist = []
212   outlist.append("Calculated Temperatures (K):")
213   outlist.append(temperatures)
214   outhandle.write(",".join(outlist)+"\n")
215   outhandle.write("\n")
216   outlist = []
217   outlist.append("Residue")
218   outlist.append("Ass. 1H")
219   outlist.append("Ass. 15N")
220   outlist.append("Notes")
221   outlist.append("1H Δδ/ΔT (ppb/K)")
222   outlist.append("1H RSS")
223   outlist.append("15N Δδ/ΔT (ppb/K)")
224   outlist.append("15N RSS")
225   outlist.append("ΔδN/ΔδH")
226   outlist.append("RSS")
227   outlist.append("") # spacer
228   for x,tempx in enumerate(temperatures.split(",")):
229     outlist.append(tempx.strip()+" K 1H")
230     outlist.append(tempx.strip()+" K 15N")
231   outlist.append("") # spacer
232   for x,tempx in enumerate(temperatures.split(",")):
233     outlist.append(tempx.strip()+" K 1H-RR")
234     outlist.append(tempx.strip()+" K 15N-RR")
235   outhandle.write(",".join(outlist)+"\n")
```

```
236
237     # get index of temperature/peak list to start with
238     start = int(start_index)-1
239
240
241     allres = [] # all residuals
242
243     # for each assigned peak
244     for peak in peaks:
245
246       log.write("Processing assignment "+peak[2]+"...")
247
248       if peak[0] == "" or peak[1] == "":
249         log.write("no assignment provided."+'\n')
250         outlist = []
251         outlist.append(peak[2]) # residue identifier
252         outlist.append(str(peak[1])) # 15N
253         outlist.append(str(peak[0])) # 1H
254         outlist.append("") # notes (blank)
255         outhandle.write(",".join(outlist)+"\n")
256
257       else:
258         log.write('\n')
259         ass_count = ass_count + 1
260
261         # find point in data from starting temperature that most closely matches
262         # the assigned peak
263         min_dist = 9999.99
264         min_point = 9999
265         for i, point in enumerate(temps[files[start]]):
266           # dist is actually distance squared; caculating the square root gains
267           # us nothing here
268           dist = (point[0]-float(peak[0]))**2 + (point[1]-float(peak[1]))**2
269           if dist < min_dist:
270             min_dist = dist
271             min_point = i
272
273         curr = temps[files[start]][min_point]
274
275         # starting point (not really a line yet)
276         lines = []
277         lines.append([curr])
278
279
280         shortlines = []
281
282         # process remaining temperatures:
283         #   Sarting point may be in the middle; go down from there first (prepending
284         #   points to candidate lines, then up afterwards (appending points).
285         for k in chain(reversed(range(0,start)), range(start+1,len(files))):
286           log.write("...working on "+files[k].split(".")[0]+'\n')
287
288           oldlines = lines
289           lines = []
290
291           # find candidate lines: previous lines extended by a point
292           # (within distance cut-off) from spectrum at the new temp
293           for j, line in enumerate(oldlines):
294             # start from either the first or last point in line
295             if k > start:
296               curr = line[len(line)-1]
297             else:
298               curr = line[0]
299
300             # calculate distances
```

```
            for i, point in enumerate(temps[files[k]]):
              # dist is actually distance squared; caculating the square root
              # gains us nothing here
              dist = (point[0]-curr[0])**2 + (point[1]-curr[1])**2
              # if within cut-off, add to list
              if dist < dist_cutoff:
                if k > start:
                  lines.append(line+[point])
                else:
                  lines.append([point]+line)

          oldlines = lines
          lines = []

          # assess linearity of candidate lines, discard obviously nonlinear
          #  (linear fitting after processing each new temp is somewhat
          #   inefficient, but probably preferable to allowing the number
          #   of lines under consideration to blow up)
          for line in oldlines:
            # unpack x and y coordinates
            n,h=zip(*line)

            # calculate coeff. of determination
            r2 = (np.corrcoef(h, n)[0,1])**2

            # linear regression
            p = np.polyfit(h,n,1)

            # calculate RSS
            rss = np.sum((np.polyval(p, h) - n) ** 2)

            # keep if linear approximation is good enough
            if rss < rss_cutoff and (increasing(h) or decreasing(h)):
              lines.append(line)

          # save shorter lines for possible consideration later on
          if k < len(files)-1:
            shortlines.append(lines)

        # check out results
        i_best = 999
        rss_min = 9999

        log.write("considering "+str(len(lines))+" full lines"+'\n')
        for i,line in enumerate(lines):
          # for std dev calculation
          ndiff = []
          hdiff = []
          for j in range(1, len(line)):
            ndiff.append(line[j][0]-line[j-1][0])
            hdiff.append(line[j][1]-line[j-1][1])

          # calculate deviations from mean spacings
          hstdev = np.std(hdiff)
          hmean = np.mean(hdiff)
          hdiff_dm = [] # list of absolute differences from the mean
          for x, diff in enumerate(hdiff):
            hdiff_dm.append(abs(diff-hmean))

          nstdev = np.std(ndiff)
          nmean = np.mean(ndiff)
          ndiff_dm = [] # list of absolute differences from the mean
          for x, diff in enumerate(ndiff):
            ndiff_dm.append(abs(diff-nmean))

```

```
366           # unpack x and y coordinates
367           n,h=zip(*line)
368           r2 = (np.corrcoef(h, n)[0,1])**2
369           p = np.polyfit(h,n,1)
370           rss = np.sum((np.polyval(p, h) - n) ** 2)
371
372           # if this set of points is more linear than the previous best,
373           # and isn't weeded out by standard deviation cut-offs or
374           # outlier checks, it becomes the leading candidate
375           if  rss < rss_min and \
376               hstdev < stdev_h_cutoff and \
377               nstdev < stdev_n_cutoff and \
378               max(hdiff_dm) <= outlier_h_cutoff * hstdev and \
379               max(ndiff_dm) <= outlier_n_cutoff * nstdev:
380             best_i = i
381             rss_min = rss
382             r2_best = r2
383             std_best_h = hstdev
384             std_best_n = nstdev
385             hmaxdm = max(hdiff_dm)
386             nmaxdm = max(ndiff_dm)
387
388         # if no full length lines (i.e. a point at each temperature)
389         # were found, consider shorter solutions
390         k = len(shortlines)-1
391         while rss_min == 9999 and k >= (short_line_limit-1):
392           lines = shortlines[k]
393
394           log.write("considering "+str(len(lines))+" of length "+str(k+2)+'\n')
395           for i,line in enumerate(lines):
396             # for std dev calculation
397             ndiff = []
398             hdiff = []
399             for j in range(1, len(line)):
400               ndiff.append(line[j][0]-line[j-1][0])
401               hdiff.append(line[j][1]-line[j-1][1])
402
403             # calculate deviations from mean spacings
404             hstdev = np.std(hdiff)
405             hmean = np.mean(hdiff)
406             hdiff_dm = [] # list of absolute differences from the mean
407             for x, diff in enumerate(hdiff):
408               hdiff_dm.append(abs(diff-hmean))
409
410             nstdev = np.std(ndiff)
411             nmean = np.mean(ndiff)
412             ndiff_dm = [] # list of absolute differences from the mean
413             for x, diff in enumerate(ndiff):
414               ndiff_dm.append(abs(diff-nmean))
415
416             # unpack x and y coordinates
417             n,h=zip(*line)
418             r2 = (np.corrcoef(h, n)[0,1])**2
419             p = np.polyfit(h,n,1)
420             rss = np.sum((np.polyval(p, h) - n) ** 2)
421
422             # if this set of points is more linear than the previous best,
423             # and isn't weeded out by standard deviation cut-offs or
424             # outlier checks, it becomes the leading candidate
425             if  rss < rss_min and \
426                 hstdev < stdev_h_cutoff and \
427                 nstdev < stdev_n_cutoff and \
428                 max(hdiff_dm) <= outlier_h_cutoff * hstdev and \
429                 max(ndiff_dm) <= outlier_n_cutoff * nstdev:
430               best_i = i
```

```
            rss_min = rss
            r2_best = r2
            std_best_h = hstdev
            std_best_n = nstdev
            hmaxdm = max(hdiff_dm)
            nmaxdm = max(ndiff_dm)

        # decrement k to look at even shorter lines
        # if the while loop doesn't terminate
        k = k - 1



    # if a good line was found, process and write output file
    if rss_min != 9999:
        log.write("done."+'\n')
        log.flush()
        # unpack 1H and 15N coordinates
        n,h=zip(*lines[best_i])

        # rereference
        refs = references.split(",")
        h_rr = []
        for l, h_raw in enumerate(h):
            h_rr.append(h_raw-float(refs[l]))

        n_rr = []
        for l, n_raw in enumerate(n):
            # calculate DSS frequency in Hz
            dss_freq = (float(refs[l])*(bf_h/1000000)) + bf_h
            # multiply by Σ ratio
            n_zero_freq = dss_freq * n_xi
            # find difference between calculated zero ppm and transmitter freq.
            n_ppm_adj = (n_zero_freq - bf_n)/bf_n*1000000
            # rereference
            n_rr.append(n_raw-n_ppm_adj)

        # calculate temperature coefficient using rereferenced shifts
        t = [float(ts) for ts in temperatures.split(",")]
        ph = np.polyfit(t[0:len(h_rr)],h_rr,1)
        pn = np.polyfit(t[0:len(n_rr)],n_rr,1)
        phn = np.polyfit(h_rr,n_rr,1)

        # calculate residuals for the purpose of estimating the standard deviation
        # (used later to help flag points for review)
        allres = np.concatenate((allres, (np.polyval(ph,t[0:len(h_rr)])-h_rr)), axis=0)

        # construct a line of output in list form
        outlist = []
        outlist.append(peak[2]) # residue identifier
        outlist.append(str(peak[1])) # 1H
        outlist.append(str(peak[0])) # 15N

        if len(h_rr) == len(t):
            outlist.append("") # notes (blank)
        else:
            outlist.append("Short line.") # notes

        outlist.append(str(ph[0]*1000)) # 1H temp coefficient in ppb/K
        outlist.append(str(np.sum((np.polyval(ph, t[0:len(h_rr)]) - h_rr) ** 2))) # RSS
        outlist.append(str(pn[0]*1000)) # 15N temp coefficient in ppb/K
        outlist.append(str(np.sum((np.polyval(pn, t[0:len(n_rr)]) - n_rr) ** 2))) # RSS
        outlist.append(str(phn[0])) # slope in the 1H - 15N plane
        outlist.append(str(np.sum((np.polyval(phn, h_rr) - n_rr) ** 2))) # RSS

```

```
496         outlist.append("") # column spacer
497
498         for l in range(len(h)):
499           outlist.append(str(h[l]))
500           outlist.append(str(n[l]))
501
502         # pad short lines so that columns in the csv output line up
503         for l in range(len(files)-len(h)):
504           outlist.append("") # column spacer
505           outlist.append("") # column spacer
506
507         outlist.append("") # column spacer
508
509         for l in range(len(h_rr)):
510           outlist.append(str(h_rr[l]))
511           outlist.append(str(n_rr[l]))
512
513         # pad short lines so that columns in the csv output line up
514         for l in range(len(files)-len(h_rr)):
515           outlist.append("") # column spacer
516           outlist.append("") # column spacer
517
518
519         # output to file
520         outhandle.write(",".join(outlist)+"\n")
521
522         sol_count = sol_count + 1
523
524       else:
525         log.write("no solution found."+'\n')
526         outlist = []
527         outlist.append(peak[2]) # residue identifier
528         outlist.append(str(peak[1])) # 15N
529         outlist.append(str(peak[0])) # 1H
530         outlist.append("No solution found.") # notes
531         outhandle.write(",".join(outlist)+"\n")
532
533       log.write('\n')
534
535
536   log.write('\n')
537   log.write("Residues: "+str(len(peaks))+'\n')
538   log.write("Assignments: "+str(ass_count)+'\n')
539   log.write("Solved: "+str(sol_count)+'\n')
540   outhandle.close()
541
542   mu, stddev = stats.norm.fit(allres)
543   if (plot_figs(stddev)==1):
544     return 1
545
546   zfile = zipfile.ZipFile('ShiftTrack.zip','w')
547   zfile.write('logfile.txt', 'logfile.txt', zipfile.ZIP_DEFLATED)
548
549   for fname in glob.glob("*.csv"):
550     zfile.write(fname, os.path.basename(fname), zipfile.ZIP_DEFLATED)
551
552   for fname in glob.glob("*.png"):
553     zfile.write(fname, os.path.basename(fname), zipfile.ZIP_DEFLATED)
554
555   zfile.close()
556
557   return 0
558
559   ## end of main_program() function definition
560
```

```
561 ######################
562
563 def email_results():
564   # email the results
565   emailText = "***DO NOT REPLY TO THIS EMAIL***\n"
566   emailText += "The 'meieringlab@gmail.com' email address is not monitored."
567   emailText += " Send questions/comments/bug reports to 'kjtrainor@uwaterloo.ca'.\n\n"
568   emailText += "Results for your job '"+projname+ \
569                "' can be found in the attached 'ShiftTrack.zip' archive."
570   emailText += " The main results file, named 'ShiftTrack.csv',"
571   emailText += " can be opened in a spreadsheet program such"
572   emailText += " as Microsoft Excel. Also included are PNG figures generated for each"
573   emailText += " peak that was successfully"
574   emailText += " followed over temperature.\n\n"
575   if jobId != 'example':
576     emailText += "Your unique job ID was '"+jobId+"'."
577     emailText += " Please quote this string if you email"
578     emailText += " 'kjtrainor@uwaterloo.ca' with a question or bug report.\n"
579
580   message = MIMEMultipart()
581   message["Subject"] = "Shift-T Server: ShiftTrack results for "+projname
582   message["From"] = "meieringlab@gmail.com"
583   message["To"] = email
584   message.attach(MIMEText(emailText, "plain"))
585
586   part = MIMEBase('application', 'octet-stream')
587   part.set_payload(open('ShiftTrack.zip', 'rb').read())
588   encoders.encode_base64(part)
589   part.add_header('Content-Disposition', 'attachment; filename="ShiftTrack.zip"')
590   message.attach(part)
591
592   server = smtplib.SMTP_SSL(host="smtp.gmail.com", port=465)
593   server.ehlo()
594   server.login("meieringlab", "password")
595   server.sendmail("meieringlab@gmail.com", email, message.as_string())
596   server.close()
597
598   ## end of email_results() function definition
599
600 ######################
601
602 def email_error():
603   # email the results
604   emailText = "***DO NOT REPLY TO THIS EMAIL***\n"
605   emailText += "The 'meieringlab@gmail.com' email address is not monitored."
606   emailText += " Send questions/comments/bug reports to 'kjtrainor@uwaterloo.ca'.\n\n"
607   emailText += "An error was encountered while processing your job '"+projname+ \
608                "'. The log file has been sent as an attachment to this email.\n\n"
609   if jobId != 'example':
610     emailText += "Your unique job ID was '"+jobId+"'."
611     emailText += " Please quote this string if you email"
612     emailText += " 'kjtrainor@uwaterloo.ca' with a question or bug report.\n"
613
614   message = MIMEMultipart()
615   message["Subject"] = "Shift-T Server: ShiftTrack results for "+projname
616   message["From"] = "meieringlab@gmail.com"
617   message["To"] = email
618   message.attach(MIMEText(emailText, "plain"))
619
620   part = MIMEBase('application', 'octet-stream')
621   part.set_payload(open('logfile.txt', 'rb').read())
622   encoders.encode_base64(part)
623   part.add_header('Content-Disposition', 'attachment; filename="logfile.txt"')
624   message.attach(part)
625
```

```
626     server = smtplib.SMTP_SSL(host="smtp.gmail.com", port=465)
627     server.ehlo()
628     server.login("meieringlab", "password")
629     server.sendmail("meieringlab@gmail.com", email, message.as_string())
630     server.close()
631 
632     ## end of email_error() function definition
633 
634   ######################
635 
636   def plot_figs(stddev):
637     # load temperatures from header
638     full_fname = directory + "/" + csvfile
639     if not(os.path.isfile(full_fname)):
640       log.write("File Not Found: "+full_fname+"\nCannot plot figures.\n")
641       return 1
642     else:
643       infile = open(full_fname)
644       for idx, line in enumerate(infile):
645         if idx == 3:
646           templist = line.strip().split(',')[1:]
647           temperatures = ','.join(templist)
648 
649     files = [f.strip() for f in peaklists.split(',')]
650 
651     refs = [ref.strip() for ref in references.split(',')]
652     ts = [temp.strip() for temp in temperatures.split(',')]
653 
654     h_cols = [t.strip()+" K 1H-RR" for t in temperatures.split(',')]
655     n_cols = [t.strip()+" K 15N-RR" for t in temperatures.split(',')]
656 
657     # initialize empty dictionaries (fill from results file)
658     n_columns = {}
659     h_columns = {}
660 
661     full_fname = directory + "/" + csvfile
662     if not(os.path.isfile(full_fname)):
663       log.write("File Not Found: "+full_fname+"\nCannot plot figures.\n")
664       return 1
665     else:
666       df = pd.read_csv(full_fname, skiprows=5)
667 
668       residues = df["Residue"]
669       if residues.dtype == np.int64:
670         numeric_ids = True
671         largest_id = residues.max()
672       else:
673         numeric_ids = False
674 
675       for i in range(len(files)):
676         n_columns[h_cols[i].split(" K")[0]] = df[n_cols[i]]
677         h_columns[n_cols[i].split(" K")[0]] = df[h_cols[i]]
678 
679       ts_int = [float(j) for j in ts]
680       for i in range(len(residues)):
681         ts_work = []
682         n = []
683         h = []
684         if not(np.isnan(n_columns[n_cols[0].split(" K")[0]][i])) \
685            and not(np.isnan(h_columns[h_cols[0].split(" K")[0]][i])):
686           for j in range(len(files)):
687             if not(np.isnan(n_columns[n_cols[j].split(" K")[0]][i])):
688               n.append(n_columns[n_cols[j].split(" K")[0]][i])
689             if not(np.isnan(h_columns[h_cols[j].split(" K")[0]][i])):
690               h.append(h_columns[h_cols[j].split(" K")[0]][i])
```

```
691             ts_work.append(ts_int[j])
692
693         # find best linear fit
694         ph = np.polyfit(ts_work,h,1)
695
696         # calculate residuals (hres)
697         [f,p, hres] = f_calc(ts_work,h)
698
699         # flag outliers
700         flag = 0
701         t_flag = []
702         h_flag = []
703         for idx, res in enumerate(hres):
704           if abs(res) > (2*stddev): # stddev calculated in main_program()
705             t_flag.append(ts_work[idx])
706             h_flag.append(h[idx])
707
708         f, (p1,p2) = plt.subplots(2, sharex=True, figsize=(10,10))
709         plt.suptitle("Residue "+str(residues[i]))
710
711         # plot 1H chemical shifts vs temperature
712         p1.scatter(ts_work,h, s=160, facecolors='none', edgecolors='goldenrod')
713
714         # 1H chemical shifts for review
715         if flag_review == 1:
716           p1.scatter(t_flag,h_flag, s=320, facecolors='none', edgecolors='red')
717
718         # plot 1H chemical shift vs. temperature line of best fit
719         p1.plot(ts_int,np.polyval(ph, ts_int), color="black", linewidth=2)
720
721         p1.set_ylabel("$\delta$ (ppm)")
722
723         # plot residuals vs temperature
724         p2.scatter(ts_work,hres, s=160, facecolors='red', edgecolors='goldenrod')
725         p2.set_ylabel("Residual (ppm)")
726         p2.set_xlabel("Temperature (K)")
727
728         if numeric_ids:
729           chars = len(str(largest_id))
730           plt.savefig("residue"+str(residues[i]).zfill(chars)+".png")
731         else:
732           plt.savefig(str(residues[i]).strip()+".png")
733
734         plt.close()
735
736   return 0
737
738   ## end of plot_figs() function definition
739
740
741 ######################
742
743 # f-test to help solve the model selection problem
744 def f_calc(t,cs):
745   lin_fit = np.polyfit(t,cs,1)
746   hres = np.polyval(lin_fit, t)-cs
747   rss_lin = np.sum((hres)**2)
748
749   quad_fit = np.polyfit(t,cs,2)
750   rss_quad = np.sum((np.polyval(quad_fit, t)-cs)**2)
751
752   f = (rss_lin-rss_quad)/(rss_quad/(len(cs)-3))
753   p = 1.0 - scipy.stats.f.cdf(f,1,9-2-1)
754   return [f,p, hres]
755
```

```
756     ## end of f_calc() function definition
757
758
759   ######################################
760   # start of function executed by Celery #
761   ######################################
762
763   @app.task
764   def pyShiftTrack(**kwargs):
765     global email
766     global jobId
767     email = kwargs['email']
768     jobId = kwargs['jobId']
769
770
771     # load config file and set up run
772
773     # these definitions are used elsewhere, but don't really
774     # need to vary; we hard code them here
775     global directory
776     global csvfile
777     directory= '.'
778     csvfile = 'ShiftTrack.csv'
779
780     # store current working directory; change to job dir
781     start_wd = os.getcwd()
782     os.chdir("/home/nmr/shiftt/ShiftTrack/jobs/"+str(jobId))
783
784     # open log
785     global log
786     log = open("logfile.txt", 'w')
787
788     # load configuration file created during job submission
789     global projname
790     global assignments
791     global peaklists
792     global references
793     global nom_temperatures
794     global start_index
795     global n_xi
796     global bf_h
797     global bf_n
798     global dist_cutoff
799     global short_line_limit
800     global flag_review
801     global calc_temp
802
803     config_file = open("STconfig.txt", "rU")
804     config = []
805     for line in config_file:
806       config.append(line.strip())
807
808     if len(config) == 13:
809       projname = config[0]
810       assignments = config[1]
811       peaklists = config[2]
812       references = config[3]
813       nom_temperatures = config[4]
814       start_index = config[5]
815
816       # 15N/1H Σ (reference compound: liq. NH3)
817       n_xi = float(config[6]) # default: 0.10132912
818
819       # base transmitter frequencies (Hz)
820       bf_h = float(config[7]) # default: 600130000
```

```
    bf_n = float(config[8]) # default: 60810645.0

    # distance cut-off, units of ppm squared (see usage below)
    #  - large values will increase runtime and possibly find unlikely
    #    "solutions" instead of reporting none found
    #  - small values will prevent finding valid solutions with larger
    #    point spacing in the 1H-15N plane
    #  - the sweet spot may depend on both the protein and the ΔT (0.25 has
    #    been found to work well for Adnectins with ΔT=5 K; 0.125 has been
    #    found to work well for hisactophilin with ΔT=2.5 K)
    #  - the value supplied via web server submissions is in units of ppm
    #    therefore we square it
    dist_cutoff = float(config[9])**2

    # optionally find lines shorter than the number of
    # temperatures at which we have data...
    # how short is too short?
    short_line_limit = int(config[10])

    flag_review = int(config[11])
    calc_temp = int(config[12])
  else:
    log.write("Aborted due to configuration file error!"+'\n')
    log.close()
    email_error()
    # restore working directory; probably unnecessary
    os.chdir(start_wd)
    return

  # start
  if jobId != "example":
    main_ret = main_program()
  else:
    main_ret = 0

  # close log
  log.close()

  # send results
  if main_ret == 0:
    email_results()
  else:
    email_error()

  # restore working directory; probably unnecessary
  os.chdir(start_wd)

  return
```

F.2 Curvalyzer

```
#!/usr/bin/python3

# Curvalyzer.py
#
# This program detects curvature in the temperature dependence of amide proton
# chemical shifts. Curvature detection is treated as a model selection problem.
# Nested linear and quadratic models are considered; the statistical
# significance of the improvement of the quadratic model (over the linear) is
# assessed using an extra sum of squares F-test. If curvature is detected, the
# likelihood that it is due to random errors is calculated via numerical
# simulation based on the distribution of experimentally observed residuals.
#
# Kyle Trainor, March 2018

# ------------
# Imports
# ------------

# CSV
import pandas as pd

# numerical python
import numpy as np

# scipy
import scipy, scipy.stats as stats

# to get command-line arguments
import sys

# to work with files and directories
import os

import datetime

from itertools import islice

# manipulate zip files
import zipfile

# get pathnames matching pattern
import glob


# for emailing results
import smtplib
from email import encoders
from email.mime.base import MIMEBase
from email.mime.text import MIMEText
from email.mime.multipart import MIMEMultipart
from email.mime.application import MIMEApplication

# plotting
import matplotlib
matplotlib.use('AGG')
import matplotlib.pyplot as plt
import pylab

from celery import Celery
import time

app = Celery('curvalyzer', backend='rpc://', broker='pyamqp://guest@localhost//')
```

```
63
64  # useful constants
65
66  # 15N/1H Σ (reference compound: liq. NH3)
67  n_xi = 0.10132912
68
69  # base transmitter frequencies (Hz)
70  bf_h = 600130000
71  bf_n = 60810645.0
72
73  ######################
74
75  def plot_figs(jobDir,curves,sim_curves):
76    # load temperatures from header
77    full_fname = os.path.join(jobDir,"Curvalyzer.csv")
78    if not(os.path.isfile(full_fname)):
79      log.write("File Not Found: "+full_fname+"\nCannot plot figures.\n")
80      return 1
81    else:
82      infile = open(full_fname)
83      for idx, line in enumerate(infile):
84        if idx == 3:
85          templist = line.strip().split(',')[1:]
86          temperatures = ','.join(templist)
87
88    ts = [temp.strip() for temp in temperatures.split(',')]
89
90    h_cols = [t.strip()+" K 1H-RR" for t in temperatures.split(',')]
91    n_cols = [t.strip()+" K 15N-RR" for t in temperatures.split(',')]
92
93    # initialize empty dictionaries (fill from results file)
94    n_columns = {}
95    h_columns = {}
96
97    if not(os.path.isfile(full_fname)):
98      log.write("File Not Found: "+full_fname+"\nCannot plot figures.\n")
99      return 1
100   else:
101     df = pd.read_csv(full_fname, skiprows=5)
102
103     residues = df["Residue"]
104     if residues.dtype == np.int64:
105       numeric_ids = True
106       largest_id = residues.max()
107     else:
108       numeric_ids = False
109
110     for i in range(len(ts)):
111       n_columns[h_cols[i].split(" K")[0]] = df[n_cols[i]]
112       h_columns[n_cols[i].split(" K")[0]] = df[h_cols[i]]
113
114
115     ts_int = [float(j) for j in ts]
116     for i in range(len(residues)):
117       print(i)
118       ts_work = []
119       n = []
120       h = []
121       if not(np.isnan(n_columns[n_cols[0].split(" K")[0]][i])) \
122          and not(np.isnan(h_columns[h_cols[0].split(" K")[0]][i])):
123         for j in range(len(ts)):
124           if not(np.isnan(n_columns[n_cols[j].split(" K")[0]][i])):
125             n.append(n_columns[n_cols[j].split(" K")[0]][i])
126           if not(np.isnan(h_columns[h_cols[j].split(" K")[0]][i])):
127             h.append(h_columns[h_cols[j].split(" K")[0]][i])
```

```
128             ts_work.append(ts_int[j])
129
130         # find best linear fit
131         ph = np.polyfit(ts_work,h,1)
132
133
134         # lazy way to calculate residuals (hres)
135         [f,p, hres, gq] = f_calc(ts_work,h)
136
137         f, (p1,p2) = plt.subplots(2, sharex=True, figsize=(10,10))
138         plt.suptitle("Residue "+str(residues[i]))
139
140         # plot 1H chemical shifts vs temperature
141         p1.scatter(ts_work,h, s=160, facecolors='none', edgecolors='goldenrod')
142
143         # plot 1H chemical shift vs. temperature line of best fit
144         p1.plot(ts_int,np.polyval(ph, ts_int), color="black", linewidth=2)
145
146         # plot quadratic if curvature is detected
147         if i in curves:
148
149           if p < 0.01:
150             # calculate p-value 2
151             # (likelihood that the curvature is due to random errors)
152             curve_count = 0
153             for sim_curve in sim_curves:
154               if abs(sim_curve) >= abs(curves[i]):
155                 curve_count = curve_count + 1
156             pv2 = curve_count/100000
157
158             if pv2 < 0.01:
159               # plot quadratic
160               pqh = np.polyfit(ts_work,h,2)
161               p1.plot(ts_int,np.polyval(pqh, ts_int), color="red", linewidth=2)
162               if pv2 < 0.0001:
163                 p1.set_title("Curvature detected: p-value 1 = "+ \
164                              "{0:.2g}".format(p)+"; p-value 2 < 0.0001")
165               else:
166                 p1.set_title("Curvature detected: p-value 1 = "+ \
167                              "{0:.2g}".format(p)+"; p-value 2 = "+"{0:.2g}".format(pv2))
168
169
170         p1.set_ylabel("$\delta$ (ppm)")
171
172         # plot residuals vs temperature
173         p2.scatter(ts_work,hres, s=160, facecolors='blue', edgecolors='blue')
174         p2.set_ylabel("Residual (ppm)")
175         p2.set_xlabel("Temperature (K)")
176
177         if numeric_ids:
178           chars = len(str(largest_id))
179           plt.savefig(jobDir+"/residue"+str(residues[i]).zfill(chars)+".png")
180         else:
181           plt.savefig(jobDir+'/'+str(residues[i]).strip()+".png")
182
183         plt.close()
184   return 0
185
186   ## end of plot_figs() function definition
187
188 ######################
189
190 def email_results(jobId,projname,email):
191   jobDir = 'Curvalyzer/jobs/'+jobId
192   # email the results
```

```
193     emailText = "***DO NOT REPLY TO THIS EMAIL***\n"
194     emailText += "The 'meieringlab@gmail.com' email address is not monitored."
195     emailText += " Send questions/comments/bug reports to 'kjtrainor@uwaterloo.ca'.\n\n"
196     emailText += "Results for your job '"+projname+ \
197                  "' can be found in the attached 'Curvalyzer.zip' archive."
198     emailText += " The main results file, named 'Curvalyzer.csv',"
199     emailText += " can be opened in a spreadsheet program such as Microsoft"
200     emailText += " Excel. Also included are PNG figures generated for each peak.\n\n"
201     emailText += "Your unique job ID was '"+jobId+"'."
202     emailText += " Please quote this string if you email"
203     emailText += " 'kjtrainor@uwaterloo.ca' with a question or bug report.\n"
204
205     message = MIMEMultipart()
206     message["Subject"] = "Shift-T Server: Curvalyzer results for "+projname
207     message["From"] = "meieringlab@gmail.com"
208     message["To"] = email
209     message.attach(MIMEText(emailText, "plain"))
210
211     part = MIMEBase('application', 'octet-stream')
212     part.set_payload(open(os.path.join(jobDir,'Curvalyzer.zip'), 'rb').read())
213     encoders.encode_base64(part)
214     part.add_header('Content-Disposition', 'attachment; filename="Curvalyzer.zip"')
215     message.attach(part)
216
217     server = smtplib.SMTP_SSL(host="smtp.gmail.com", port=465)
218     server.ehlo()
219     server.login("meieringlab", "password")
220     server.sendmail("meieringlab@gmail.com", email, message.as_string())
221     server.close()
222
223     ## end of email_results() function definition
224
225 ######################
226
227 def email_error(jobId,projname,email):
228     jobDir = 'Curvalyzer/jobs/'+jobId
229
230     # email the results
231     emailText = "***DO NOT REPLY TO THIS EMAIL***\n"
232     emailText += "The 'meieringlab@gmail.com' email address is not monitored."
233     emailText += " Send questions/comments/bug reports to 'kjtrainor@uwaterloo.ca'.\n\n"
234     emailText += "An error was encountered while processing your job '"+projname+ \
235                  "'. The log file has been sent as an attachment to this email.\n\n"
236     if jobId != 'example':
237       emailText += "Your unique job ID was '"+jobId+"'."
238       emailText += " Please quote this string if you email"
239       emailText += " 'kjtrainor@uwaterloo.ca' with a question or bug report.\n"
240
241     message = MIMEMultipart()
242     message["Subject"] = "Shift-T Server: Curvalyzer results for "+projname
243     message["From"] = "meieringlab@gmail.com"
244     message["To"] = email
245     message.attach(MIMEText(emailText, "plain"))
246
247     part = MIMEBase('application', 'octet-stream')
248     part.set_payload(open(os.path.join(jobDir,'logfile.txt'), 'rb').read())
249     encoders.encode_base64(part)
250     part.add_header('Content-Disposition', 'attachment; filename="logfile.txt"')
251     message.attach(part)
252
253     server = smtplib.SMTP_SSL(host="smtp.gmail.com", port=465)
254     server.ehlo()
255     server.login("meieringlab", "password")
256     server.sendmail("meieringlab@gmail.com", email, message.as_string())
257     server.close()
```

```
258
259   ## end of email_error() function definition
260
261 ######################
262
263 def f_calc(t,cs):
264   lin_fit = np.polyfit(t,cs,1)
265   hres = np.polyval(lin_fit, t)-cs
266   rss_lin = np.sum((hres)**2)
267
268   quad_fit = np.polyfit(t,cs,2)
269   rss_quad = np.sum((np.polyval(quad_fit, t)-cs)**2)
270
271   f = (rss_lin-rss_quad)/(rss_quad/(len(cs)-3))
272   p = 1.0 - scipy.stats.f.cdf(f,1,9-2-1)
273   return [f,p, hres, quad_fit[0]]
274
275   ## end of f_calc() function definition
276
277 ######################
278
279 def isFloat(string):
280     try:
281         float(string)
282         return True
283     except ValueError:
284         return False
285
286   ## end of isFloat() function definition
287
288 ####################################
289 # start of function executed by Celery #
290 ####################################
291
292 @app.task
293 def pyCurvalyzer(**kwargs):
294   email = kwargs['email']
295   jobName = kwargs['job']
296   jobId = kwargs['jobId']
297
298   # initialize empty dictionaries (fill from results file)
299   n_columns = {}
300   h_columns = {}
301
302
303   jobDir = 'Curvalyzer/jobs/'+jobId
304   jobfile = open(os.path.join(jobDir,'job.txt'), 'w')
305   jobfile.write(email+'\n')
306   jobfile.write(jobName+'\n')
307   jobfile.write(jobId+'\n')
308   jobfile.write('{:%Y-%m-%d %H:%M:%S}'.format(datetime.datetime.now())+'\n')
309   jobfile.close()
310
311   # open log
312   global log
313   log = open(os.path.join(jobDir,'logfile.txt'), 'w')
314
315
316   header=open(os.path.join(jobDir,"curvalyzer_input.csv"), 'rU')
317   header_lines = list(islice(header, 4))
318   projname = jobName
319   references = \
320     ','.join(list(filter(None, header_lines[1].strip().split(',')[1:])))
321   nom_temperatures = \
322     ','.join(list(filter(None, header_lines[2].strip().split(',')[1:])))
```

```
323   temps = \
324     ','.join(list(filter(None, header_lines[3].strip().split(',')[1:])))
325
326   # make sure that number of reference peaks matches number of temperatures
327   if len(references.split(",")) != len(temps.split(",")):
328     log.write("Error: "+ \
329               "The number of reference peaks must match the number of peak lists\n")
330     log.close()
331     email_error(jobId, jobName, email)
332     return
333
334   # make sure that the reference peaks are all numbers
335   for ref in references.split(","):
336     ref = ref.strip()
337     if not(isFloat(ref)):
338       log.write("Error: "+"One of the reference peaks is not a number\n")
339       log.close()
340       email_error(jobId, jobName, email)
341       return
342   refs = references.split(',')
343
344   # make sure that the temperatures are all numbers
345   for idx, temp in enumerate(nom_temperatures.split(",")):
346     temp = temp.strip()
347     if not(isFloat(temp)):
348       log.write("Error: "+"One of the temperatures is not a number\n")
349       log.close()
350       email_error(jobId, jobName, email)
351       return
352
353   templist = list()
354   for idx, temp in enumerate(temps.split(",")):
355     temp = temp.strip()
356     if not(isFloat(temp)):
357       log.write("Error: "+"One of the temperatures is not a number\n")
358       log.close()
359       email_error(jobId, jobName, email)
360       return
361     else:
362       templist.append(format(float(temp), '.1f'))
363
364   temperatures = ', '.join(templist)
365
366   h_cols = [t.strip()+" K 1H" for t in temperatures.split(',')]
367   n_cols = [t.strip()+" K 15N" for t in temperatures.split(',')]
368
369
370   # read chemical shift data
371   df = pd.read_csv(os.path.join(jobDir,"curvalyzer_input.csv"), skiprows=5)
372
373   # determine residue naming convention
374   residues = df["Residue"]
375   if residues.dtype == np.int64:
376     numeric_ids = True
377     largest_id = residues.max()
378   else:
379     numeric_ids = False
380
381   # load chemical shifts
382   for i in range(len(temperatures.split(','))):
383     n_columns[h_cols[i].split(" K")[0]] = df[n_cols[i]]
384     h_columns[n_cols[i].split(" K")[0]] = df[h_cols[i]]
385
386   peaks=list(zip(df["Ass. 15N"],df["Ass. 1H"],df["Residue"]))
387
```

```
388     # open output file handle
389     csvfile = os.path.join(jobDir, 'Curvalyzer.csv')
390     outhandle = open(csvfile, "w")
391
392     # write headers
393     outlist = []
394     outlist.append("Project: "+projname)
395     outhandle.write(",".join(outlist)+"\n")
396     outlist = []
397     outlist.append("Reference Peaks (ppm):")
398     outlist.append(references.strip())
399     outhandle.write(",".join(outlist)+"\n")
400     outlist = []
401     outlist.append("Nominal Temperatures (K):")
402     outlist.append(nom_temperatures.strip())
403     outhandle.write(",".join(outlist)+"\n")
404     outlist = []
405     outlist.append("Calculated Temperatures (K):")
406     outlist.append(temperatures.strip())
407     outhandle.write(",".join(outlist)+"\n")
408     outhandle.write("\n")
409     outlist = []
410     outlist.append("Residue")
411     outlist.append("Ass. 1H")
412     outlist.append("Ass. 15N")
413     outlist.append("Notes")
414     outlist.append("1H Δδ/ΔT (ppb/K)")
415     outlist.append("1H RSS")
416     outlist.append("15N Δδ/ΔT (ppb/K)")
417     outlist.append("15N RSS")
418     outlist.append("") # spacer
419     for x,tempx in enumerate(temperatures.split(",")):
420       outlist.append(tempx.strip()+" K 1H")
421       outlist.append(tempx.strip()+" K 15N")
422     outlist.append("") # spacer
423     for x,tempx in enumerate(temperatures.split(",")):
424       outlist.append(tempx.strip()+" K 1H-RR")
425       outlist.append(tempx.strip()+" K 15N-RR")
426     outhandle.write(",".join(outlist)+"\n")
427
428
429     ccount = 0
430     tcount = 0
431     scount = 0
432     cres = []
433     ncres = []
434     curves = {}
435     n_dict = {}
436     h_dict = {}
437     n_rr_dict = {}
438     h_rr_dict = {}
439     pn_dict = {}
440     ph_dict = {}
441     ts_work_dict = {}
442
443
444     t = [float(ts) for ts in temperatures.split(",")]
445     for i in range(len(residues)):
446       n = []
447       h = []
448       ts_work = []
449       if not(np.isnan(n_columns[n_cols[0].split(" K")[0]][i])) \
450           and not(np.isnan(h_columns[h_cols[0].split(" K")[0]][i])):
451         for j in range(len(temperatures.split(','))):
452           if not(np.isnan(n_columns[n_cols[j].split(" K")[0]][i])):
```

```
453             n.append(n_columns[h_cols[j].split(" K")[0]][i])
454           if not(np.isnan(h_columns[h_cols[j].split(" K")[0]][i])):
455             h.append(h_columns[n_cols[j].split(" K")[0]][i])
456             ts_work.append(t[j])
457 
458       if len(h) > 0:
459         # recalc temperature coefficients etc.
460 
461         # rereference
462         refs = references.split(",")
463         h_rr = []
464         for l, h_raw in enumerate(h):
465           h_rr.append(h_raw-float(refs[l]))
466 
467         n_rr = []
468         for l, n_raw in enumerate(n):
469           # calculate DSS frequency in Hz
470           dss_freq = (float(refs[l])*(bf_h/1000000)) + bf_h
471           # multiply by Σ ratio
472           n_zero_freq = dss_freq * n_xi
473           # find difference between calculated zero ppm and transmitter freq.
474           n_ppm_adj = (n_zero_freq - bf_n)/bf_n*1000000
475           # rereference
476           n_rr.append(n_raw-n_ppm_adj)
477 
478         # calculate temperature coefficient using rereferenced shifts
479         ph = np.polyfit(t[0:len(h_rr)],h_rr,1)
480         pn = np.polyfit(t[0:len(n_rr)],n_rr,1)
481 
482         # debug output
483         print("RES:",peaks[i][2])
484 
485 
486         # ignore short lines; otherwise, calculate residuals
487         # and sort into curved ('cres') and non ('ncres')
488         if len(t)>len(h_rr):
489           print("\tshort.")
490           scount = scount + 1
491         else:
492           [f,p, hres, quad] = f_calc(t,h_rr)
493           flag = 0
494           for k in range(len(t)):
495             [fs,ps, hres_s, quad_s] = f_calc(t[0:k]+t[k+1:],h_rr[0:k]+h_rr[k+1:])
496             if ps >= 0.01:
497               flag=1
498 
499           if p<0.01 and not(flag):
500             print("\tcurved.")
501             ccount=ccount+1
502             curves[i] = quad
503             cres = np.concatenate((cres,hres),axis=0)
504           else:
505             ncres = np.concatenate((ncres,hres),axis=0)
506 
507           tcount=tcount+1
508 
509           print()
510       n_dict[i] = n
511       h_dict[i] = h
512       n_rr_dict[i] = n_rr
513       h_rr_dict[i] = h_rr
514       pn_dict[i] = pn
515       ph_dict[i] = ph
516       ts_work_dict[i] = ts_work
517 
```

```
518   # fit residuals (non-curved only!) to t-distribution
519   param = stats.t.fit(ncres)
520   # debug output
521   print("non-curved residuals: t-distribution fit parameters:")
522   print("\tdf = "+str(list(param)[0]))
523   print("\tloc = "+str(param[1]))
524   print("\tscale = "+str(param[2]))
525   x = np.linspace(ncres.min(), ncres.max(), 100)
526
527   # plot fitted t-distribution overlaid on histogram of non-curved residuals
528   p = stats.t.pdf(x, loc=param[1], scale=param[2], df=param[0])
529   histo, bin_edges = np.histogram(ncres, bins='auto', normed=False)
530   number_of_bins = len(bin_edges)-1
531   scaling_factor = len(ncres)*(ncres.max()-ncres.min())/number_of_bins
532   plt.close()
533   plt.plot(x, scaling_factor*p, 'k', linewidth=2)
534   plt.hist(ncres, bins='auto', normed=False)
535   title = "Non-curved Residuals: Histogram and t-Distribution Fit"
536   plt.title(title)
537   plt.savefig(jobDir+"/residuals_t-distribution.png")
538
539   # perform numerical simulation
540   #  - calculate 100000 sets of 'fake' residuals
541   #    (i.e., drawn from above t-distribution)
542   #  - for any sets of 'fake' residuals that test
543   #    positive for curvature, record the quadratic
544   #    coefficient (for comparison with those from
545   #    'real' sets of residuals in the plot_figs
546   #    function)
547   sim_curves = []
548   for i in range(100000):
549
550     # monitor progress (debug purposes only)
551     if i%1000 == 0:
552       print(i)
553
554     # simulate residuals
555     r = list(stats.t.rvs(param[0],param[1],param[2], 9))
556
557     # check simulated residuals for curvature
558     [f, p, hres, quad] = f_calc(t,r)
559
560     flag = 0
561     for k in range(len(t)):
562       [fs,ps, hres_s,q] = f_calc(t[0:k]+t[k+1:],r[0:k]+r[k+1:])
563       if ps >= 0.01:
564         flag=1
565
566     if flag and p<0.01:
567       sim_curves.append(quad)
568
569   # debug output
570   print(sim_curves)
571   print(len(sim_curves), len(sim_curves)/100000)
572
573
574   # generate per-residue CSV output and PNG figures
575   for i in range(len(residues)):
576     n = n_dict[i]
577     h = h_dict[i]
578     n_rr = n_rr_dict[i]
579     h_rr = h_rr_dict[i]
580     pn = pn_dict[i]
581     ph = ph_dict[i]
582     ts_work = ts_work_dict[i]
```

```
583       if len(h) > 0:
584         # construct a line of output in list form
585         outlist = []
586         outlist.append(str(peaks[i][2])) # residue identifier
587         outlist.append(str(peaks[i][1])) # 1H
588         outlist.append(str(peaks[i][0])) # 15N
589 
590         if len(h_rr) == len(t):
591           if i in curves:
592             outlist.append("Curvature detected.") # notes
593           else:
594             outlist.append("") # notes (blank)
595         else:
596           outlist.append("Short line.") # notes
597 
598         # 1H temp coefficient in ppb/K
599         outlist.append(str(ph[0]*1000))
600         # RSS
601         outlist.append(str(np.sum((np.polyval(ph, t[0:len(h_rr)]) - h_rr) ** 2)))
602         # 15N temp coefficient in ppb/K
603         outlist.append(str(pn[0]*1000))
604         # RSS
605         outlist.append(str(np.sum((np.polyval(pn, t[0:len(n_rr)]) - n_rr) ** 2)))
606 
607         outlist.append("") # column spacer
608 
609         for l in range(len(h)):
610           outlist.append(str(h[l]))
611           outlist.append(str(n[l]))
612 
613         # pad short lines so that columns in the csv output line up
614         for l in range(len(temperatures.split(','))-len(h)):
615           outlist.append("") # column spacer
616           outlist.append("") # column spacer
617 
618         outlist.append("") # column spacer
619 
620         for l in range(len(h_rr)):
621           outlist.append(str(h_rr[l]))
622           outlist.append(str(n_rr[l]))
623 
624         # pad short lines so that columns in the csv output line up
625         for l in range(len(temperatures.split(','))-len(h_rr)):
626           outlist.append("") # column spacer
627           outlist.append("") # column spacer
628 
629 
630       elif np.isnan(peaks[i][1]) or np.isnan(peaks[i][0]):
631         outlist = []
632         outlist.append(str(peaks[i][2])) # residue identifier
633 
634       else:
635         outlist = []
636         outlist.append(str(peaks[i][2])) # residue identifier
637         outlist.append(str(peaks[i][1])) # 15N
638         outlist.append(str(peaks[i][0])) # 1H
639         outlist.append("No solution found.") # notes
640 
641       outhandle.write(",".join(outlist)+"\n")
642 
643 
644     outhandle.close()
645 
646     # debug output
647     print(ccount,tcount, scount)
```

```
648
649   if (plot_figs(jobDir,curves,sim_curves)!=0):
650     log.close()
651     email_error(jobId, jobName, email)
652     return
653
654   zfname = os.path.join(jobDir,'Curvalyzer.zip')
655   zfile = zipfile.ZipFile(zfname,'w')
656   zfile.write(csvfile, 'Curvalyzer.csv', zipfile.ZIP_DEFLATED)
657
658   for fname in glob.glob(jobDir+"/*.png"):
659     zfile.write(fname, os.path.basename(fname), zipfile.ZIP_DEFLATED)
660
661   zfile.close()
662   email_results(jobId, jobName, email)
```

Appendix G

Hydrogen Bond Dynamics: Supplemental Information

Table G.1: Linear (Pearson) correlations between simulated GB1 $r_{HO'}$ distances

Donor-Acceptor	**8 - 54**	**9 - 12**	**14 - 7**	**16 - 5**	**29 - 25**	**54 - 6**
8 - 54	1.00	0.24	0.10	0.06	0.04	0.44
9 - 12		1.00	0.24	0.07	0.10	0.32
14 - 7			1.00	0.09	0.03	0.40
16 - 5				1.00	0.09	-0.02
29 - 25					1.00	0.08
54 - 6						1.00

[1]Correlation coefficients ≥ 0.1 indicated by yellow highlight

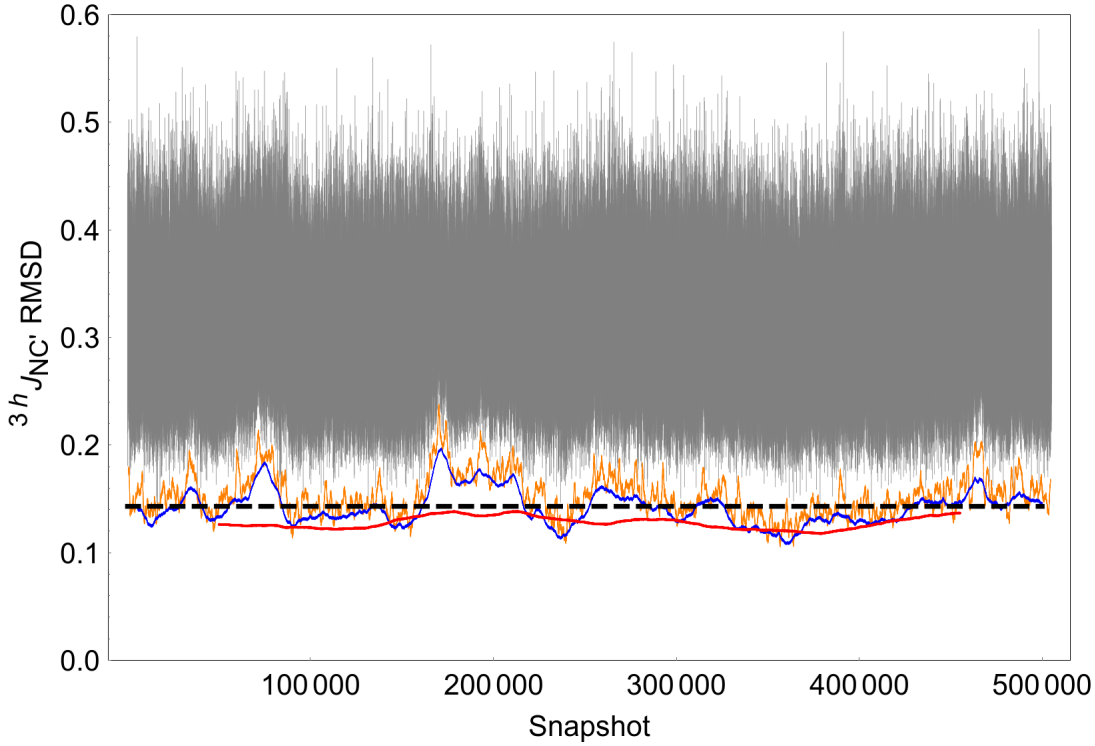


Figure G.1: Sliding window averages of the RMSDs (relative to experimentally measured values) of GB1 $^{3h}J_{NC'}$ couplings calculated by applying Eq. 8.4 to hydrogen bond geometries from snapshots of a 100 ns MD simulation (one snapshot every 200 fs). Gray line: no averaging; orange line: 1 000 snapshots averaged; blue line: 10 000 snapshots averaged; red line: 100 000 snapshots averaged. For reference, the dashed black line shows the RMSD of GB1 $^{3h}J_{NC'}$ couplings calculated by applying Eq. 8.4 to hydrogen bond geometries extracted from the 1PGB crystal structure.

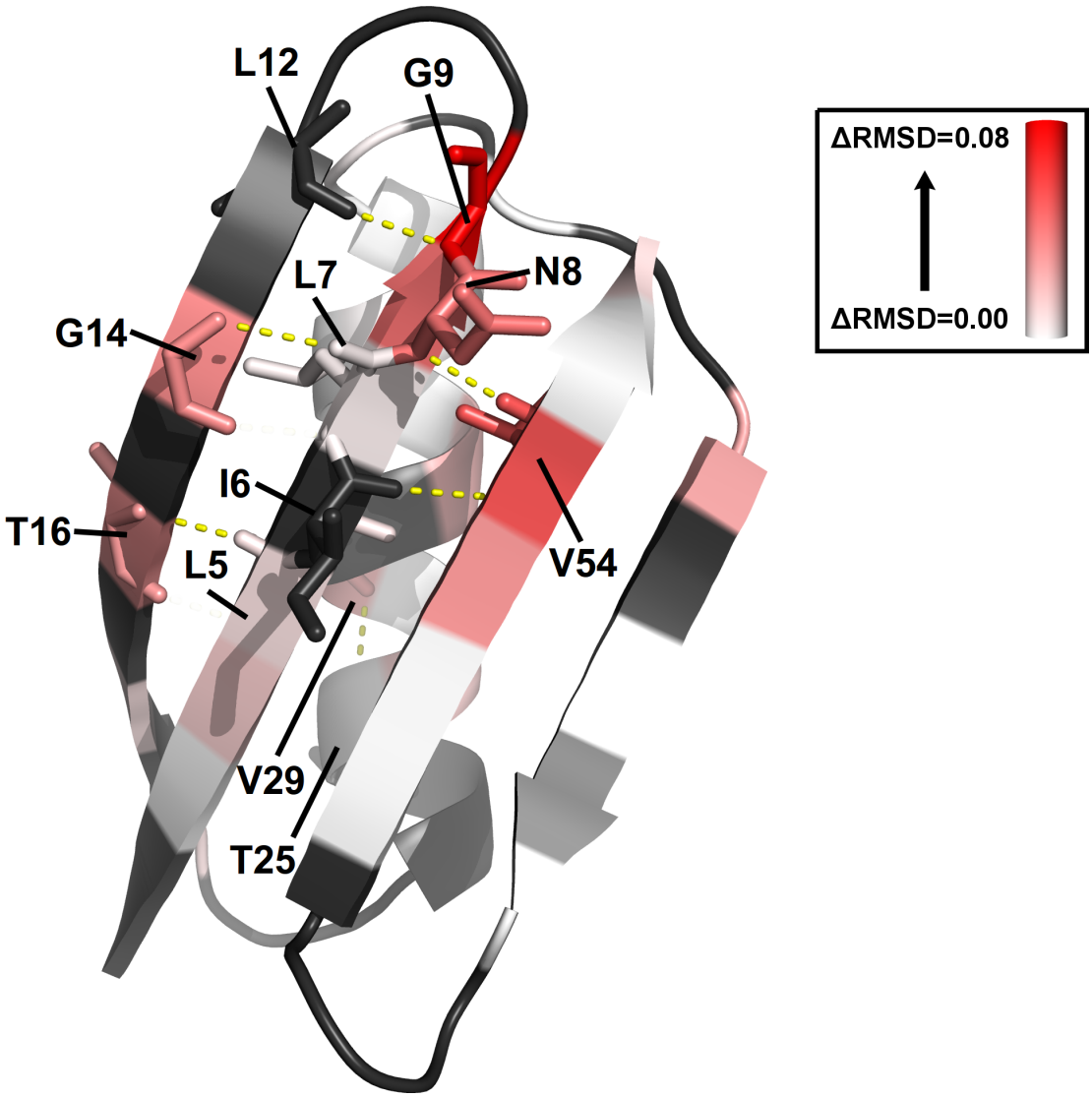


Figure G.2: GB1 ${}^{3h}J_{NC'}$ RMSD improvements from averaging over longer (100 ns, rather than 2 ns) MD simulations projected onto the 1PGB PDB structure (coloured by donor residue). Black: N/A; white to red: increasing improvement. Hydrogen bonds (dashed yellow lines) corresponding to the six ${}^{3h}J_{NC'}$ averages showing the greatest RMSD improvement (>0.03) as well as the side chains of the donor/acceptor residues involved are shown.

Appendix H

Amide Proton Temperature Coefficients: Supplemental Information

Table H.1: **Adnectin amide proton linear temperature coefficients.**

Residue #[2]	2° Struct.[3]	Temperature Coefficients (ppb/K)[1]				
		pWT[4]	Parent[5]	L78I[6]	V75R[6]	L18V/ Y88F[6]
1		-5.30	-6.48	-6.47	-6.52	-6.43
2		-8.43	N/S	-8.13	-7.99	-7.91
3		N/S	N/S	-4.98	-4.71	-4.96
4		-2.72	-7.28	-7.53	-7.45	-7.24
5						
6	E (β)	N/S	N/S	-9.07	-8.39	-9.11
7	E (β)	-8.72	N/S	-9.85	N/S	N/S
8	E (β)	2.75	0.70	0.67	-0.20	0.22
9	E (β)	-3.96	-5.48	-5.68	-4.43	N/A
10	E (β)	-5.63	N/S	N/S	N/S	N/S
11	E (β)	N/S	-5.91	-5.94	-4.92	N/S
12	E (β)	-0.87	-0.10	N/S	-0.59	N/A
13	E (β)	-9.71	-8.37	-8.55	-8.09	N/A
14	E (β)	N/S	-3.11	-3.06	-2.75	N/A
15	T (turn)					
16	T (turn)	-0.50	0.14	0.08	-0.20	0.41
17	E (β)	-0.42	-0.23	-0.30	-0.24	-1.27
18	E (β)	-3.24	-2.84	-2.88	-2.76	-3.15
19	E (β)	-2.45	-2.56	-2.66	-2.85	-2.17
20	E (β)	-4.25	-4.42	-4.57	-4.10	-4.39
21	E (β)	-3.73	-2.88	-2.99	-3.16	-2.68
22	E (β)	-2.84	-4.07	-4.01	-3.38	-3.95
23	E (β)	N/S	-2.78	-3.06	-2.36	-2.82

Table H.1: **Adnectin amide proton linear temperature coefficients.**

Residue #[2]	2° Struct.[3]	Temperature Coefficients (ppb/K)[1]				
		pWT[4]	Parent[5]	L78I[6]	V75R[6]	L18V/ Y88F[6]
24		-8.46	-5.74	-5.89	-5.73	-5.78
25						
26	S (bend)	-8.72	N/A	N/A	N/A	N/A
27	S (bend)	-2.97	N/A	N/A	N/A	N/A
28		-9.03	N/A	N/A	N/A	N/A
29		-9.61	N/A	N/A	N/S	N/A
30	S (bend)	N/S	N/A	N/A	N/S	N/A
31	E (β)	-2.19	N/A	N/A	-8.52	N/A
32	E (β)	-4.31	N/S	N/S	-5.62	N/A
33	E (β)	N/S	N/A	N/A	-8.05	N/A
34	E (β)	-1.56	-2.08	0.03	-0.29	-0.47
35	E (β)	-5.55	-4.86	-5.03	-3.66	-4.73
36	E (β)	-2.43	-2.88	-2.71	-3.26	-2.88
37	E (β)	-1.26	-1.52	-1.57	-1.65	-1.53
38	E (β)	-5.99	N/S	-5.85	-5.76	-5.85
39	T (turn)	-5.05	-5.04	-5.16	-4.92	-5.28
40	T (turn)	-9.82	-9.86	-10.00	-9.82	-10.09
41		-4.23	-3.97	-4.04	-4.16	-4.10
42		-6.45	-6.81	N/A	-6.63	-6.93
43	S (bend)	-4.78	-4.65	-4.78	-4.84	-4.74
44						
45		-8.56	-8.00	-8.15	-8.35	N/S
46	E (β)	N/S	N/S	-3.43	-3.35	-3.49
47	E (β)	N/S	-6.81	N/A	N/S	-6.93
48	E (β)	-3.37	-3.39	-3.48	-3.10	-2.97
49	E (β)	-5.99	-4.69	-4.37	-4.96	-4.62
50	E (β)	-0.26	N/S	N/S	-2.14	N/A
51	E (β)					
52	T (turn)	-1.39	N/A	N/A	N/A	N/A
53	T (turn)	-2.81	N/A	N/A	N/A	N/A
54		N/S	-2.84	-2.77	-3.37	-2.85
55		-6.12	-4.73	-4.86	-5.63	-4.72
56	E (β)	-0.74	-1.02	-1.16	-1.00	-1.15
57	E (β)	N/S	N/S	-8.06	N/S	-7.28
58	E (β)	-6.89	N/S	N/S	-6.45	-5.25
59	E (β)	-3.93	-3.74	-3.97	-4.46	-4.49
60		-8.16	-8.61	-7.98	-8.39	-6.98
61	S (bend)	-7.28	-7.06	-7.20	-7.10	-7.13
62		N/A	-7.31	-6.47	-5.98	-6.44
63		N/S	N/S	N/S	-2.98	N/A
64	T (turn)					
65	T (turn)	2.71	3.41	2.02	2.38	3.30
66		-2.77	-2.89	-3.07	-3.02	-2.45
67		-6.67	-6.26	-6.21	-6.92	-6.41

Table H.1: **Adnectin amide proton linear temperature coefficients.**

Residue #[2]	2° Struct.[3]	Temperature Coefficients (ppb/K)[1]				
		pWT[4]	Parent[5]	L78I[6]	V75R[6]	L18V/ Y88F[6]
68	E (β)	N/S	-3.42	-3.50	-3.46	-4.80
69	E (β)	-3.06	-2.98	-3.10	-2.96	-2.98
70	E (β)	-4.83	-4.58	-4.84	-4.29	-4.49
71	E (β)	-2.13	-1.88	-1.87	-2.04	-2.32
72	E (β)	-1.36	-2.10	-1.85	-2.60	-1.98
73	E (β)	N/S	N/S	N/S	1.04	N/A
74	E (β)	-3.31	N/A	N/A	-1.00	N/A
75	E (β)	-4.00	N/A	N/A	-5.25	N/A
76		N/S	N/A	N/A	-7.20	N/A
77		-5.99	N/A	N/A	N/A	N/A
78		-6.36	N/A	N/A	N/A	N/A
79	S (bend)	-8.71	N/A	N/A	N/A	N/A
80	S (bend)	-6.89	N/A	N/A	N/A	N/A
81	S (bend)	-8.69	N/A	N/A	N/A	N/A
82		N/A	N/A	N/A	N/A	N/A
83 (NA)		-8.38				
84 (NA)	B (β)	-10.11				
85 (NA)		-16.07				
86 (NA)		N/S				
87 (83)						
88 (84)	E (β)	-0.56	N/A	N/A	-2.53	N/A
89 (85)	E (β)	-6.91	-7.31	-7.34	-7.65	-7.28
90 (86)	E (β)	-2.05	-2.13	-2.13	-2.01	-3.14
91 (87)	E (β)	-7.18	N/S	-6.97	N/S	-7.69
92 (88)	E (β)	-3.45	-2.97	-3.03	-3.36	-3.01
93 (89)		-8.29	N/S	-8.09	-8.17	-6.75
94 (90)		-0.60	0.33	0.20	-0.30	1.04

[1]Proline residues indicated by gray background

[2]Pseudo-wild type ^{10}Fn3 numbering (Adnectin numbering in brackets where different)
[3]2° structure from DSSP (1FNF); blue indicates intra-β-sheet hydrogen bonds
[4]Pseudo-wild type ^{10}Fn3: D80E mutation (in the FG loop) relative to wild type
[5]Parent Adnectin: differs from pWT ^{10}Fn3 in the BC, DE, and FG loops (Chapter 1)
[6]Mutation relative to the 'Parent' Adnectin amino acid sequence

Appendix I

Amide Proton Curvature: Supplemental Information

I.1 Curvature and Secondary Structure

Table I.1: **Adnectin amide proton curvature.**

Residue #[1]	2° Struct.[2]	Curved Amide Proton Temperature Dependences				
		pWT[3]	Parent[4]	L78I[5]	V75R[5]	L18V/ Y88F[5]
1						
2						
3						✓
4		✓			✓	✓
5						
6	E (β)				✓	✓
7	E (β)	✓				
8	E (β)	✓	✓	✓		✓
9	E (β)	✓	✓			
10	E (β)	✓				
11	E (β)				✓	
12	E (β)		✓			
13	E (β)		✓	✓	✓	
14	E (β)			✓	✓	
15	T (turn)					
16	T (turn)					
17	E (β)					
18	E (β)				✓	
19	E (β)	✓	✓	✓	✓	✓
20	E (β)				✓	
21	E (β)		✓			
22	E (β)				✓	

Table I.1: **Adnectin amide proton curvature.**

Residue #[1]	2° Struct.[2]	Curved Amide Proton Temperature Dependences				
		pWT[3]	Parent[4]	L78I[5]	V75R[5]	L18V/ Y88F[5]
23	E (β)				✓	
24		✓				
25						
26	S (bend)	✓				
27	S (bend)	✓				
28		✓				
29		✓				
30	S (bend)					
31	E (β)					
32	E (β)				✓	
33	E (β)					
34	E (β)				✓	
35	E (β)				✓	
36	E (β)	✓	✓		✓	✓
37	E (β)					
38	E (β)				✓	✓
39	T (turn)	✓	✓	✓	✓	✓
40	T (turn)	✓	✓	✓	✓	✓
41			✓	✓	✓	✓
42						
43	S (bend)	✓	✓	✓	✓	✓
44						
45						
46	E (β)			✓	✓	✓
47	E (β)					
48	E (β)					✓
49	E (β)					
50	E (β)					
51	E (β)					
52	T (turn)	✓				
53	T (turn)					
54			✓	✓		
55						✓
56	E (β)					
57	E (β)					
58	E (β)				✓	
59	E (β)	✓			✓	
60				✓	✓	
61	S (bend)	✓	✓	✓	✓	✓
62				✓		✓
63						
64	T (turn)					
65	T (turn)	✓			✓	
66						

Table I.1: **Adnectin amide proton curvature.**

Residue #[1]	2° Struct.[2]	Curved Amide Proton Temperature Dependences				
		pWT[3]	Parent[4]	L78I[5]	V75R[5]	L18V/ Y88F[5]
67				✓		✓
68	E (β)		✓			✓
69	E (β)					
70	E (β)		✓			
71	E (β)					
72	E (β)				✓	
73	E (β)					
74	E (β)	✓				
75	E (β)				✓	
76						
77						
78						
79	S (bend)					
80	S (bend)	✓				
81	S (bend)	✓				
82						
83 (NA)		✓				
84 (NA)	B (β)					
85 (NA)		✓				
86 (NA)						
87 (83)						
88 (84)	E (β)	✓			✓	✓
89 (85)	E (β)					✓
90 (86)	E (β)					
91 (87)	E (β)					
92 (88)	E (β)				✓	
93 (89)		✓			✓	
94 (90)					✓	

[1]Pseudo-wild type ^{10}Fn3 numbering (Adnectin numbering in brackets where different)
[2]2° structure from DSSP (1FNF); blue indicates intra-β-sheet hydrogen bonds
[3]Pseudo-wild type ^{10}Fn3: D80E mutation (in the FG loop) relative to wild type
[4]Parent Adnectin: differs from pWT ^{10}Fn3 in the BC, DE, and FG loops (Chapter 1)
[5]Mutation relative to the 'Parent' Adnectin amino acid sequence

I.2 Bounds on Curvature Model Parameters

Table I.2: Bounds on curvature model (Eqs. 10.10, 10.13-10.15) parameters from fits to experimental data

	Res.	%[1]	ΔH[2]	ΔS[3]	m_a[4]	m_b[4]	$\delta_A(T_{ref})$[5]	$\delta_B(T_{ref})$[5]
pWT	4	0.16	-23.9 to -17.2	-0.072 to -0.051	-4.3 to -4.1	-8.9 to -1.1	7.4 to 7.5	7.4 to 8.0
	7	0.36	-14.5 to -5.8	-0.041 to -0.013	-7.2 to -5.6	-9.5 to 0.8	9.0 to 9.1	6.5 to 8.7
	8	0.18	-68.2 to 53.8	-0.236 to 0.187	-8.0 to 2.3	-7.3 to 2.4	7.2 to 7.2	7.2 to 7.2
	9	0.96	-11.4 to -2.2	-0.031 to -0.004	-4.9 to -0.2	-9.9 to 1.9	9.2 to 9.5	6.3 to 9.4
	10	0.97	-14.3 to 12.6	-0.049 to 0.044	-10.4 to 1.8	-10.8 to 0.9	8.5 to 8.9	6.4 to 8.6
	19	0.18	-58.1 to -46.5	-0.177 to -0.141	-3.1 to -3.1	-9.1 to -5.5	8.6 to 8.6	8.8 to 8.9
	24	0.12	-7.3 to -3.5	-0.019 to -0.008	-5.4 to -1.8	-9.6 to -3.3	8.9 to 9.1	6.4 to 8.2
	26	2.76	-52.0 to -9.9	-0.163 to -0.027	-9.5 to -9.1	-11.0 to -0.7	8.7 to 8.7	8.4 to 9.0
	27	0.40	-29.6 to -18.1	-0.089 to -0.054	-4.4 to -4.1	-7.2 to -0.9	7.2 to 7.2	7.2 to 7.6
	28	0.03	-34.7 to -23.2	-0.118 to -0.080	-9.1 to -9.0	-9.5 to -9.4	8.3 to 8.3	8.3 to 8.3
	29	0.20	-9.9 to -2.8	-0.028 to -0.007	-7.1 to 0.1	-9.4 to -1.2	8.8 to 9.2	6.8 to 8.2
	36	0.64	-34.7 to 32.0	-0.120 to 0.112	-5.3 to -2.7	-7.8 to -2.8	8.8 to 8.9	8.8 to 8.9
	39	0.30	-10.6 to -4.4	-0.029 to -0.009	-3.5 to -1.7	-9.7 to -1.2	9.0 to 9.1	6.5 to 8.7
	40	0.31	-16.0 to -7.4	-0.050 to -0.020	-8.3 to -6.2	-9.9 to 0.8	9.1 to 9.2	7.9 to 9.1
	43	1.20	-28.8 to -8.0	-0.087 to -0.021	-6.0 to -5.3	-8.6 to 0.4	7.8 to 7.9	7.8 to 9.4
	52	0.91	-44.2 to 43.3	-0.153 to 0.151	-4.6 to -1.7	-7.1 to -1.7	7.8 to 7.8	7.8 to 7.8
	59	4.35	-58.2 to 50.3	-0.196 to 0.175	-6.9 to -3.4	-8.9 to -4.1	9.1 to 9.1	9.0 to 9.1
	61	3.54	-90.2 to -50.9	-0.300 to -0.174	-8.8 to -6.8	-7.5 to -7.5	8.6 to 8.6	8.7 to 8.7
	65	2.06	-61.3 to -25.4	-0.187 to -0.075	2.0 to 2.2	-9.5 to 2.7	7.6 to 7.6	7.6 to 8.2
	74	0.25	-15.9 to -10.0	-0.054 to -0.033	-8.9 to -3.1	-7.7 to -6.1	8.8 to 9.0	9.1 to 9.2
	80	3.74	-37.3 to 25.5	-0.127 to 0.089	-9.8 to 1.1	-9.8 to 2.6	8.6 to 9.1	6.3 to 8.8
	81	0.59	-31.7 to -17.9	-0.099 to -0.056	-10.1 to -9.7	-10.1 to -6.6	8.3 to 8.3	8.3 to 8.4
	83	0.19	-50.0 to -37.2	-0.170 to -0.127	-8.8 to -7.8	-8.8 to -8.8	8.4 to 8.4	8.4 to 8.4
	85	0.02	-10.7 to -8.6	-0.034 to -0.027	-5.1 to -0.9	-9.2 to -2.0	9.1 to 9.2	7.7 to 8.4
	88	4.00	-57.8 to -19.9	-0.179 to -0.058	-1.3 to -1.0	-7.6 to 1.0	7.8 to 7.8	7.7 to 8.2
	93	0.11	-32.3 to -20.3	-0.110 to -0.069	-10.0 to -7.1	-9.6 to -9.4	7.7 to 7.7	7.8 to 7.8
Parent	8	1.59	-40.6 to -21.7	-0.118 to -0.058	-0.1 to 0.2	-6.6 to 2.3	7.3 to 7.3	7.4 to 9.6
	9	5.94	-93.9 to -15.1	-0.300 to -0.045	-5.2 to -4.7	-5.9 to -1.5	9.1 to 9.1	8.9 to 9.1
	12	1.89	-24.7 to -5.6	-0.070 to -0.013	-2.5 to -0.6	-7.4 to 1.8	7.6 to 7.6	7.7 to 9.6
	19	2.76	-20.7 to 16.6	-0.061 to 0.058	-5.2 to -0.8	-5.3 to 1.9	8.5 to 8.7	8.5 to 9.6
	21	2.16	-40.5 to -12.4	-0.123 to -0.038	-3.1 to -2.2	-7.8 to 0.0	8.8 to 8.8	8.7 to 9.0
	36	2.16	-40.5 to -12.4	-0.123 to -0.038	-3.1 to -2.2	-7.8 to 0.0	8.8 to 8.8	8.7 to 9.0
	39	0.61	-19.2 to -8.9	-0.056 to -0.021	-4.0 to -3.2	-9.3 to -0.2	9.0 to 9.0	6.6 to 8.9
	40	0.21	-12.7 to -6.0	-0.034 to -0.013	-8.4 to -7.3	-9.0 to -0.6	9.2 to 9.2	6.4 to 8.3
	41	4.42	-65.1 to -14.8	-0.200 to -0.042	-4.4 to -4.2	-8.3 to 2.0	8.0 to 8.0	7.8 to 9.2
	43	1.93	-15.9 to 17.0	-0.054 to 0.060	-9.9 to 0.5	-8.6 to 1.5	7.3 to 8.0	7.6 to 9.6
	54	4.12	-91.4 to 18.0	-0.300 to 0.063	-9.1 to 1.6	-9.8 to 1.5	6.9 to 7.8	7.5 to 9.7
	61	2.46	-46.2 to -16.6	-0.145 to -0.049	-6.7 to -6.2	-7.1 to 0.2	8.7 to 8.7	8.3 to 8.7
	68	3.83	-63.6 to -26.1	-0.197 to -0.080	-3.0 to -2.8	-4.2 to 2.3	9.4 to 9.4	9.1 to 9.4
	70	2.06	-12.7 to 8.5	-0.035 to 0.031	-8.1 to -0.1	-9.6 to 1.5	9.3 to 10.0	6.5 to 9.5

Table I.2: Bounds on curvature model (Eqs. 10.10, 10.13-10.15) parameters from fits to experimental data

	Res.	%[1]	ΔH^2	ΔS^3	m_a[4]	m_b[4]	$\delta_A(T_{ref})^5$	$\delta_B(T_{ref})^5$
L78I	8	0.01	-39.8 to -35.3	-0.134 to -0.119	1.0 to 1.4	1.3 to 1.4	7.3 to 7.3	7.2 to 7.2
	13	0.72	-23.0 to -8.8	-0.067 to -0.021	-7.7 to -6.8	-9.8 to 0.2	8.6 to 8.6	6.5 to 8.6
	14	0.71	-9.2 to 4.5	-0.025 to 0.016	-10.5 to 2.8	-10.7 to 1.8	8.6 to 9.7	6.1 to 9.3
	19	0.62	-57.3 to -33.3	-0.195 to -0.115	-1.7 to 1.5	-2.3 to -2.1	8.6 to 8.7	8.6 to 8.6
	39	0.34	-18.5 to -12.9	-0.052 to -0.032	-4.2 to -3.9	-9.1 to 0.8	9.0 to 9.0	6.6 to 8.8
	40	0.65	-13.8 to -2.5	-0.039 to -0.006	-8.5 to 0.5	-10.1 to 1.3	9.2 to 9.7	6.3 to 8.9
	41	2.54	-36.4 to -8.1	-0.109 to -0.026	-4.7 to -1.4	-7.6 to 1.5	8.0 to 8.1	7.7 to 9.2
	43	0.73	-14.4 to -5.3	-0.041 to -0.011	-7.0 to -5.5	-6.9 to 1.1	7.8 to 7.9	7.8 to 9.5
	46	0.81	-9.6 to 11.5	-0.033 to 0.042	-10.1 to -2.2	-8.4 to 1.2	8.2 to 9.0	6.4 to 8.3
	54	1.60	-58.5 to 21.9	-0.198 to 0.077	-2.3 to 0.5	-2.5 to 1.1	7.6 to 7.6	7.6 to 7.6
	60	0.23	-91.4 to -82.9	-0.300 to -0.273	-7.2 to -6.9	-9.3 to -9.2	8.5 to 8.5	8.6 to 8.6
	61	2.54	-17.3 to 18.5	-0.056 to 0.065	-8.6 to -0.1	-8.3 to 0.6	8.5 to 9.1	6.7 to 8.9
	62	2.57	-93.9 to -40.8	-0.300 to -0.130	-5.9 to -5.9	-7.5 to -7.3	8.4 to 8.4	8.4 to 8.4
	67	0.80	-55.0 to -31.7	-0.168 to -0.098	-6.9 to -6.8	-8.8 to -4.7	8.1 to 8.1	8.1 to 8.3
V75R	4	0.12	-5.3 to 4.5	-0.017 to 0.016	-9.0 to -4.6	-5.3 to 0.6	7.9 to 8.6	7.4 to 8.6
	6	0.62	-61.4 to -38.6	-0.191 to -0.119	-9.2 to -9.0	-10.2 to -8.8	8.5 to 8.5	8.6 to 8.6
	11	0.31	-11.2 to -4.9	-0.031 to -0.010	-3.4 to -2.0	-9.3 to -1.0	9.0 to 9.1	6.5 to 8.8
	13	0.43	-49.4 to -34.9	-0.157 to -0.110	-7.6 to -7.5	-8.4 to -7.8	8.6 to 8.6	8.6 to 8.7
	14	0.32	-14.7 to 12.0	-0.048 to 0.042	-6.0 to -1.1	-7.8 to -3.3	8.7 to 9.1	8.7 to 9.1
	18	2.05	-83.8 to -38.4	-0.265 to -0.117	-2.6 to -2.6	-2.8 to -1.2	8.7 to 8.7	8.6 to 8.7
	19	0.60	-62.0 to -48.3	-0.192 to -0.146	-3.2 to -3.2	-7.8 to -3.0	8.7 to 8.7	8.7 to 8.9
	20	1.06	-33.8 to -13.0	-0.114 to -0.042	-3.9 to -2.9	-4.3 to -3.5	9.1 to 9.1	9.0 to 9.1
	22	0.96	-16.8 to -6.2	-0.049 to -0.015	-4.3 to -2.2	-9.6 to -1.2	7.9 to 8.0	6.8 to 8.2
	23	1.63	-26.3 to 22.5	-0.086 to 0.079	-4.1 to -1.8	-9.9 to -2.2	7.9 to 8.1	7.7 to 8.1
	32	0.85	-33.4 to -13.4	-0.098 to -0.041	-5.6 to -5.1	-8.3 to 0.5	9.7 to 9.7	7.3 to 9.8
	34	0.53	-5.4 to 12.0	-0.018 to 0.045	-6.5 to 0.4	-8.4 to -1.0	8.4 to 8.8	8.4 to 9.4
	35	0.36	-23.0 to -13.1	-0.075 to -0.045	-7.8 to -3.7	-5.5 to -4.7	9.3 to 9.4	9.5 to 9.5
	36	0.46	-22.3 to 20.6	-0.075 to 0.072	-4.9 to -3.0	-6.9 to -3.7	8.8 to 8.9	8.8 to 8.9
	38	1.32	-33.2 to -19.2	-0.108 to -0.061	-6.8 to -5.3	-6.1 to -5.6	9.1 to 9.1	9.0 to 9.1
	39	0.31	-11.2 to -4.9	-0.031 to -0.010	-3.4 to -2.0	-9.3 to -1.0	9.0 to 9.1	6.5 to 8.8
	40	0.04	-4.7 to -2.6	-0.012 to -0.007	-5.4 to 0.7	-8.8 to -6.0	9.3 to 9.7	7.2 to 8.3
	41	3.76	-62.1 to -10.0	-0.190 to -0.027	-4.4 to -4.2	-9.1 to 1.9	8.0 to 8.0	7.9 to 9.1
	43	0.26	-8.0 to -2.3	-0.022 to -0.007	-8.4 to -4.6	-6.7 to 0.3	7.8 to 8.0	7.6 to 9.3
	46	0.12	-34.6 to -25.3	-0.118 to -0.087	-9.1 to -6.6	-4.5 to -4.2	8.2 to 8.2	8.3 to 8.3
	58	2.86	-83.7 to -34.7	-0.270 to -0.119	-8.9 to -5.8	-7.0 to -6.5	8.5 to 8.5	8.5 to 8.5
	59	0.97	-27.5 to 27.8	-0.095 to 0.097	-6.6 to -4.3	-9.7 to -4.9	9.1 to 9.2	9.0 to 9.2
	60	0.62	-61.4 to -38.6	-0.191 to -0.119	-9.2 to -9.0	-10.2 to -8.8	8.5 to 8.5	8.6 to 8.6
	61	0.05	-25.5 to -19.6	-0.086 to -0.066	-7.2 to -6.7	-7.5 to -7.4	8.6 to 8.6	8.7 to 8.7
	65	0.17	-11.2 to -5.0	-0.031 to -0.012	-3.9 to -0.7	-8.4 to 1.8	7.3 to 7.5	8.3 to 9.7
	72	1.18	-43.4 to -23.5	-0.139 to -0.073	-1.9 to -1.5	-2.9 to -0.7	8.9 to 8.9	8.8 to 8.9
	75	0.04	-10.6 to -8.8	-0.028 to -0.022	-1.4 to -0.8	-8.2 to -3.0	8.9 to 8.9	6.4 to 7.9
	84	0.46	-24.4 to 20.3	-0.084 to 0.071	-6.6 to -3.2	-9.3 to -3.5	8.4 to 8.6	8.4 to 8.6
	88	0.19	-44.7 to 40.5	-0.152 to 0.141	-5.6 to -3.6	-7.6 to -3.6	9.5 to 9.5	9.5 to 9.5
	89	0.05	-34.7 to -28.4	-0.118 to -0.095	-9.8 to -7.6	-8.7 to -8.6	7.6 to 7.6	7.7 to 7.7
	90	1.79	-23.1 to 12.6	-0.069 to 0.045	-6.8 to 1.3	-9.5 to 1.1	8.4 to 9.0	8.3 to 9.4

Table I.2: Bounds on curvature model (Eqs. 10.10, 10.13-10.15) parameters from fits to experimental data

	Res.	%[1]	ΔH^2	ΔS^3	m_a[4]	m_b[4]	$\delta_A(T_{ref})$[5]	$\delta_B(T_{ref})$[5]
L18V-Y88F	3	1.18	-16.3 to -3.1	-0.044 to -0.006	-5.5 to -2.1	-8.9 to 0.1	8.2 to 8.4	6.5 to 8.5
	4	1.14	-28.9 to 14.0	-0.093 to 0.049	-7.7 to -4.3	-7.3 to -1.3	8.0 to 8.2	8.0 to 8.3
	6	0.15	-47.9 to -40.6	-0.146 to -0.123	-8.4 to -8.3	-4.6 to 0.3	8.5 to 8.5	8.1 to 8.3
	8	0.15	-12.3 to -8.0	-0.033 to -0.019	-1.7 to -1.0	-6.2 to -0.2	7.3 to 7.3	7.8 to 9.2
	19	0.98	-15.5 to -3.2	-0.043 to -0.007	-4.6 to -2.7	-4.7 to 1.6	8.5 to 8.6	8.6 to 9.6
	36	0.10	-38.4 to -33.6	-0.117 to -0.102	-2.3 to -2.2	-1.7 to 0.7	8.9 to 8.9	8.7 to 8.8
	38	0.55	-21.1 to -8.4	-0.060 to -0.018	-5.5 to -5.3	-7.5 to 0.6	9.0 to 9.1	7.0 to 9.0
	39	0.18	-15.5 to -11.3	-0.043 to -0.028	-4.3 to -4.1	-6.7 to 0.1	9.0 to 9.0	6.8 to 8.7
	40	0.24	-10.4 to -3.2	-0.028 to -0.007	-8.2 to -3.5	-8.9 to 0.2	9.1 to 9.4	6.5 to 8.4
	41	0.12	-26.2 to -14.8	-0.076 to -0.042	-4.5 to -4.4	-7.2 to 0.1	8.0 to 8.0	7.9 to 8.3
	43	0.50	-8.1 to 8.7	-0.027 to 0.031	-8.6 to -0.4	-8.4 to 0.0	7.4 to 8.0	7.6 to 9.2
	46	0.38	-27.1 to -19.8	-0.083 to -0.064	-2.8 to -2.0	-5.2 to 0.0	8.2 to 8.2	8.0 to 8.3
	48	1.69	-92.3 to -44.7	-0.300 to -0.138	-2.9 to -2.7	-3.2 to -2.0	9.1 to 9.1	9.1 to 9.2
	55	3.79	-22.9 to 23.0	-0.079 to 0.081	-9.3 to 0.1	-10.0 to 2.3	6.9 to 7.9	7.1 to 9.4
	61	0.06	-19.9 to -15.8	-0.059 to -0.046	-6.5 to -6.3	-4.7 to -2.2	8.6 to 8.6	8.3 to 8.5
	62	0.09	-25.9 to -21.9	-0.087 to -0.074	-8.3 to -6.3	-7.4 to -7.3	8.4 to 8.4	8.4 to 8.4
	68	0.91	-14.1 to -3.7	-0.040 to -0.007	-3.9 to -0.9	-9.6 to -0.5	9.5 to 9.6	6.3 to 9.5
	88	1.02	-37.7 to -10.8	-0.121 to -0.037	-5.3 to -2.9	-4.0 to -3.3	9.4 to 9.4	9.4 to 9.5
	89	1.14	-56.9 to -36.6	-0.182 to -0.115	-6.1 to -6.0	-7.4 to -6.1	7.7 to 7.7	7.7 to 7.8

[1]Percentage of the total number of fits (45075) with RMSE within 5% of the minimum
[2]Range observed in fits with RMSE within 5% of the minimum (kcal·mol^{-1})
[3]Range observed in fits with RMSE within 5% of the minimum (kcal·mol^{-1}·K^{-1})
[4]Range observed in fits with RMSE within 5% of the minimum (ppb/K)
[5]Range observed in fits with RMSE within 5% of the minimum (ppm)

I.3 pWT Adnectin Curvalyzer Results

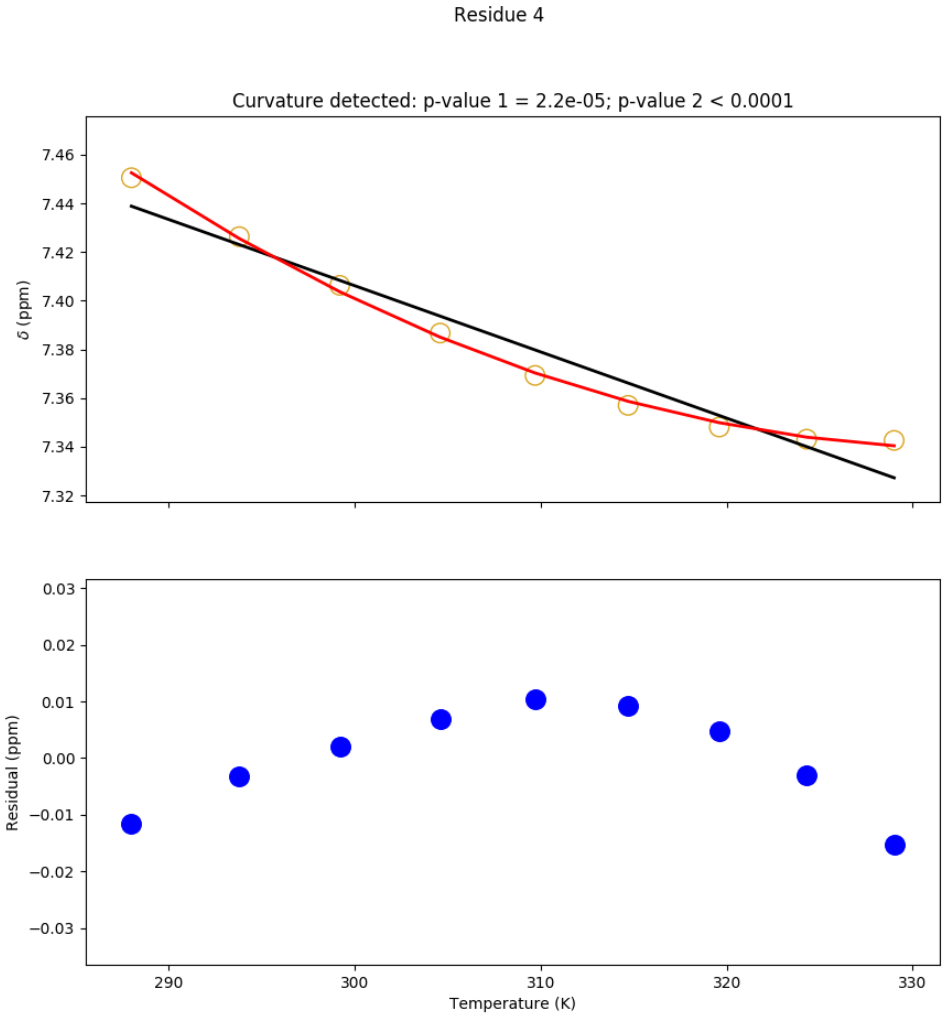


Figure I.1: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 4. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

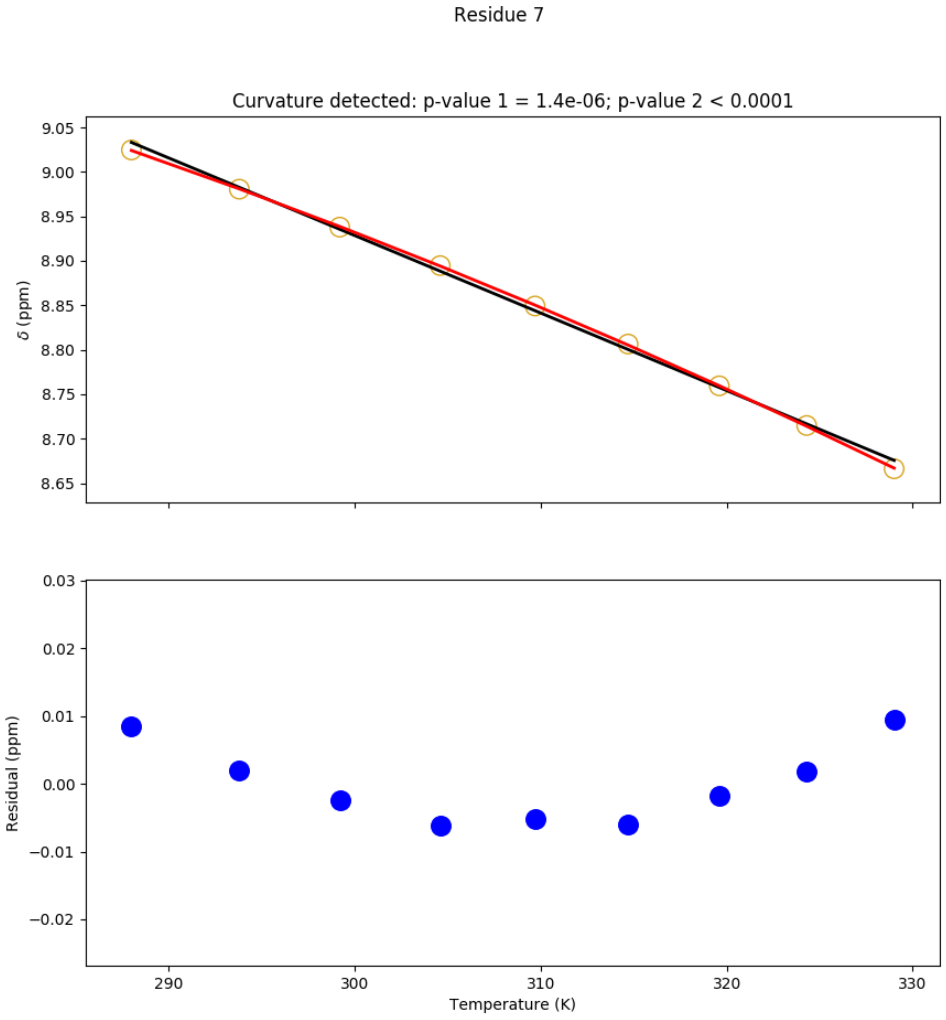


Figure I.2: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 7. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

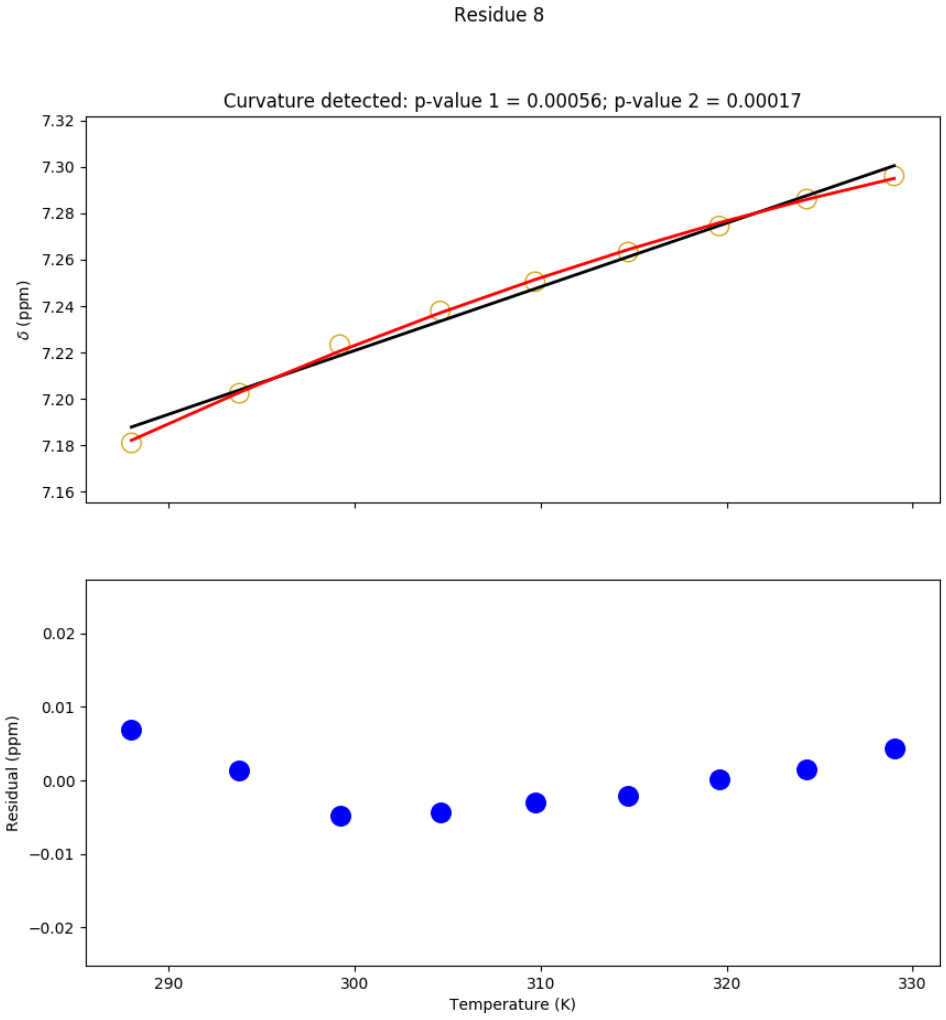


Figure I.3: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 8. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

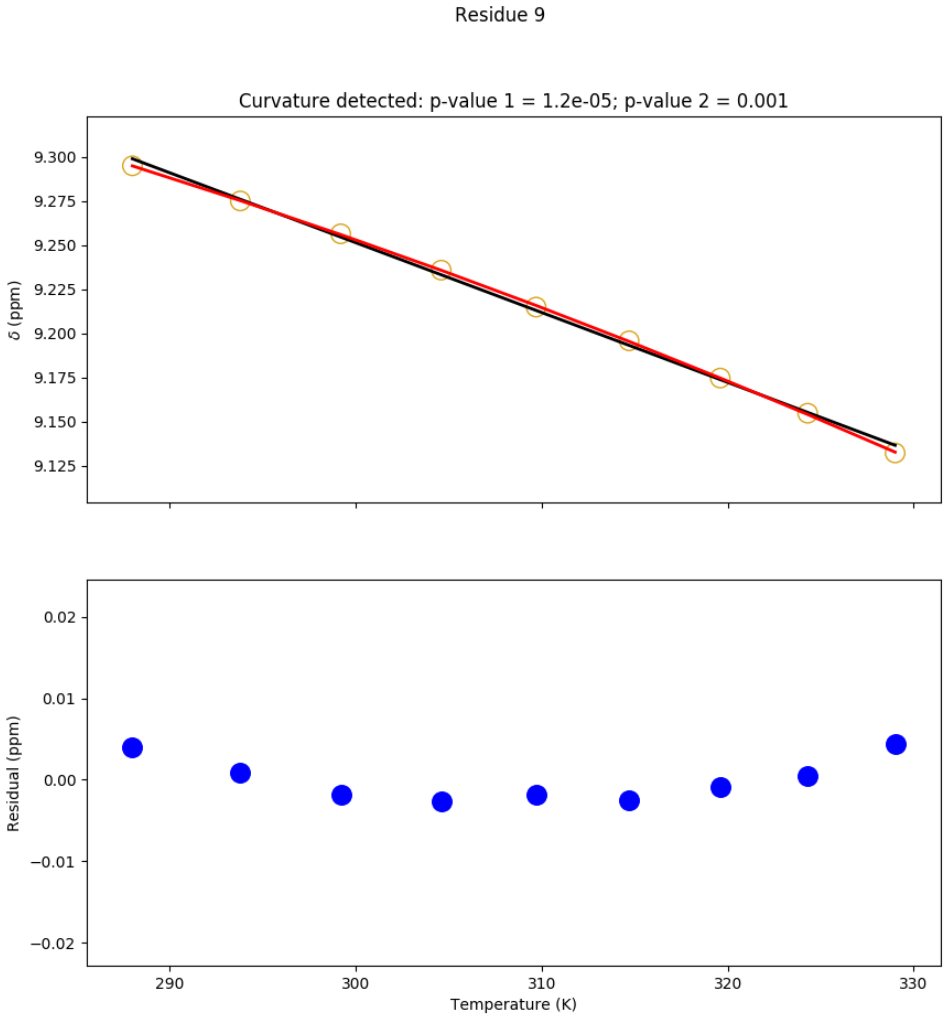


Figure I.4: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 9. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

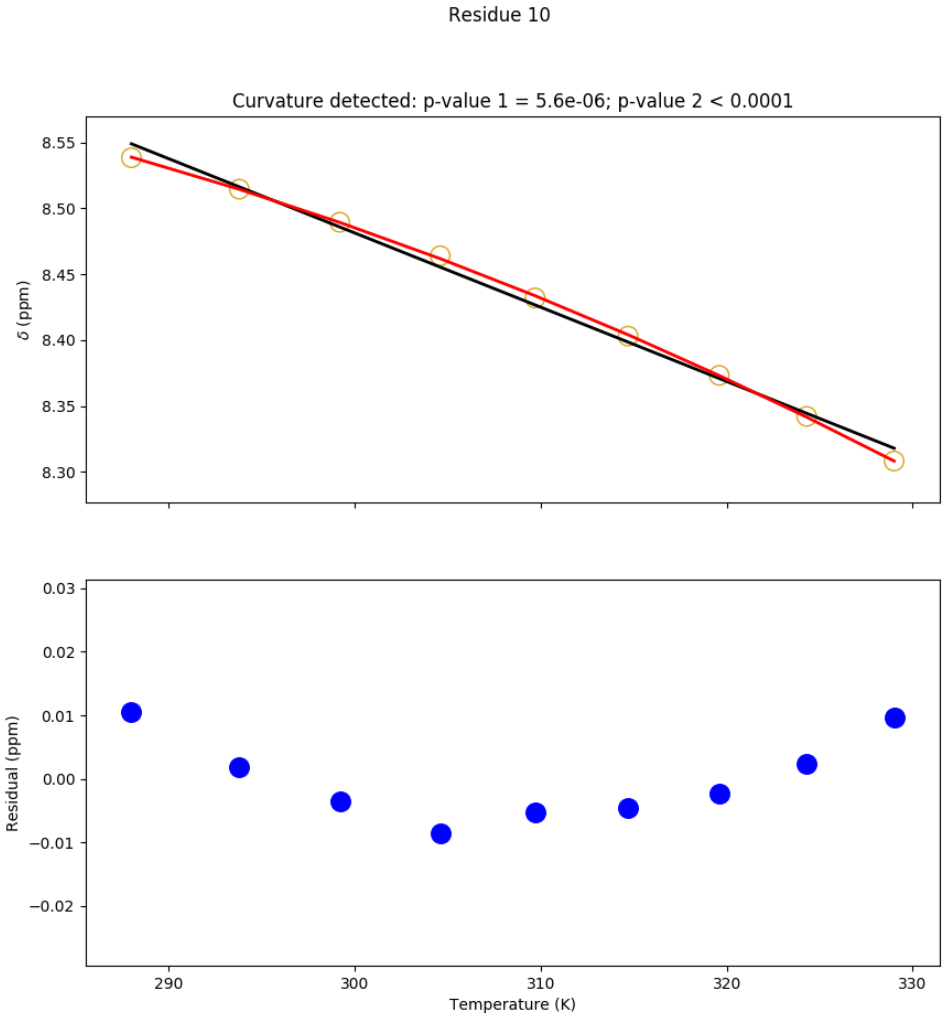


Figure I.5: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 10. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

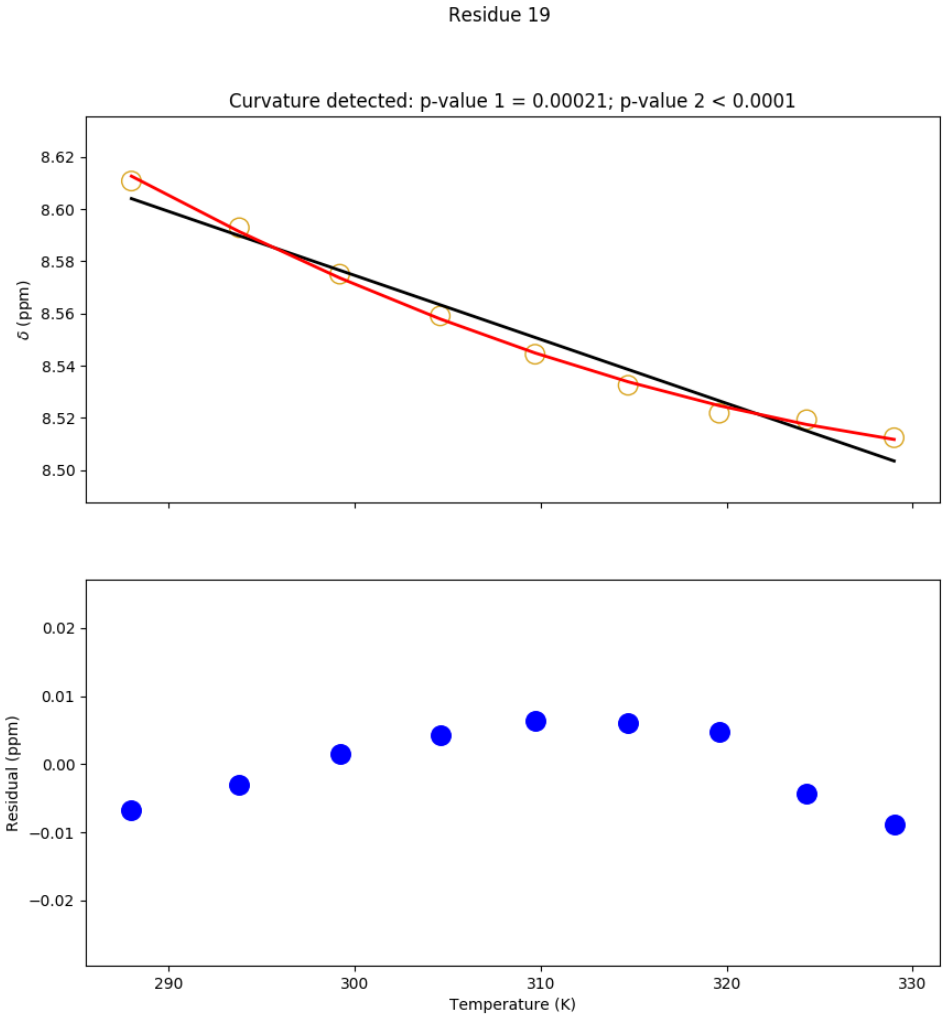


Figure I.6: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 19. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

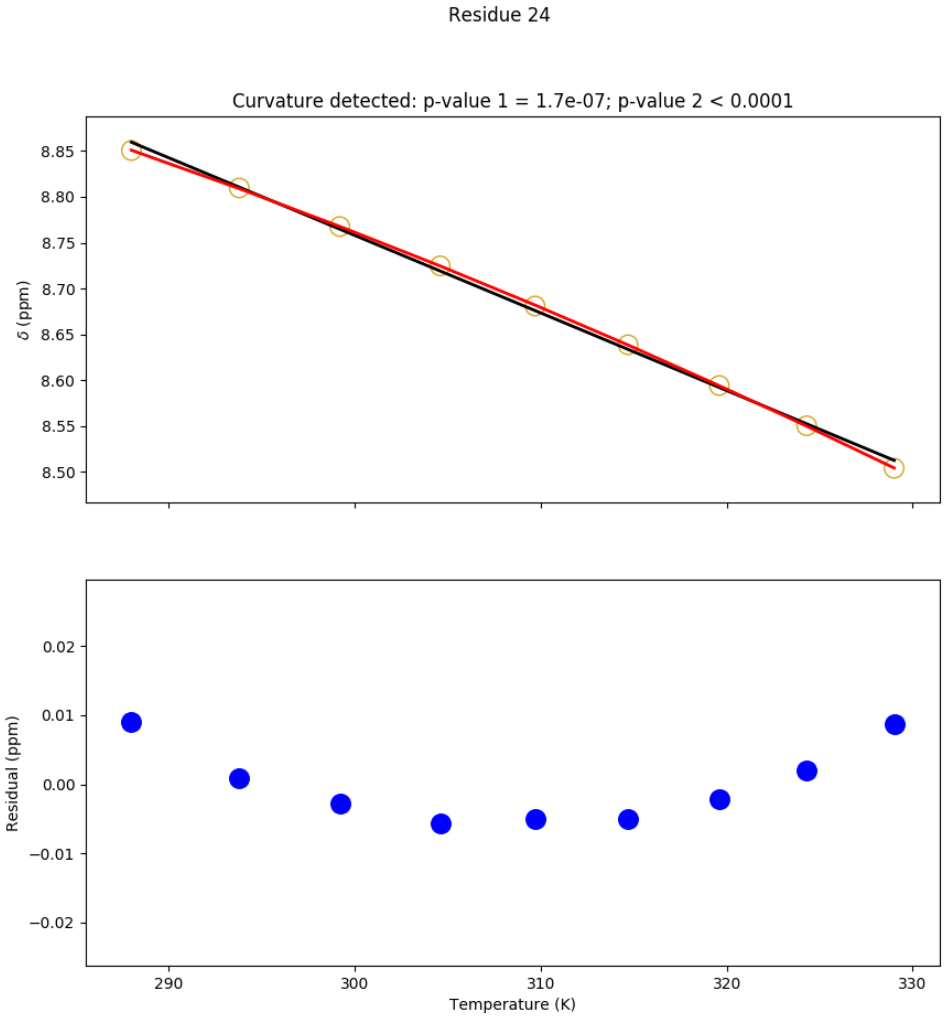


Figure I.7: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 24. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

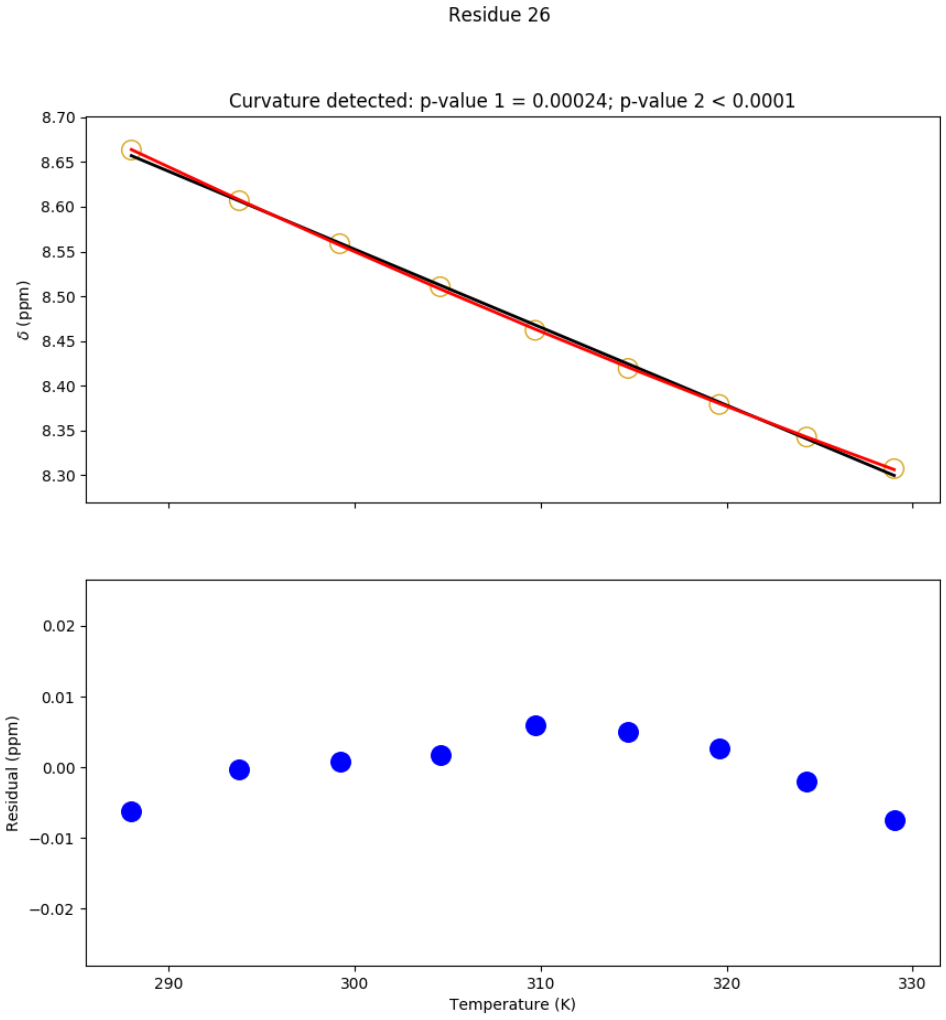


Figure I.8: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 26. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

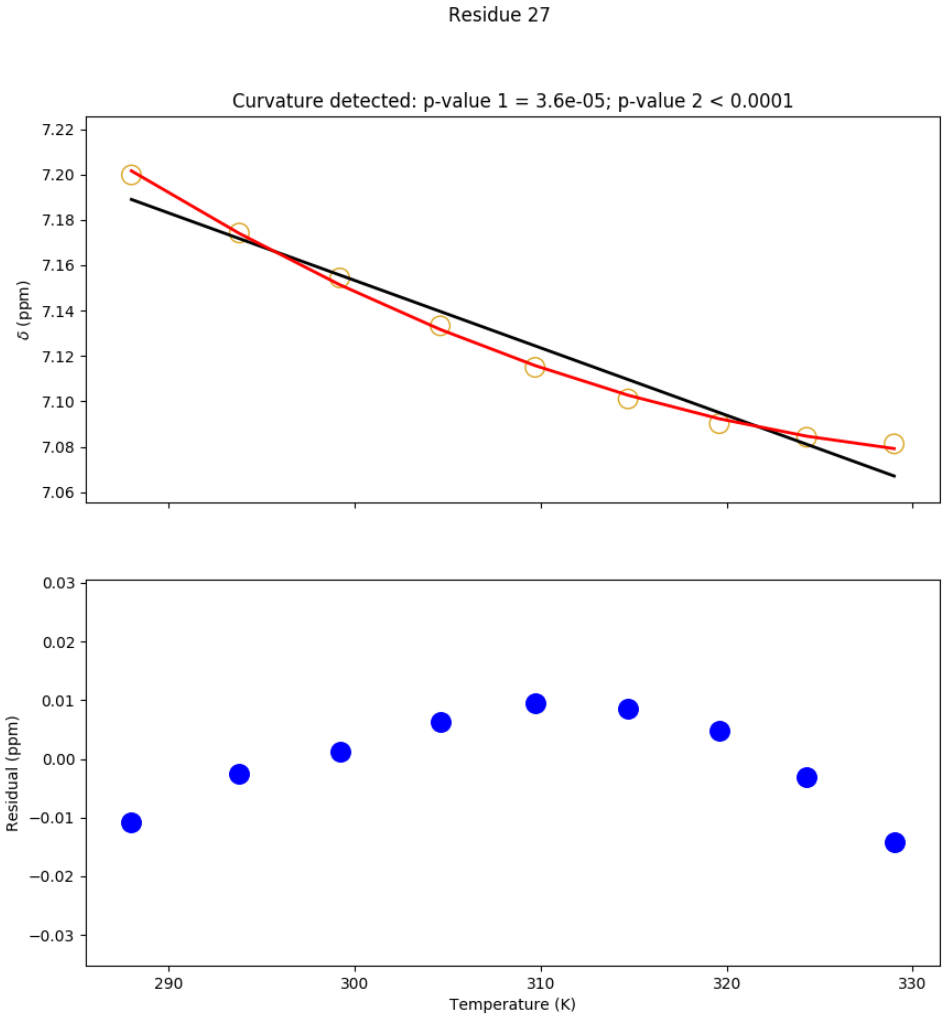


Figure I.9: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 27. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

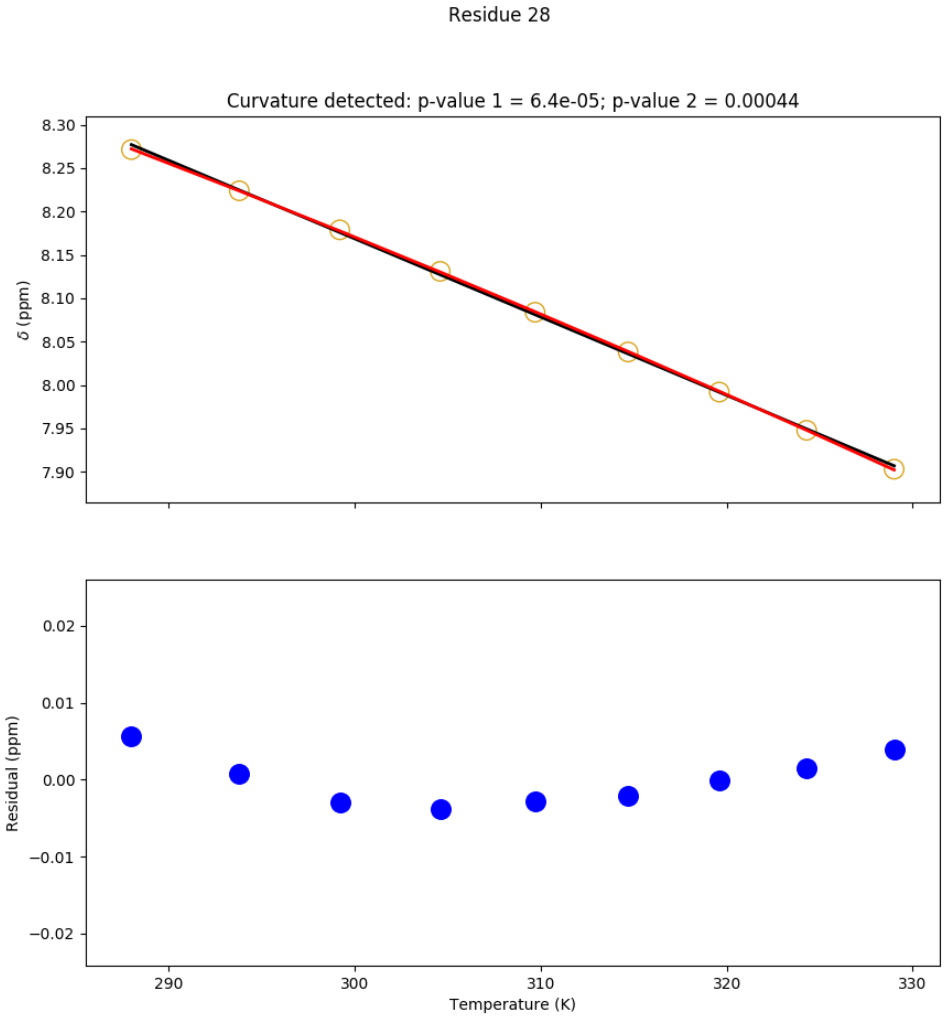


Figure I.10: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 28. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

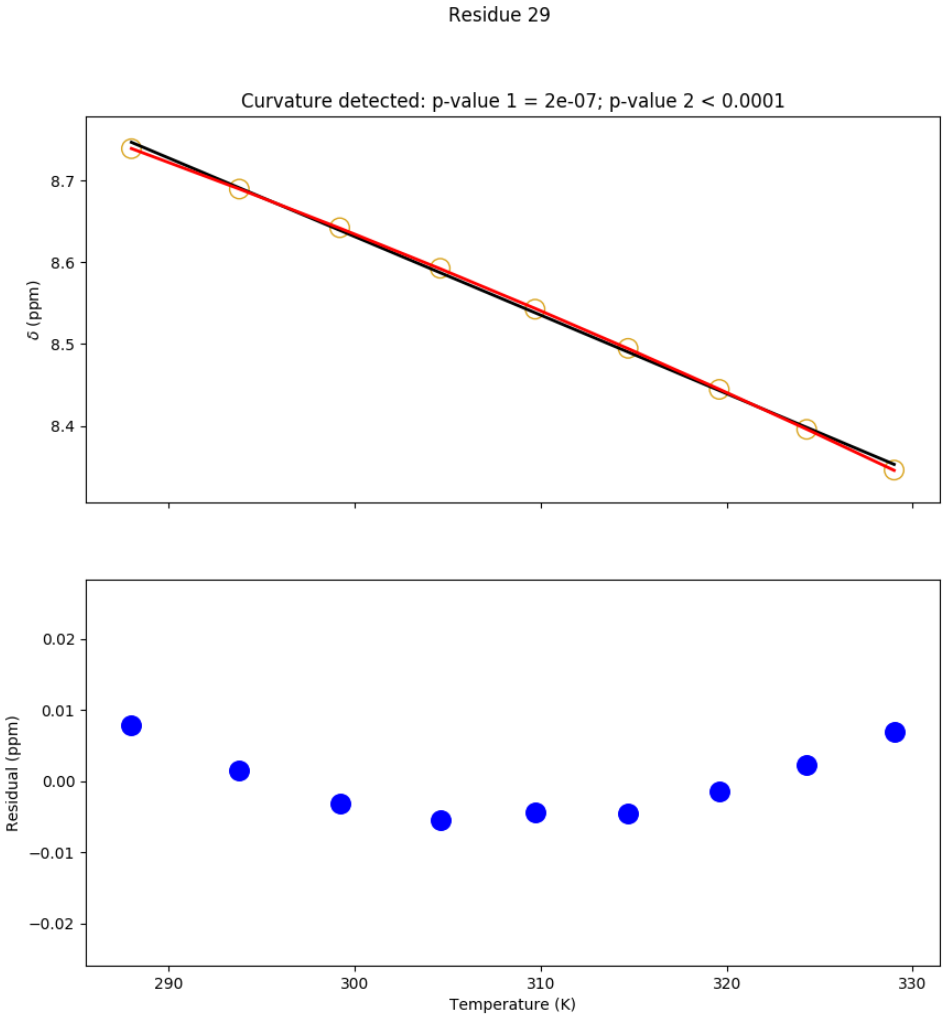


Figure I.11: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 29. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

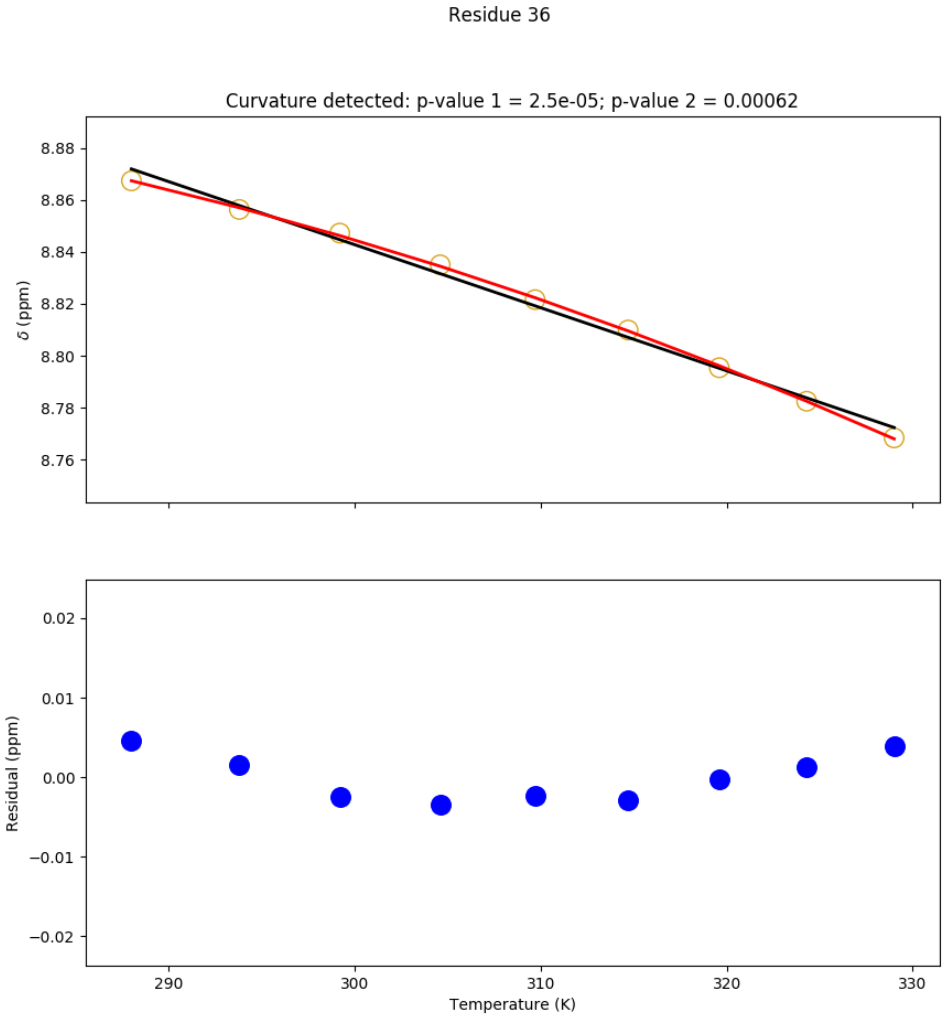


Figure I.12: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 36. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

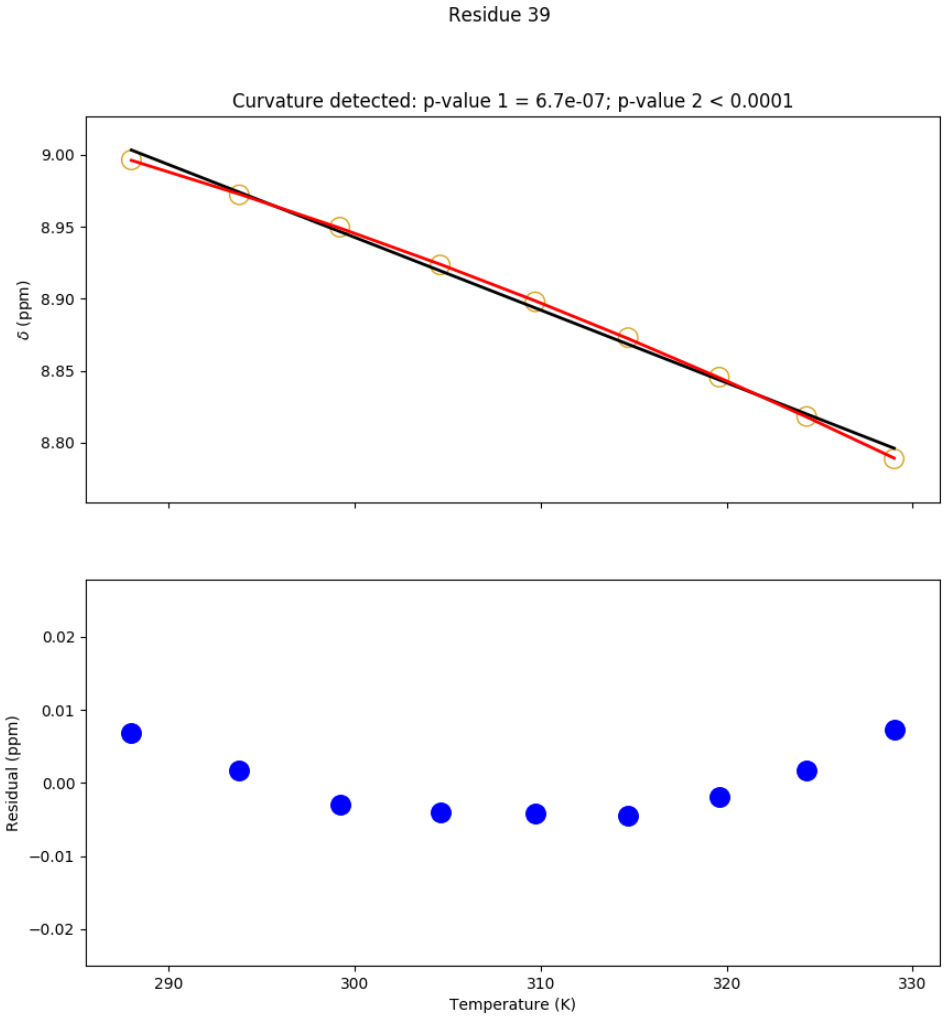


Figure I.13: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 39. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

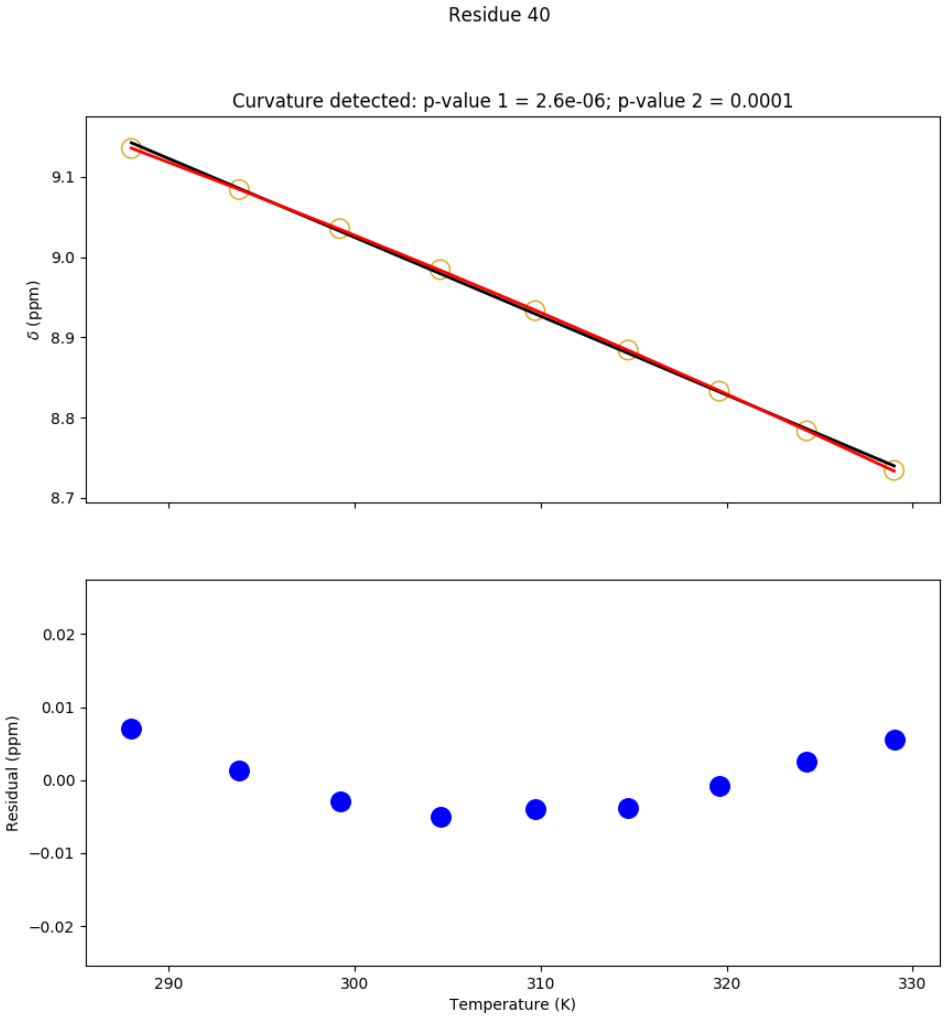


Figure I.14: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 40. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

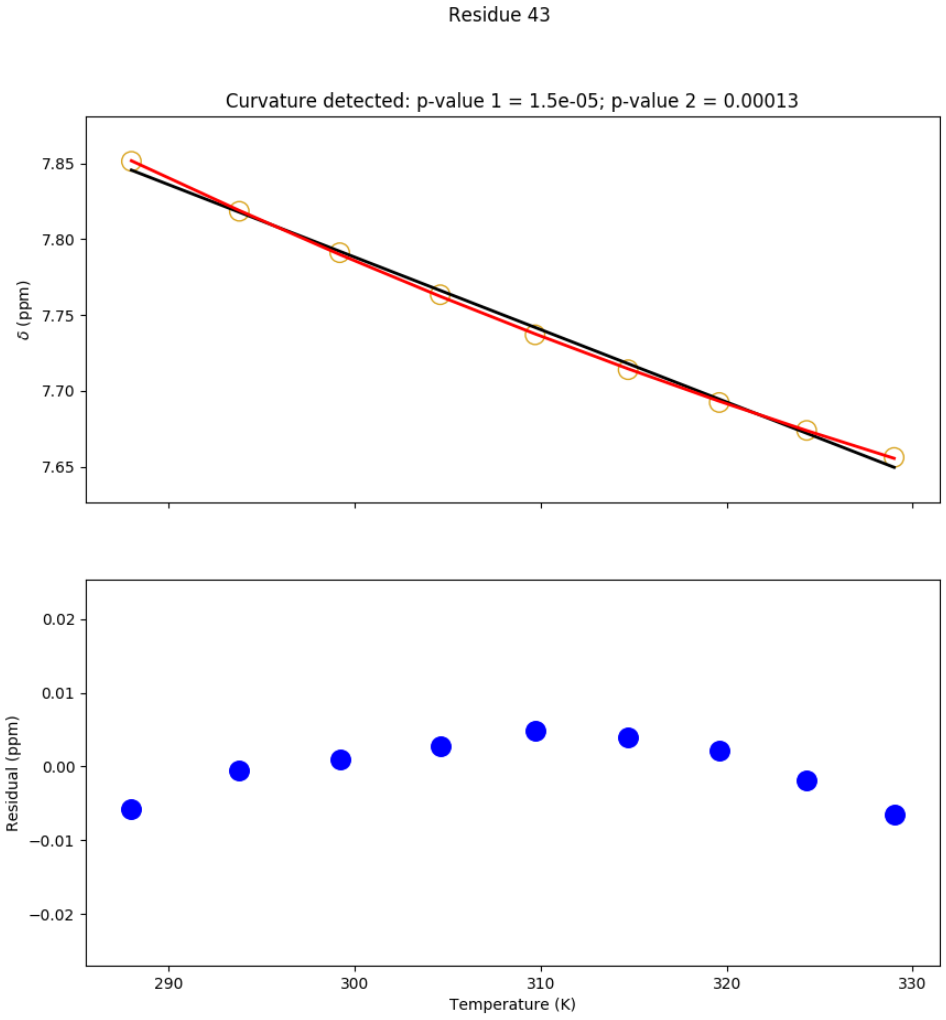


Figure I.15: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 43. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

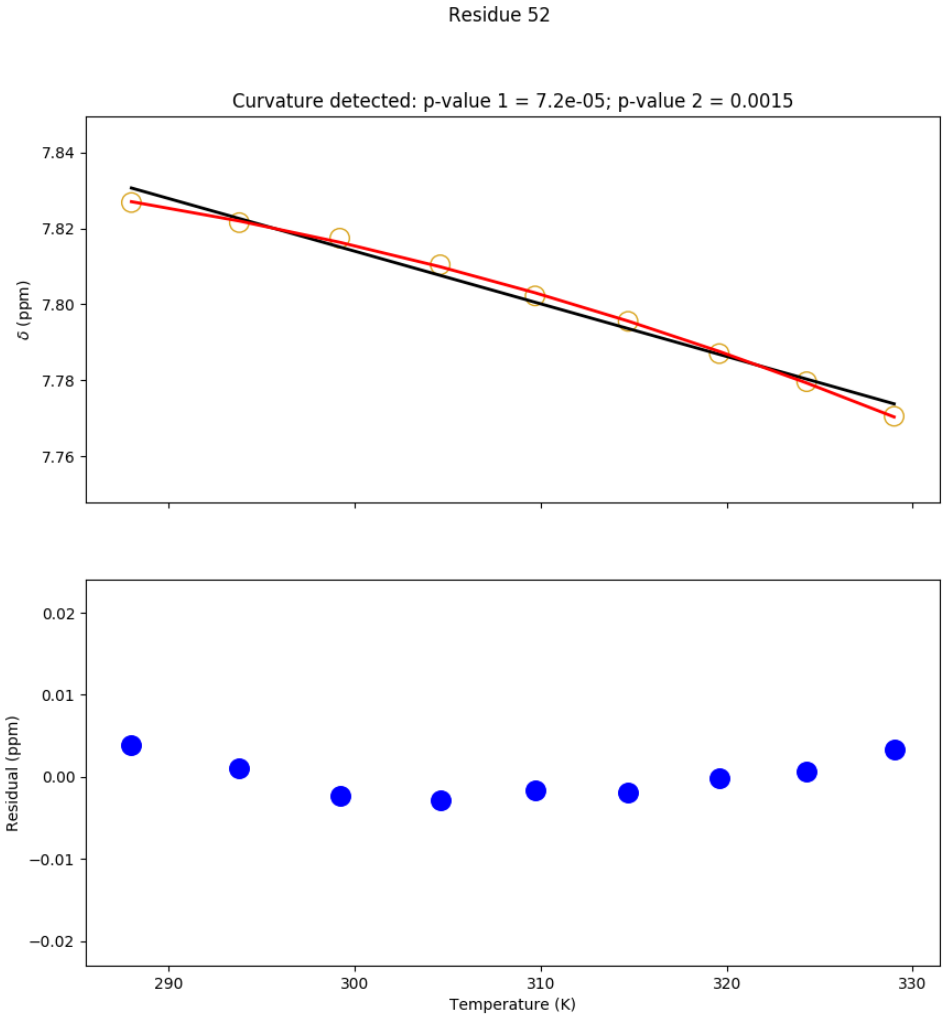


Figure I.16: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 52. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

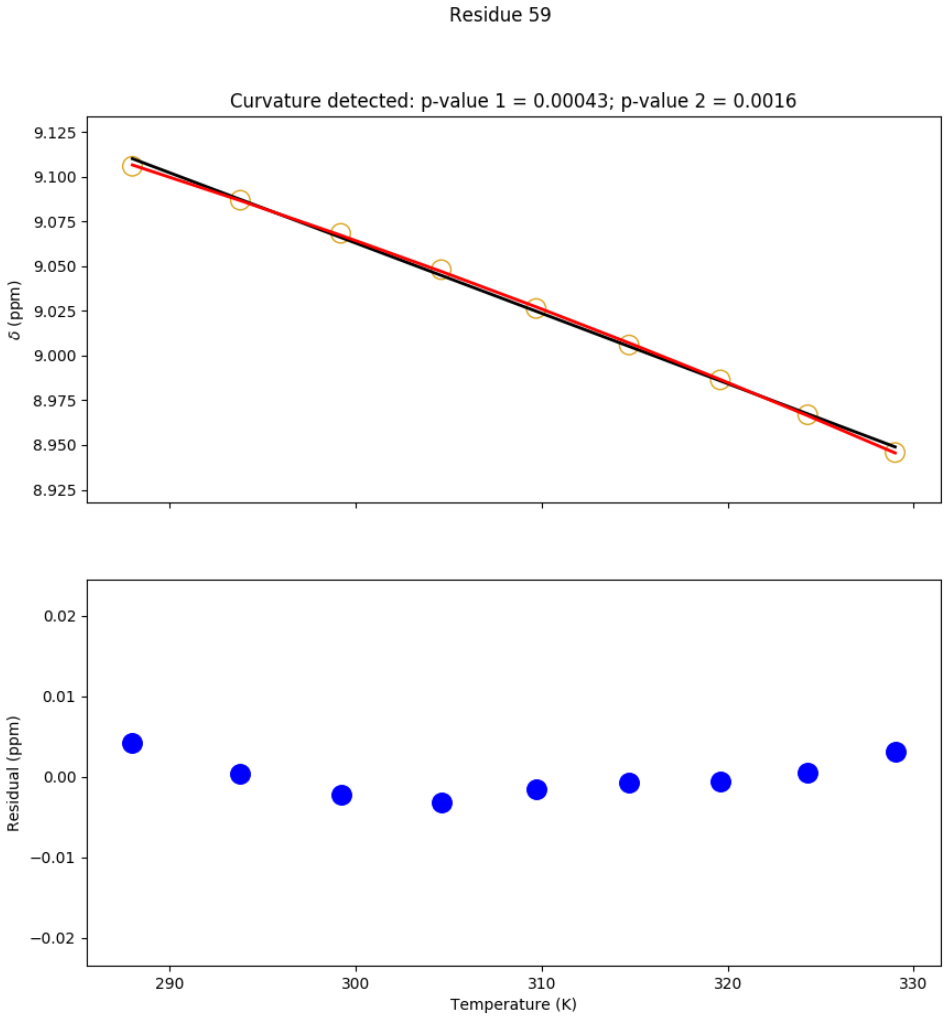


Figure I.17: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 59. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

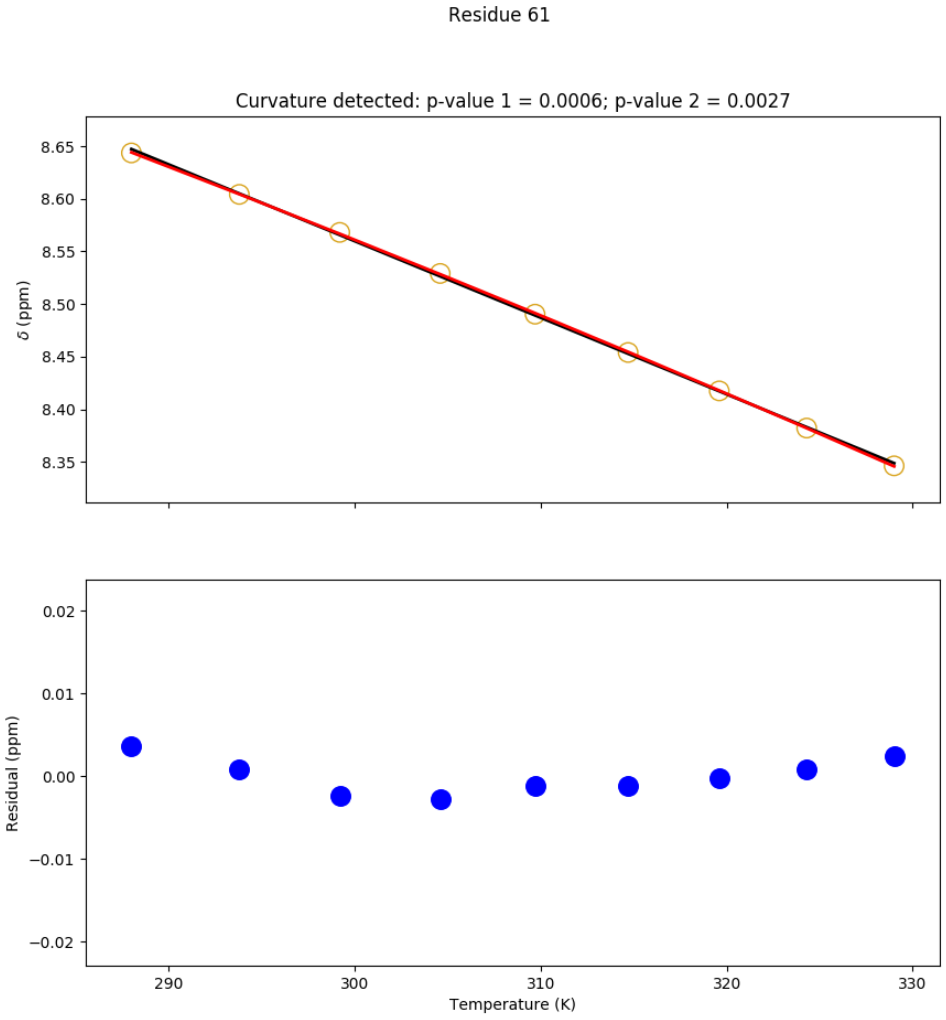


Figure I.18: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 61. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

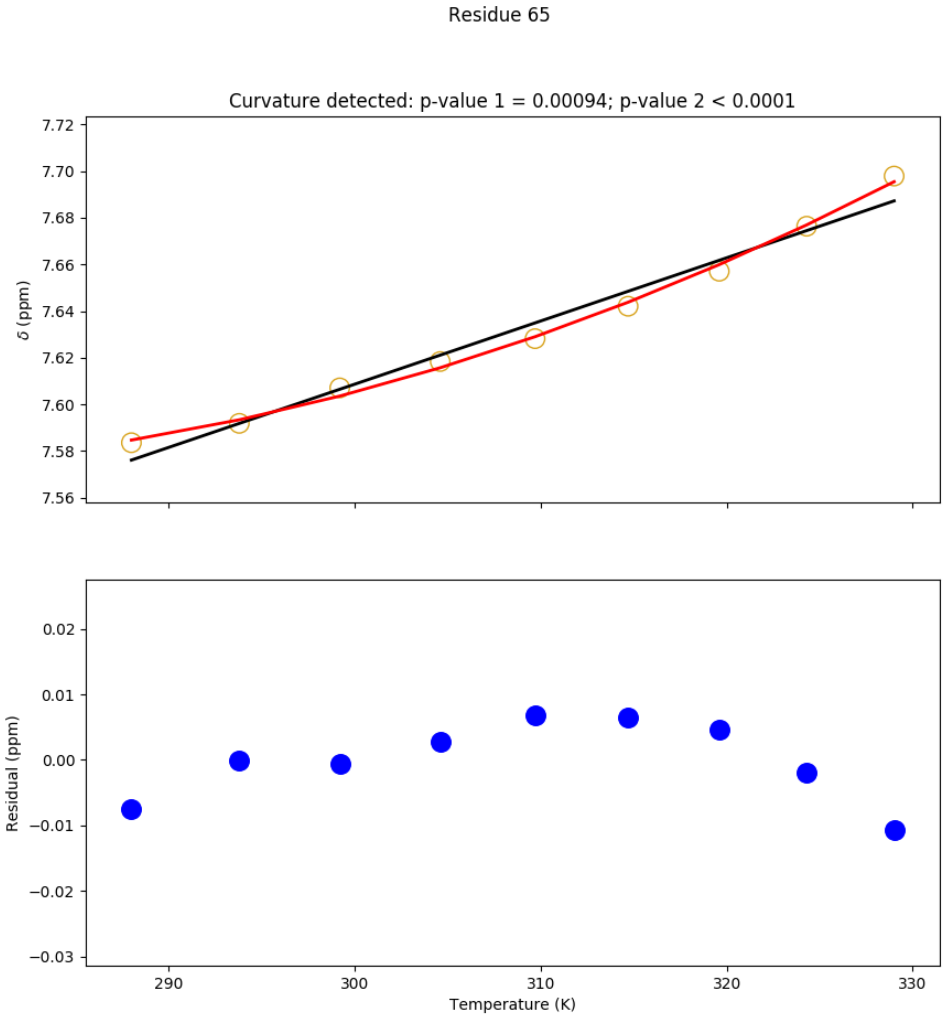


Figure I.19: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 65. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

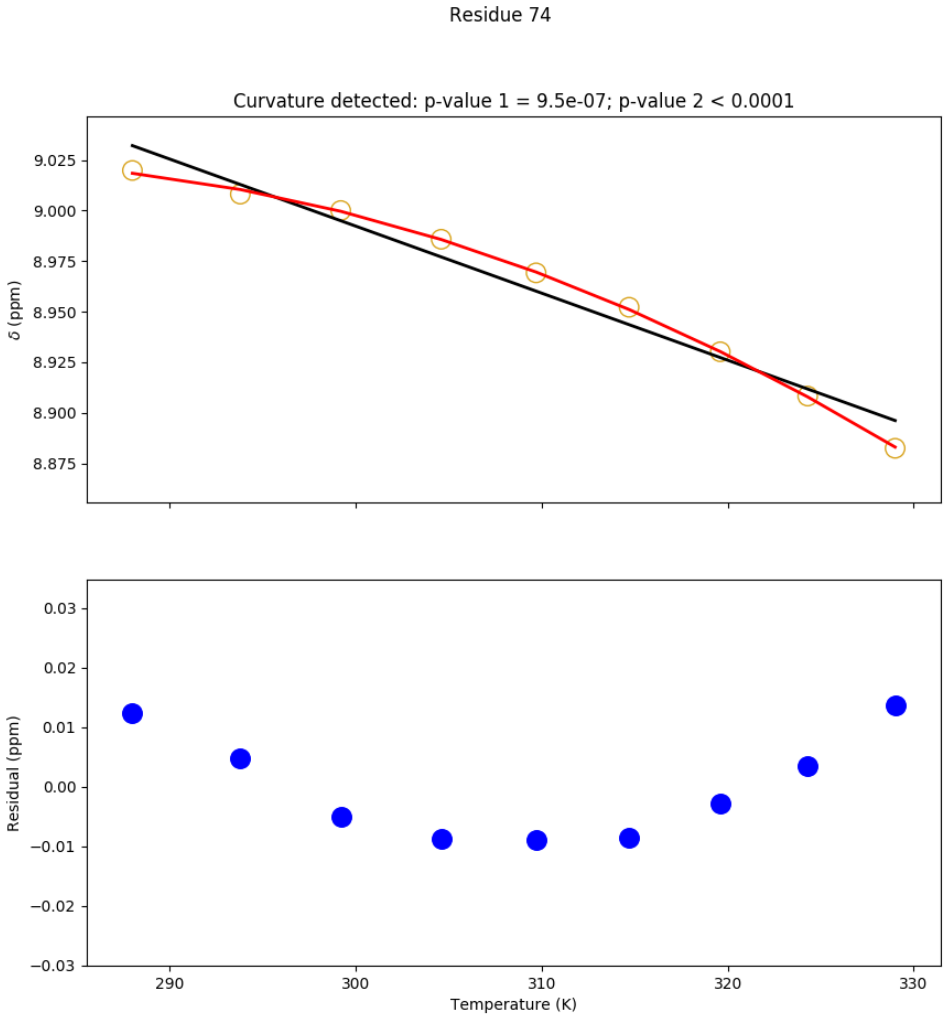


Figure I.20: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 74. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

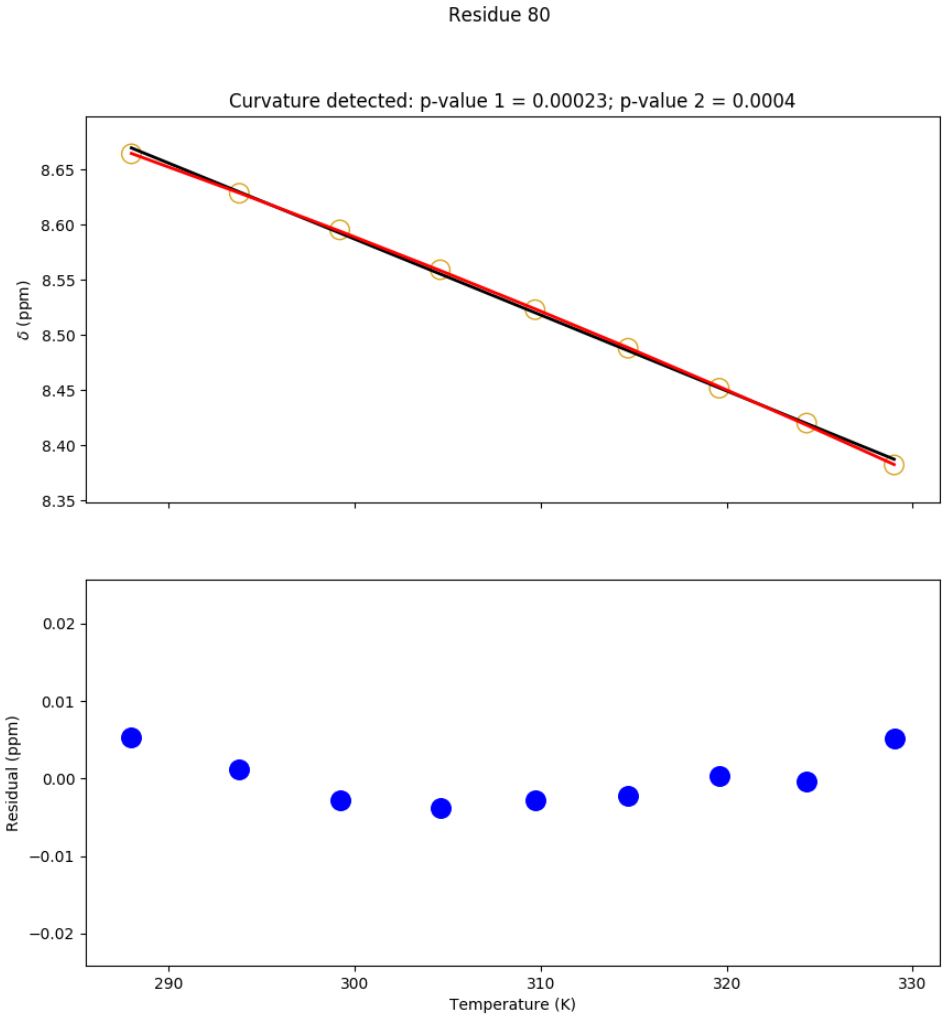


Figure I.21: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 80. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

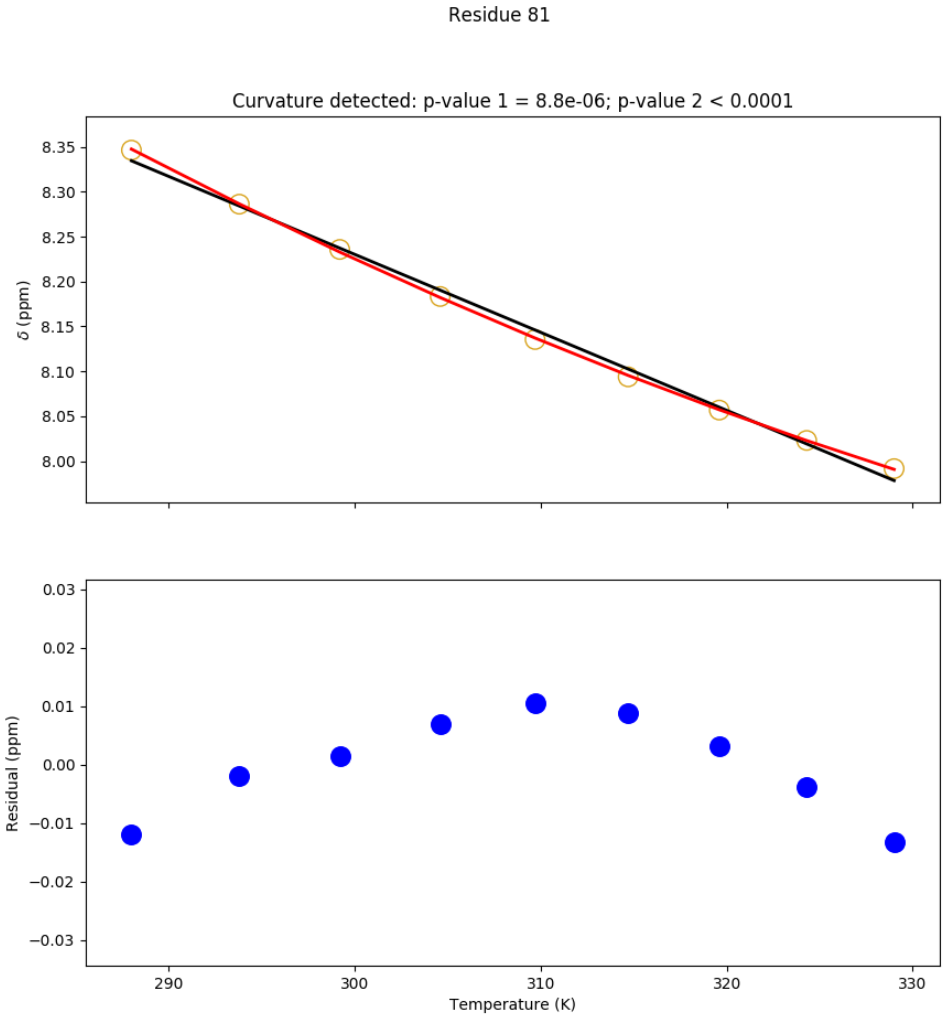


Figure I.22: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 81. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

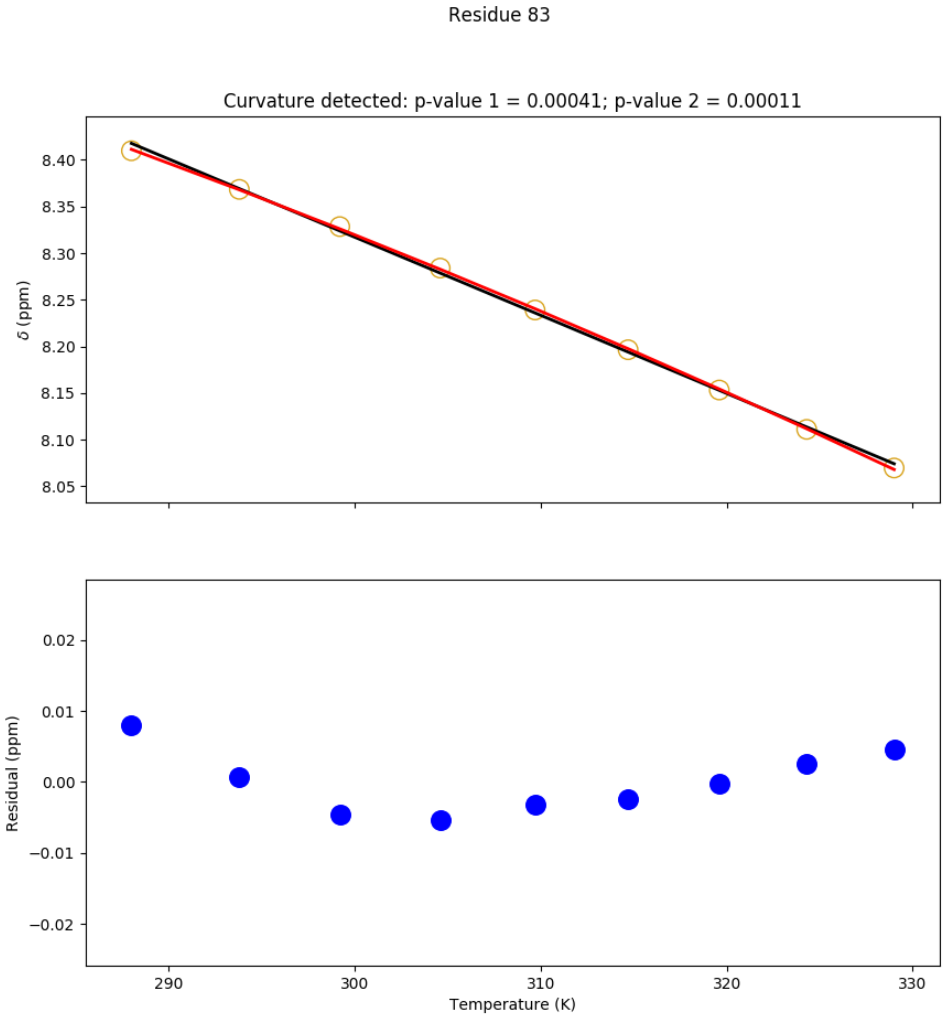


Figure I.23: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 83. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

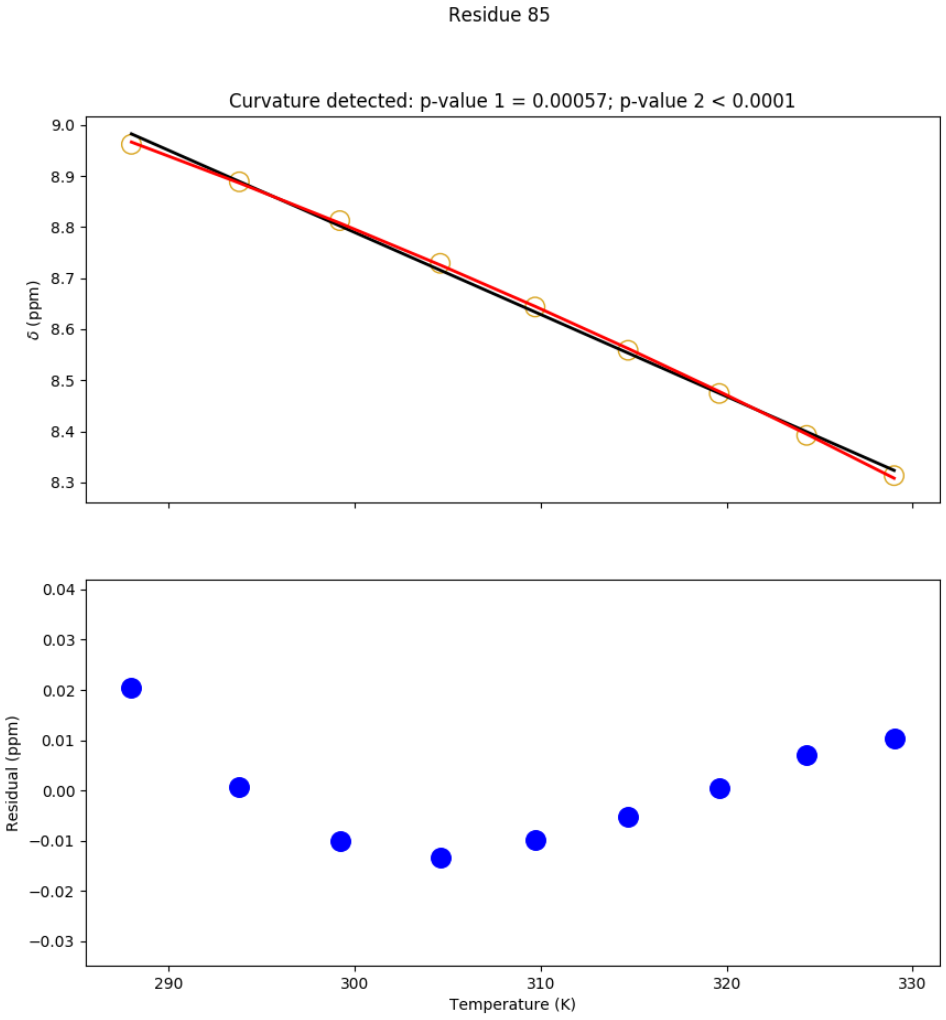


Figure I.24: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 85. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

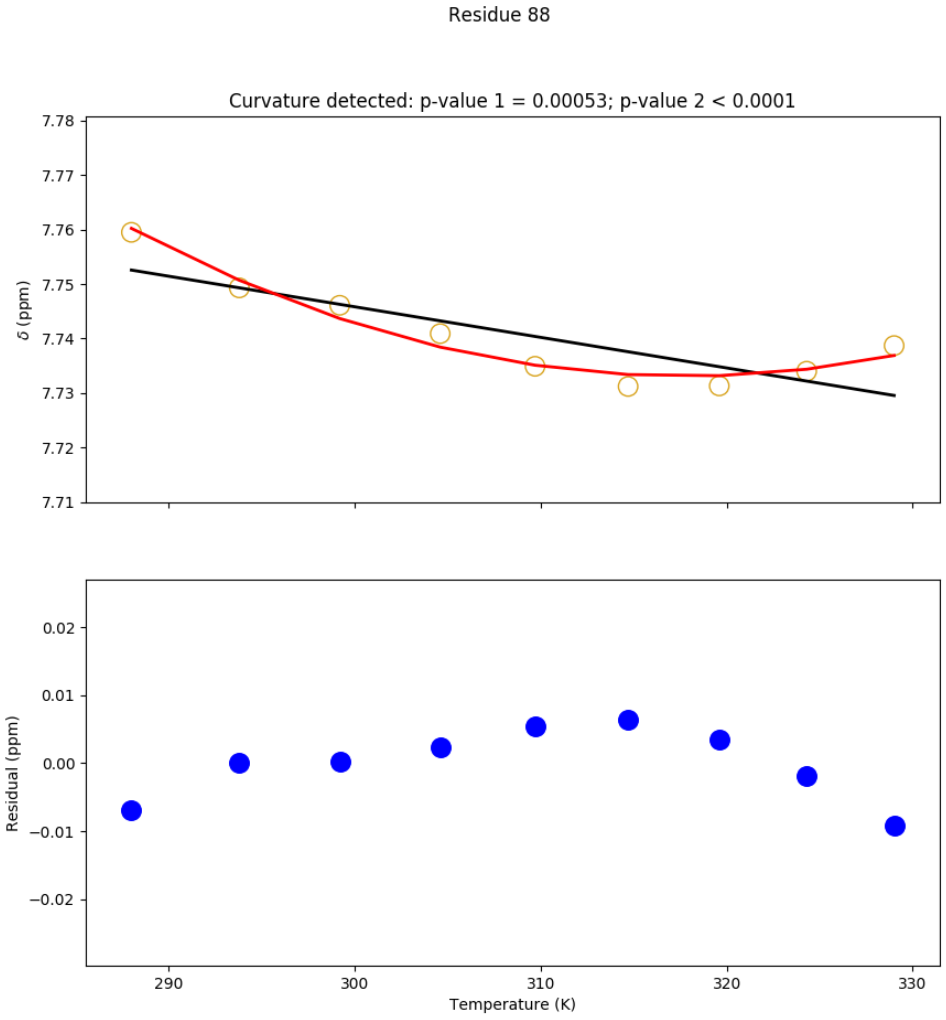


Figure I.25: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 88. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

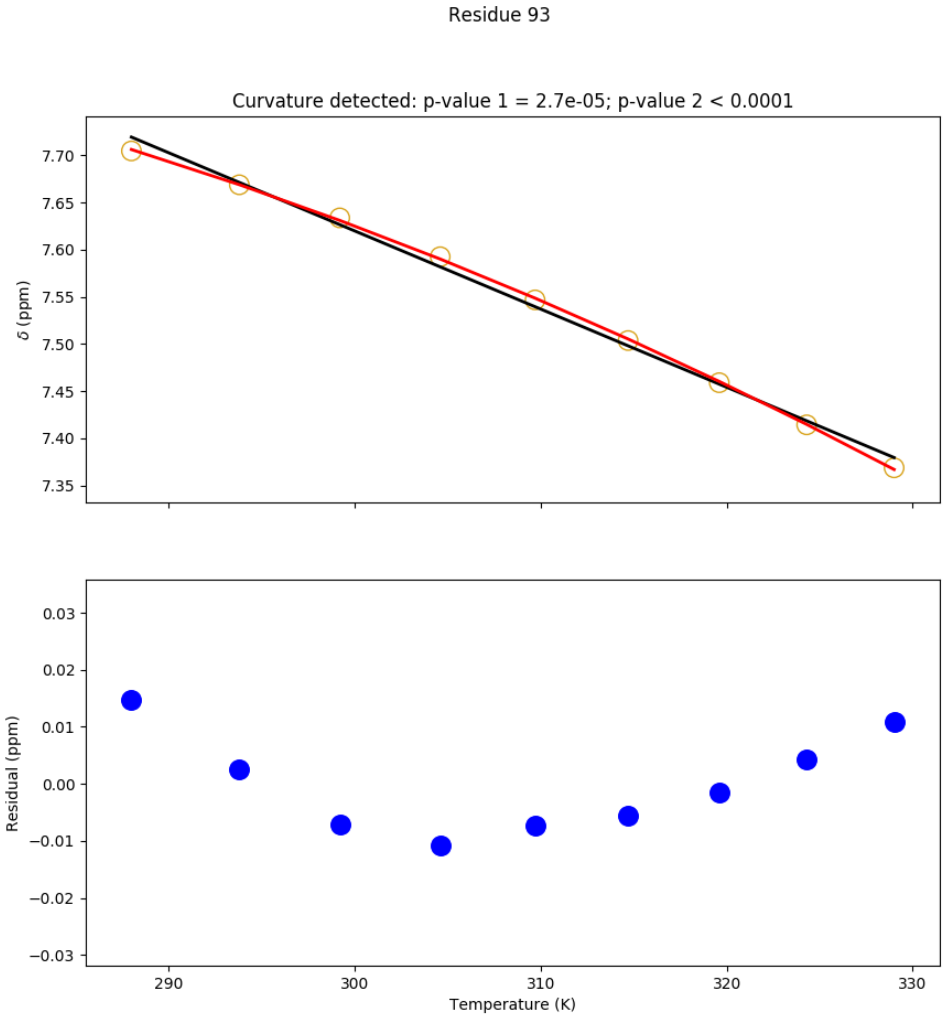


Figure I.26: Temperature dependence of the chemical shift of the pWT Adnectin amide proton from residue 93. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

I.4 pWT Adnectin Curvature Modelling

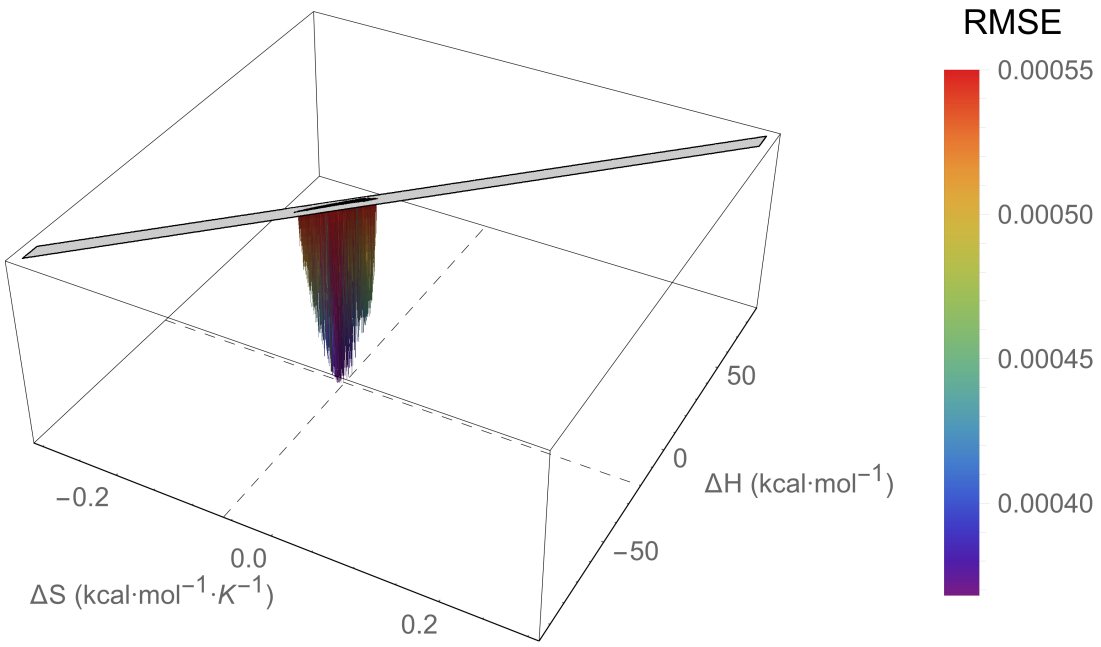


Figure I.27: An RMSE landscape from fitting pWT Adnectin residue 4 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

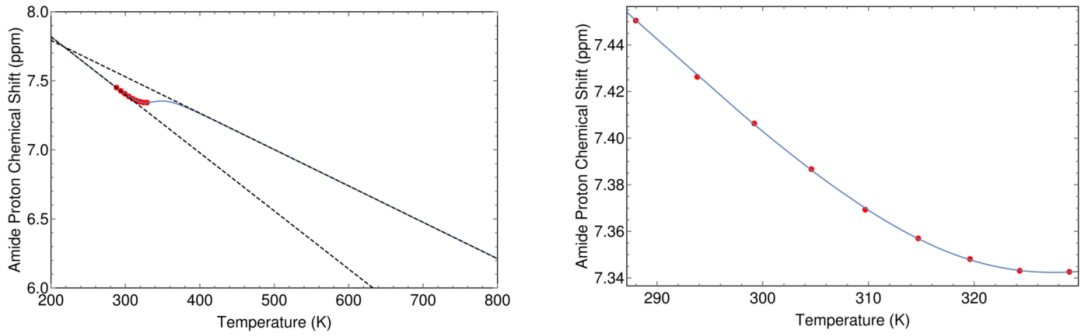


Figure I.28: The minimum RMSE fit of pWT Adnectin residue 4 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

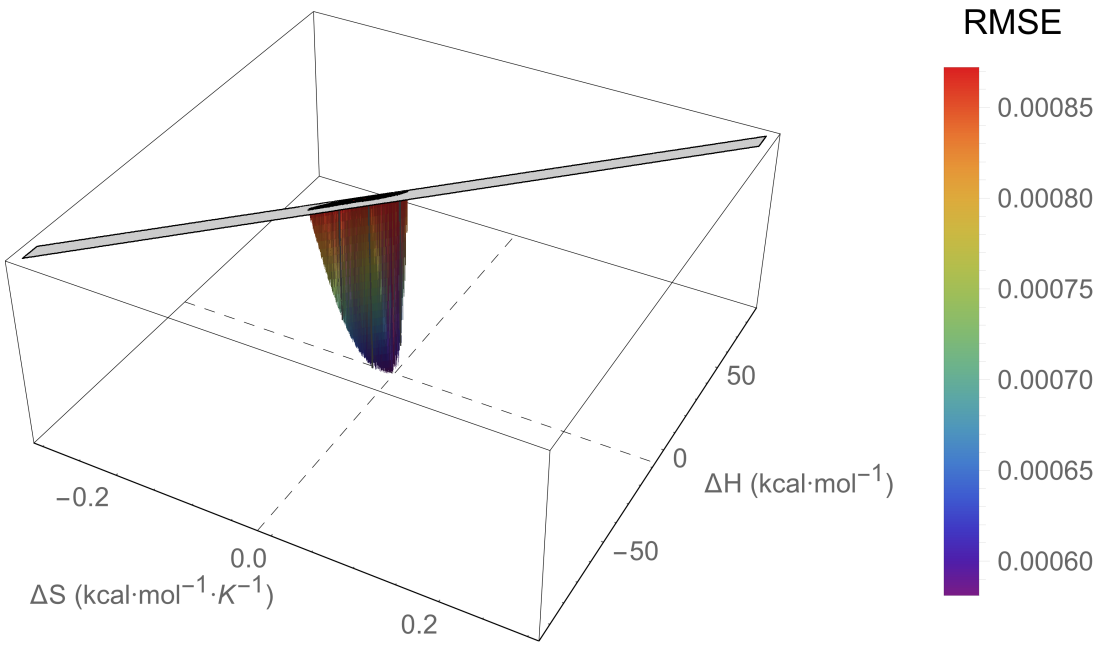


Figure I.29: An RMSE landscape from fitting pWT Adnectin residue 7 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

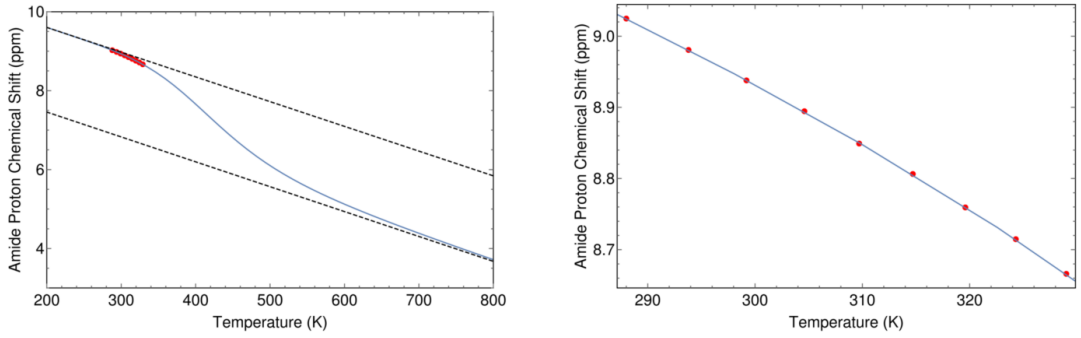


Figure I.30: The minimum RMSE fit of pWT Adnectin residue 7 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

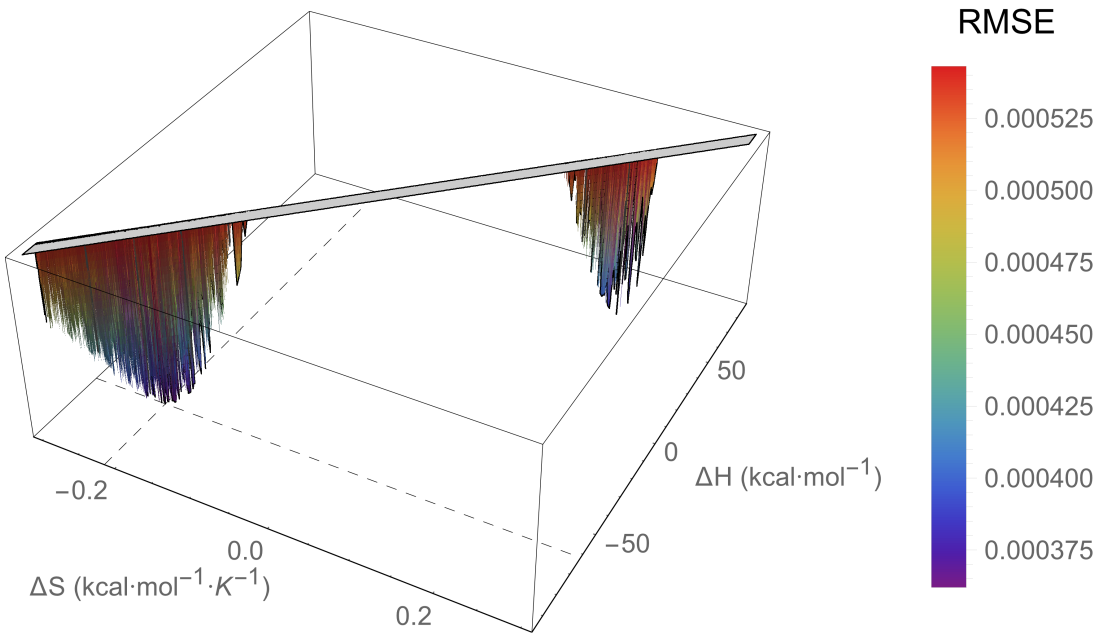


Figure I.31: An RMSE landscape from fitting pWT Adnectin residue 8 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

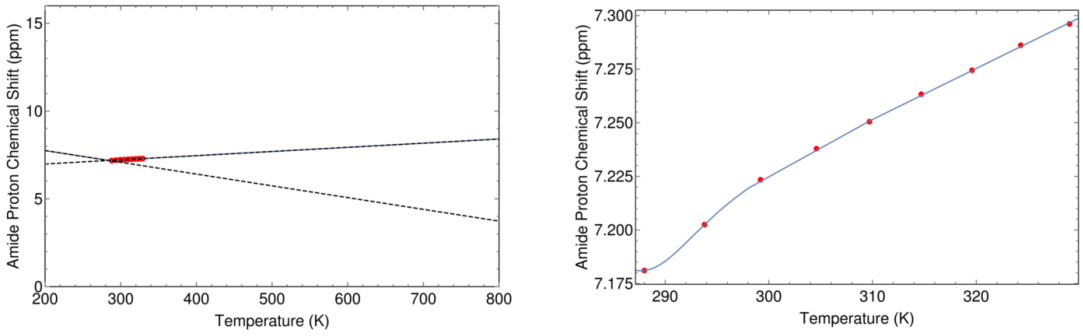


Figure I.32: The minimum RMSE fit of pWT Adnectin residue 8 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

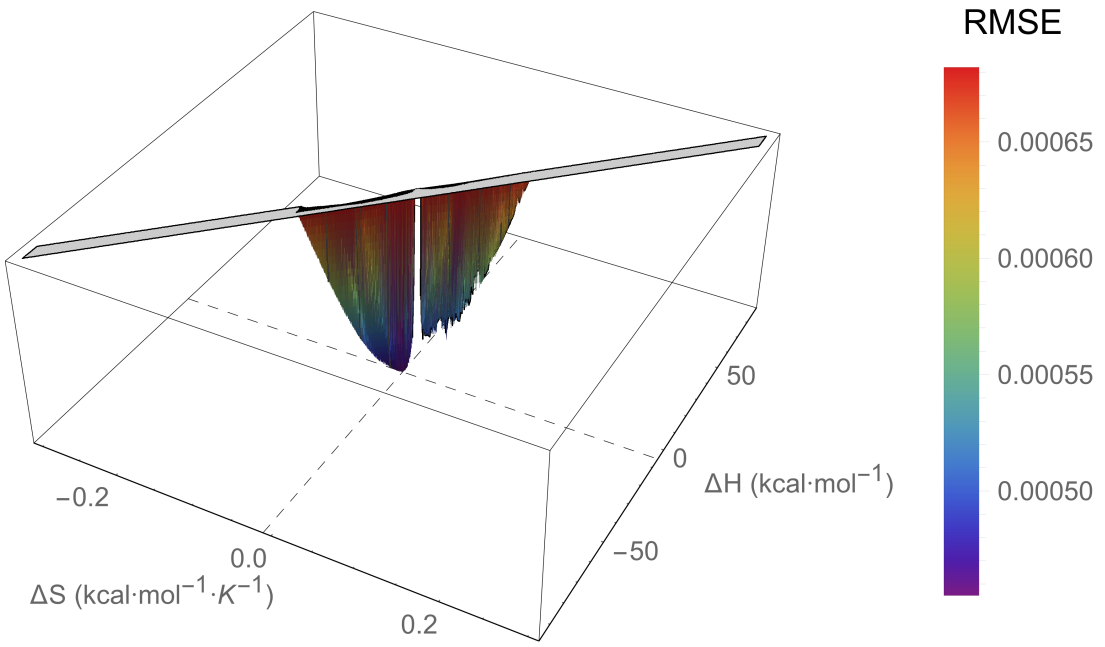


Figure I.33: An RMSE landscape from fitting pWT Adnectin residue 9 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE$_{\text{min}}$) are indicated in gray.

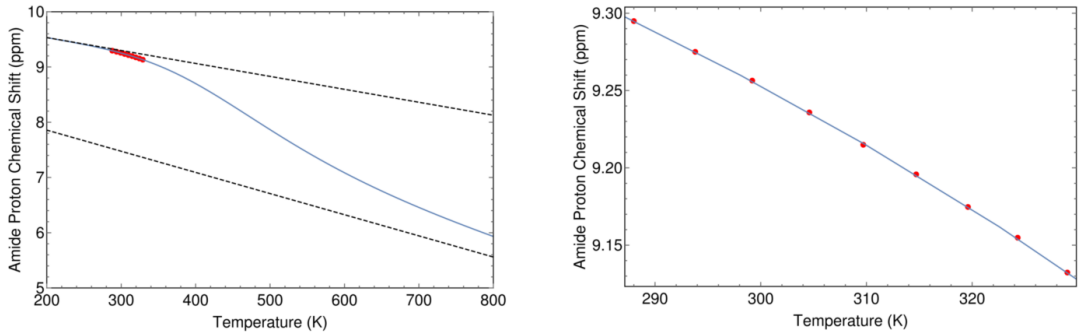


Figure I.34: The minimum RMSE fit of pWT Adnectin residue 9 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

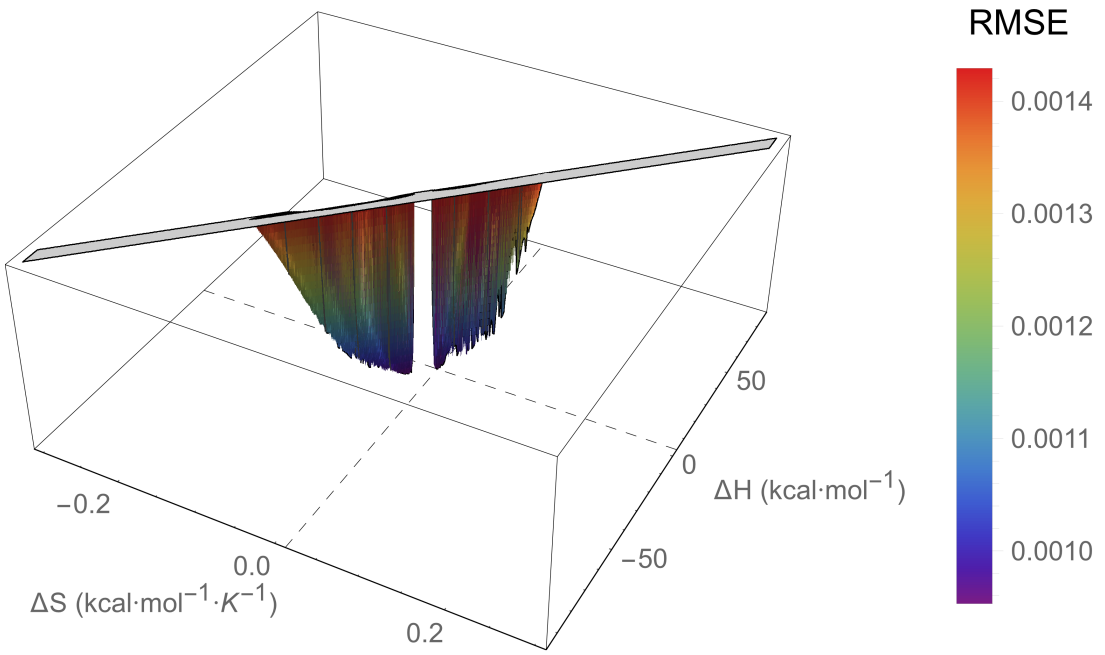


Figure I.35: An RMSE landscape from fitting pWT Adnectin residue 10 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

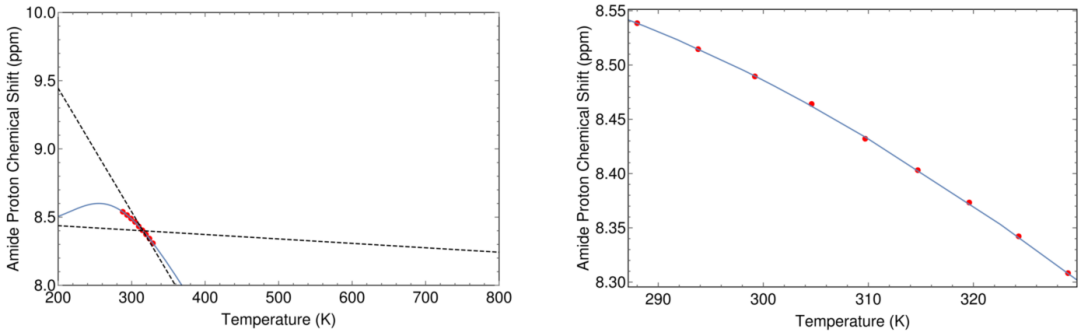


Figure I.36: The minimum RMSE fit of pWT Adnectin residue 10 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

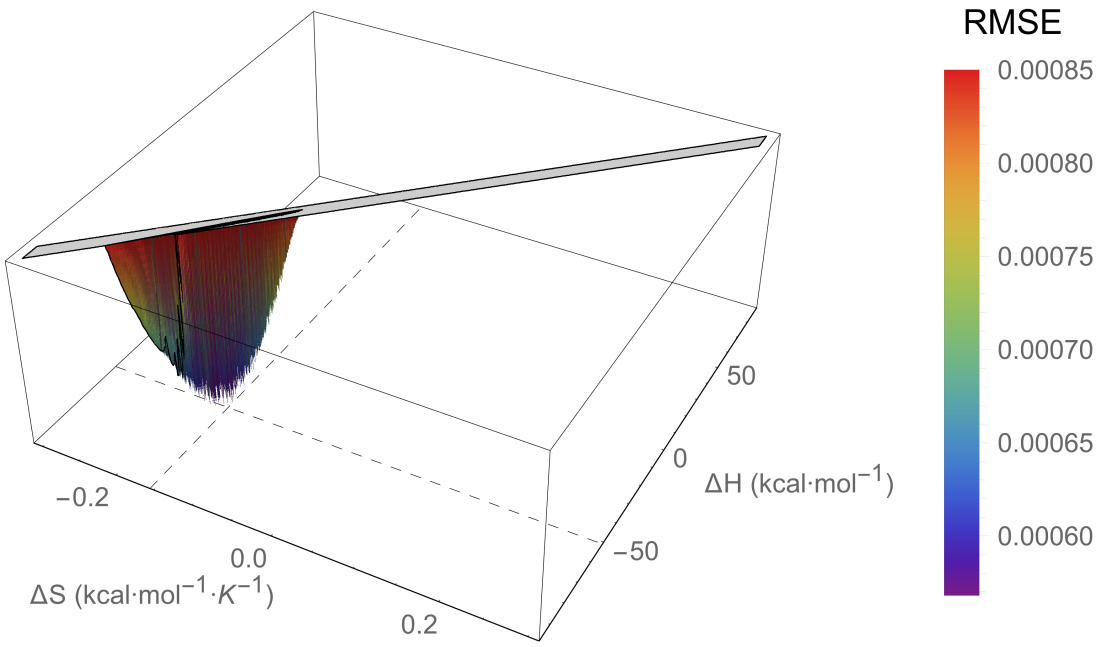


Figure I.37: An RMSE landscape from fitting pWT Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

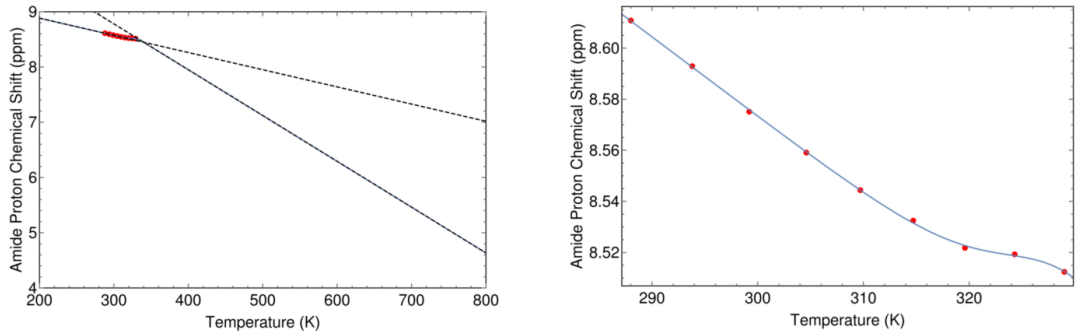


Figure I.38: The minimum RMSE fit of pWT Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

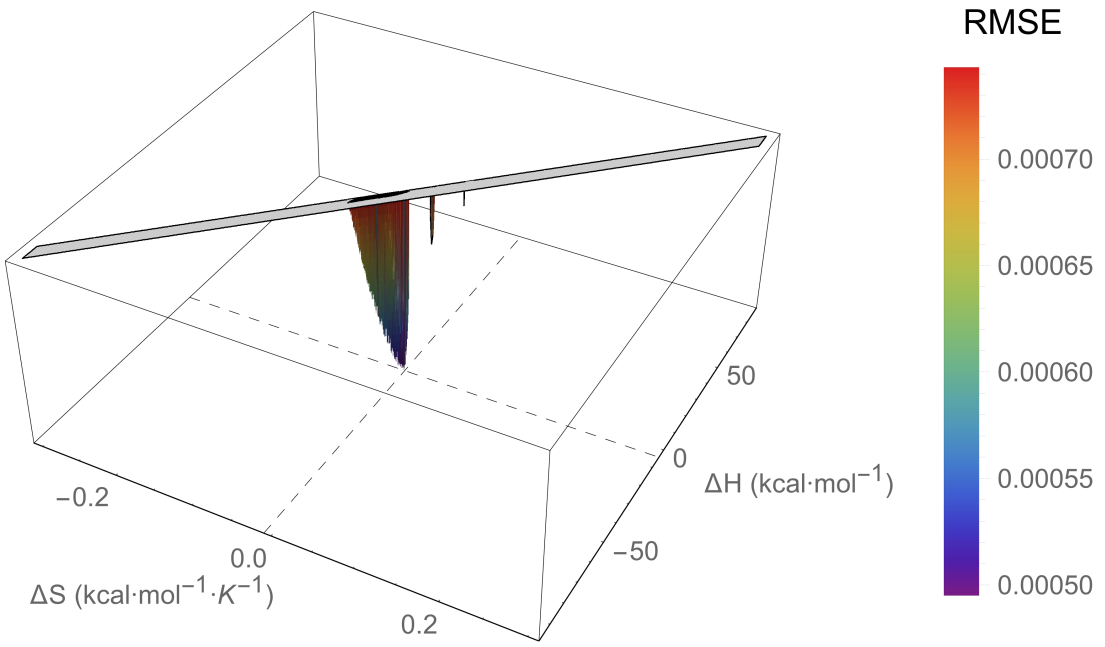


Figure I.39: An RMSE landscape from fitting pWT Adnectin residue 24 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

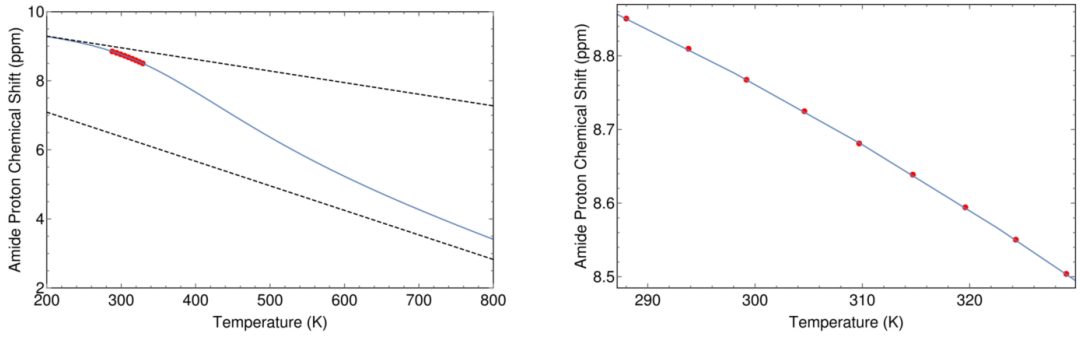


Figure I.40: The minimum RMSE fit of pWT Adnectin residue 24 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

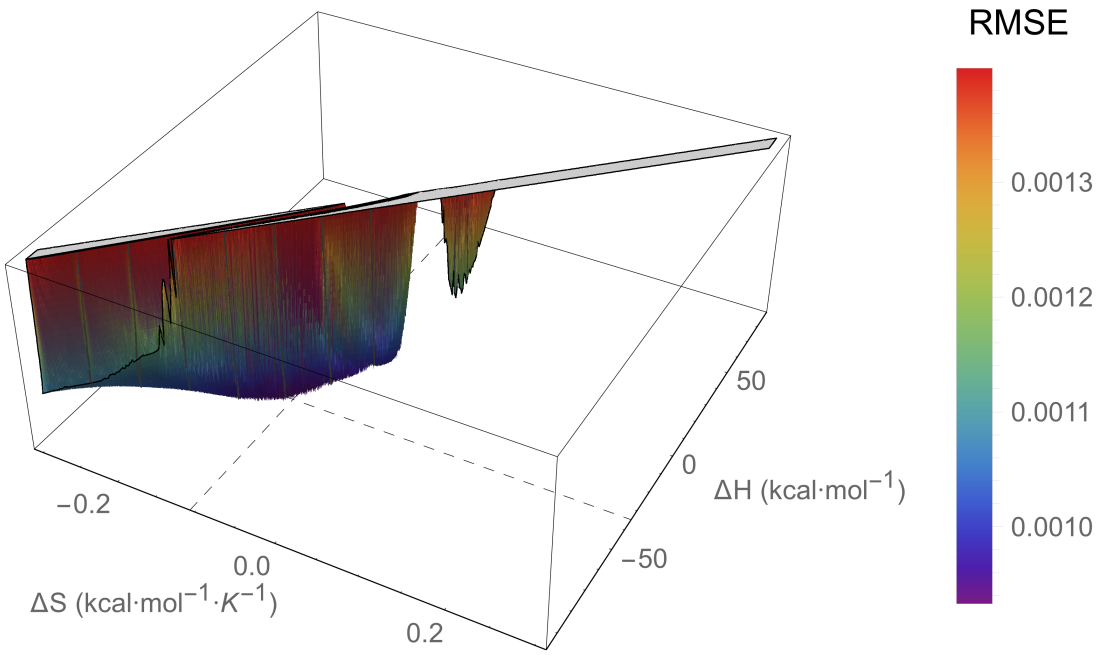


Figure I.41: An RMSE landscape from fitting pWT Adnectin residue 26 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

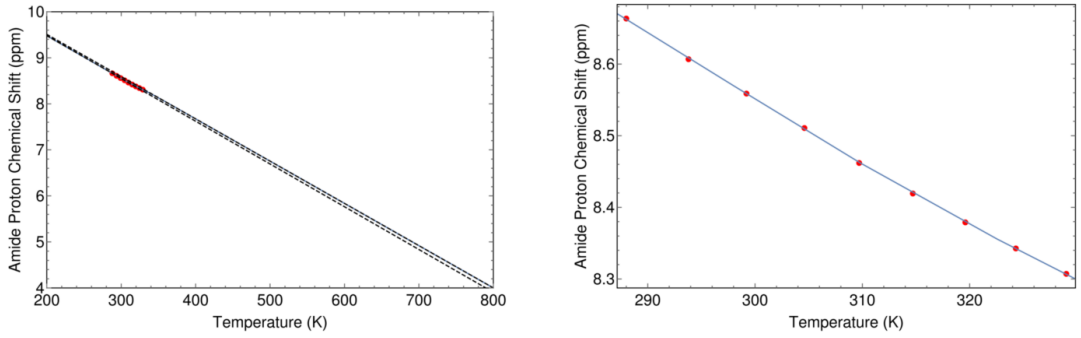


Figure I.42: The minimum RMSE fit of pWT Adnectin residue 26 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

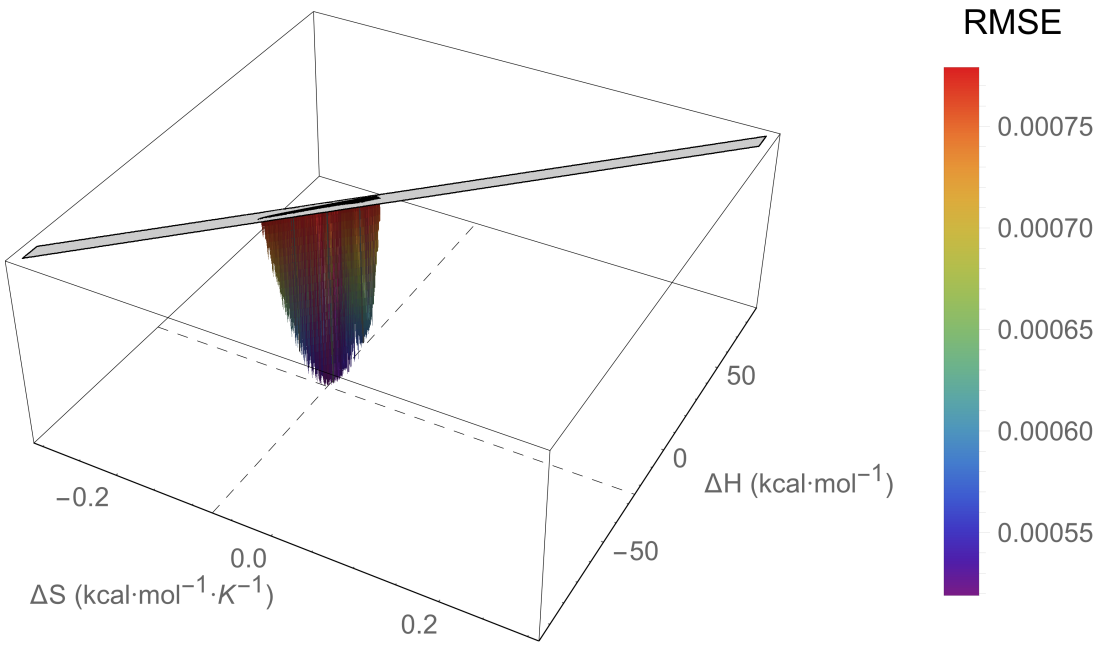


Figure I.43: An RMSE landscape from fitting pWT Adnectin residue 27 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

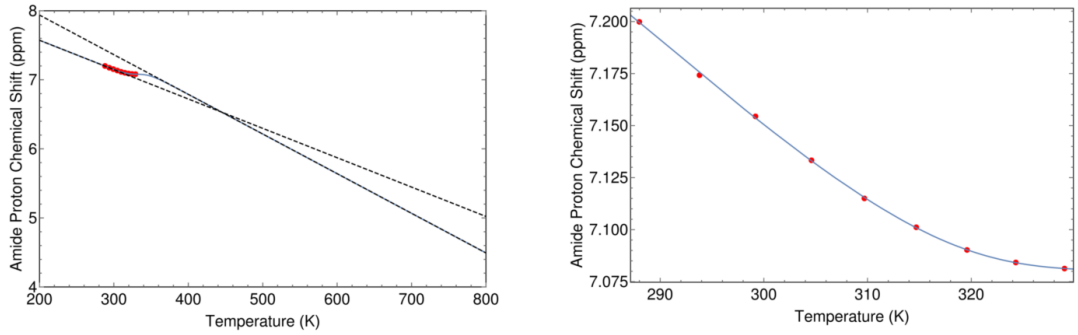


Figure I.44: The minimum RMSE fit of pWT Adnectin residue 27 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

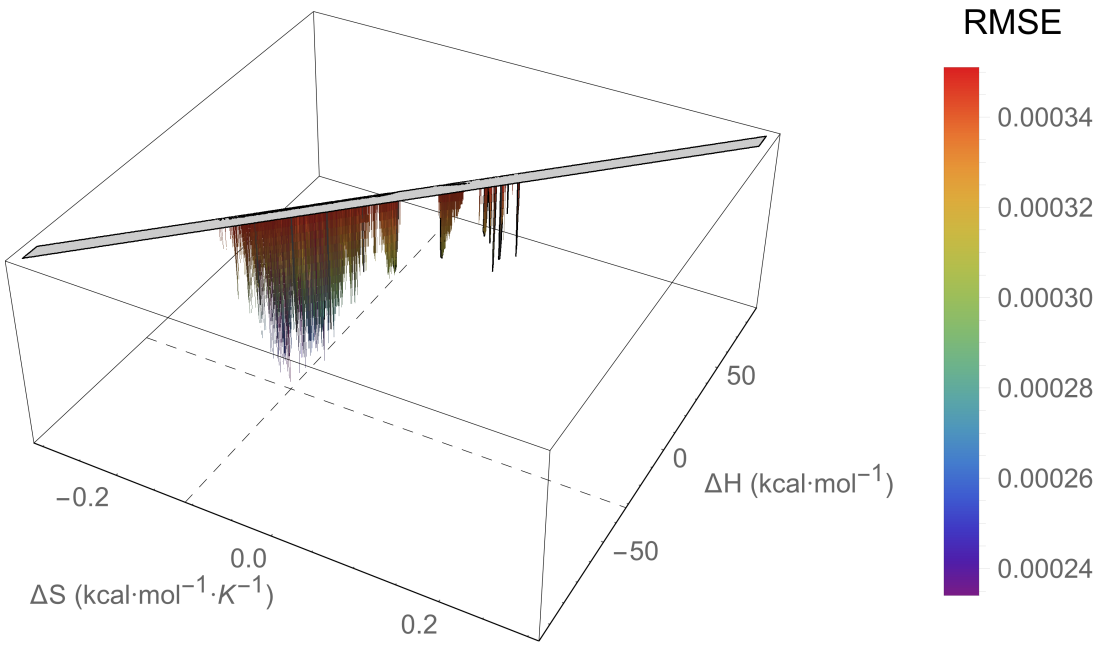


Figure I.45: An RMSE landscape from fitting pWT Adnectin residue 28 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

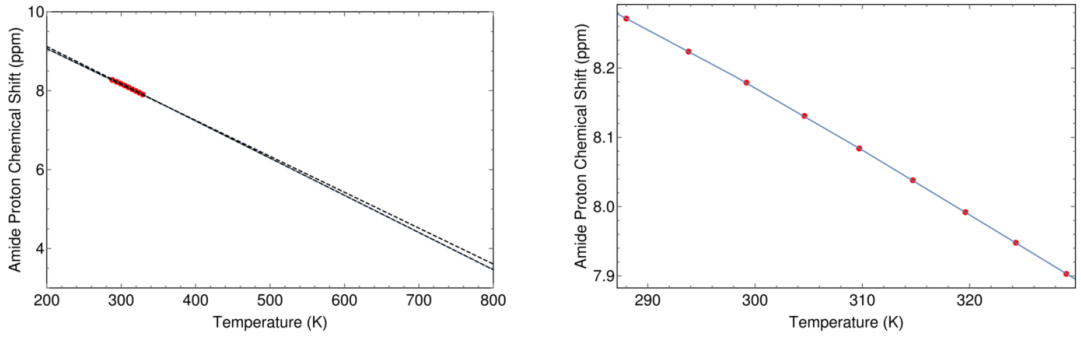


Figure I.46: The minimum RMSE fit of pWT Adnectin residue 28 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

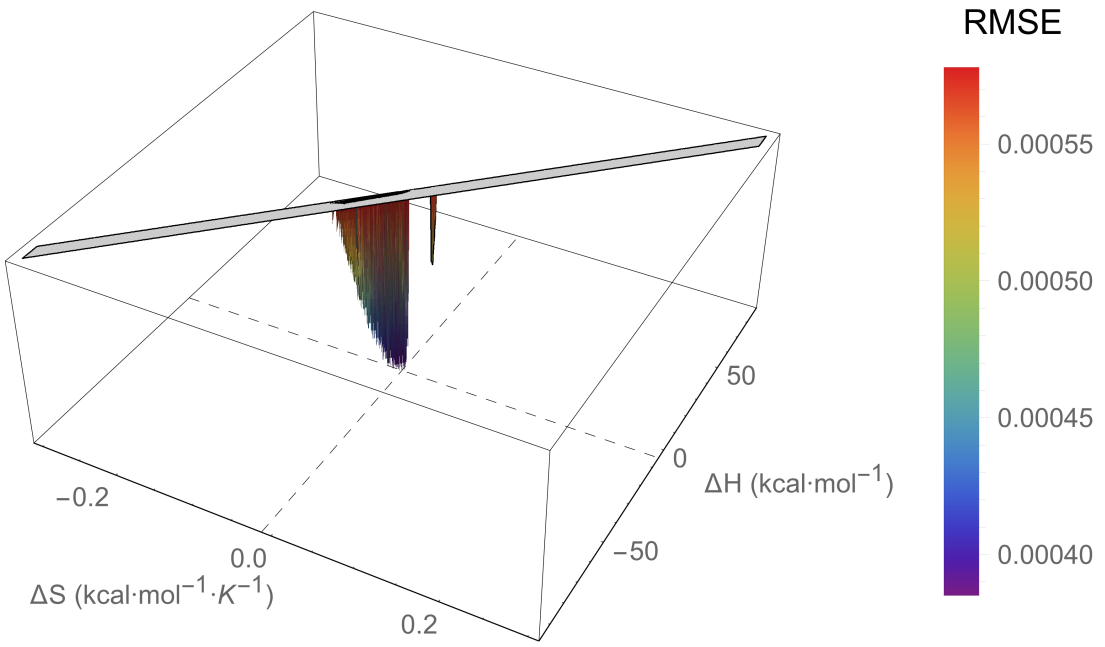


Figure I.47: An RMSE landscape from fitting pWT Adnectin residue 29 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE$_{\text{min}}$) are indicated in gray.

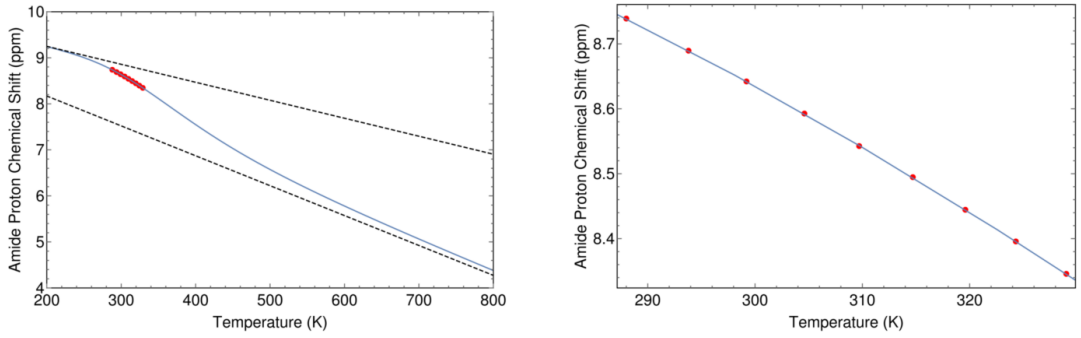


Figure I.48: The minimum RMSE fit of pWT Adnectin residue 29 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

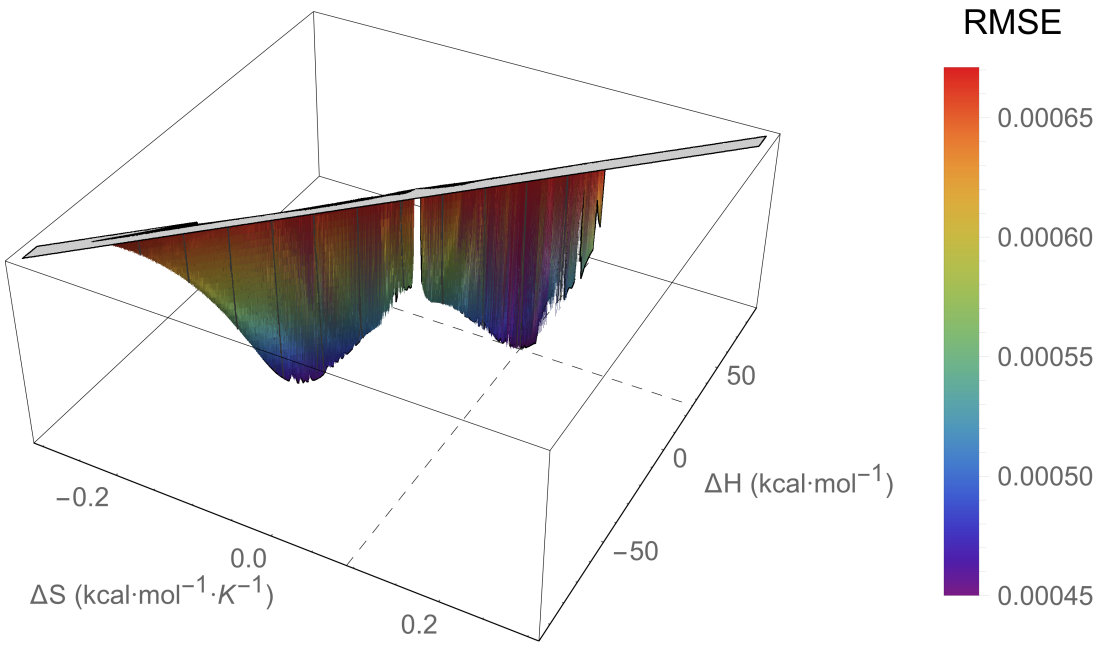


Figure I.49: An RMSE landscape from fitting pWT Adnectin residue 36 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

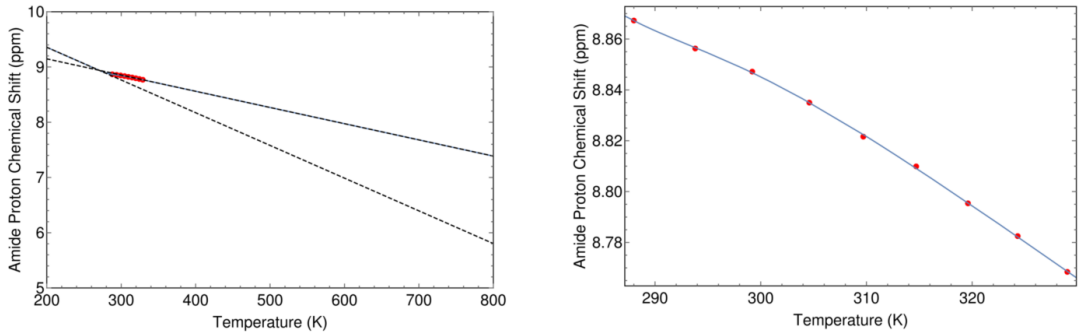


Figure I.50: The minimum RMSE fit of pWT Adnectin residue 36 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

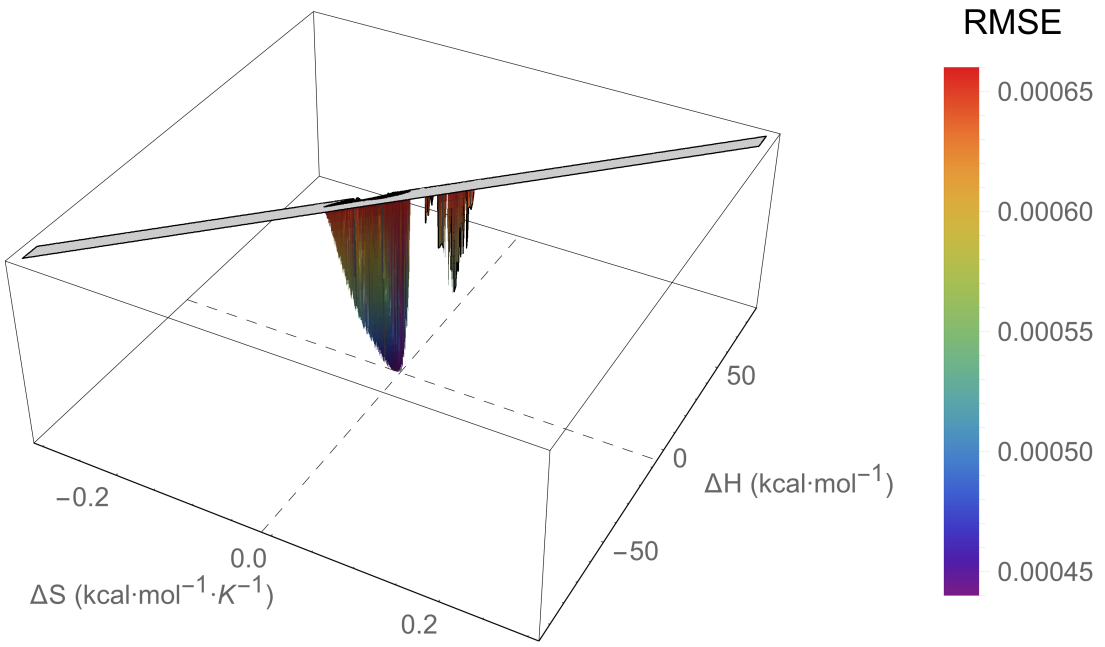


Figure I.51: An RMSE landscape from fitting pWT Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

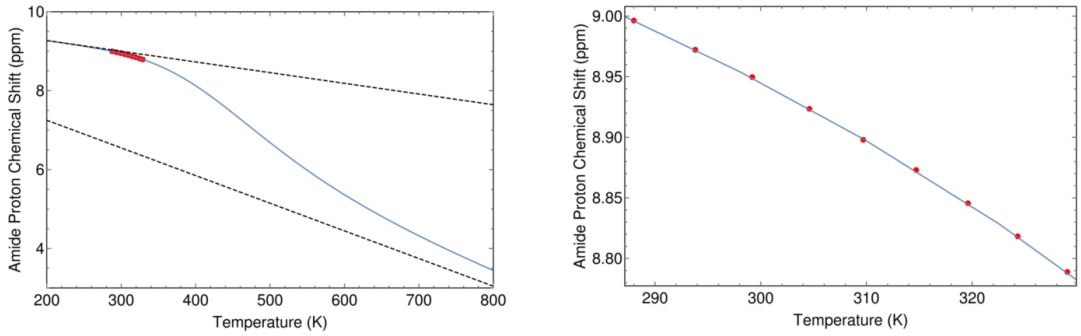


Figure I.52: The minimum RMSE fit of pWT Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

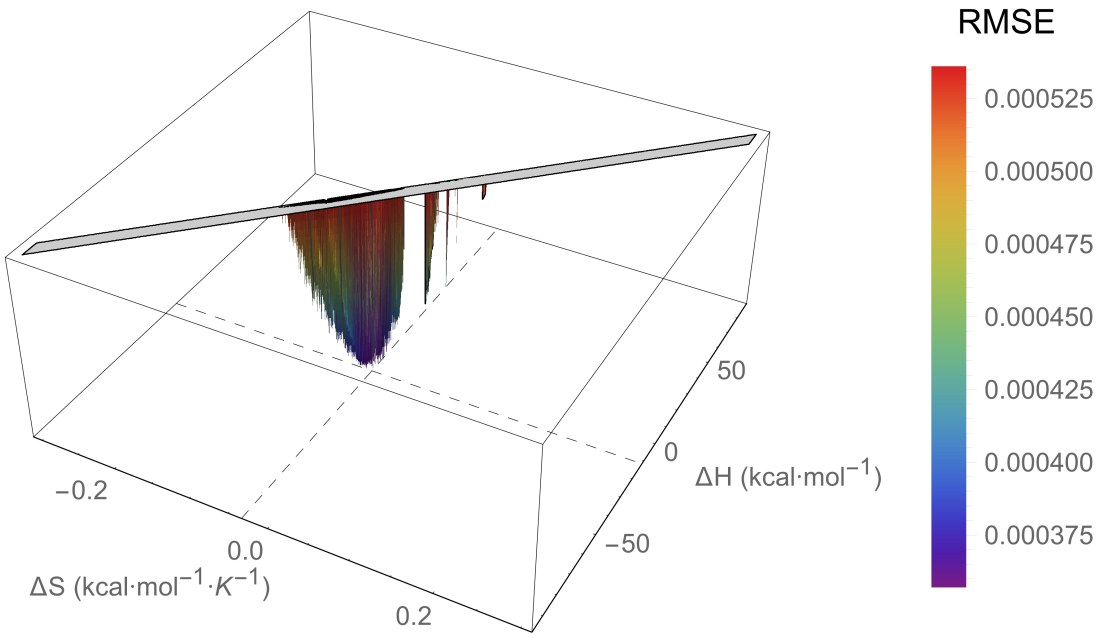


Figure I.53: An RMSE landscape from fitting pWT Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

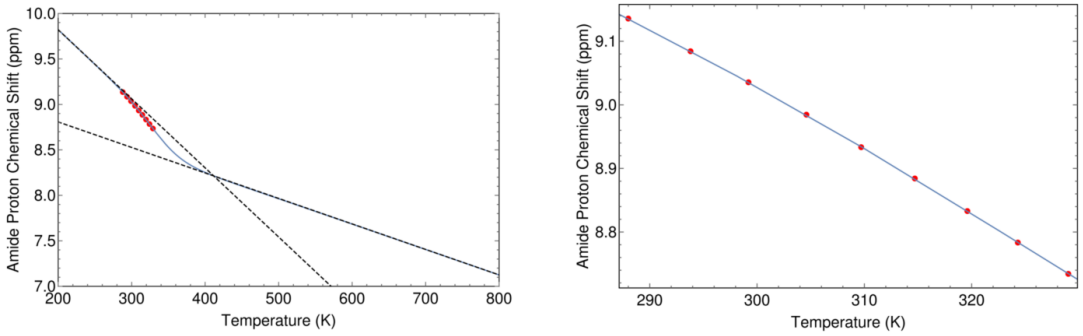


Figure I.54: The minimum RMSE fit of pWT Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

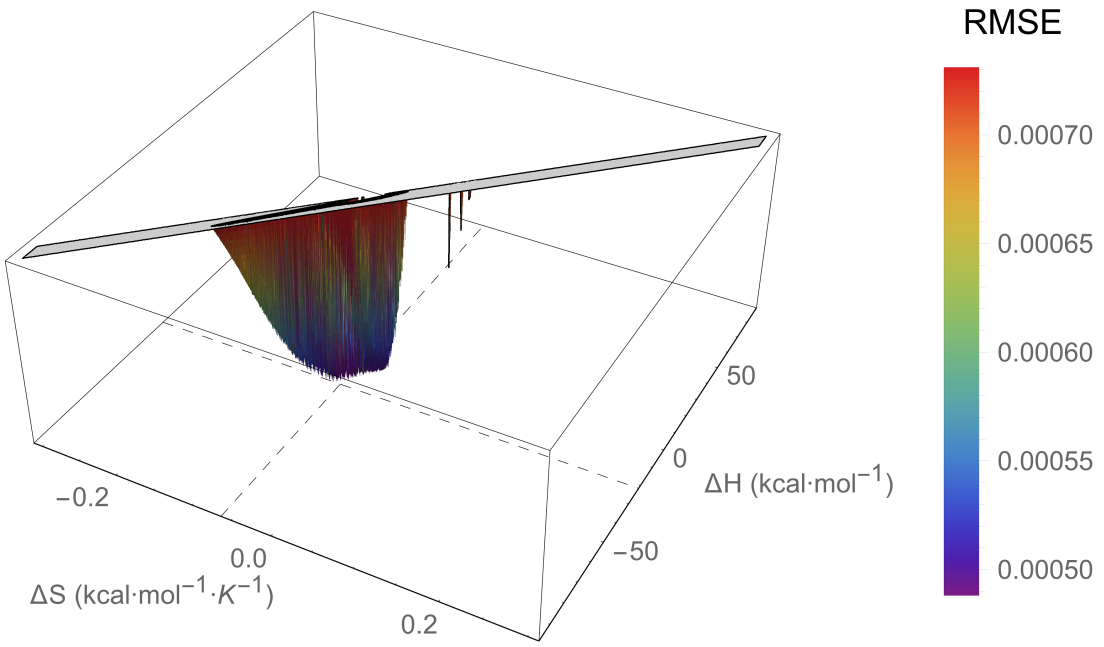


Figure I.55: An RMSE landscape from fitting pWT Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

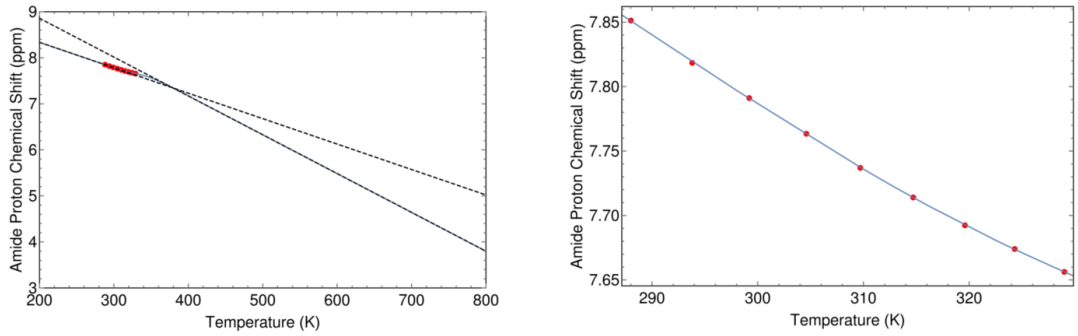


Figure I.56: The minimum RMSE fit of pWT Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

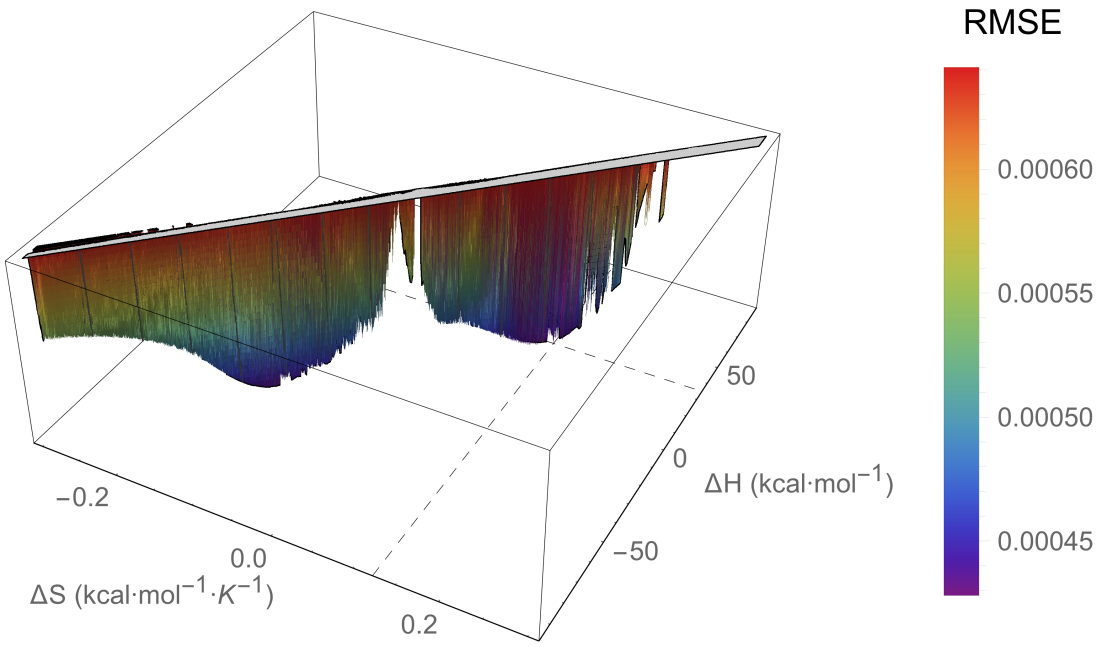


Figure I.57: An RMSE landscape from fitting pWT Adnectin residue 52 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

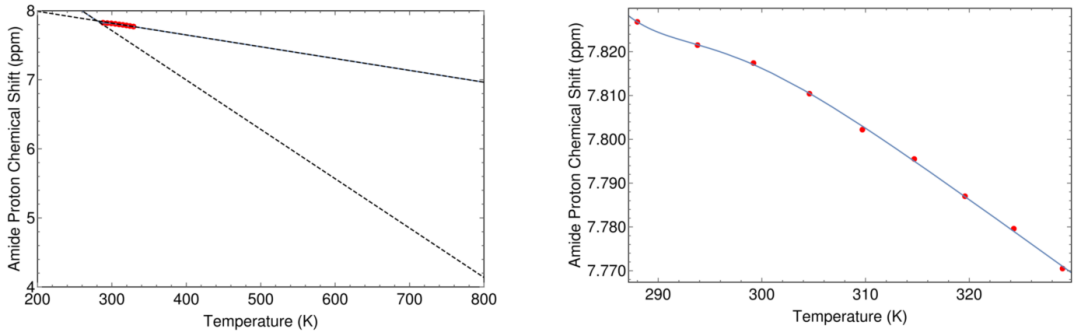


Figure I.58: The minimum RMSE fit of pWT Adnectin residue 52 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

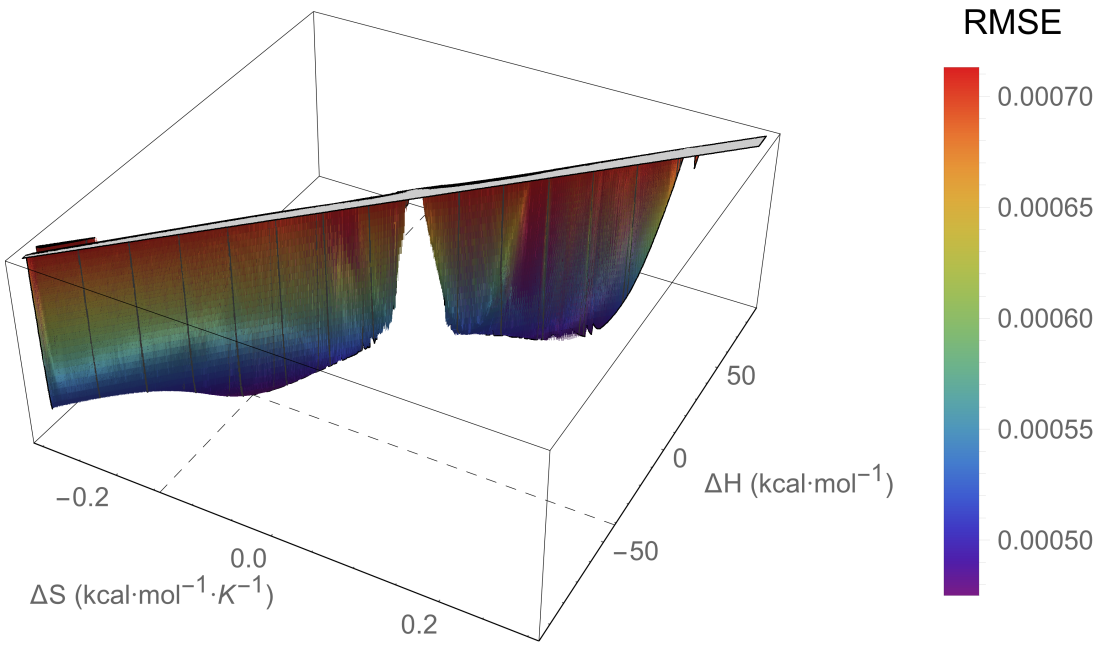


Figure I.59: An RMSE landscape from fitting pWT Adnectin residue 59 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

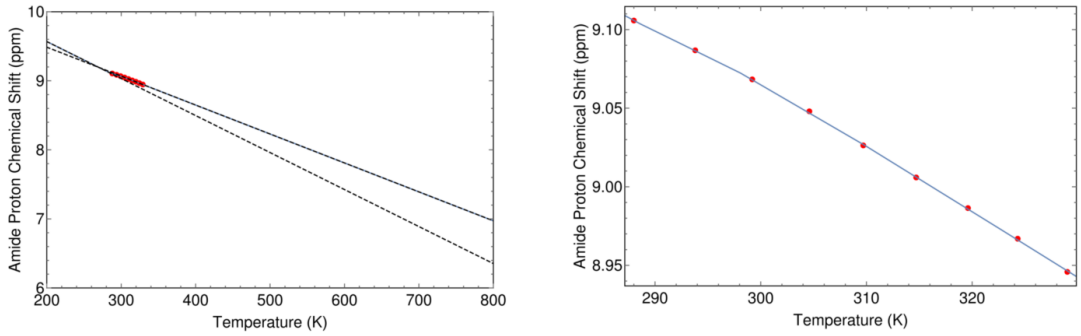


Figure I.60: The minimum RMSE fit of pWT Adnectin residue 59 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

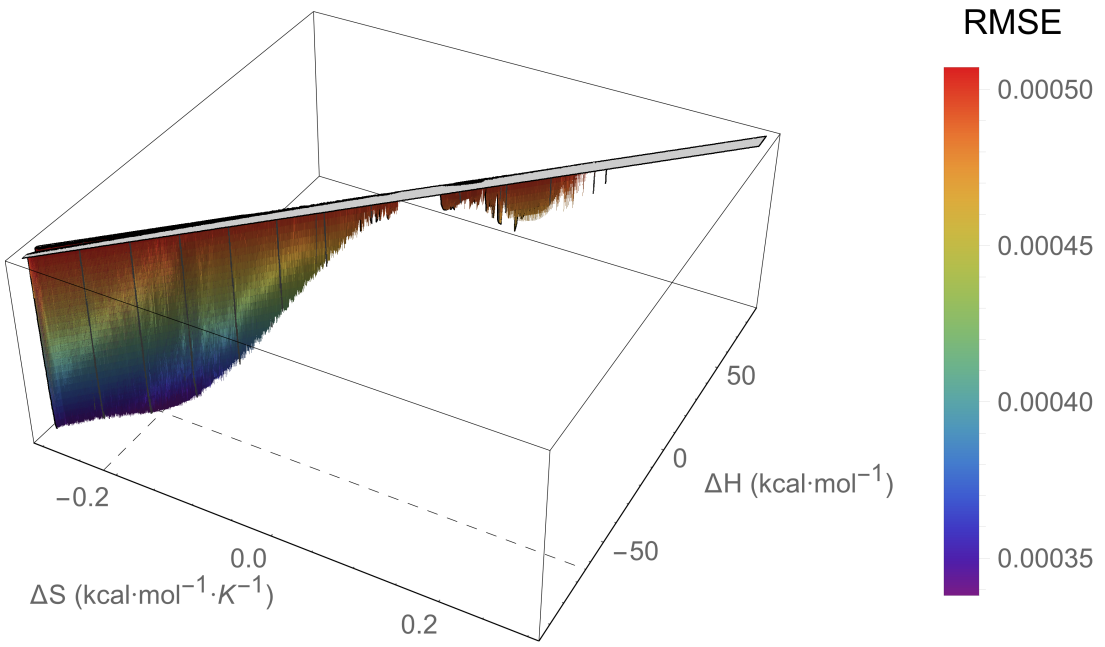


Figure I.61: An RMSE landscape from fitting pWT Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

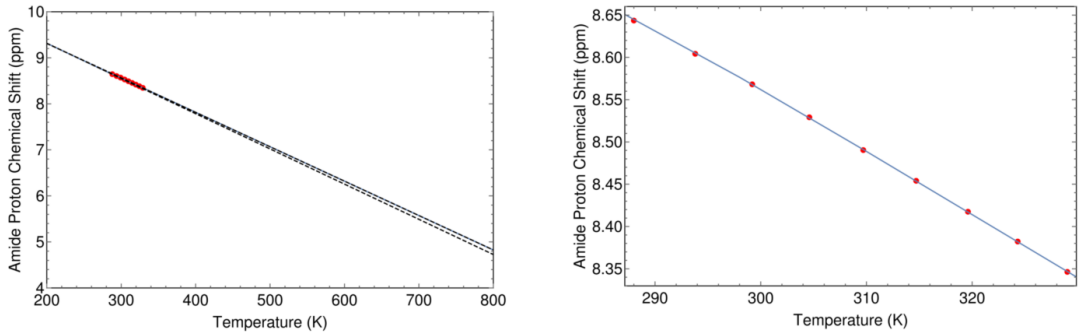


Figure I.62: The minimum RMSE fit of pWT Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

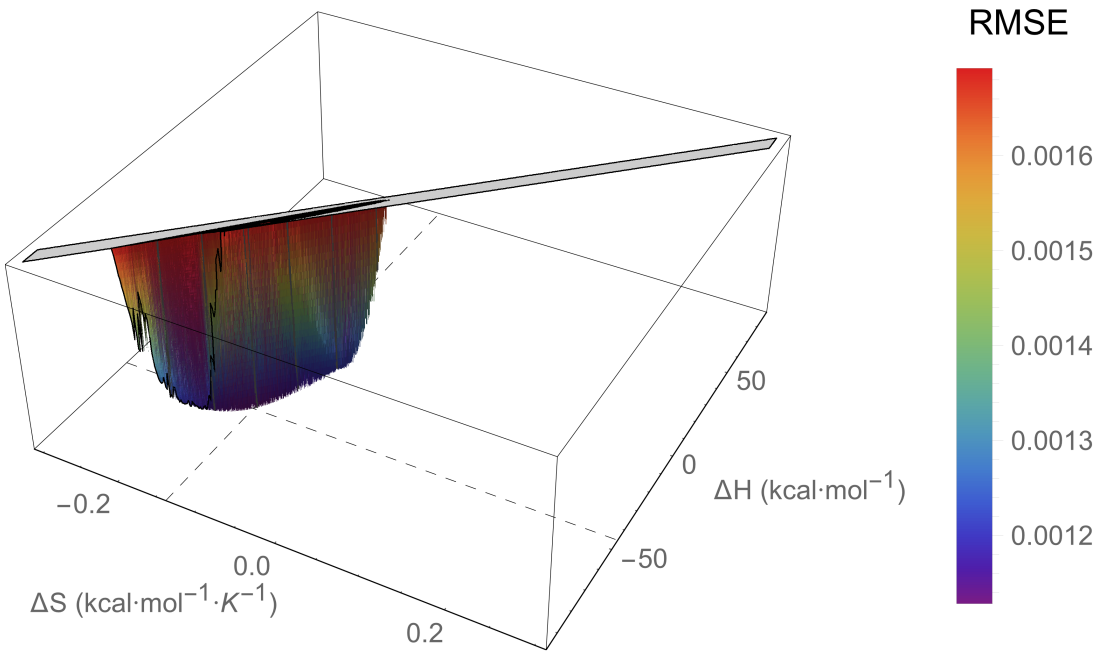


Figure I.63: An RMSE landscape from fitting pWT Adnectin residue 65 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

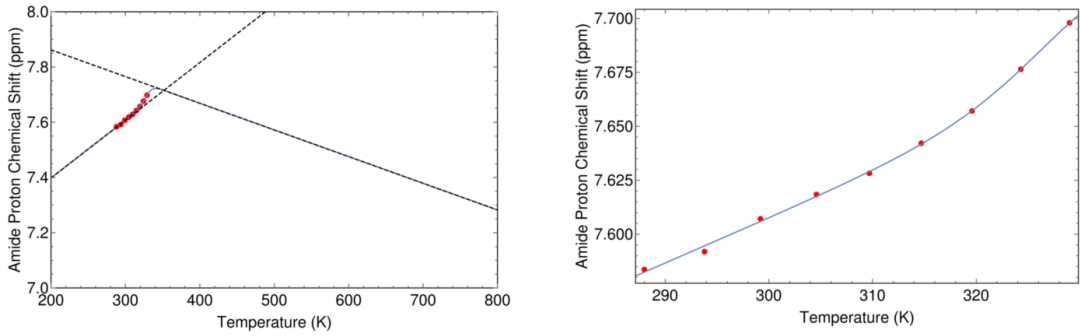


Figure I.64: The minimum RMSE fit of pWT Adnectin residue 65 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

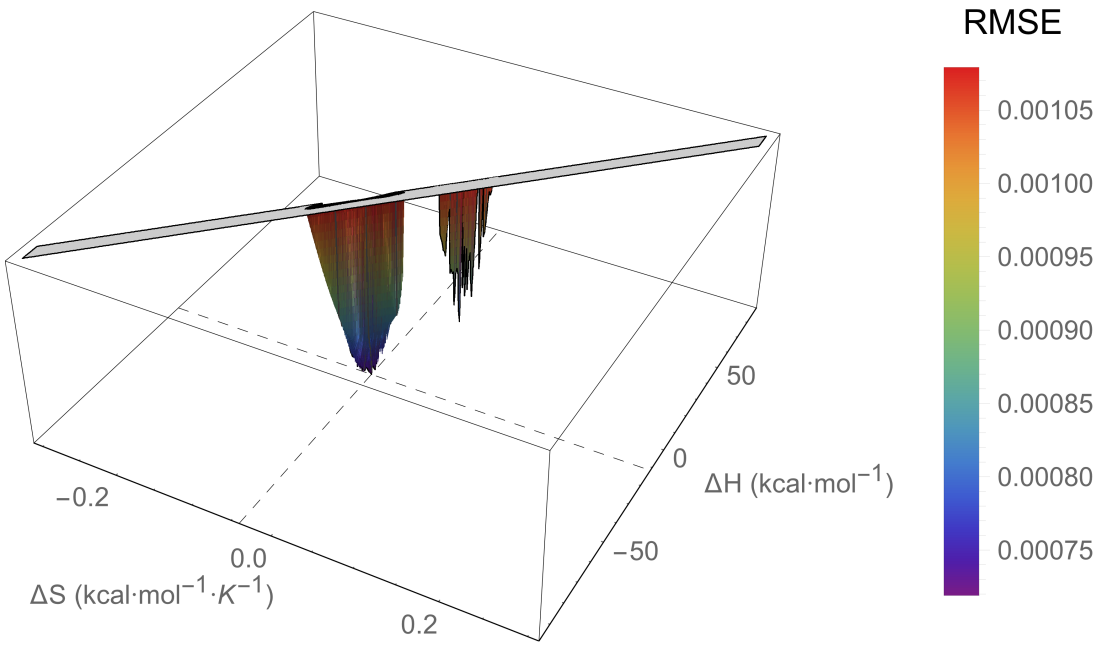


Figure I.65: An RMSE landscape from fitting pWT Adnectin residue 74 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

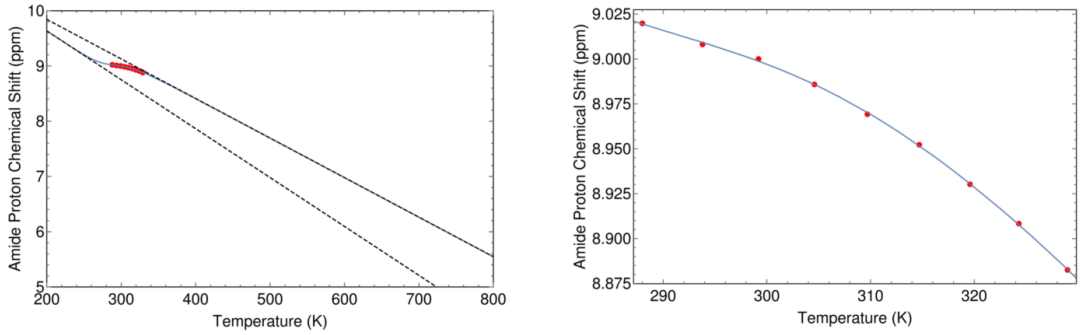


Figure I.66: The minimum RMSE fit of pWT Adnectin residue 74 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

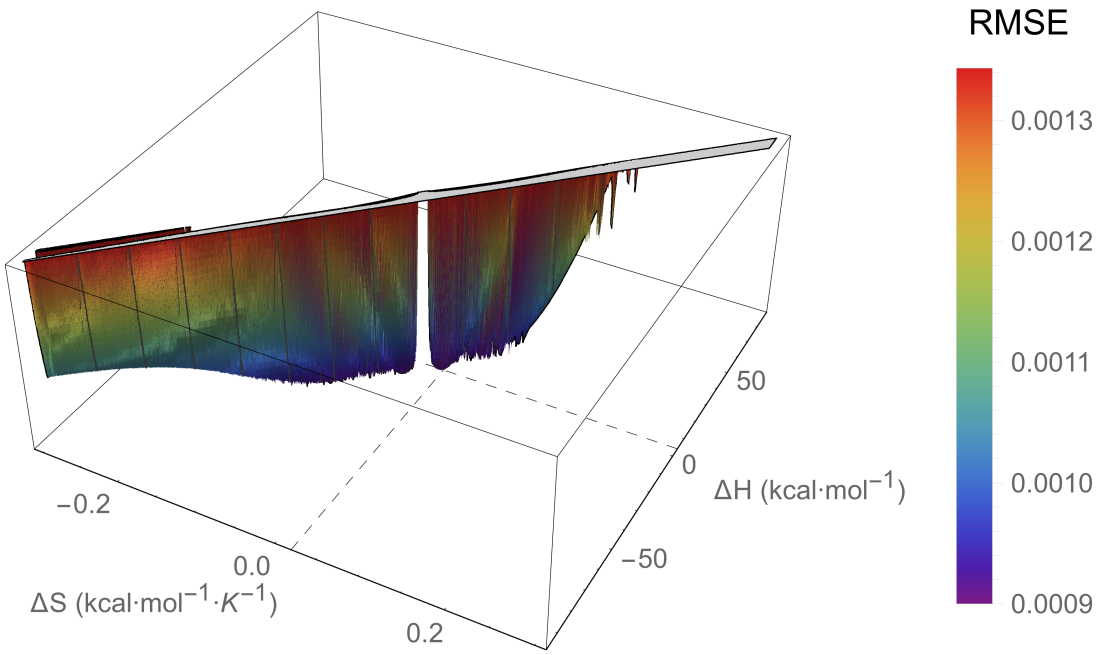


Figure I.67: An RMSE landscape from fitting pWT Adnectin residue 80 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

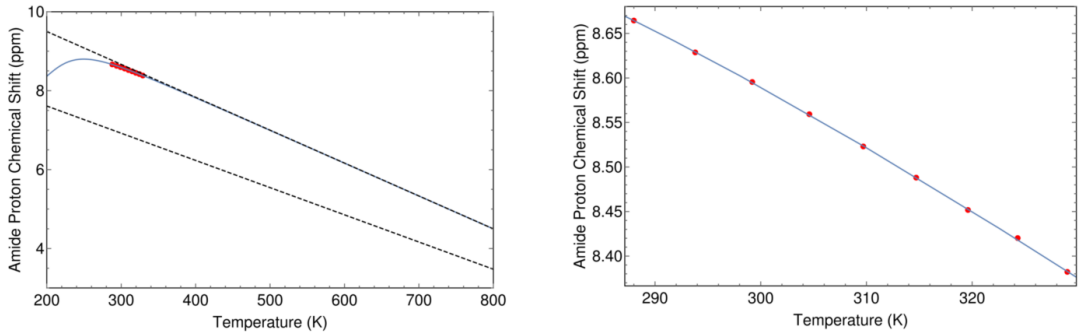


Figure I.68: The minimum RMSE fit of pWT Adnectin residue 80 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

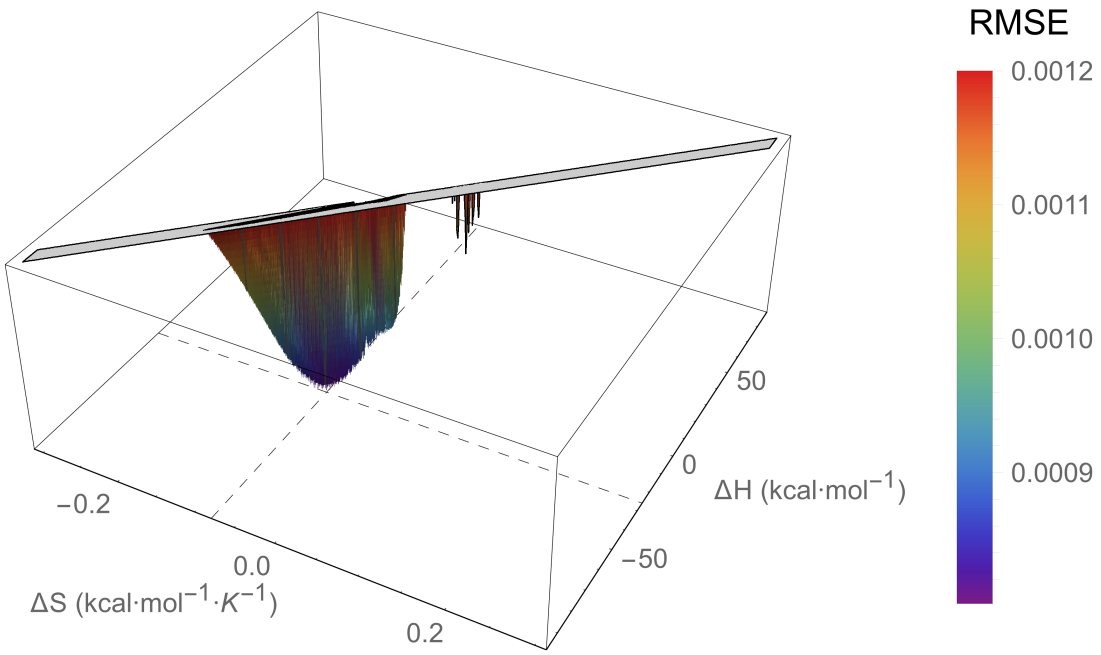


Figure I.69: An RMSE landscape from fitting pWT Adnectin residue 81 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

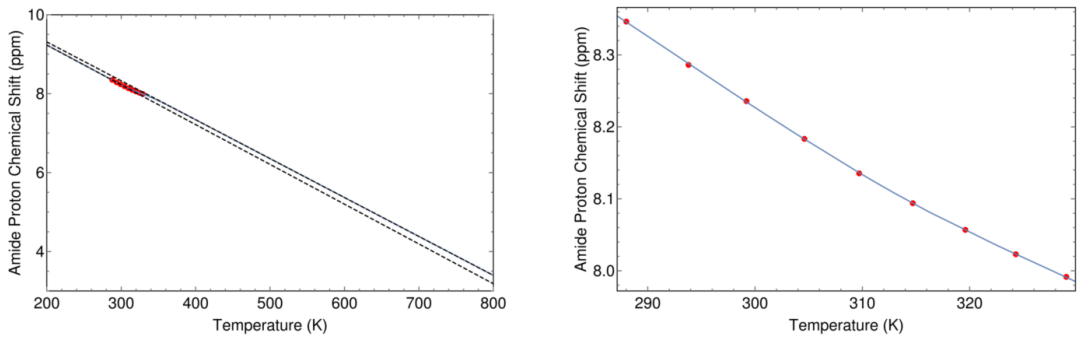


Figure I.70: The minimum RMSE fit of pWT Adnectin residue 81 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

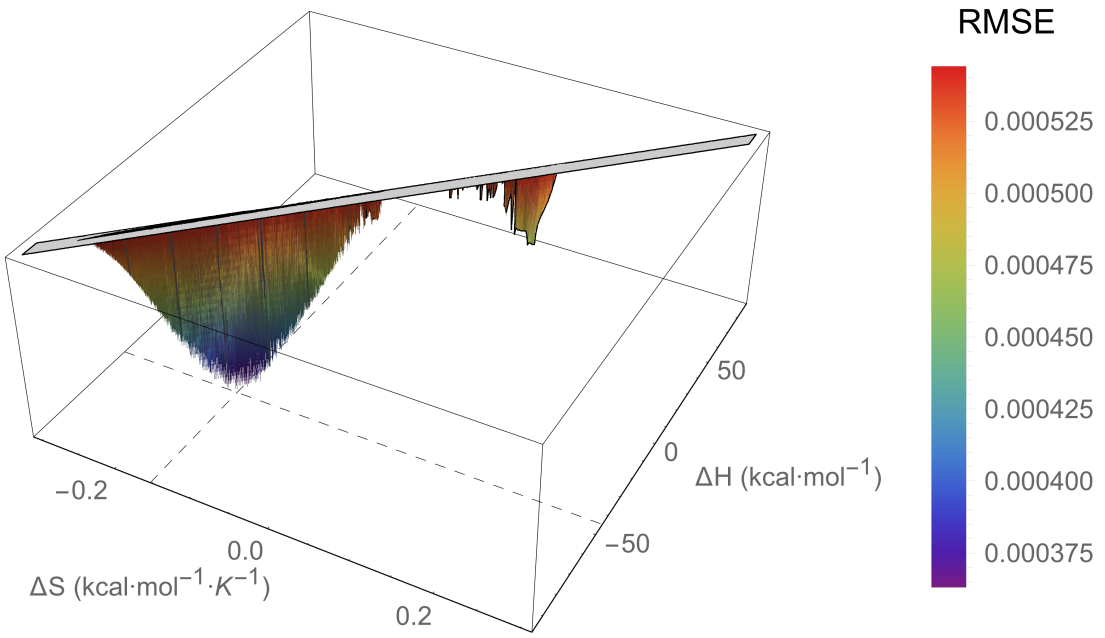


Figure I.71: An RMSE landscape from fitting pWT Adnectin residue 83 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

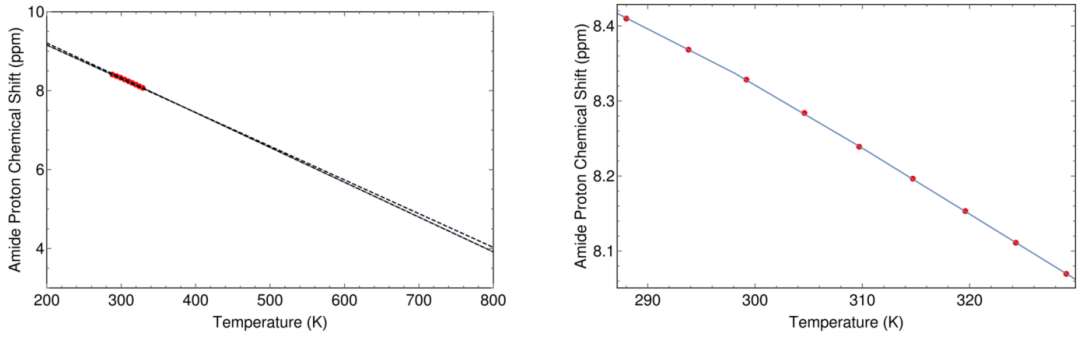


Figure I.72: The minimum RMSE fit of pWT Adnectin residue 83 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

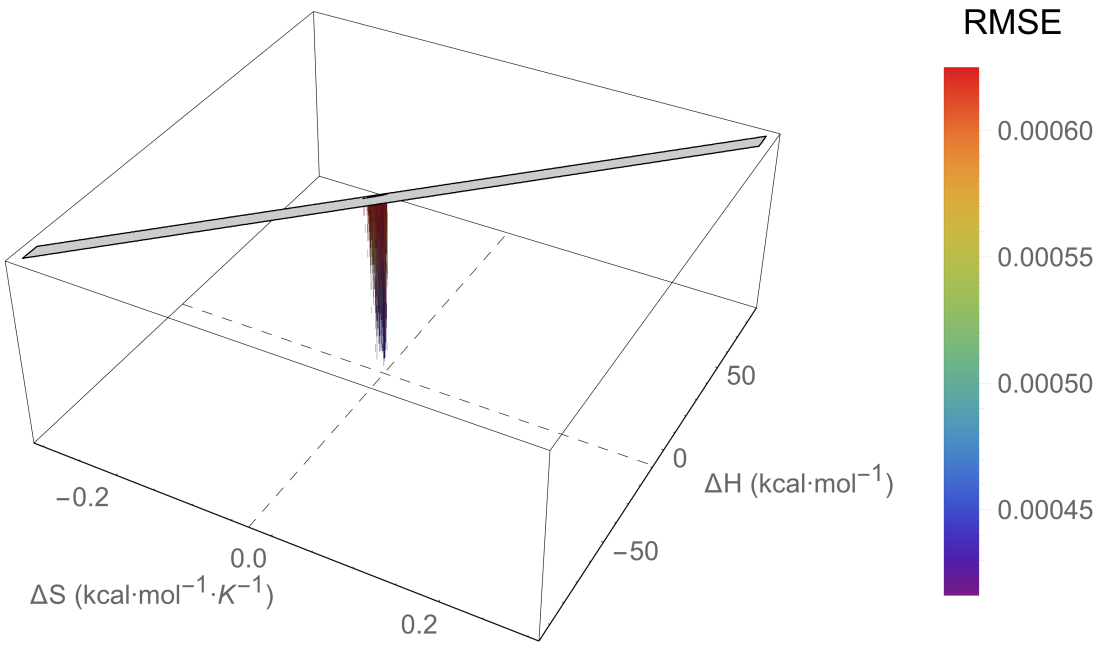


Figure I.73: An RMSE landscape from fitting pWT Adnectin residue 85 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE$_{\text{min}}$) are indicated in gray.

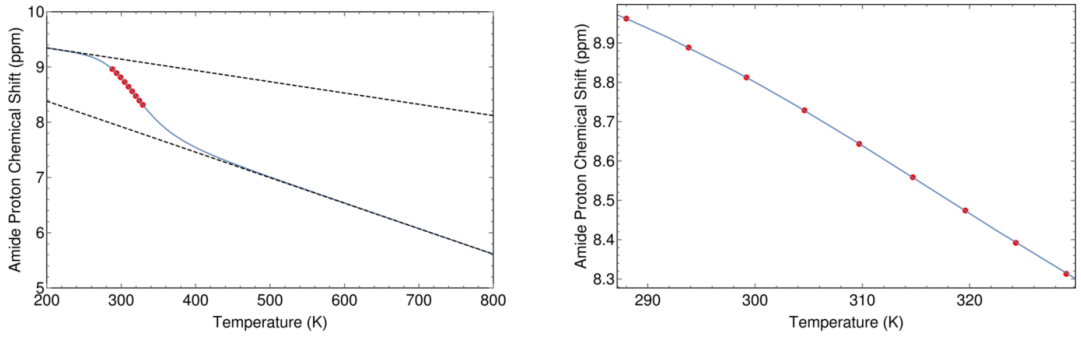


Figure I.74: The minimum RMSE fit of pWT Adnectin residue 85 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

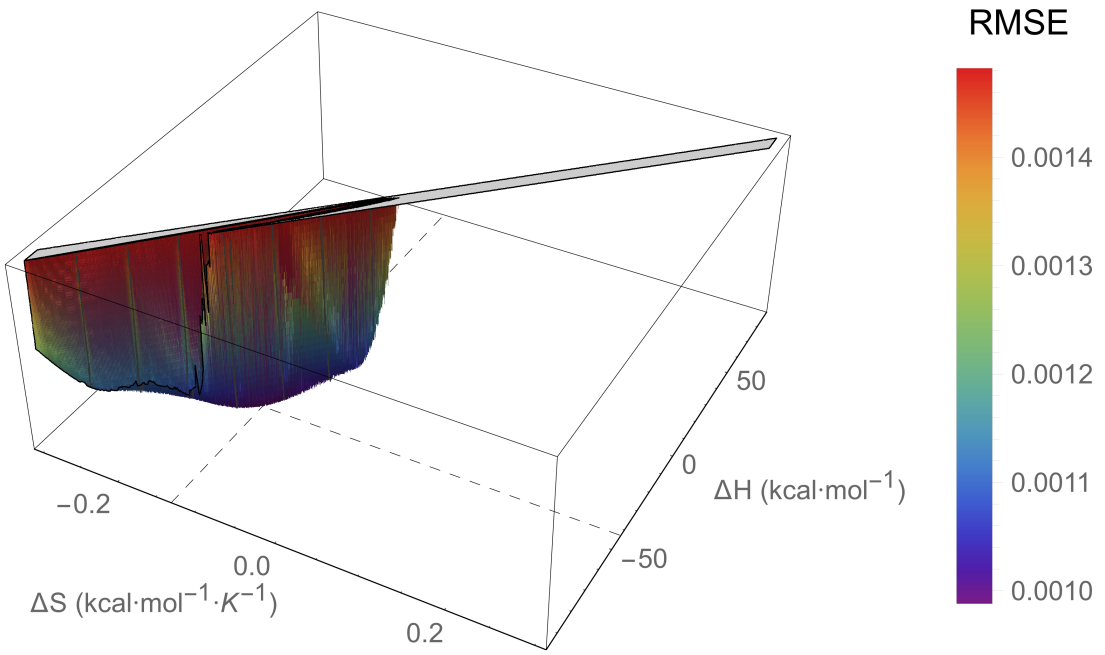


Figure I.75: An RMSE landscape from fitting pWT Adnectin residue 88 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

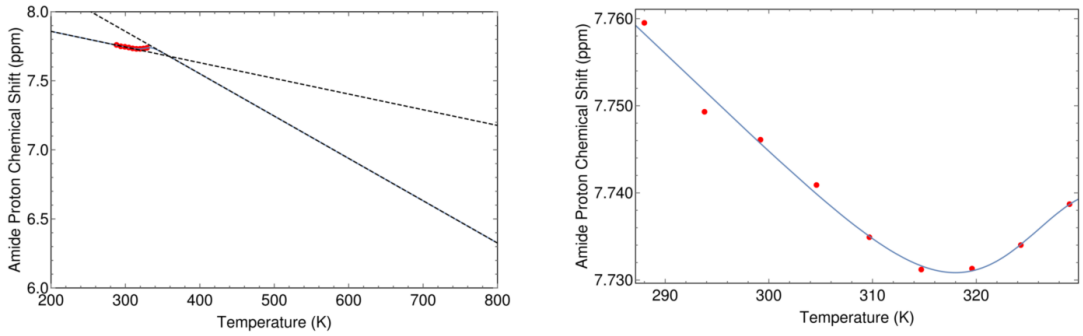


Figure I.76: The minimum RMSE fit of pWT Adnectin residue 88 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

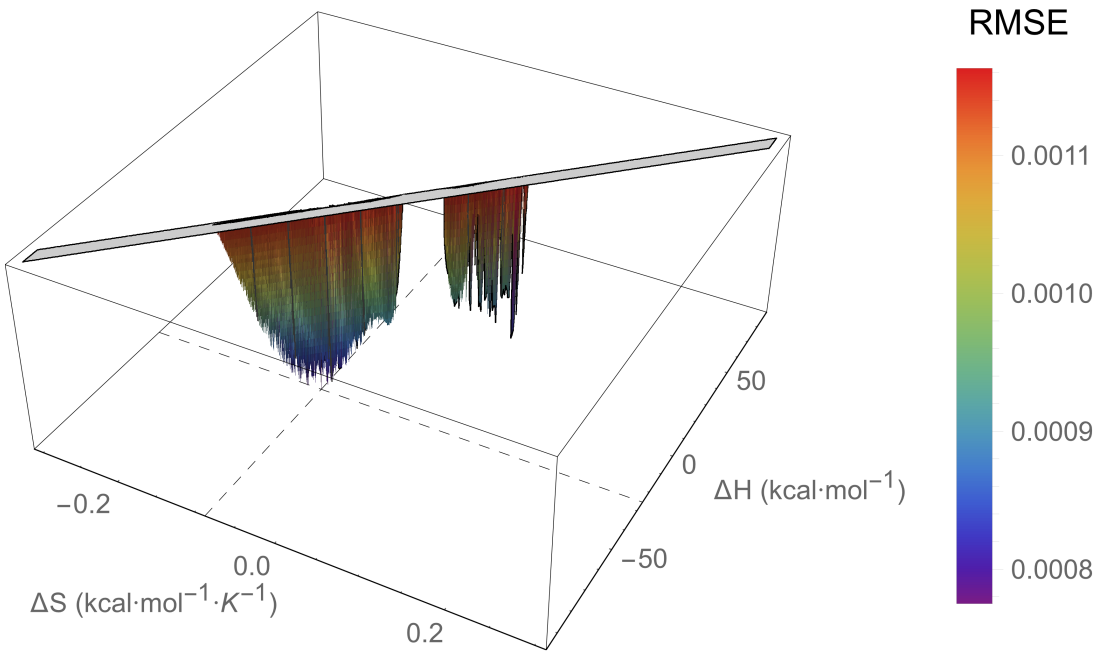


Figure I.77: An RMSE landscape from fitting pWT Adnectin residue 93 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

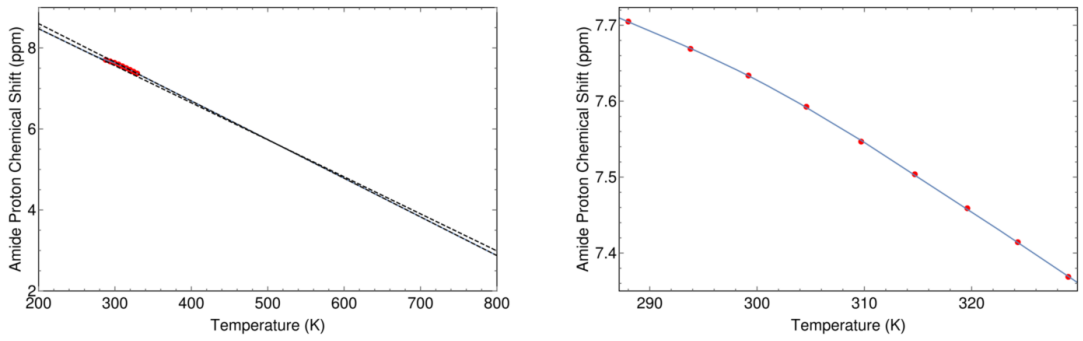


Figure I.78: The minimum RMSE fit of pWT Adnectin residue 93 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

I.5 Parent Adnectin Curvalyzer Results

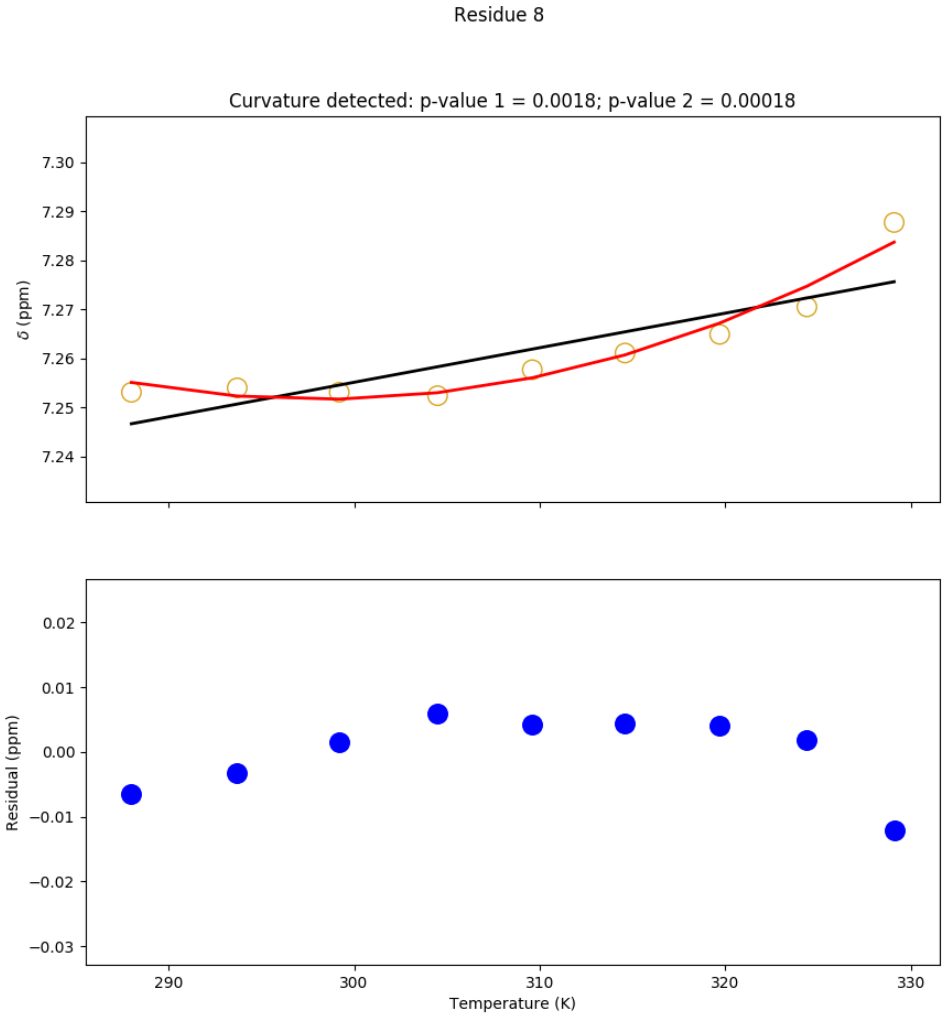


Figure I.79: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 8. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

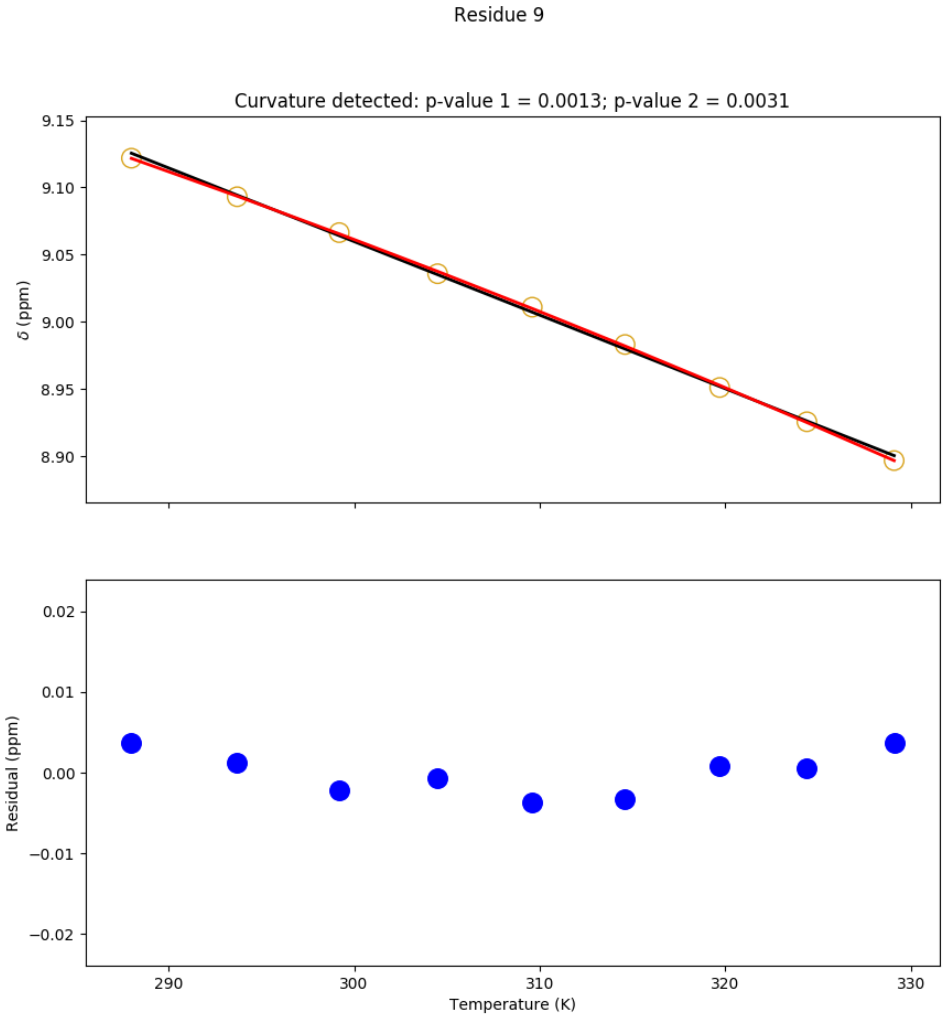


Figure I.80: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 9. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

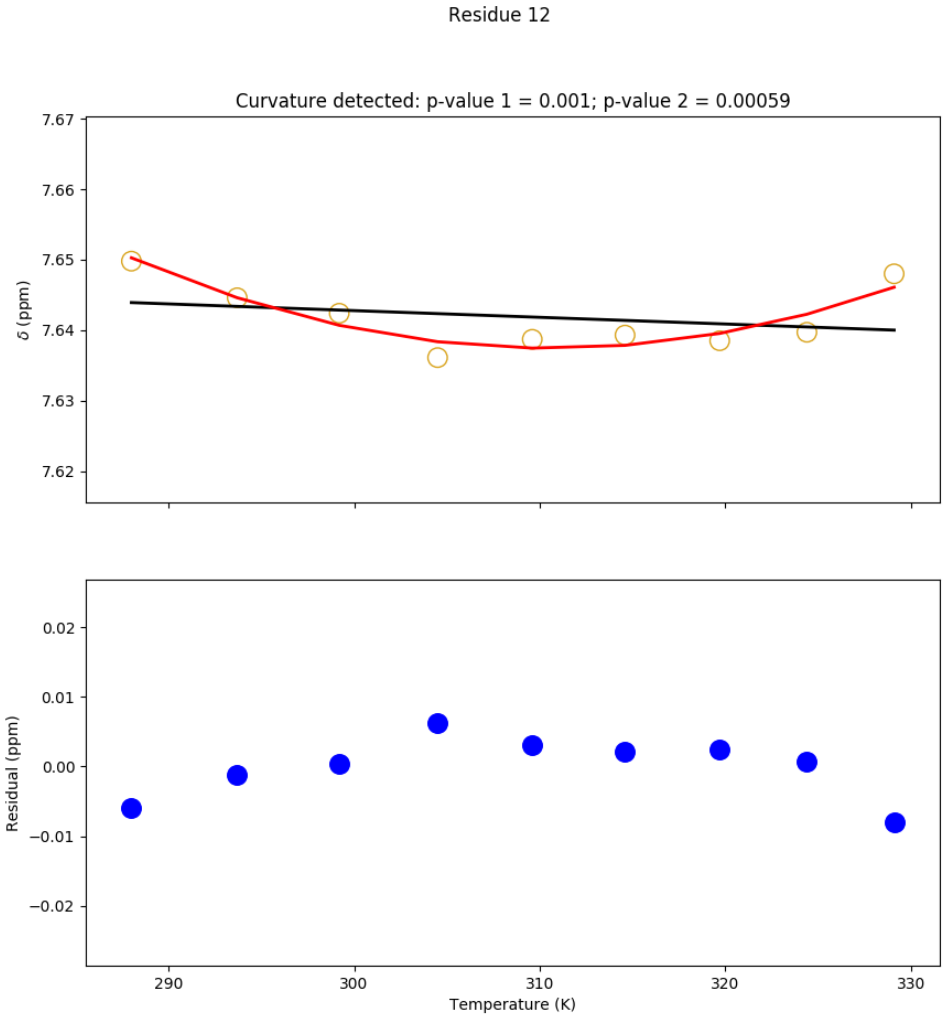


Figure I.81: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 12. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

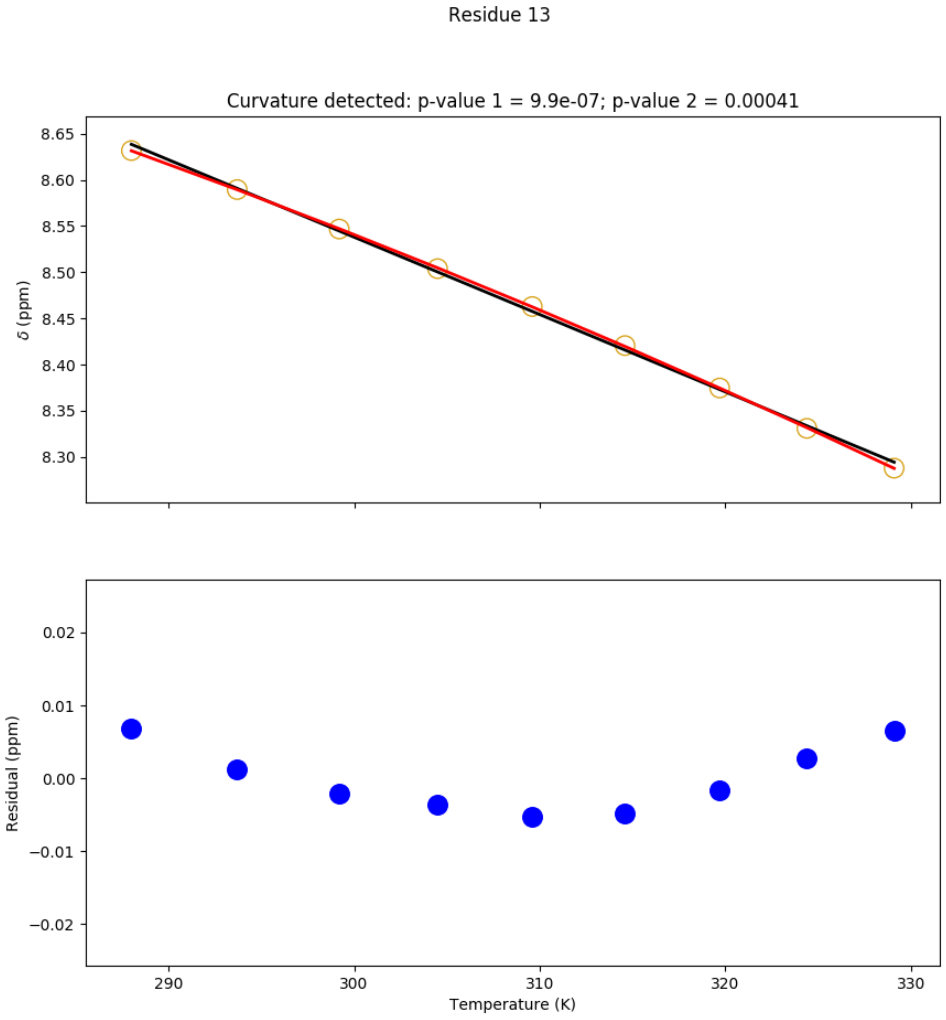


Figure I.82: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 13. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

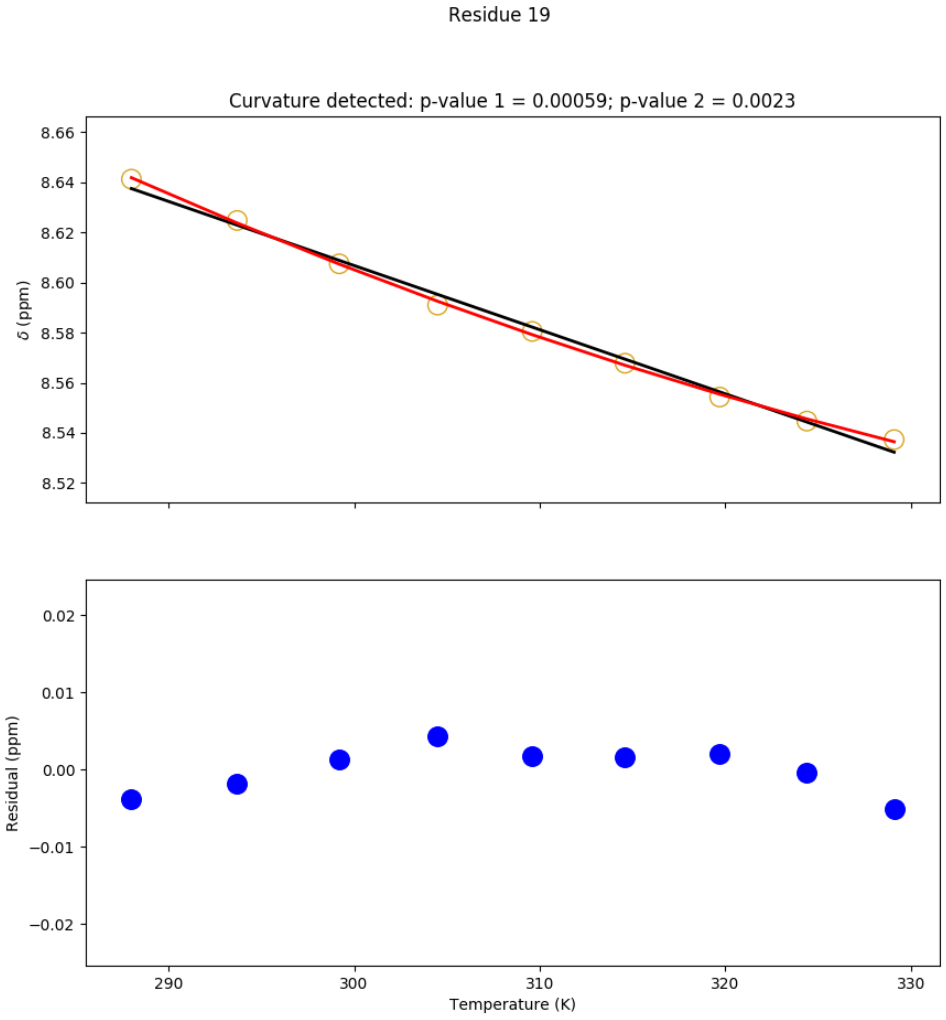


Figure I.83: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 19. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

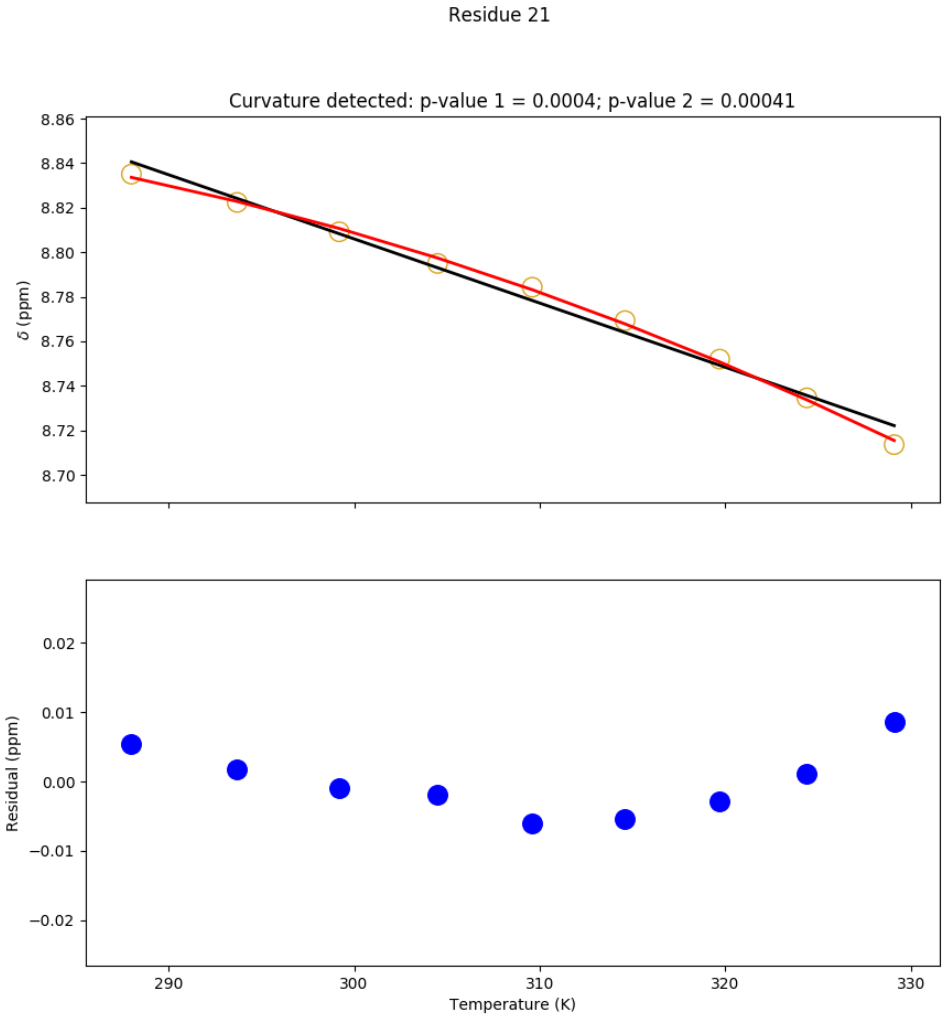


Figure I.84: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 21. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

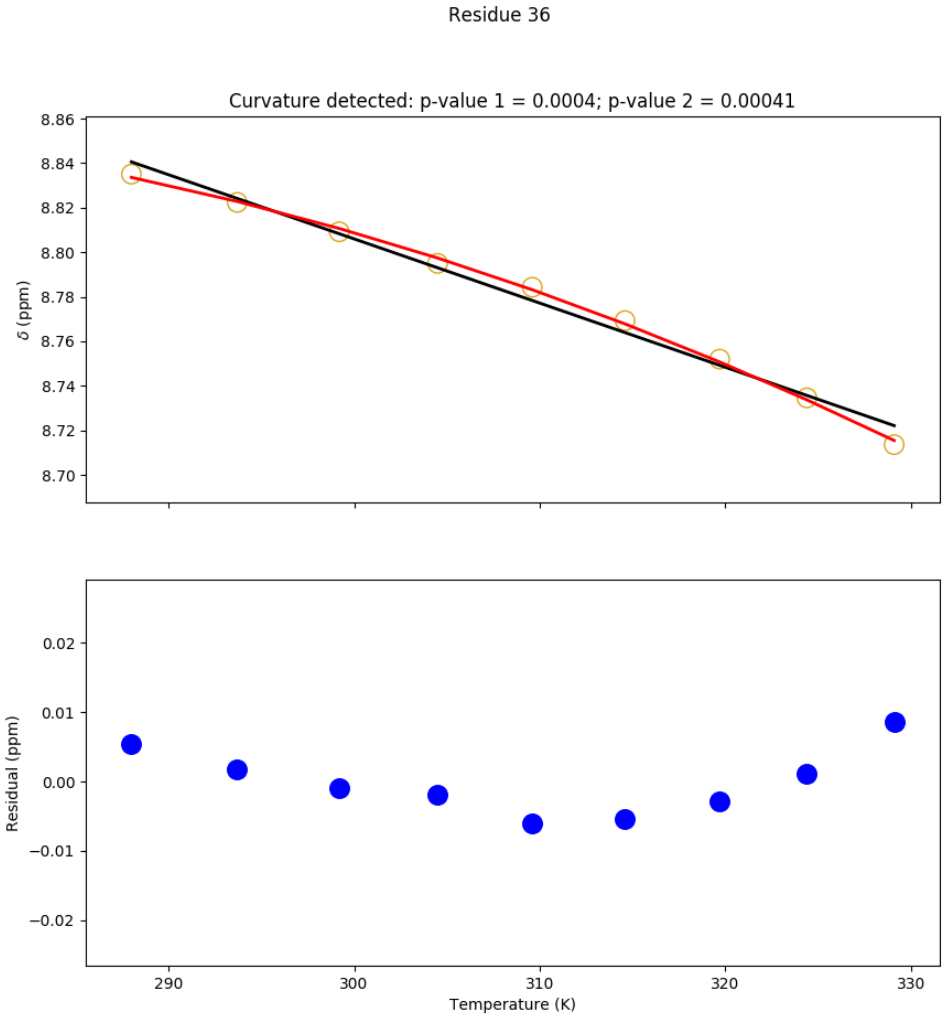


Figure I.85: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 36. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

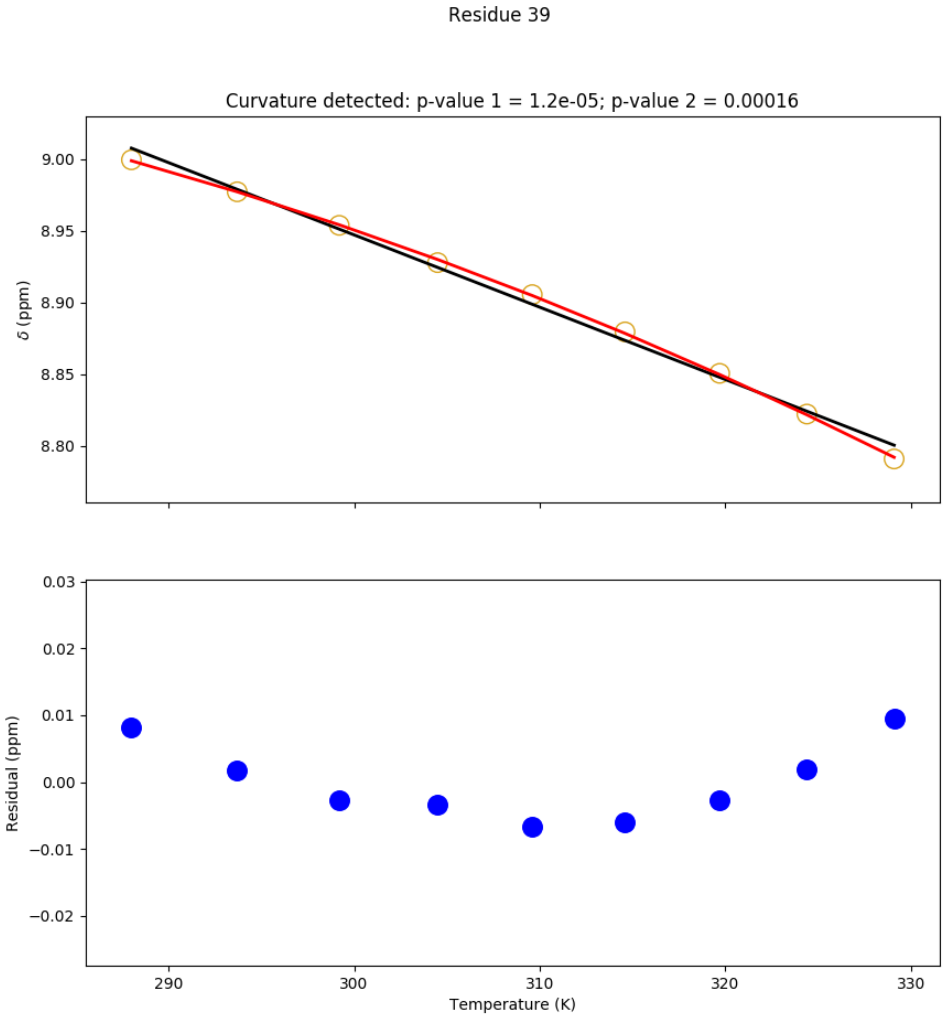


Figure I.86: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 39. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

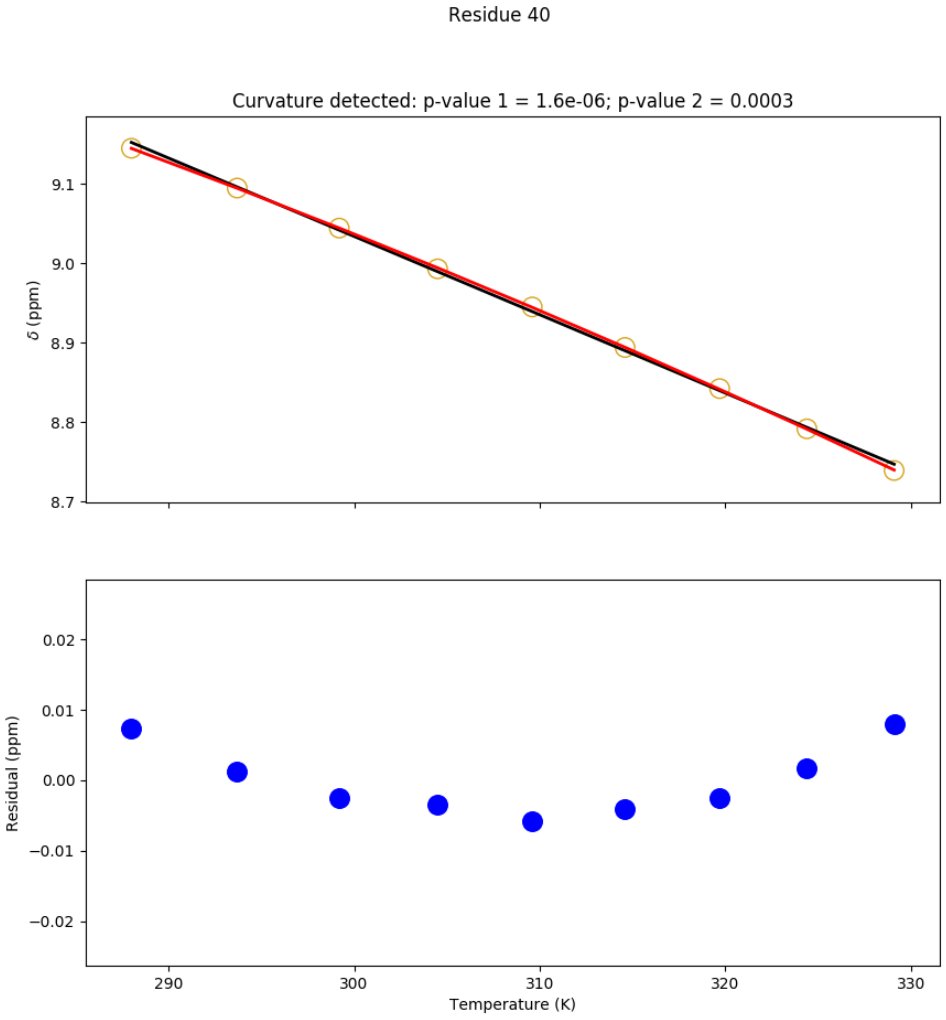


Figure I.87: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 40. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

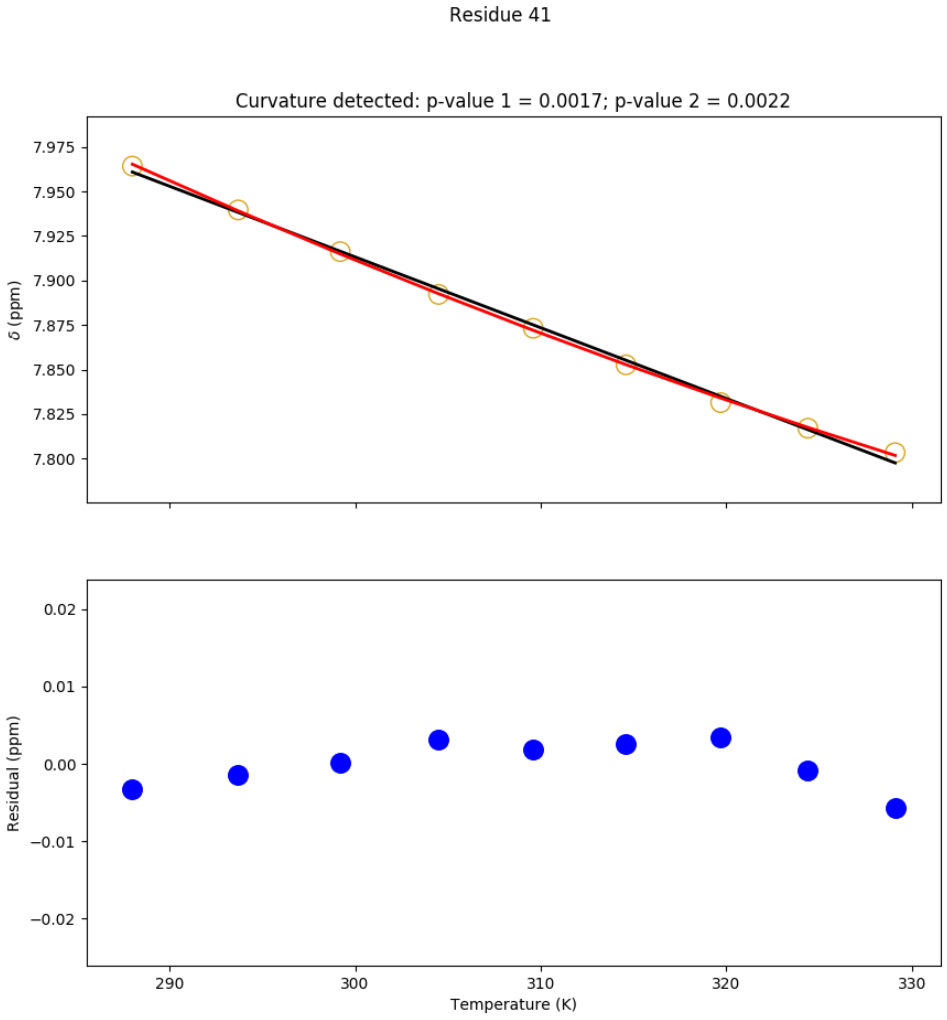


Figure I.88: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 41. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

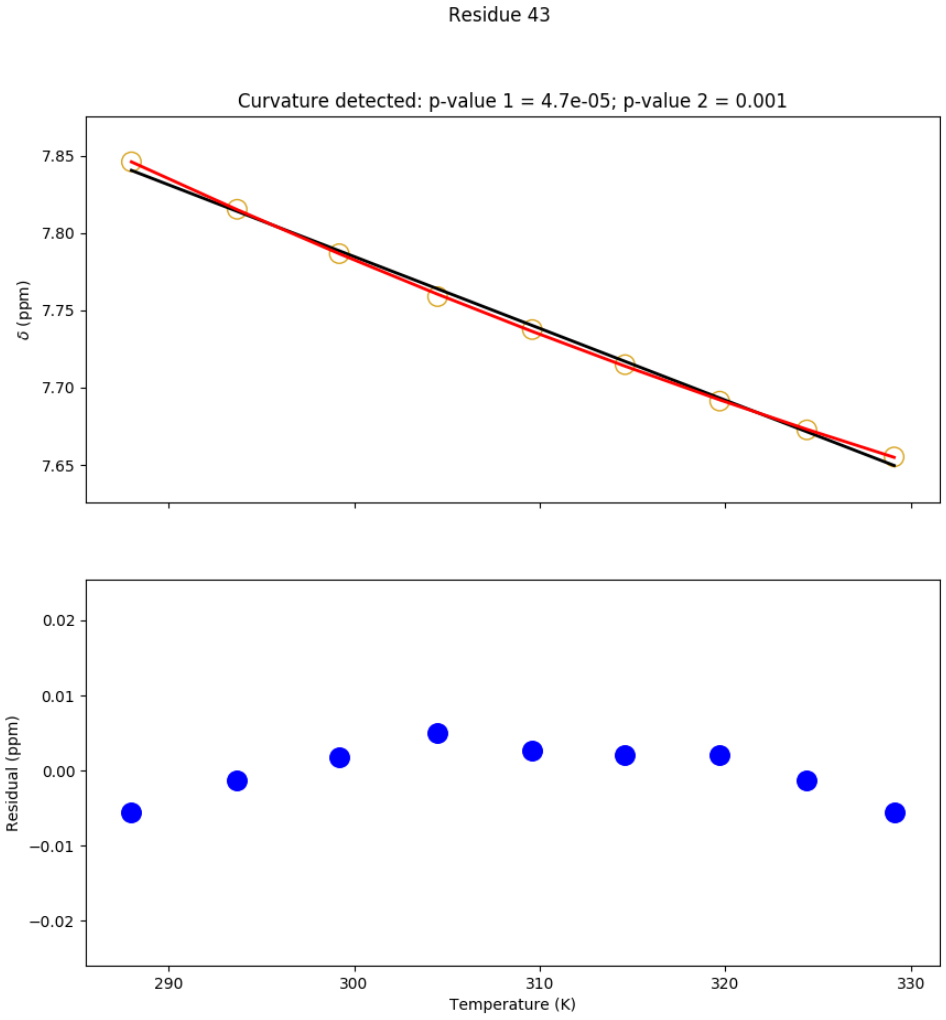


Figure I.89: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 43. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

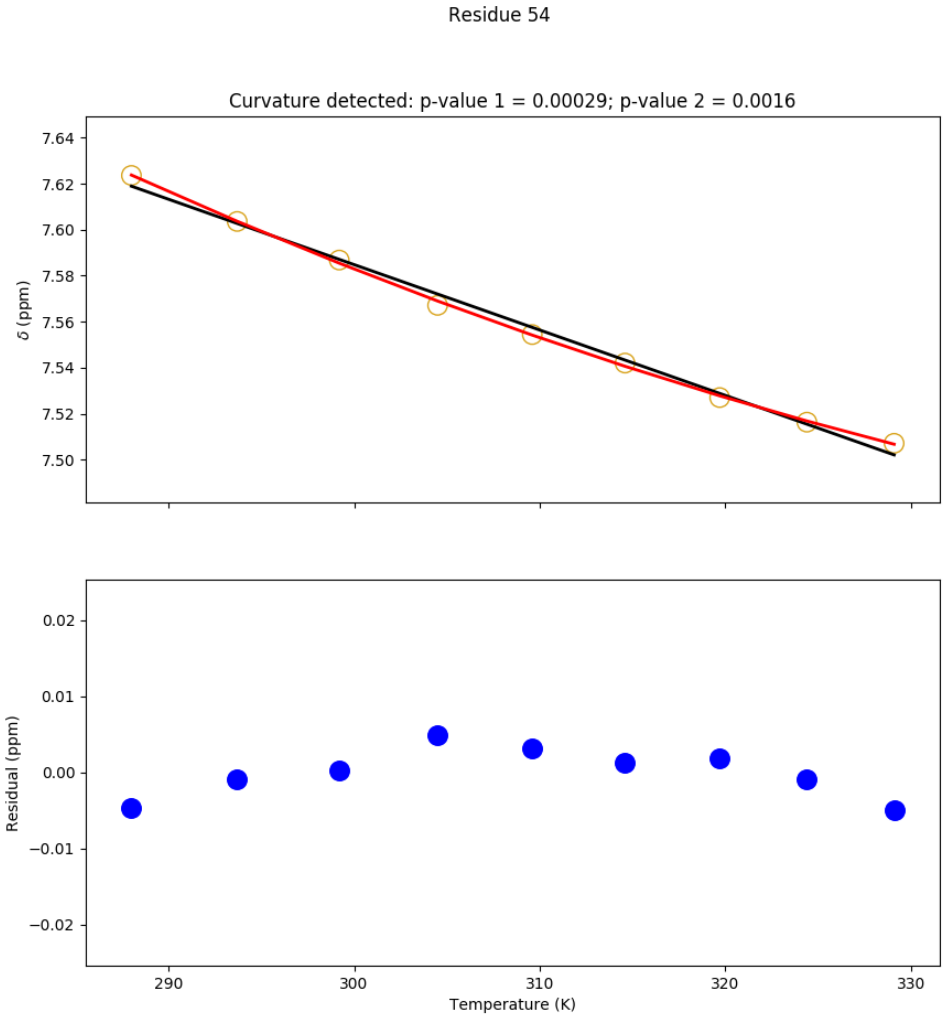


Figure I.90: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 54. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

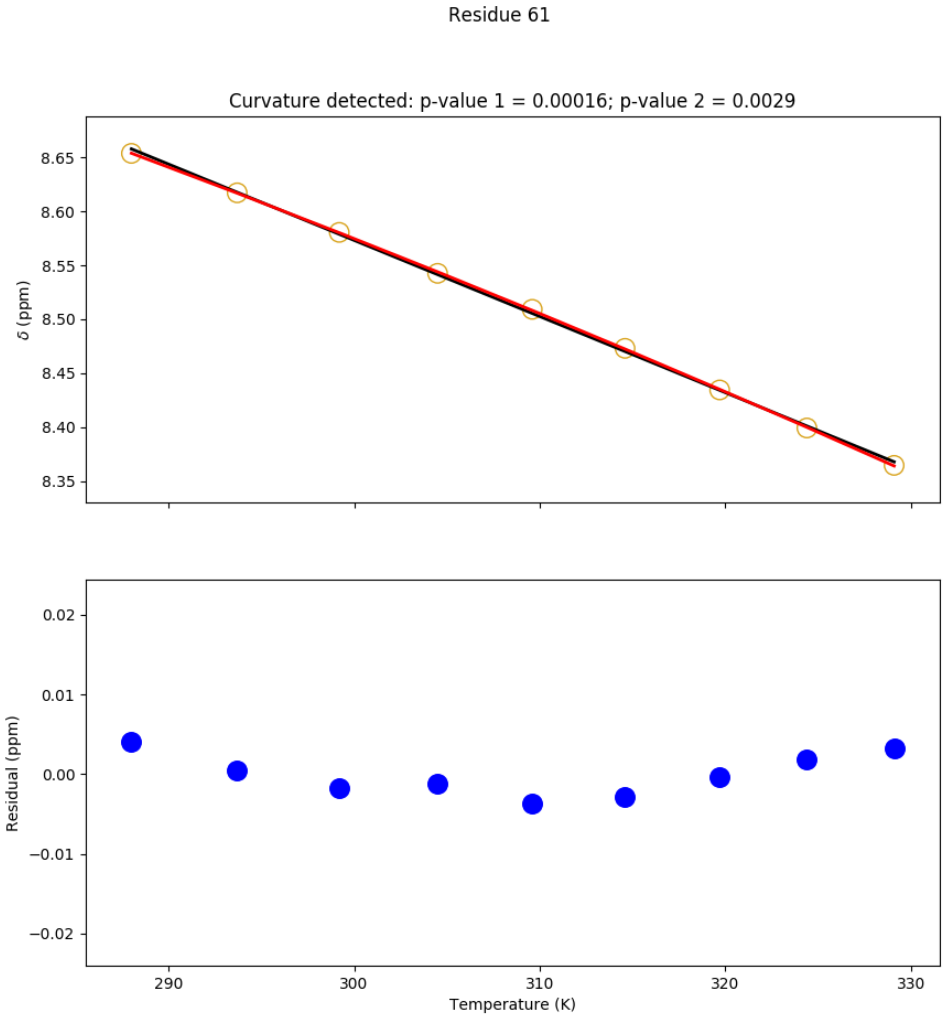


Figure I.91: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 61. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

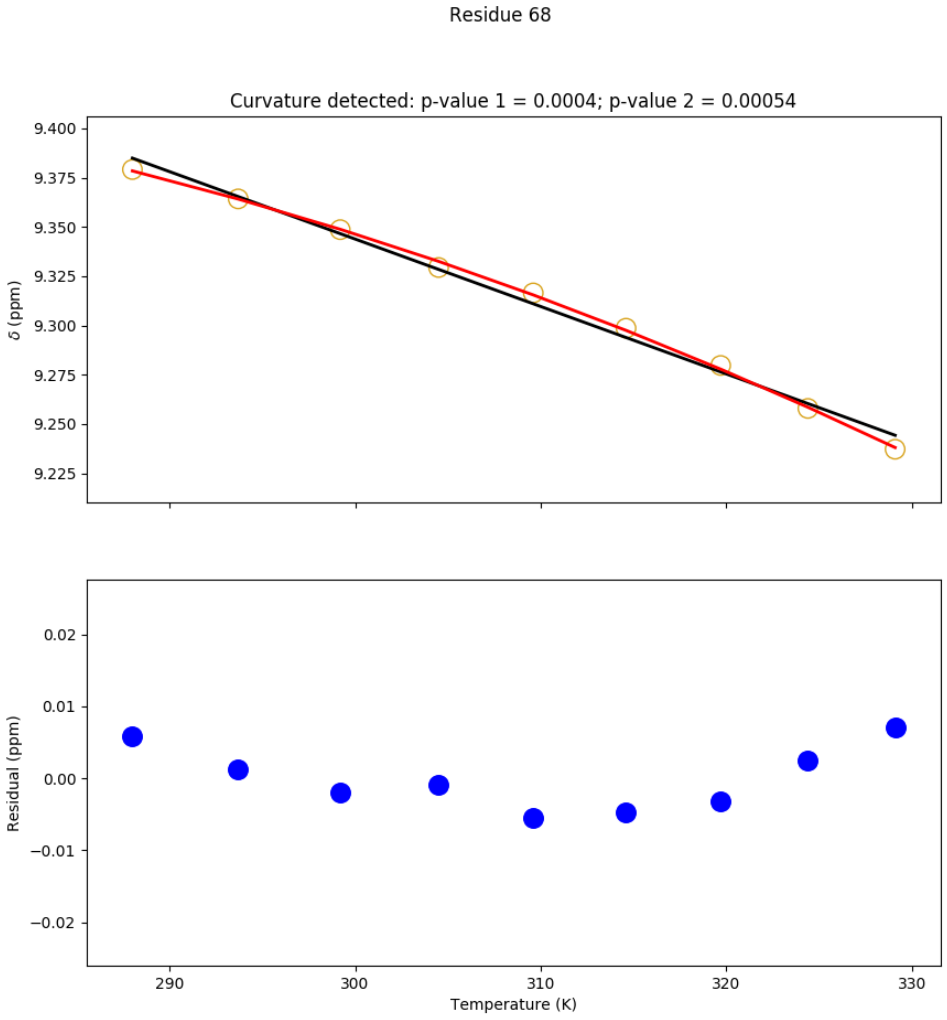


Figure I.92: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 68. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

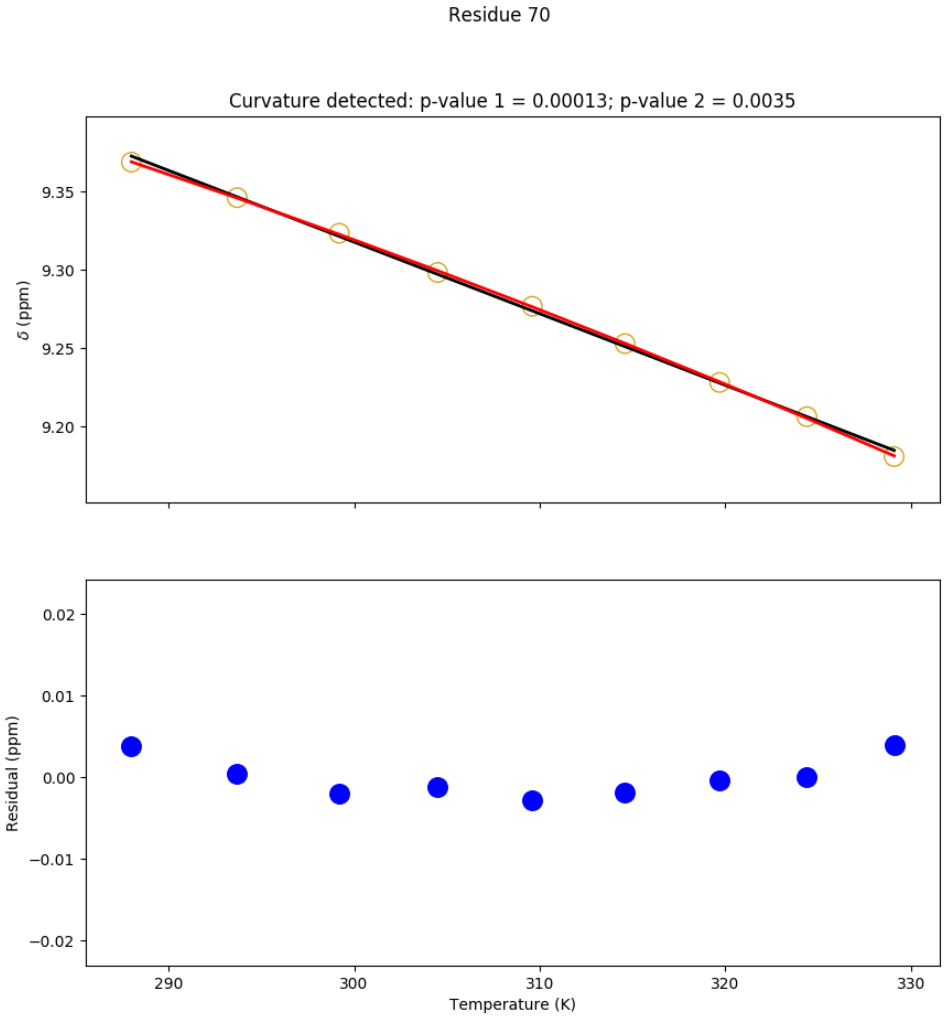


Figure I.93: Temperature dependence of the chemical shift of the Parent Adnectin amide proton from residue 70. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

I.6 Parent Adnectin Curvature Modelling

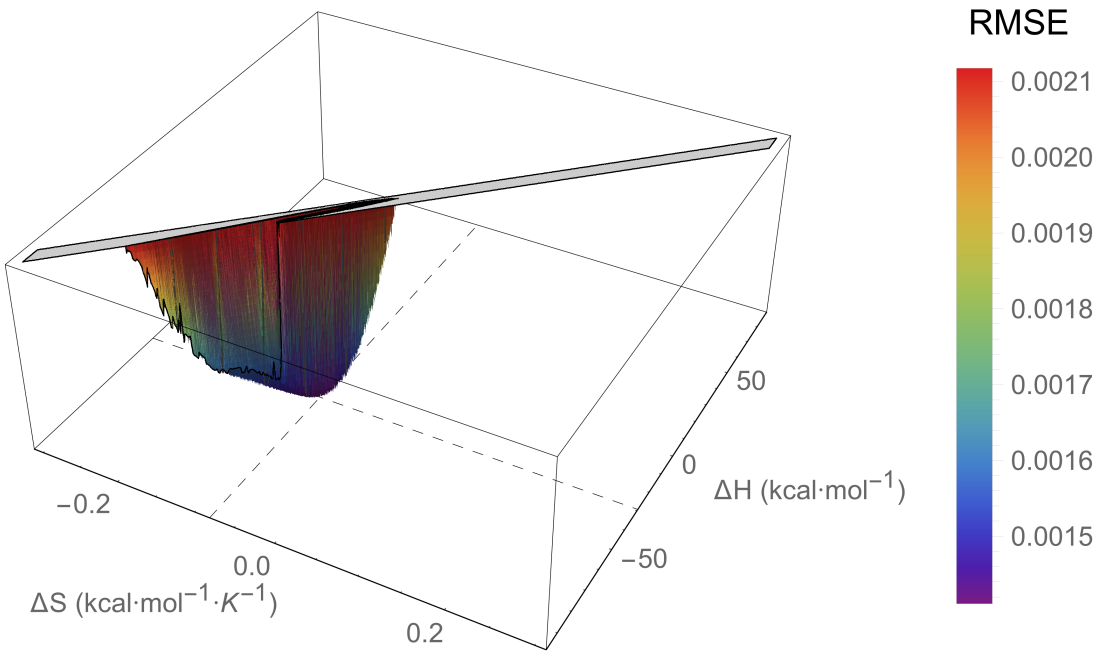


Figure I.94: An RMSE landscape from fitting Parent Adnectin residue 8 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

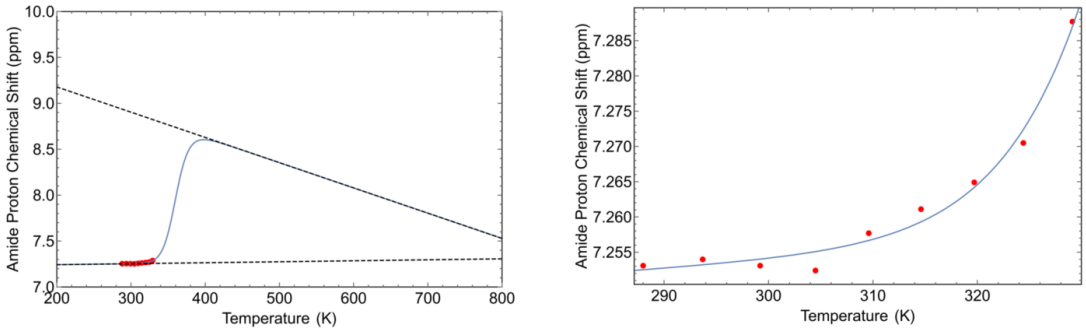


Figure I.95: The minimum RMSE fit of Parent Adnectin residue 8 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

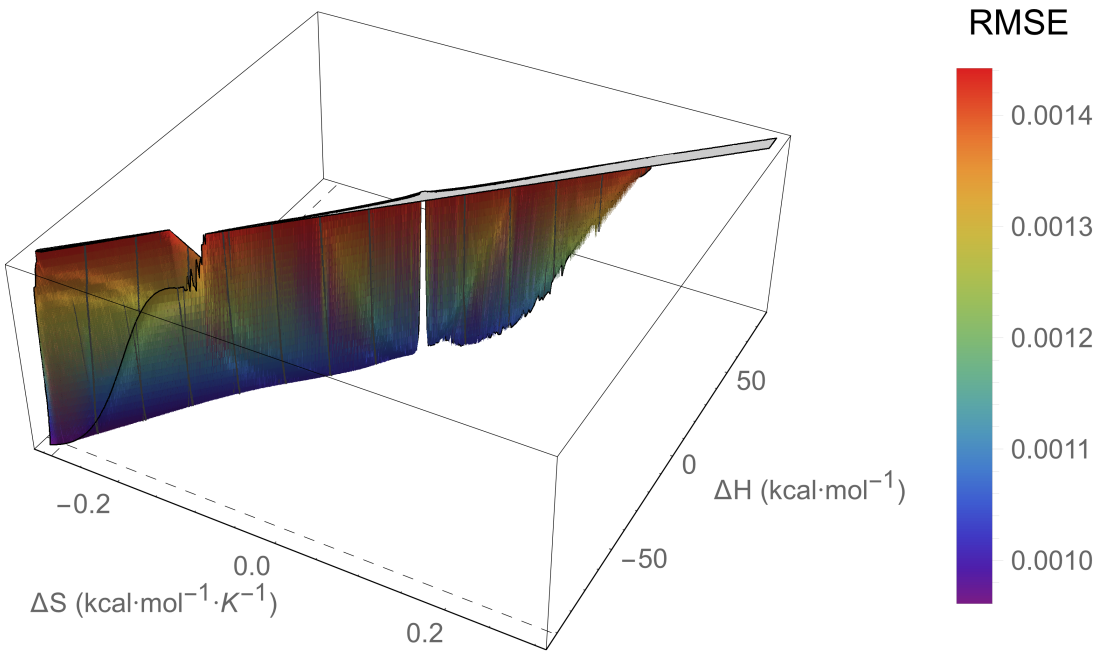


Figure I.96: An RMSE landscape from fitting Parent Adnectin residue 9 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

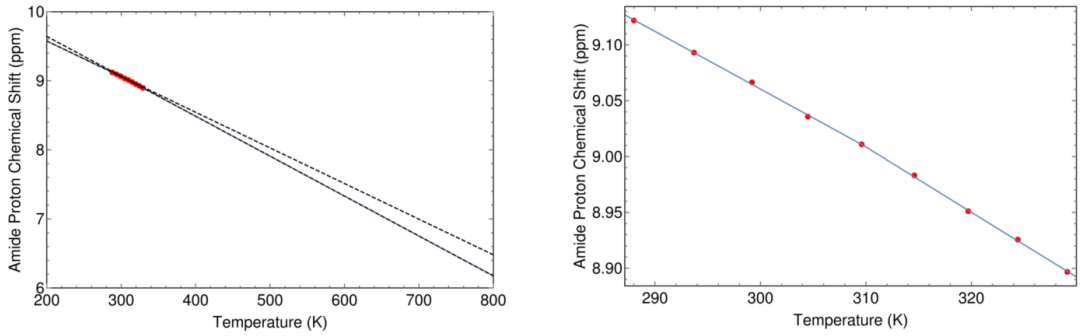


Figure I.97: The minimum RMSE fit of Parent Adnectin residue 9 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

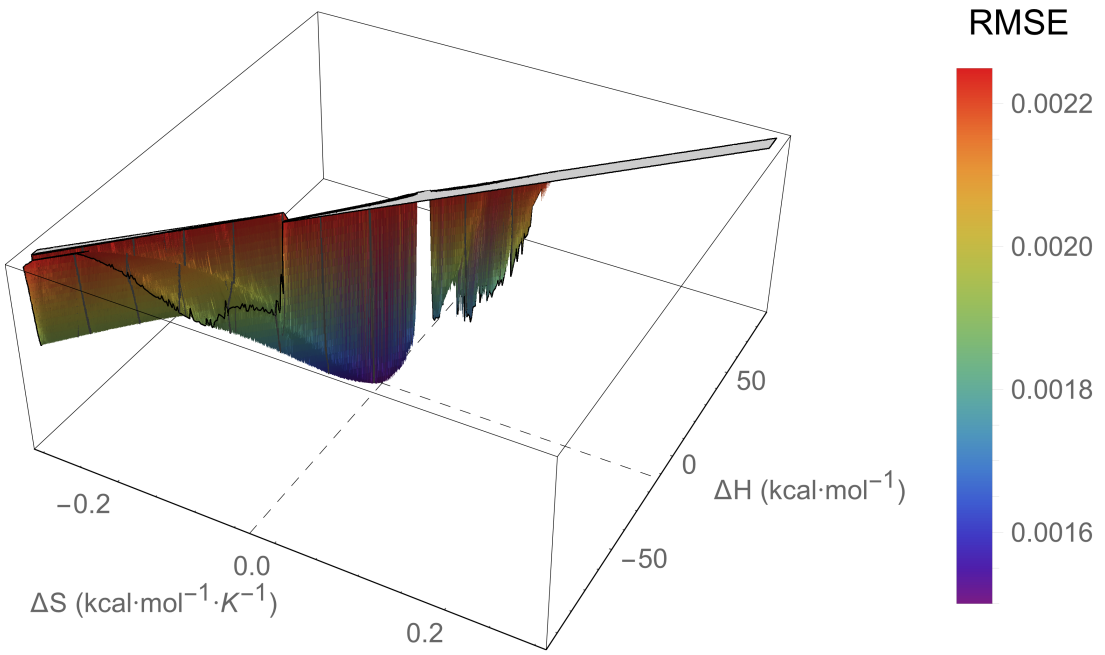


Figure I.98: An RMSE landscape from fitting Parent Adnectin residue 12 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

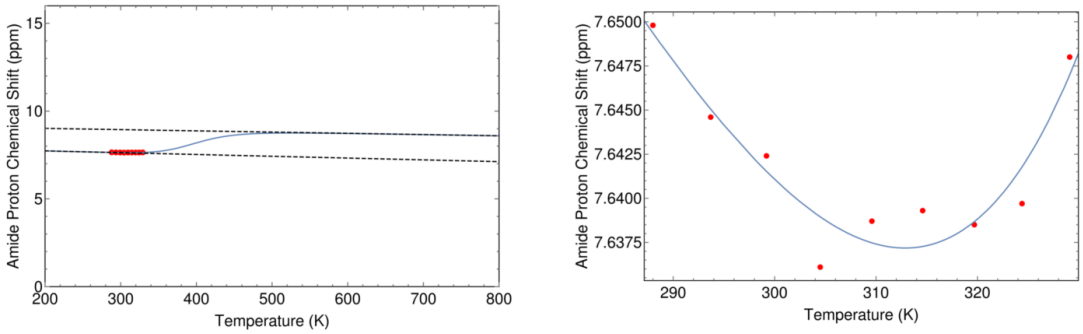


Figure I.99: The minimum RMSE fit of Parent Adnectin residue 12 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

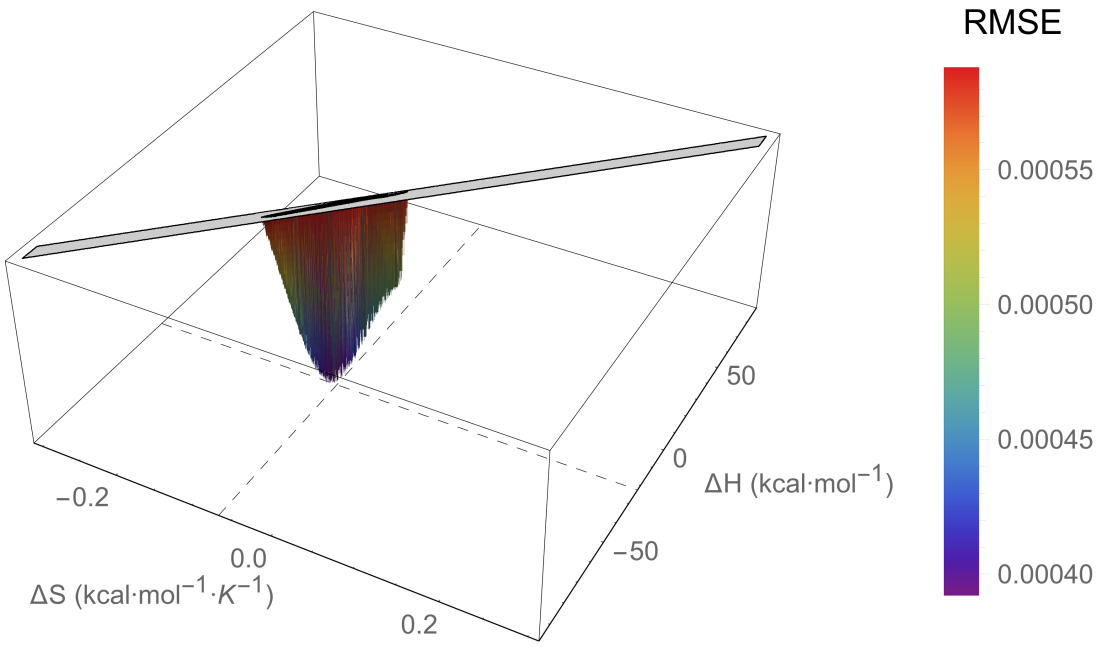


Figure I.100: An RMSE landscape from fitting Parent Adnectin residue 13 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

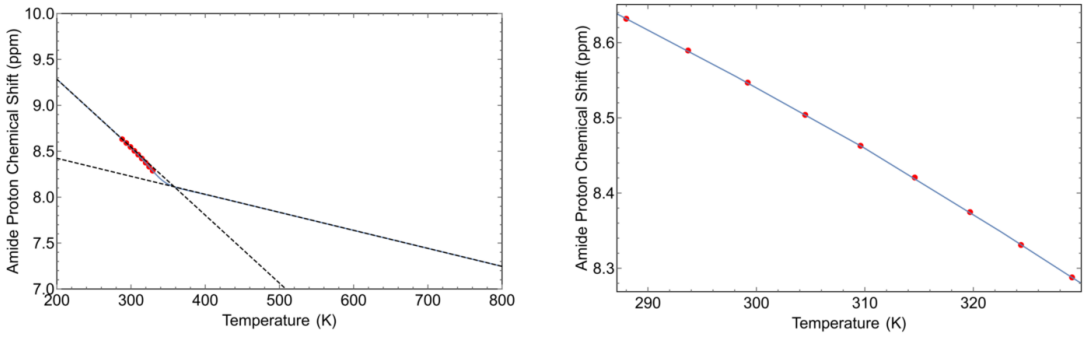


Figure I.101: The minimum RMSE fit of Parent Adnectin residue 13 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

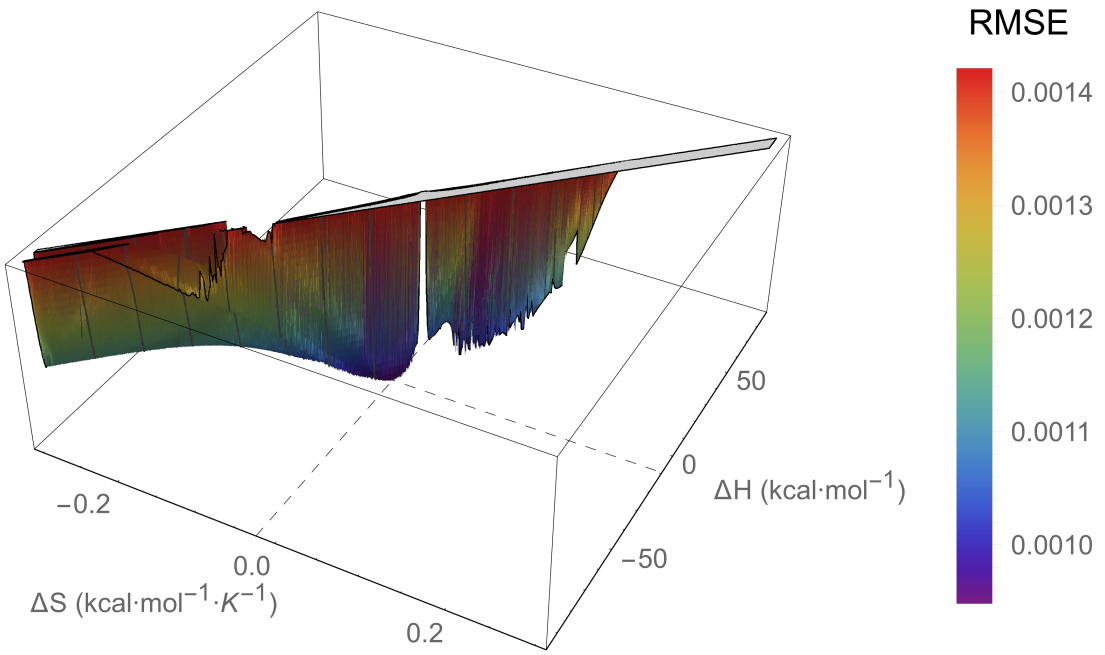


Figure I.102: An RMSE landscape from fitting Parent Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

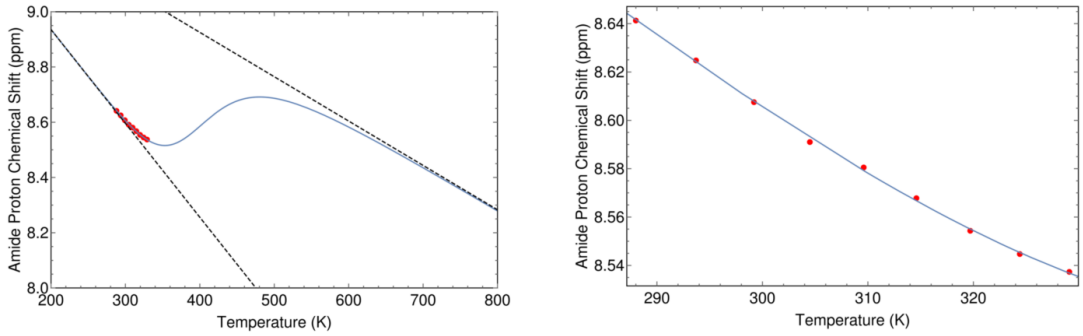


Figure I.103: The minimum RMSE fit of Parent Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

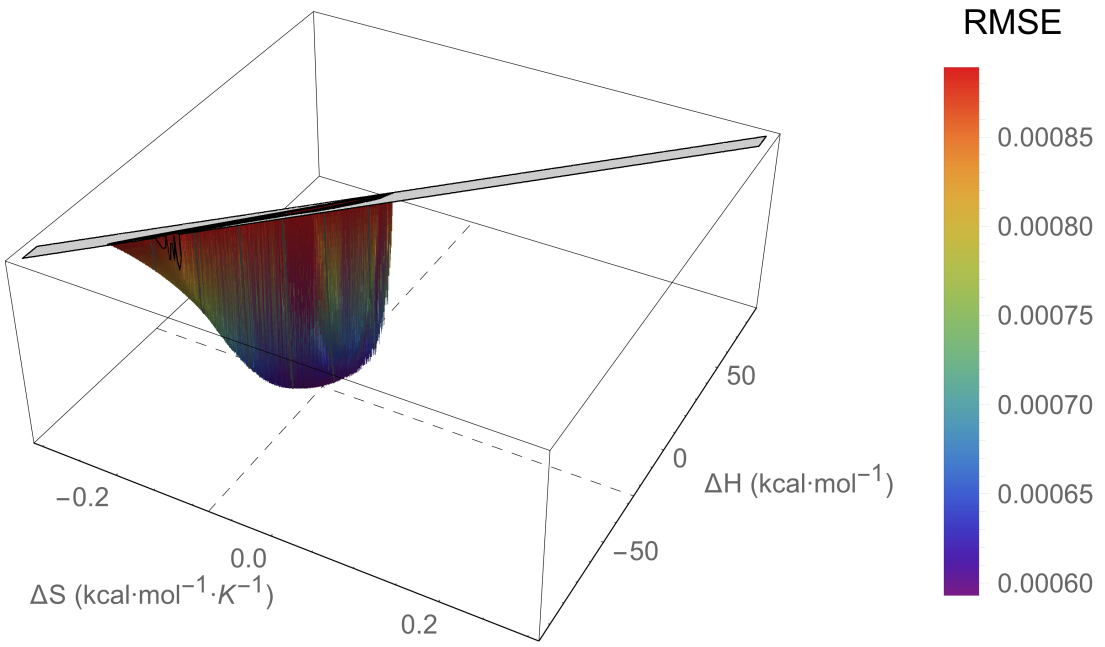


Figure I.104: An RMSE landscape from fitting Parent Adnectin residue 21 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

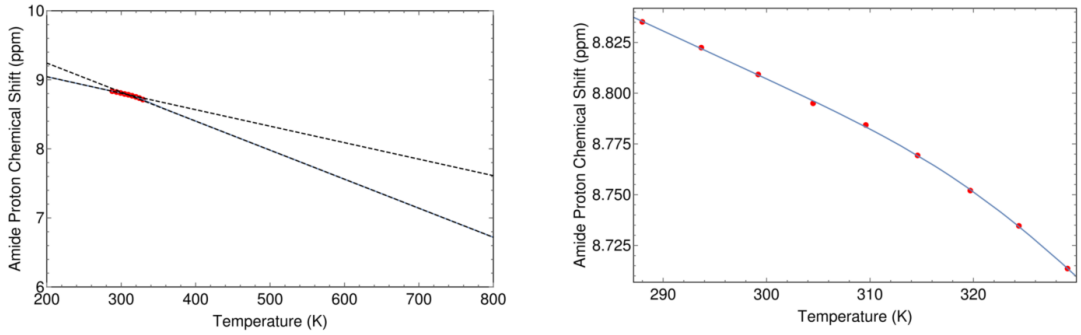


Figure I.105: The minimum RMSE fit of Parent Adnectin residue 21 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

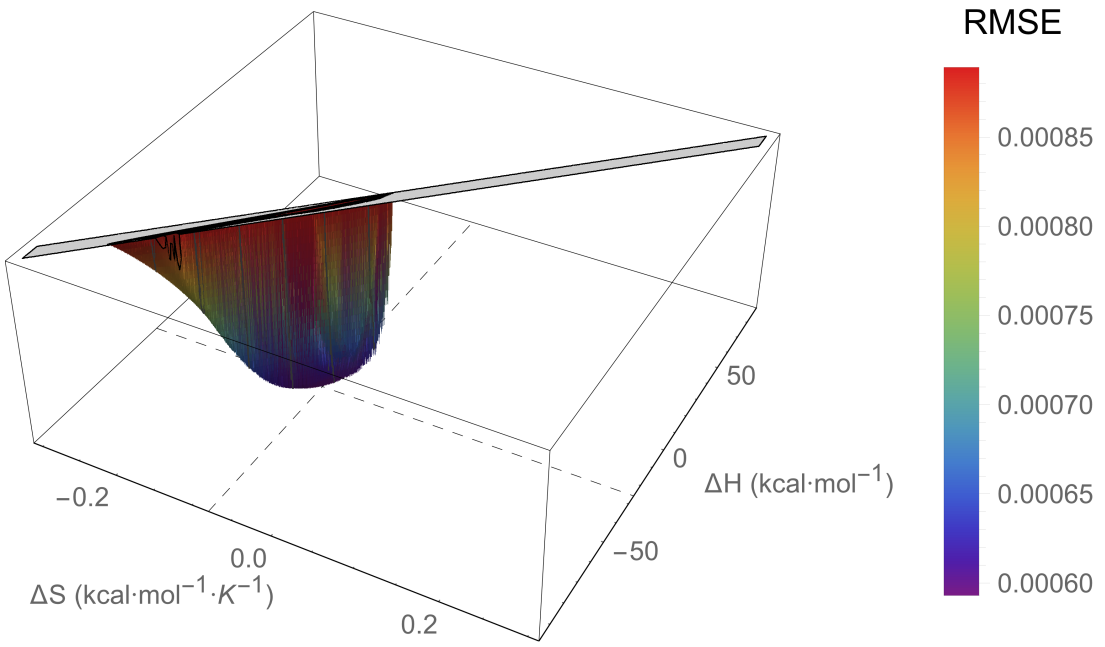


Figure I.106: An RMSE landscape from fitting Parent Adnectin residue 36 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

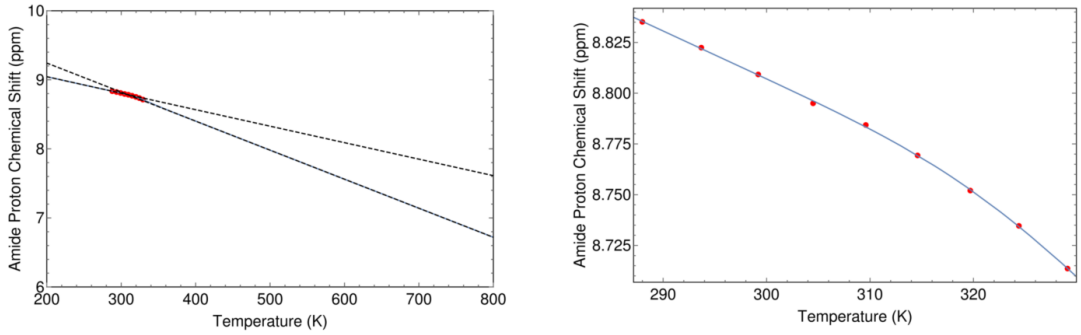


Figure I.107: The minimum RMSE fit of Parent Adnectin residue 36 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

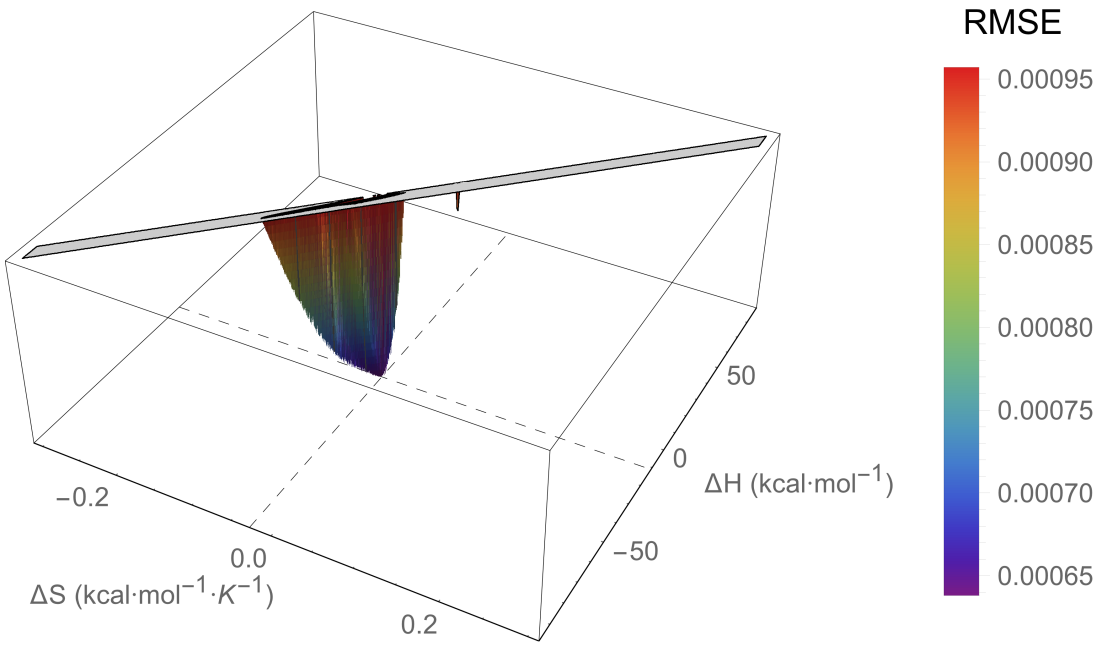


Figure I.108: An RMSE landscape from fitting Parent Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

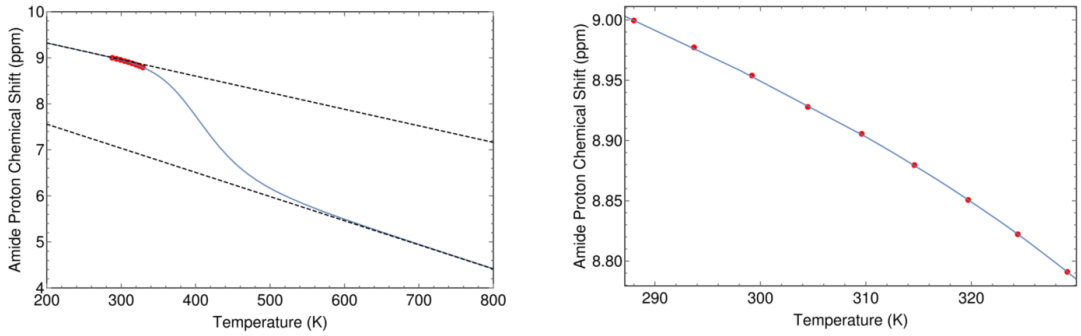


Figure I.109: The minimum RMSE fit of Parent Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

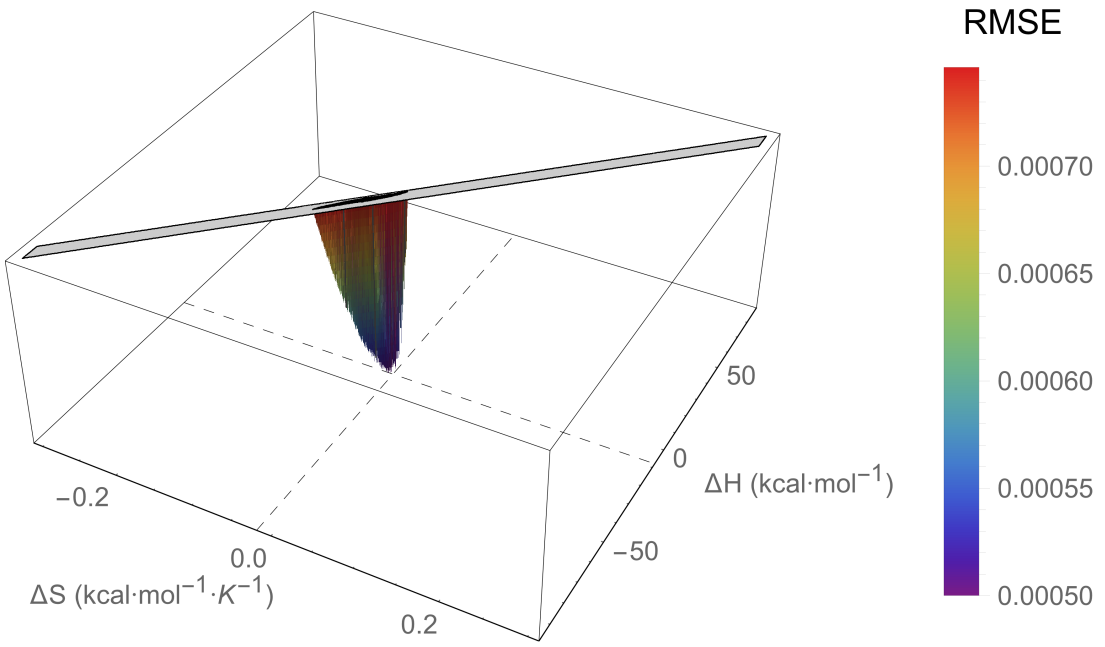


Figure I.110: An RMSE landscape from fitting Parent Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

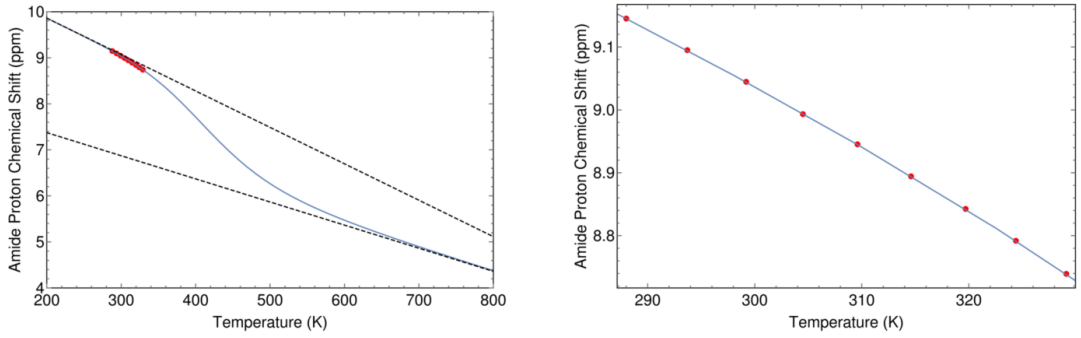


Figure I.111: The minimum RMSE fit of Parent Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

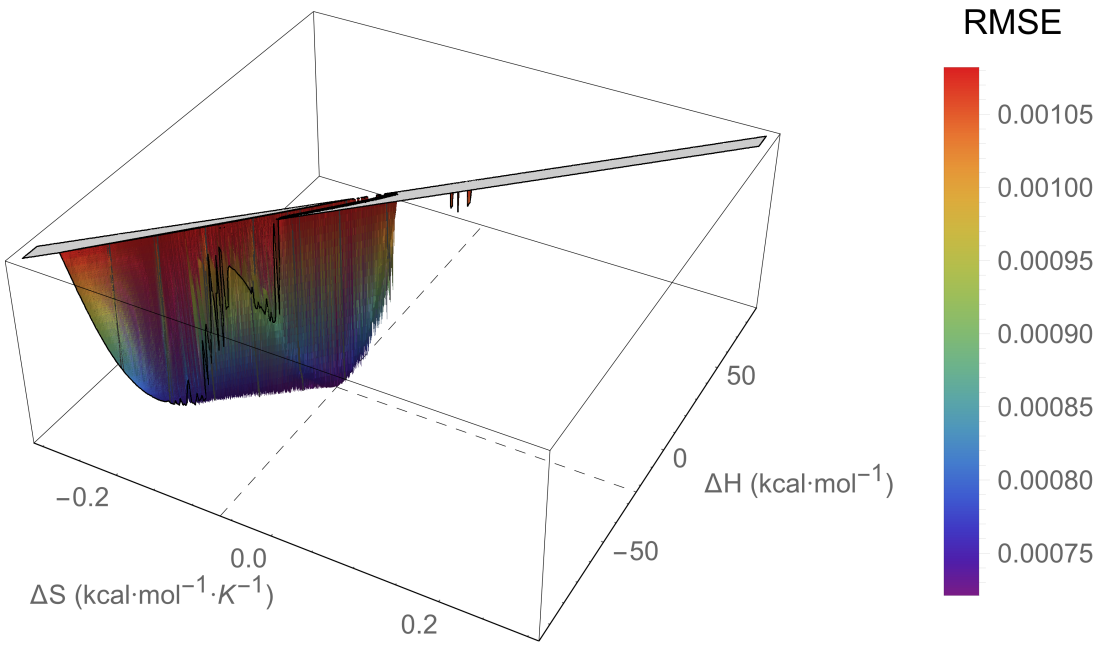


Figure I.112: An RMSE landscape from fitting Parent Adnectin residue 41 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

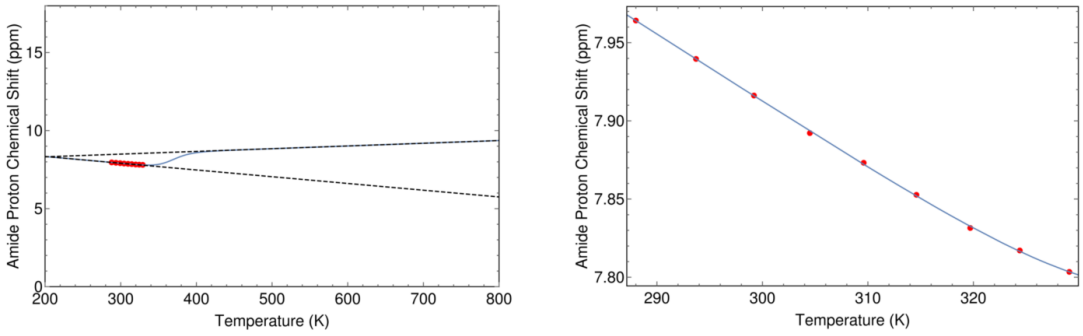


Figure I.113: The minimum RMSE fit of Parent Adnectin residue 41 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

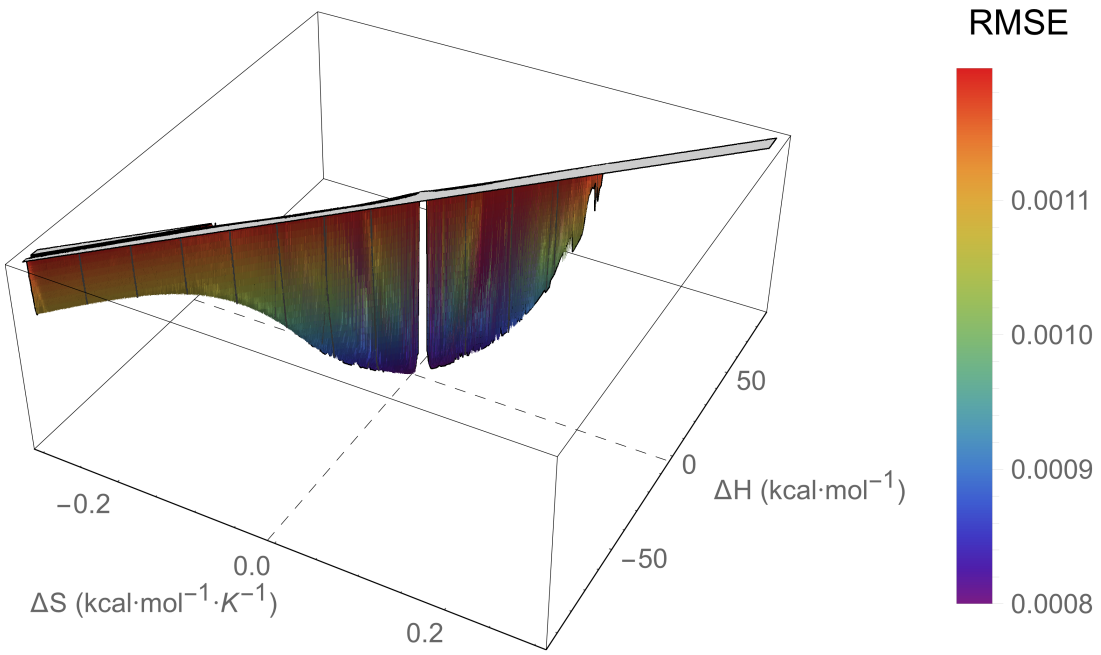


Figure I.114: An RMSE landscape from fitting Parent Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE_{min}) are indicated in gray.

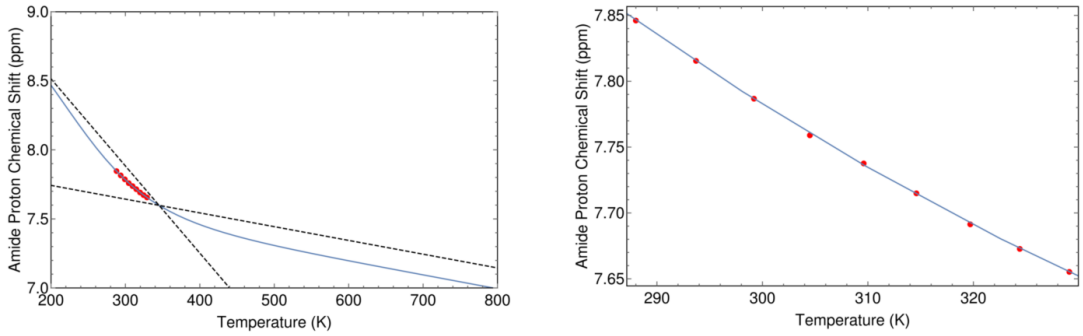


Figure I.115: The minimum RMSE fit of Parent Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

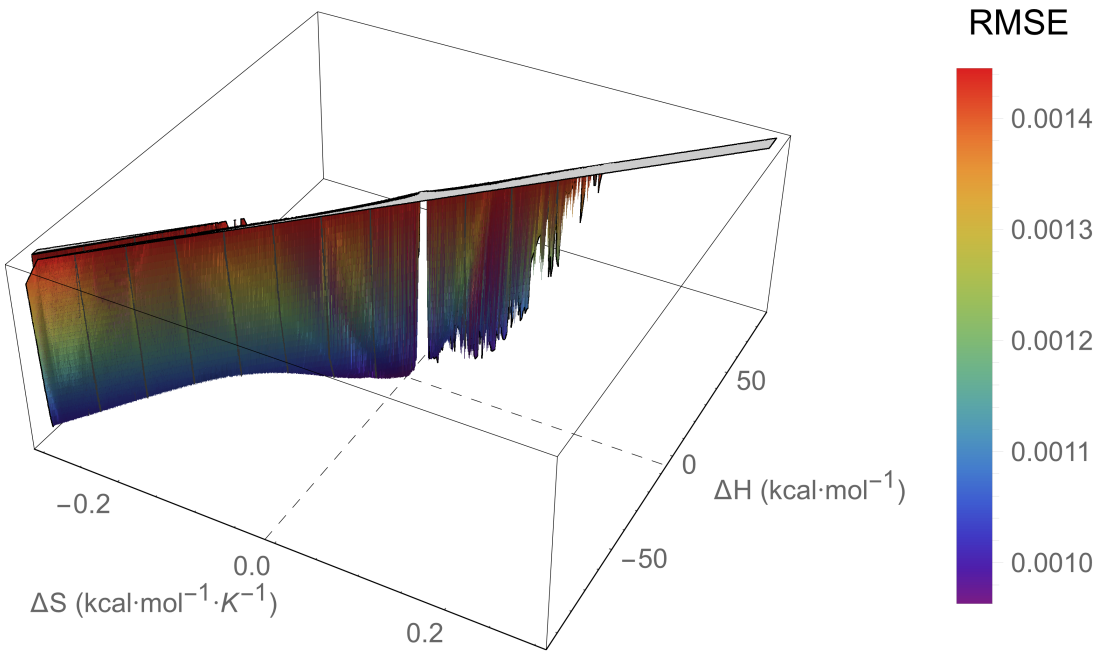


Figure I.116: An RMSE landscape from fitting Parent Adnectin residue 54 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE_{min}) are indicated in gray.

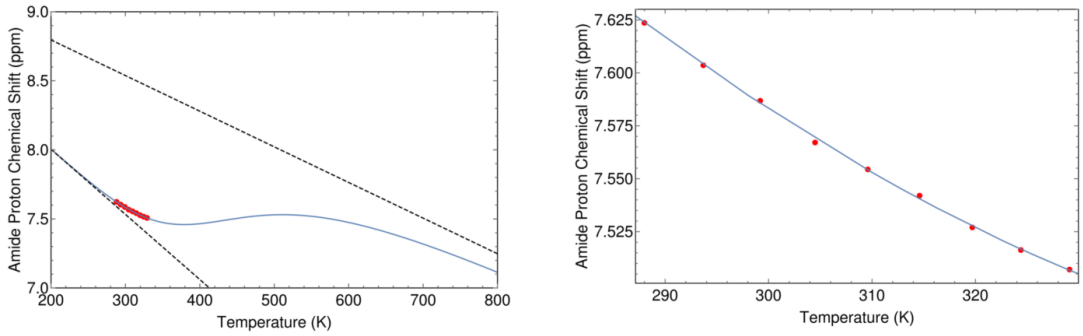


Figure I.117: The minimum RMSE fit of Parent Adnectin residue 54 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

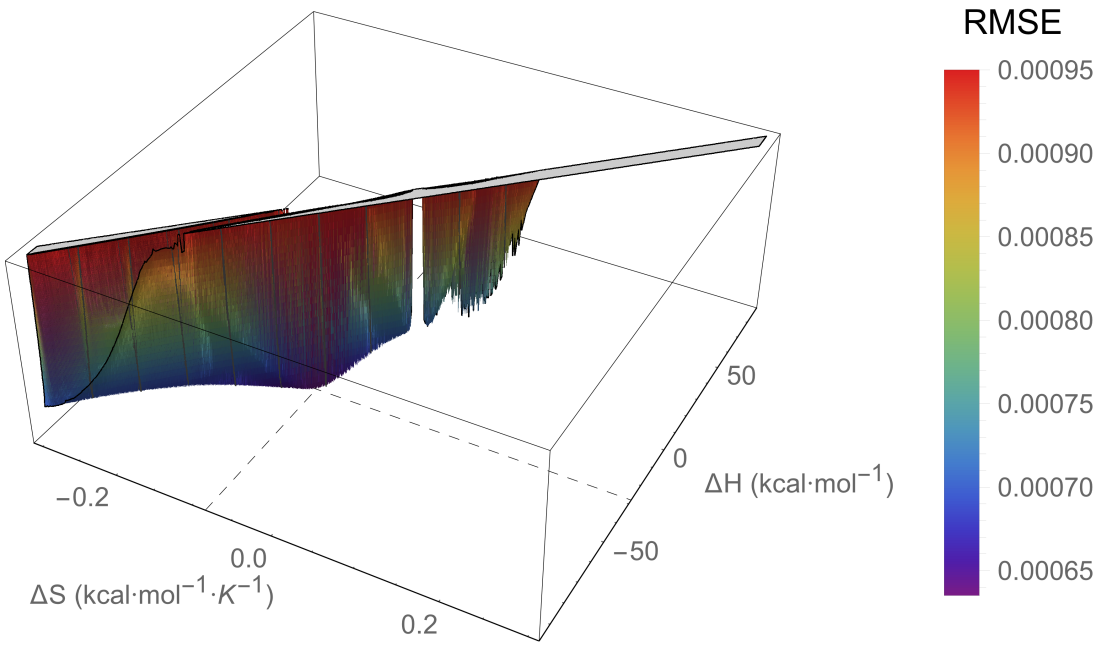


Figure I.118: An RMSE landscape from fitting Parent Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

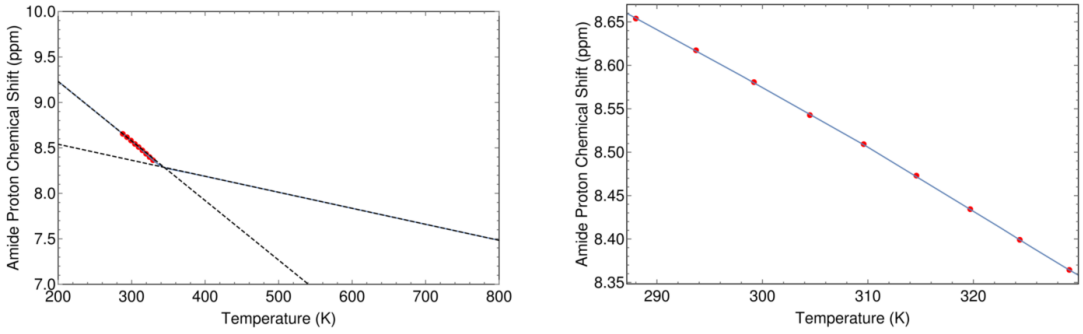


Figure I.119: The minimum RMSE fit of Parent Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

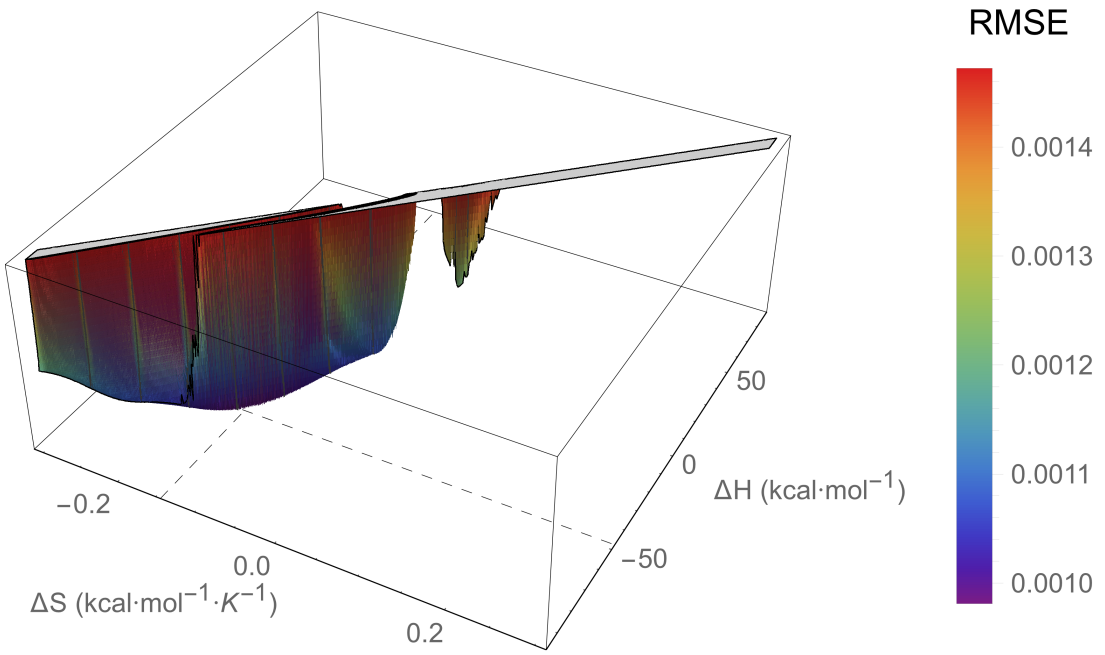


Figure I.120: An RMSE landscape from fitting Parent Adnectin residue 68 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

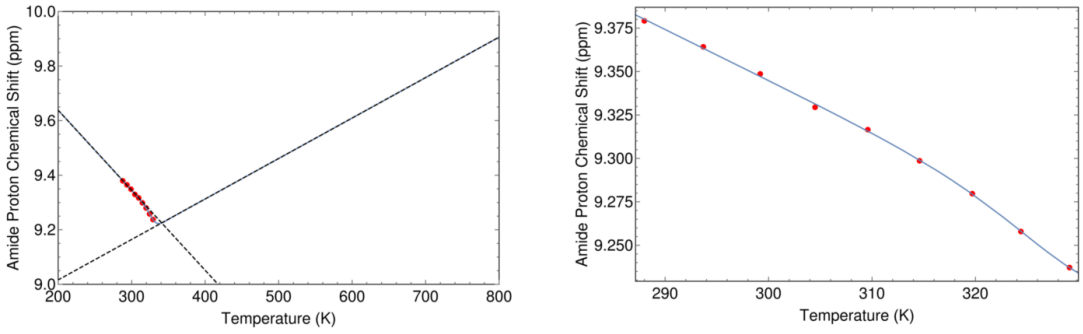


Figure I.121: The minimum RMSE fit of Parent Adnectin residue 68 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

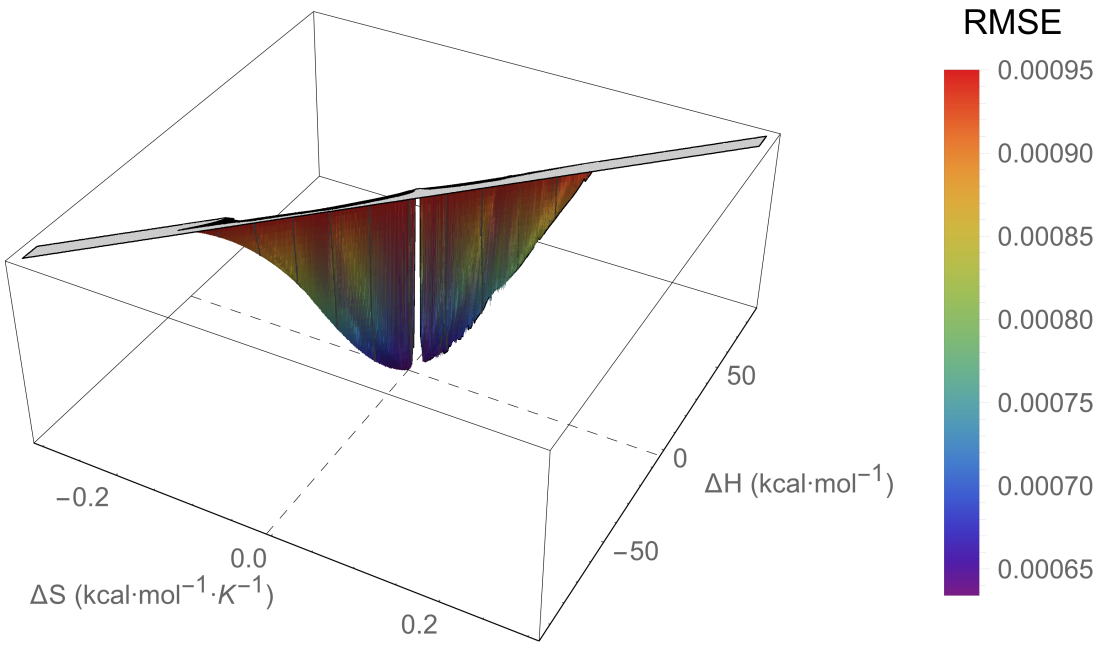


Figure I.122: An RMSE landscape from fitting Parent Adnectin residue 70 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

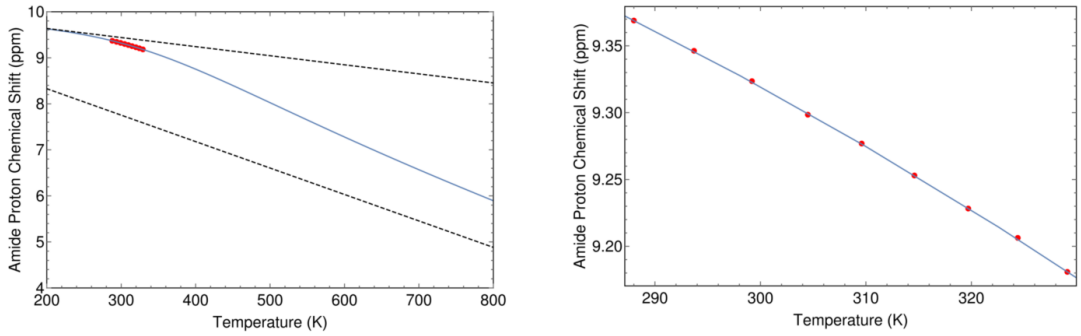


Figure I.123: The minimum RMSE fit of Parent Adnectin residue 70 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

I.7 L78I Adnectin Curvalyzer Results

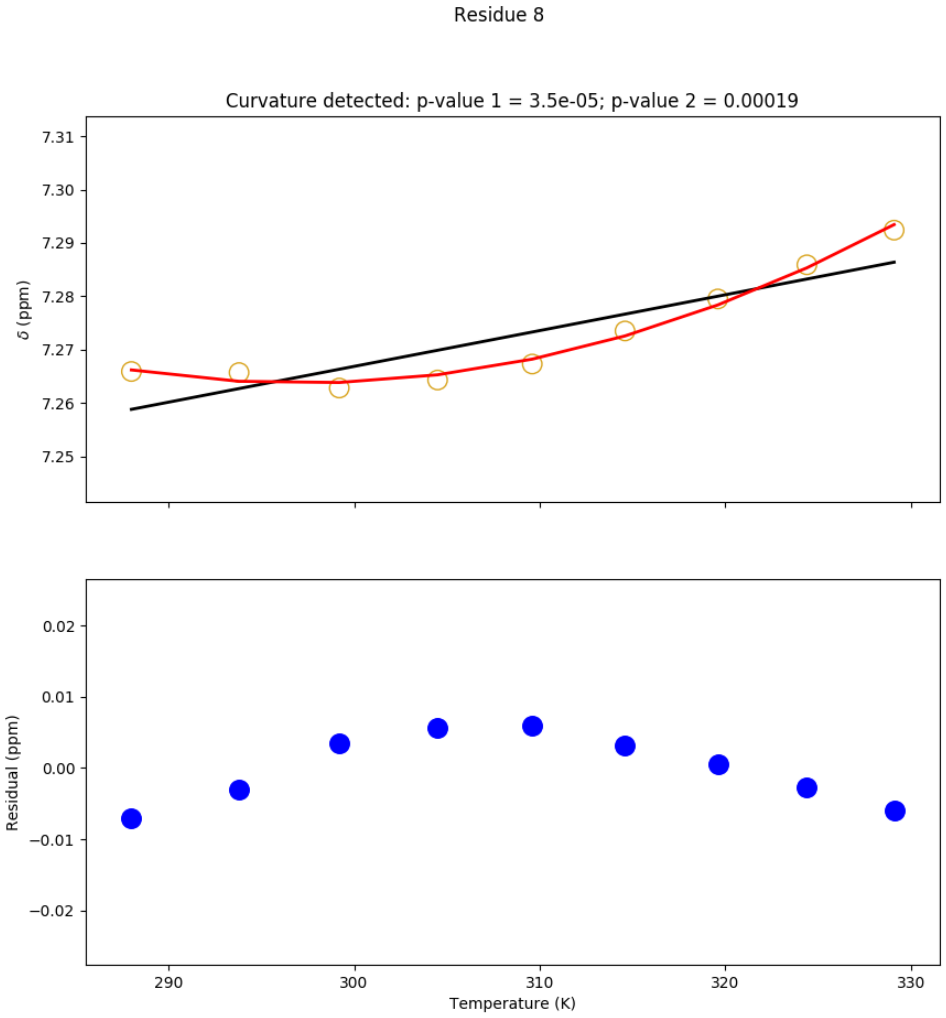


Figure I.124: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 8. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

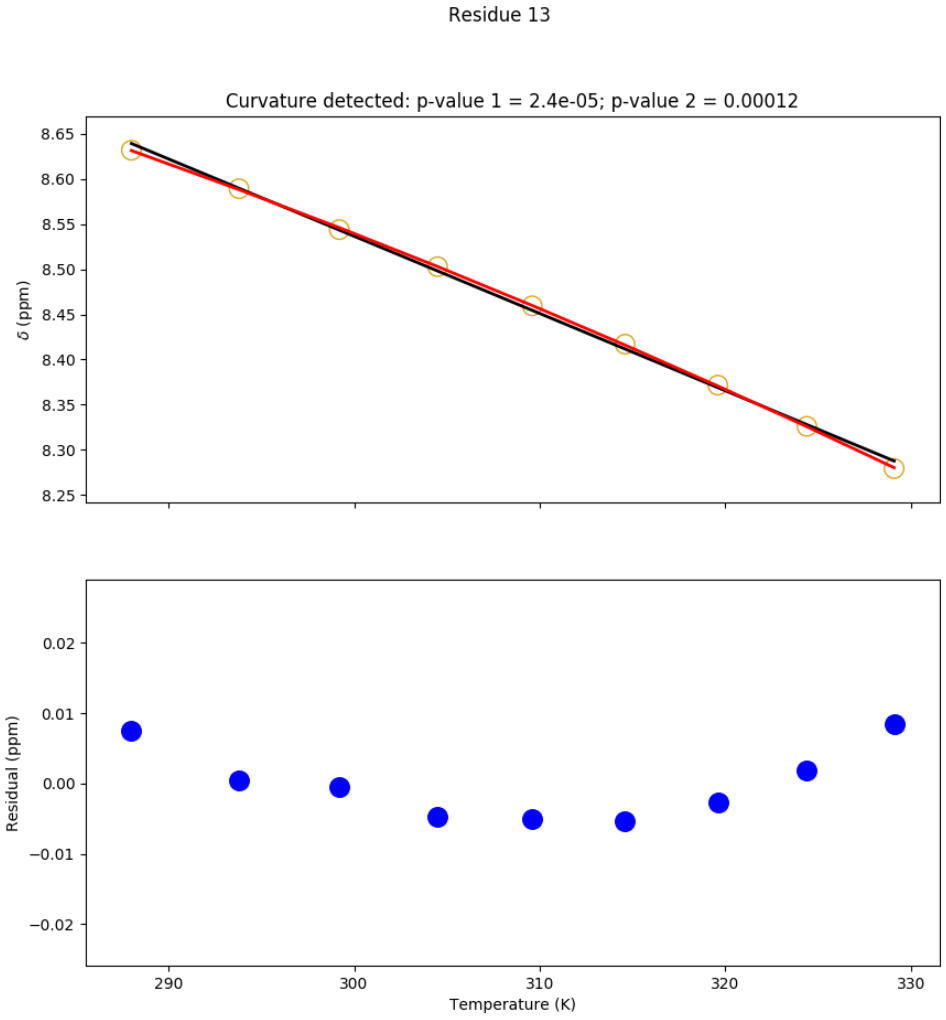


Figure I.125: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 13. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

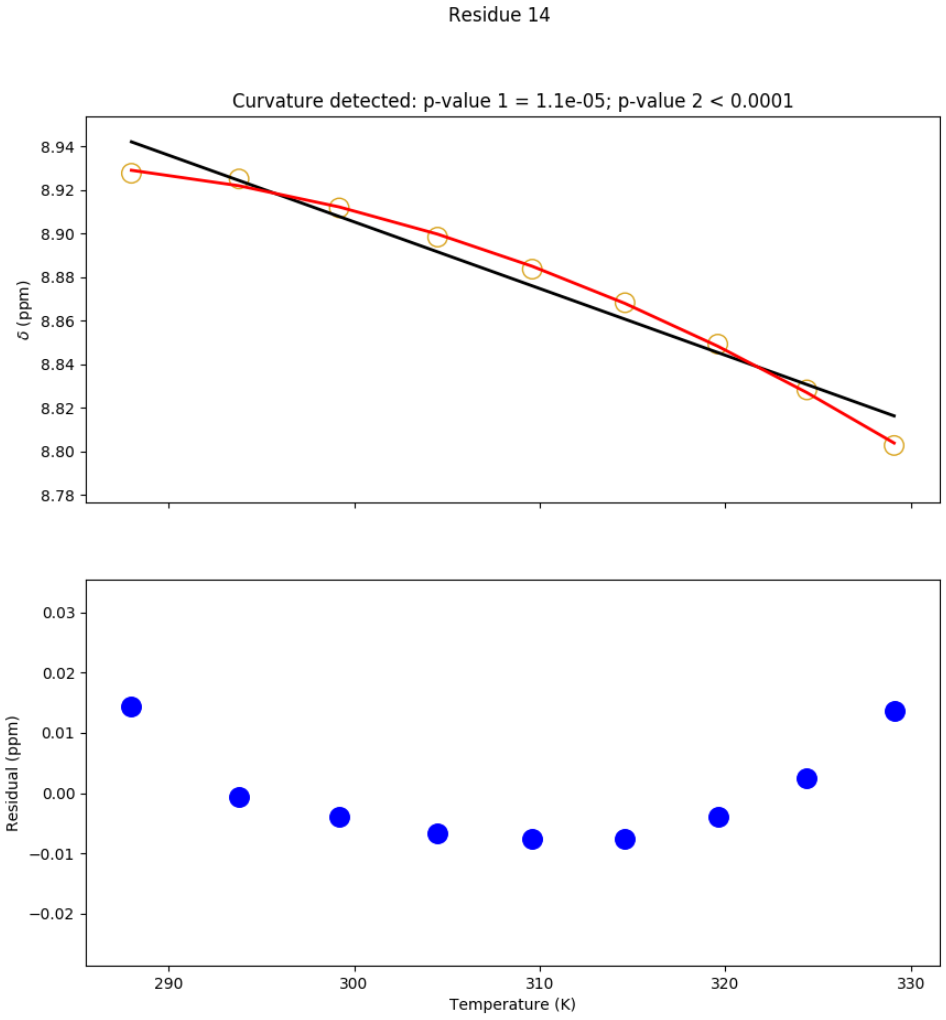


Figure I.126: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 14. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

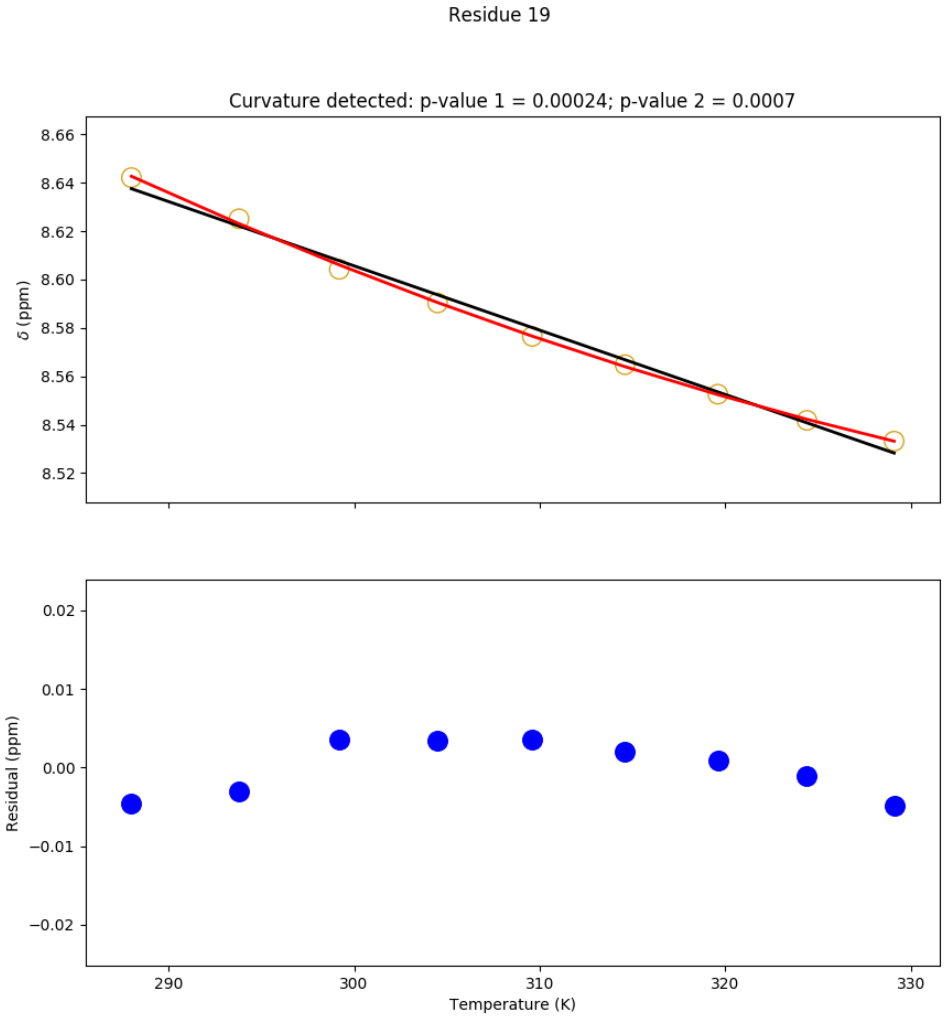


Figure I.127: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 19. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

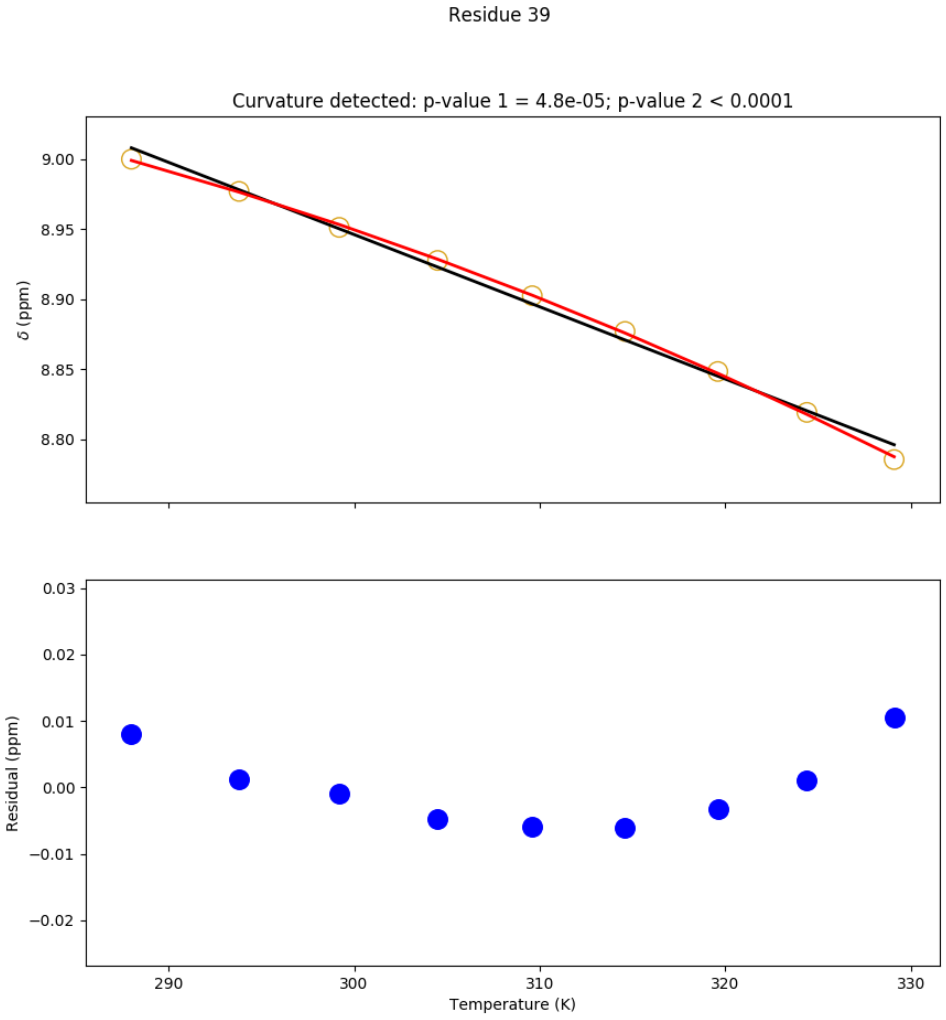


Figure I.128: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 39. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

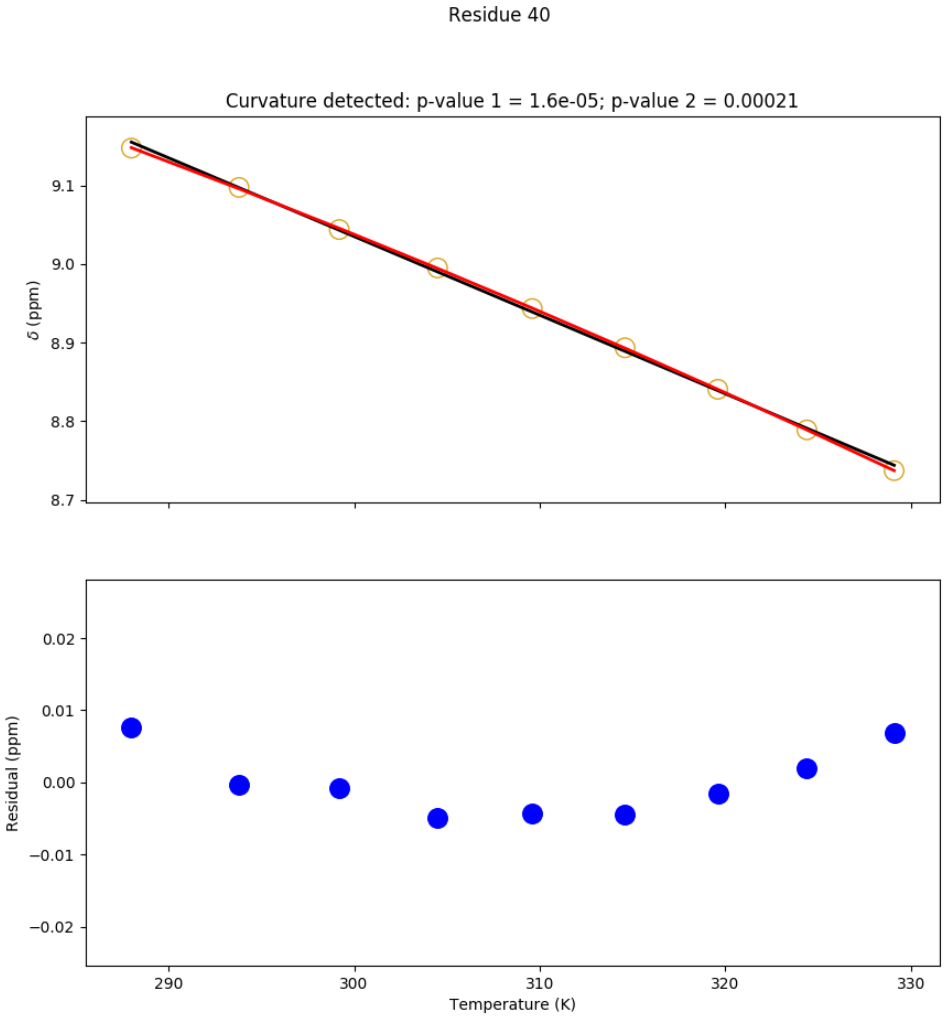


Figure I.129: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 40. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

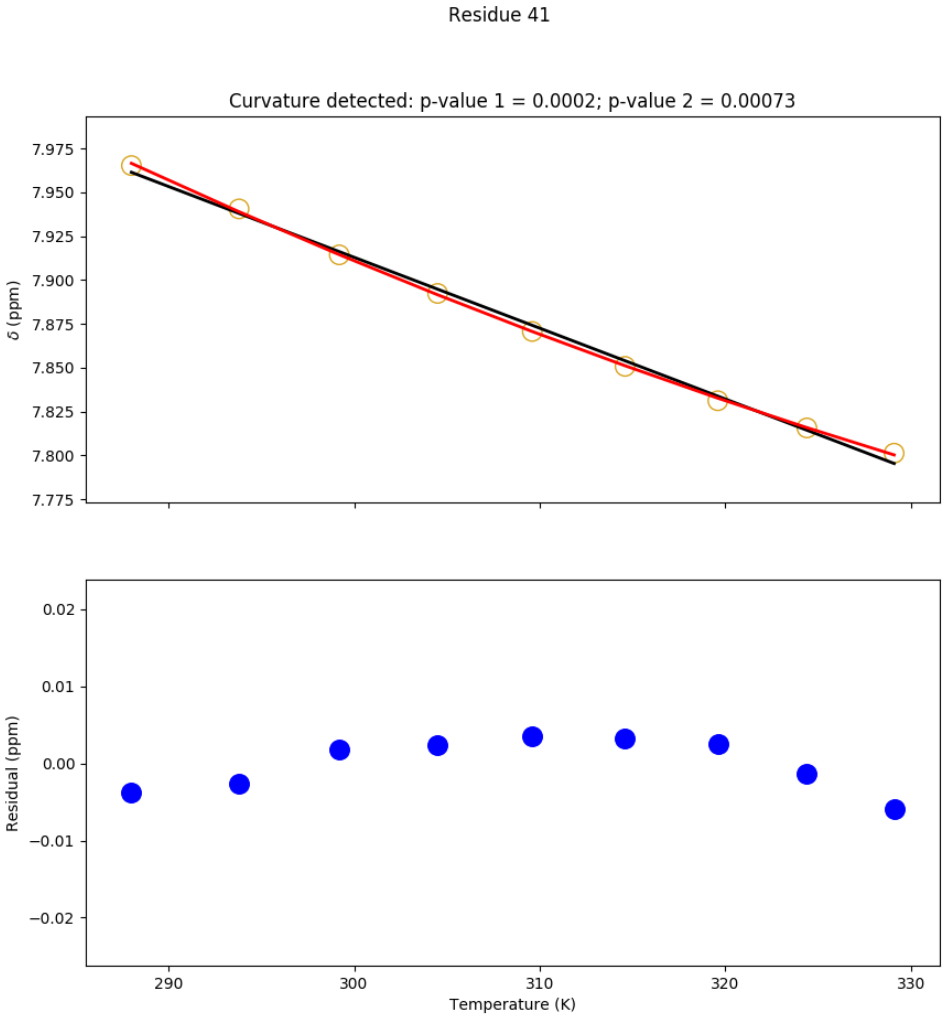


Figure I.130: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 41. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

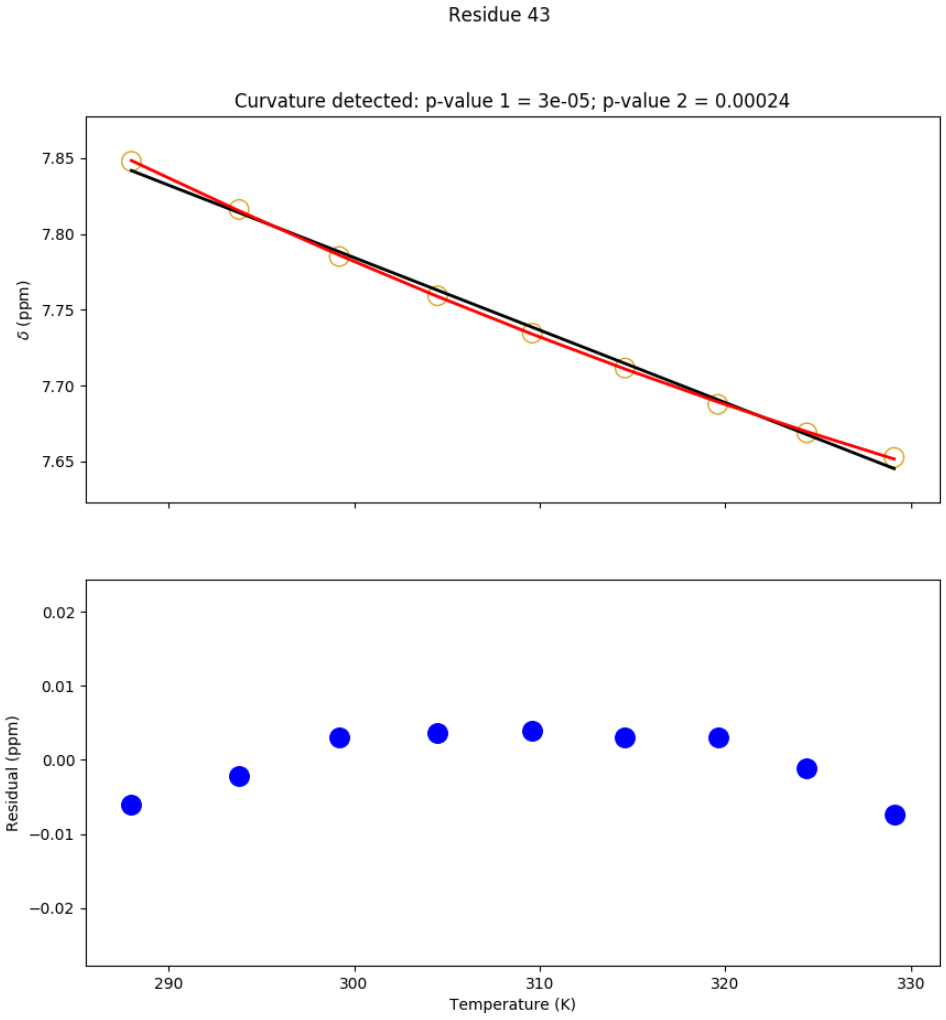


Figure I.131: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 43. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

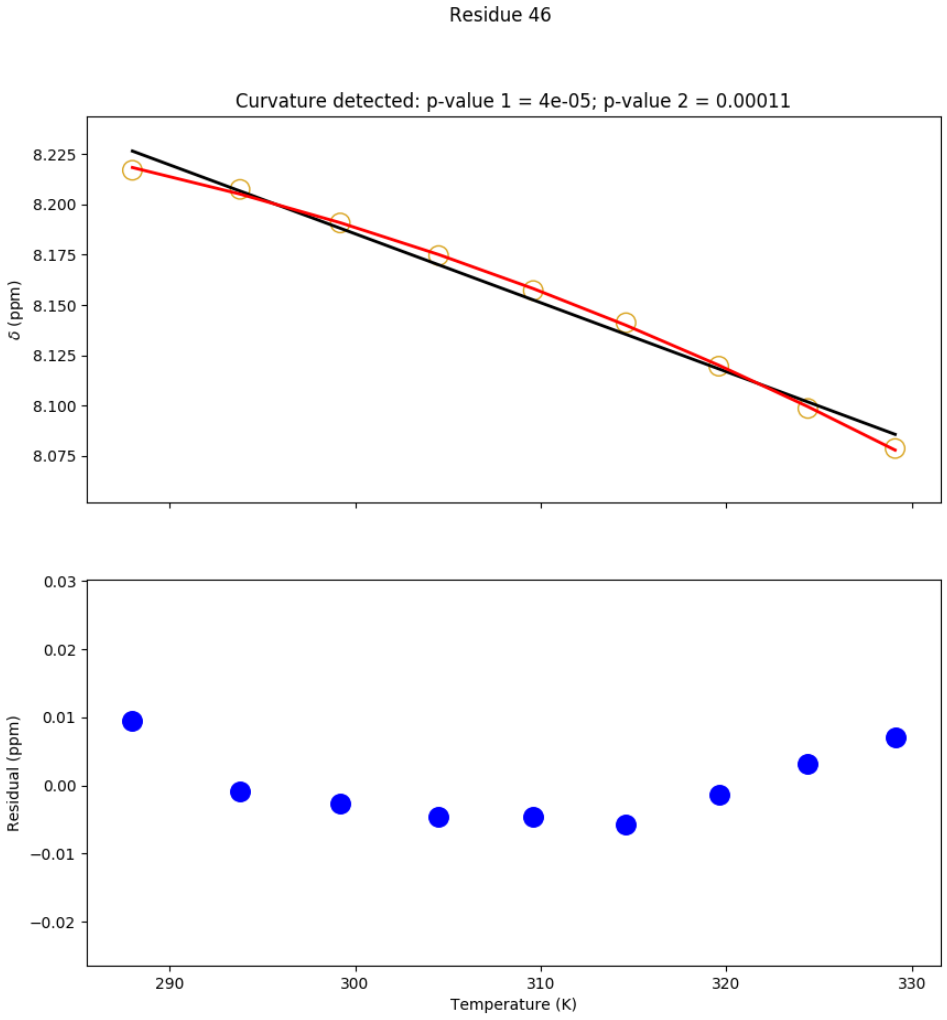


Figure I.132: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 46. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

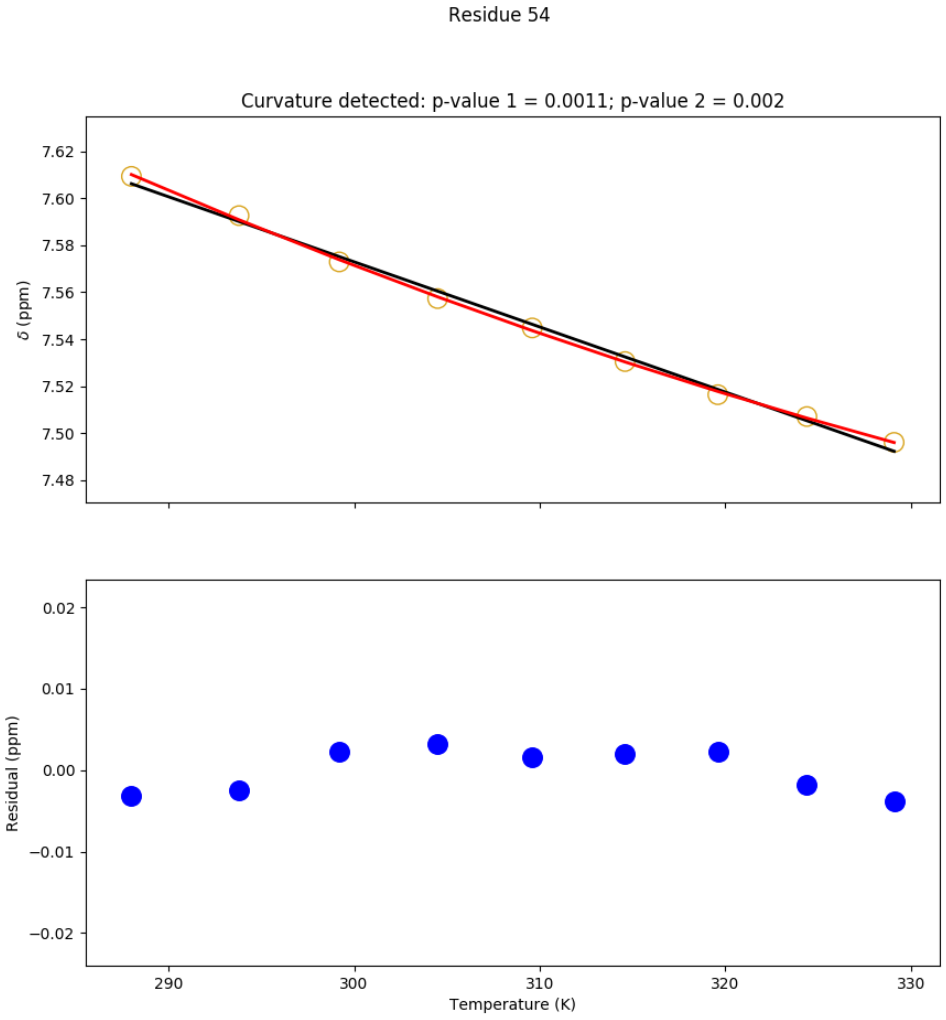


Figure I.133: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 54. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

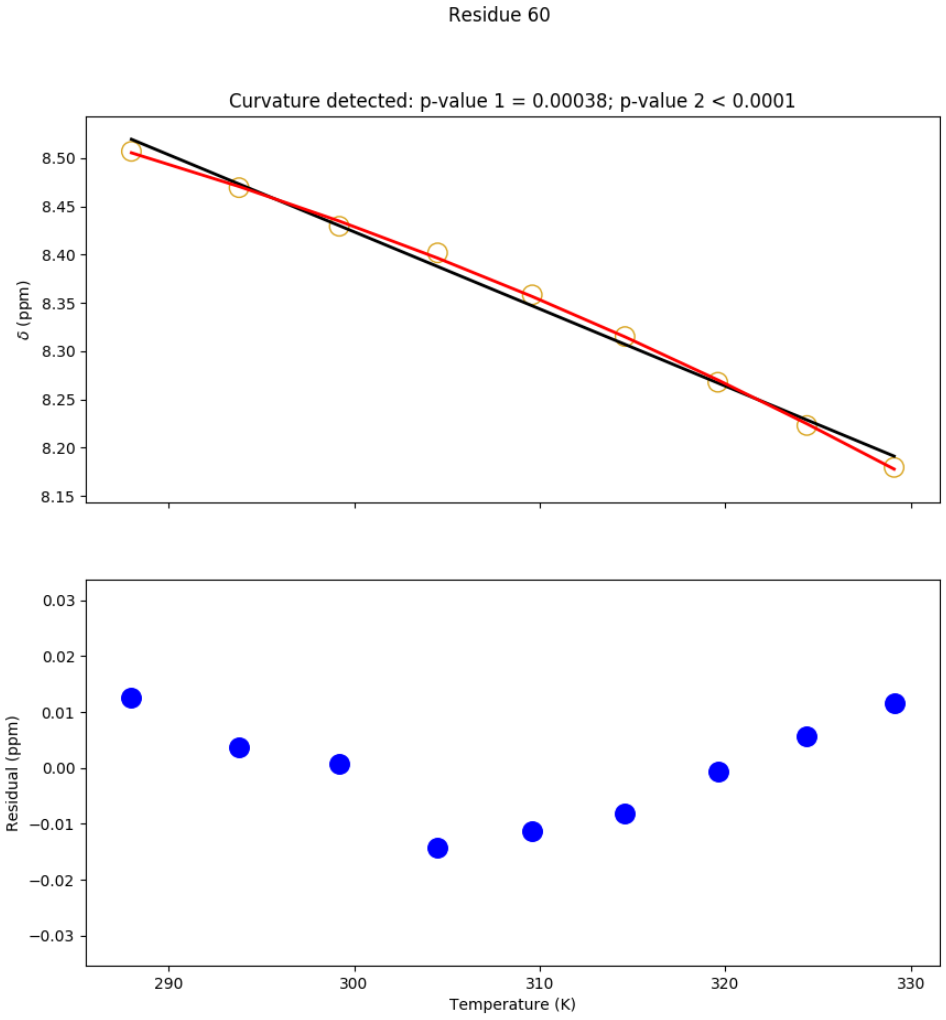


Figure I.134: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 60. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

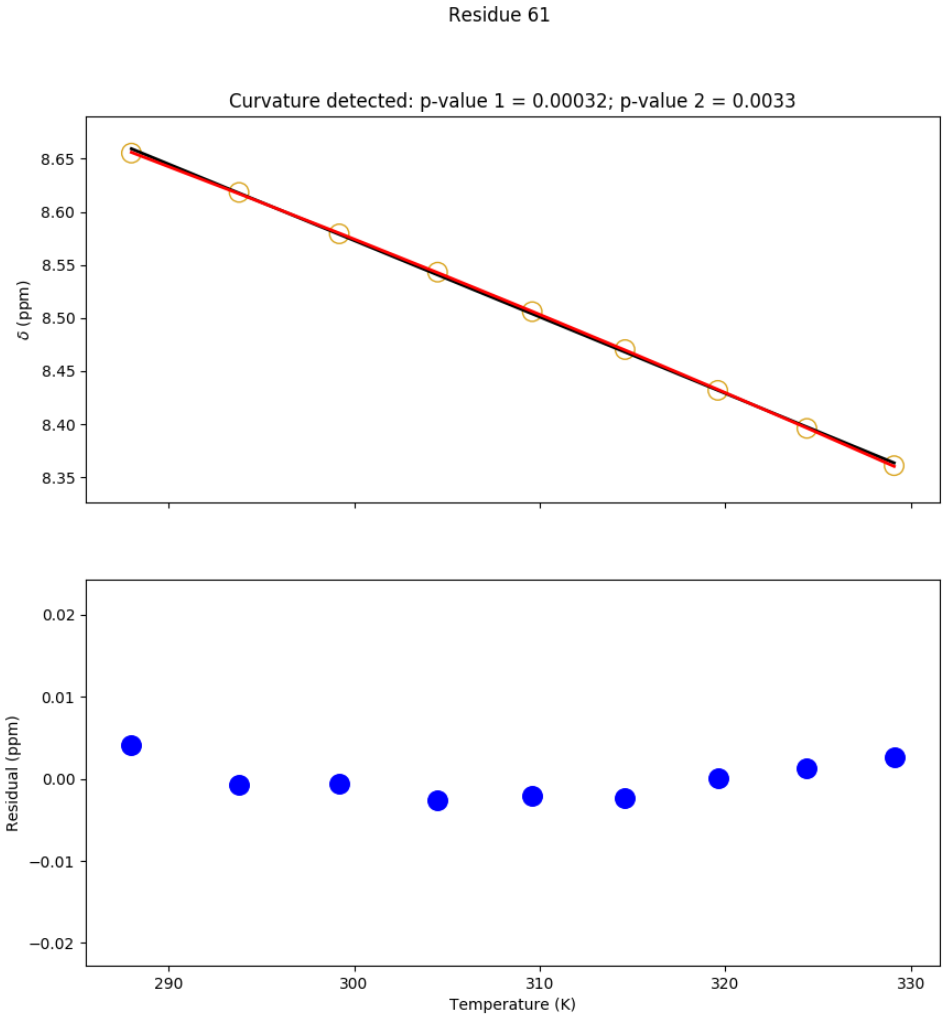


Figure I.135: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 61. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

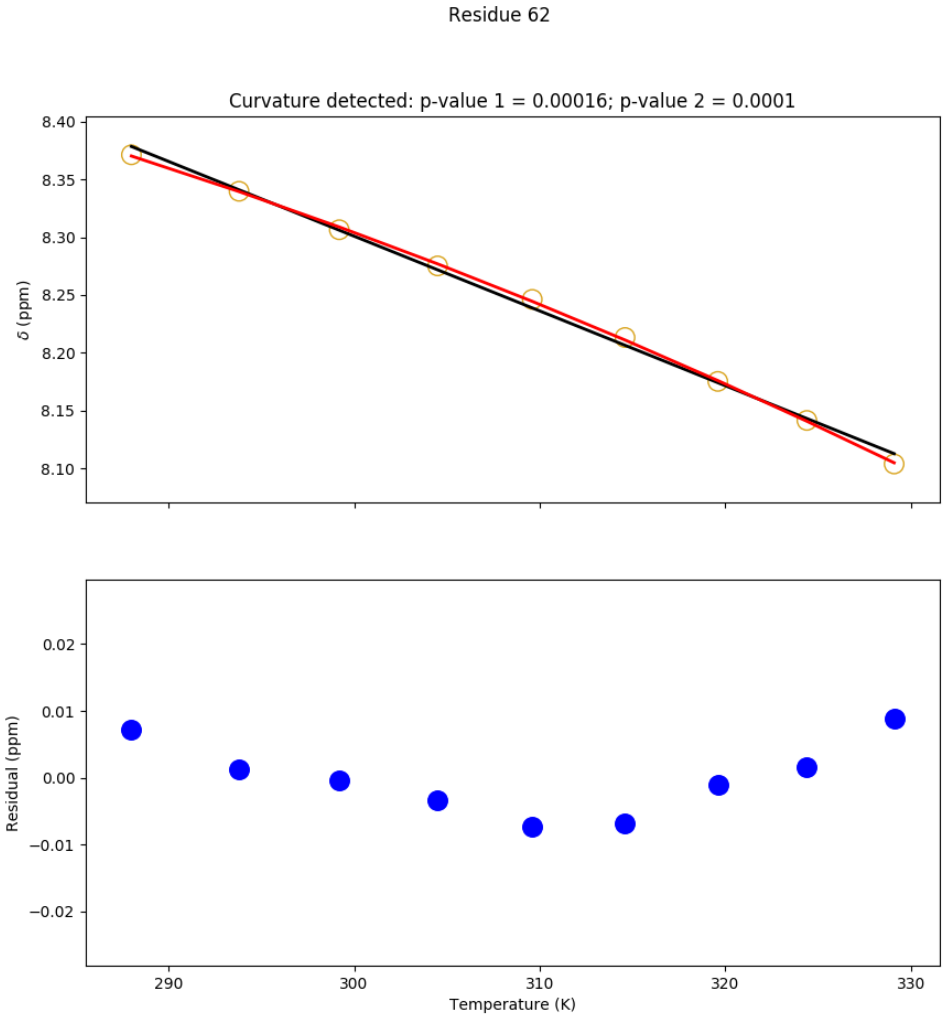


Figure I.136: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 62. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

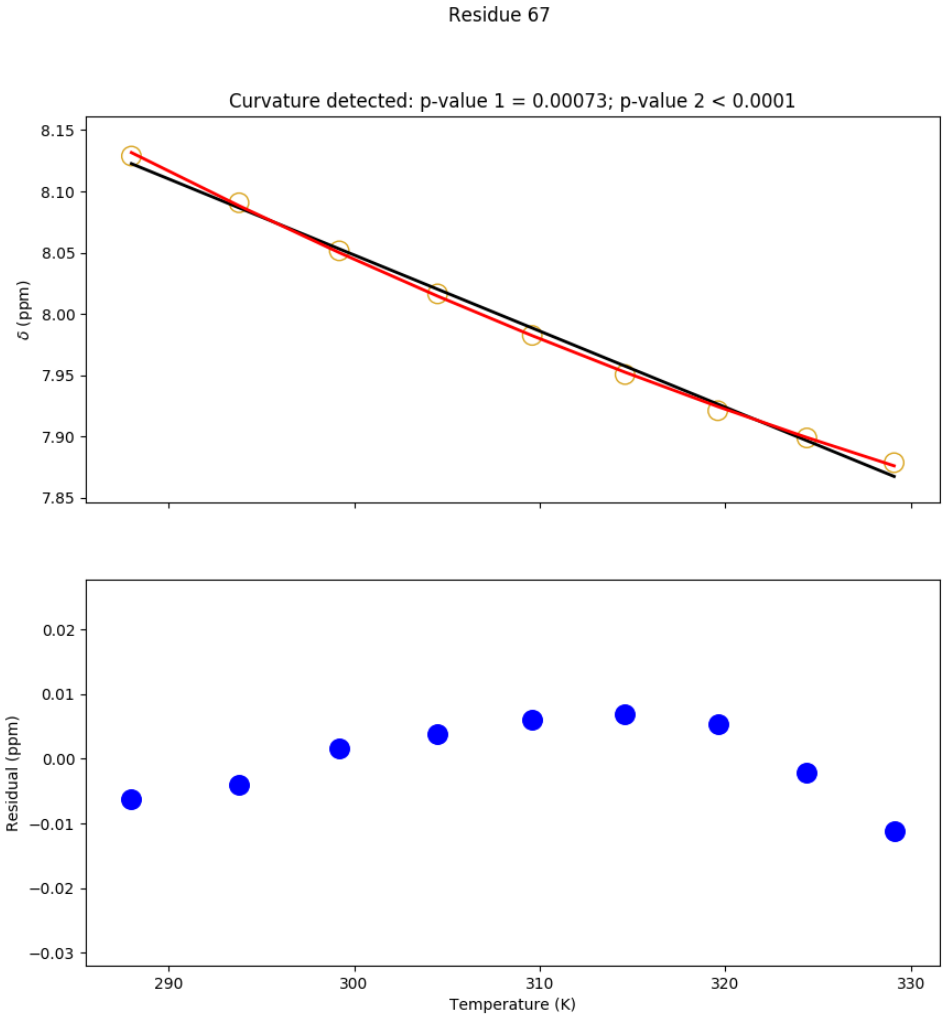


Figure I.137: Temperature dependence of the chemical shift of the L78I Adnectin amide proton from residue 67. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

I.8 L78I Adnectin Curvature Modelling

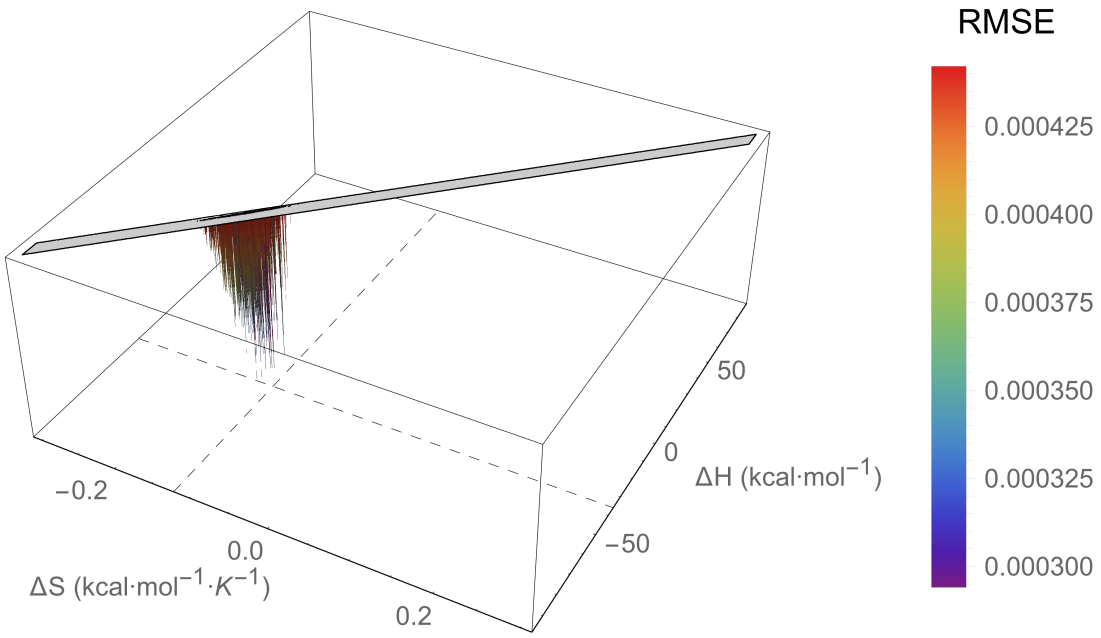


Figure I.138: An RMSE landscape from fitting L78I Adnectin residue 8 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

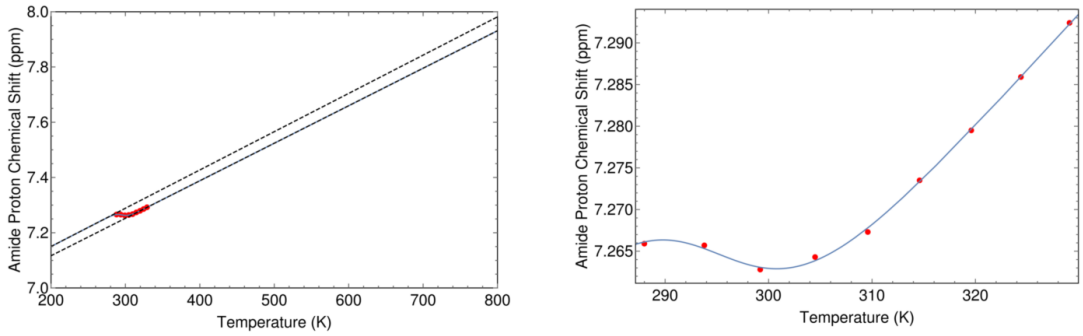


Figure I.139: The minimum RMSE fit of L78I Adnectin residue 8 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

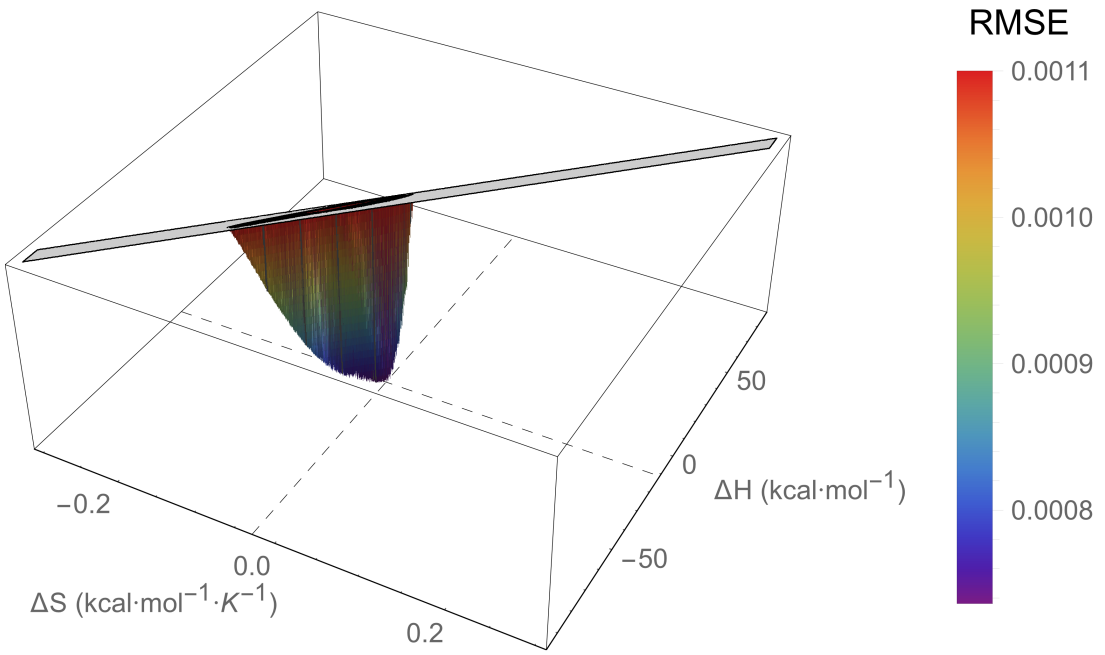


Figure I.140: An RMSE landscape from fitting L78I Adnectin residue 13 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

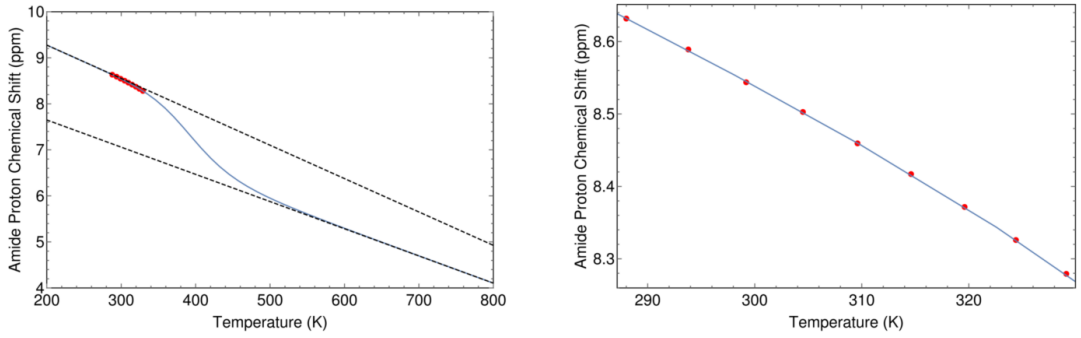


Figure I.141: The minimum RMSE fit of L78I Adnectin residue 13 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

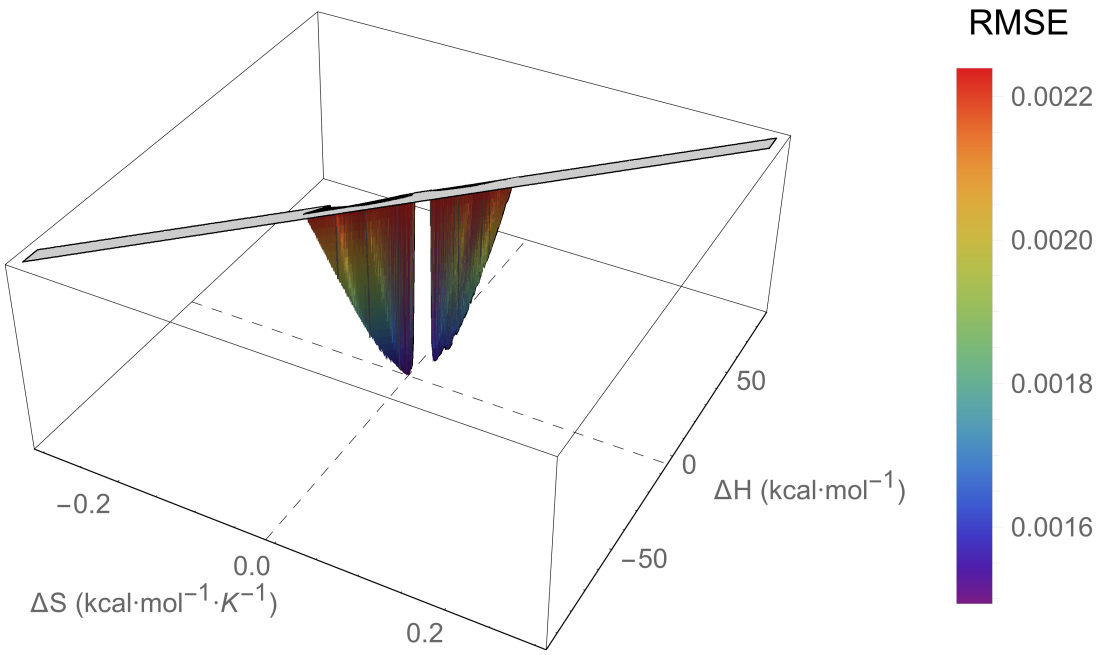


Figure I.142: An RMSE landscape from fitting L78I Adnectin residue 14 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

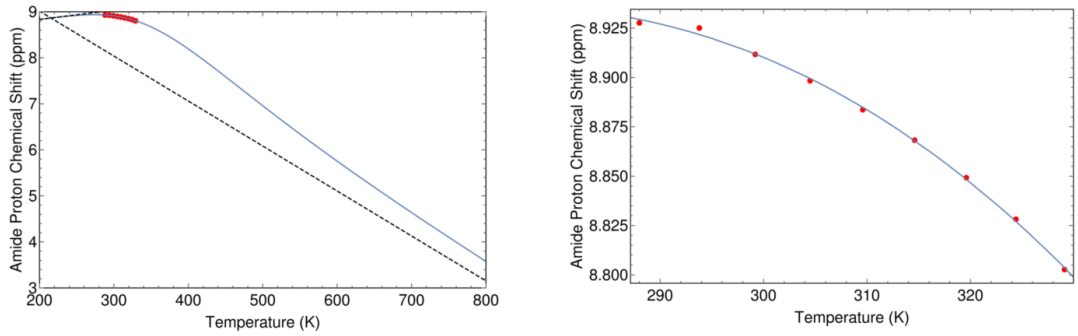


Figure I.143: The minimum RMSE fit of L78I Adnectin residue 14 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

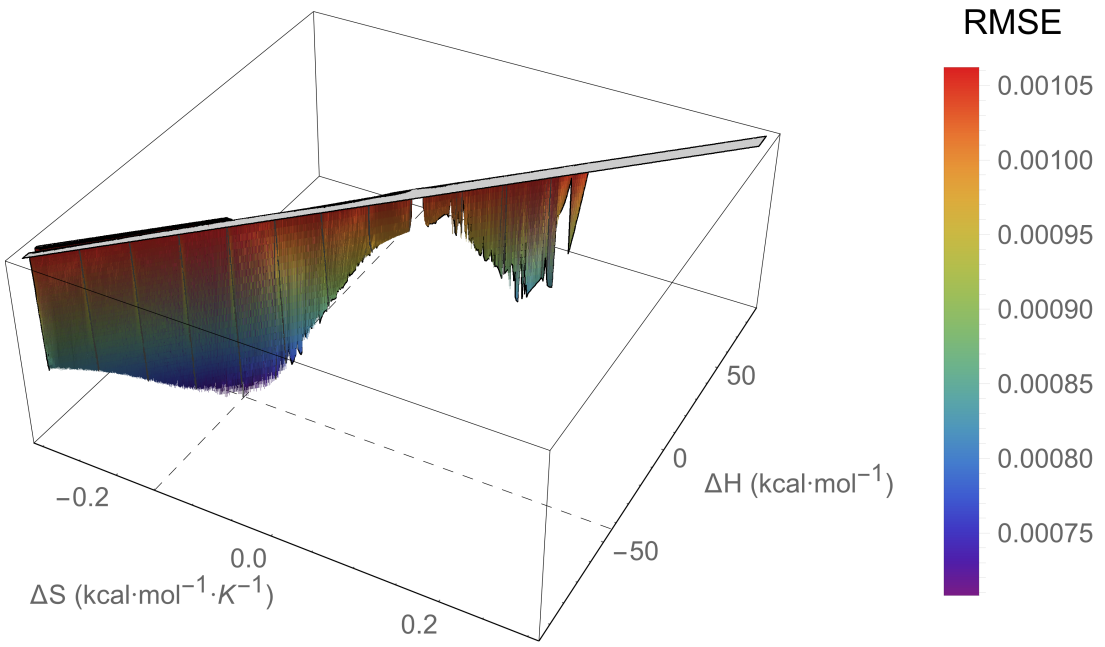


Figure I.144: An RMSE landscape from fitting L78I Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

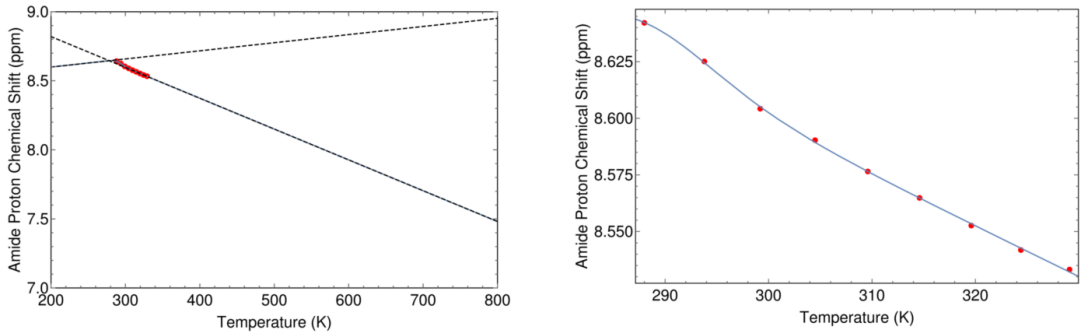


Figure I.145: The minimum RMSE fit of L78I Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

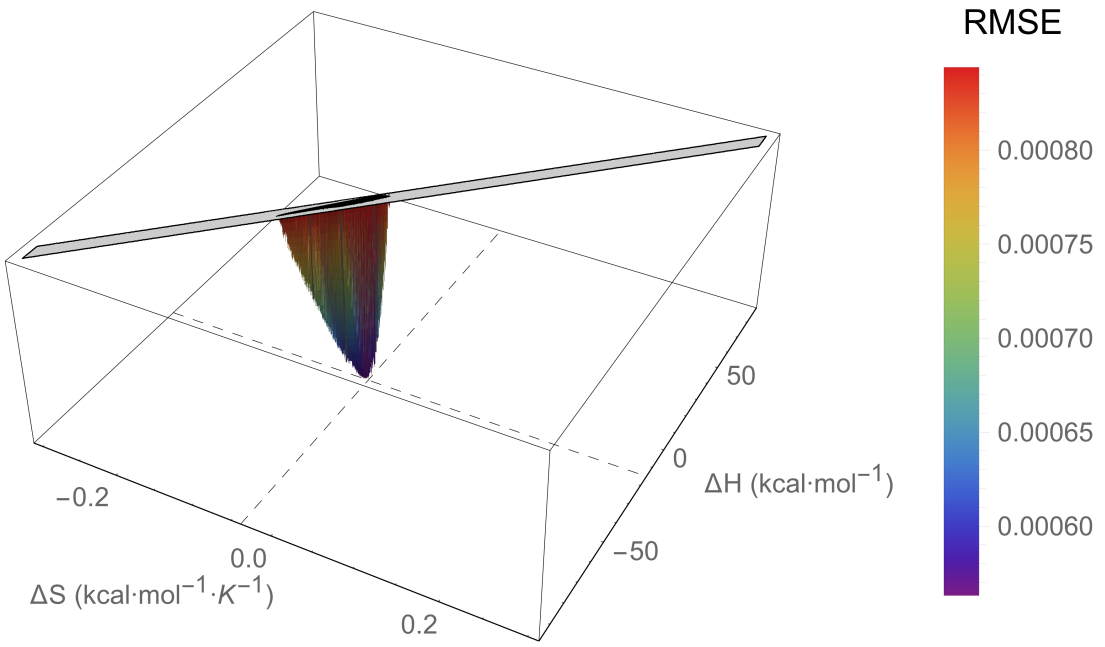


Figure I.146: An RMSE landscape from fitting L78I Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

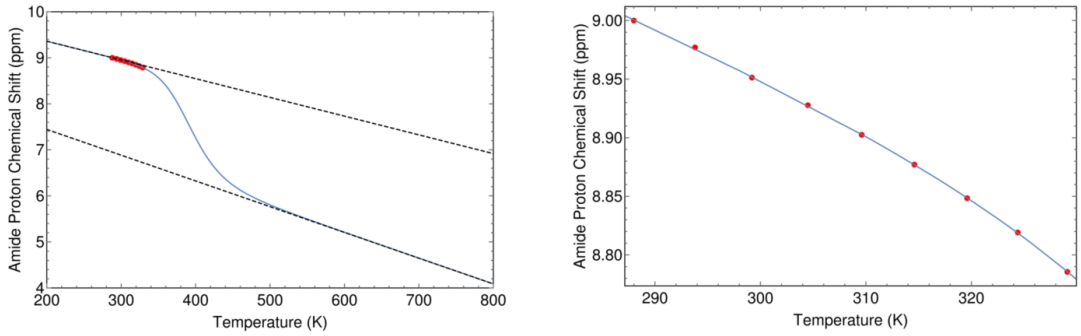


Figure I.147: The minimum RMSE fit of L78I Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

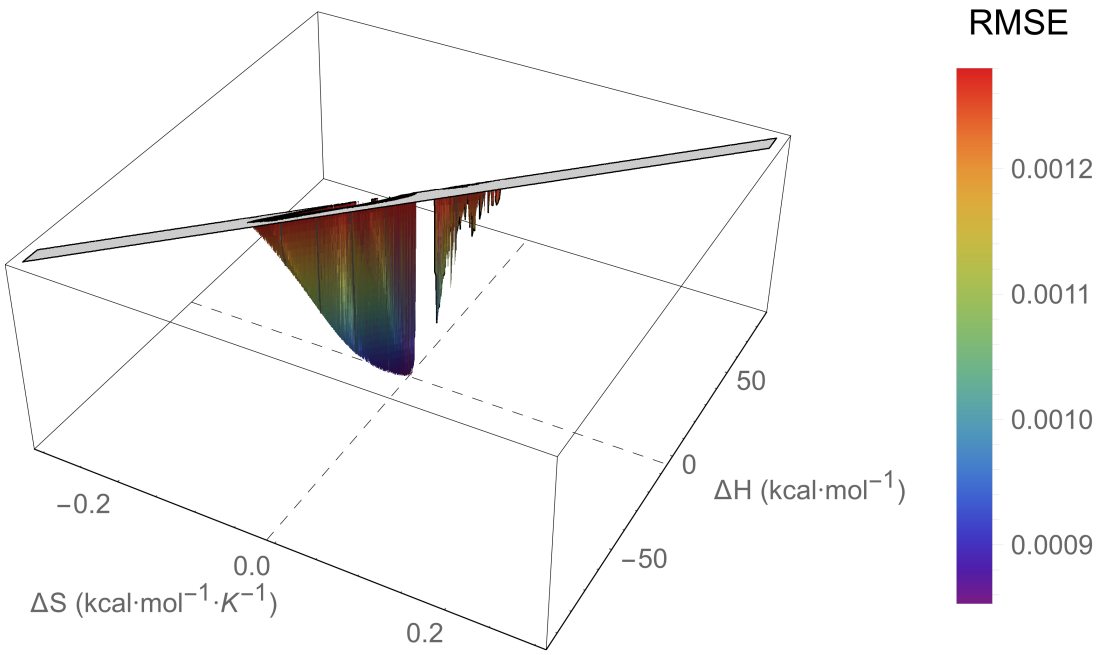


Figure I.148: An RMSE landscape from fitting L78I Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

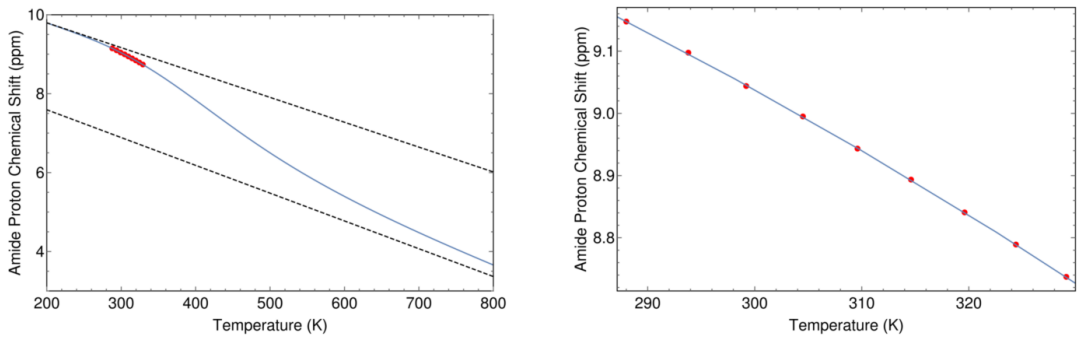


Figure I.149: The minimum RMSE fit of L78I Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

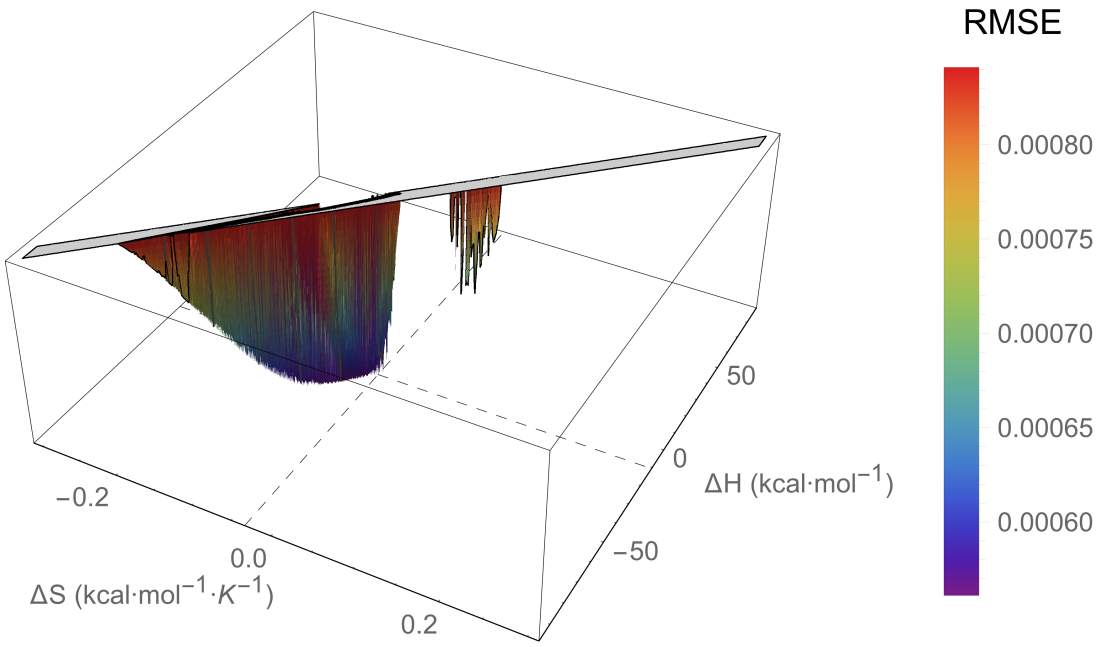


Figure I.150: An RMSE landscape from fitting L78I Adnectin residue 41 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

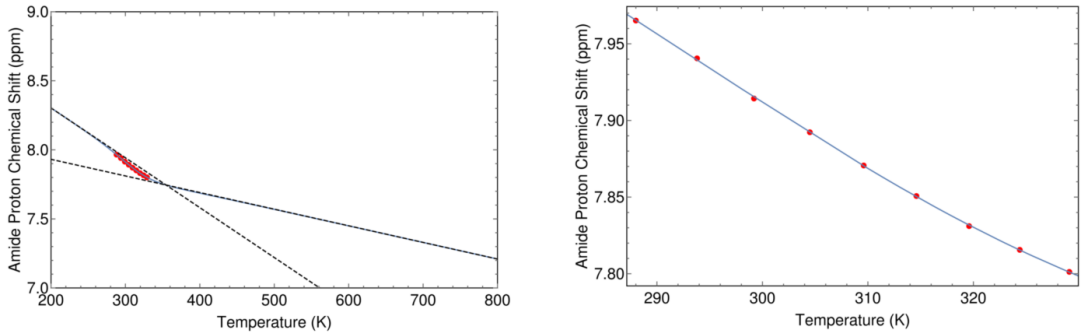


Figure I.151: The minimum RMSE fit of L78I Adnectin residue 41 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

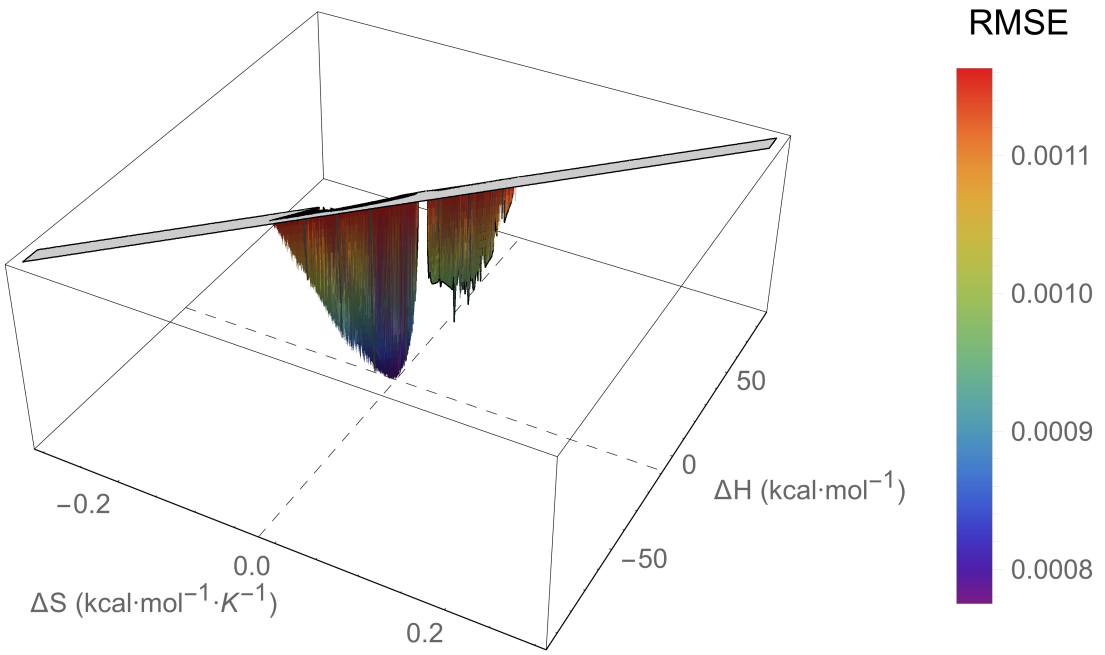


Figure I.152: An RMSE landscape from fitting L78I Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

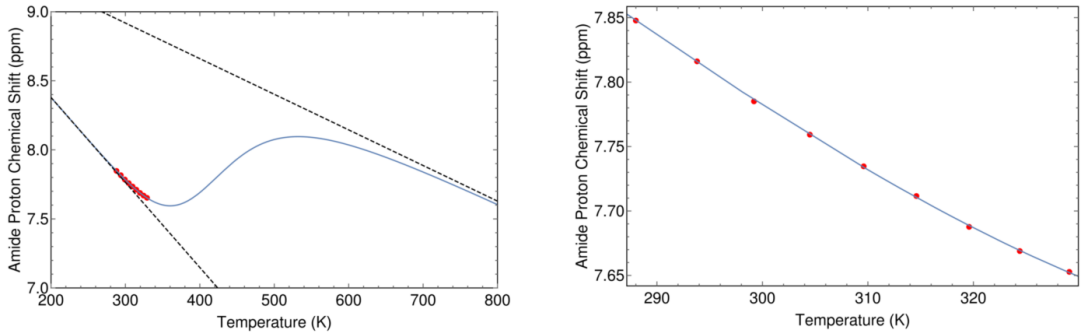


Figure I.153: The minimum RMSE fit of L78I Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

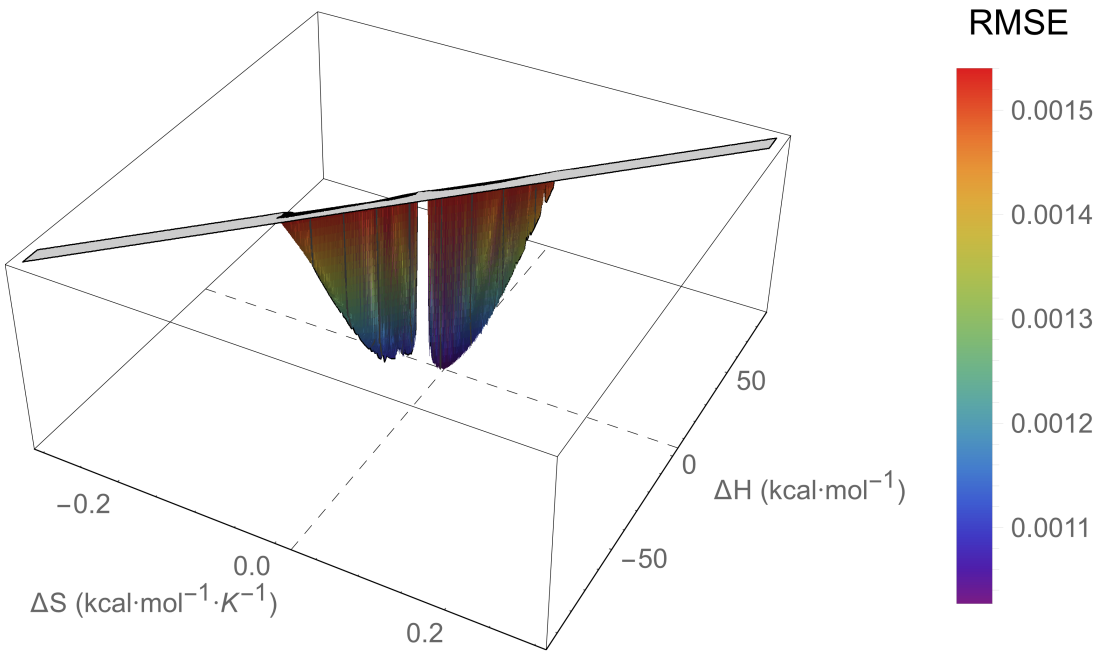


Figure I.154: An RMSE landscape from fitting L78I Adnectin residue 46 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

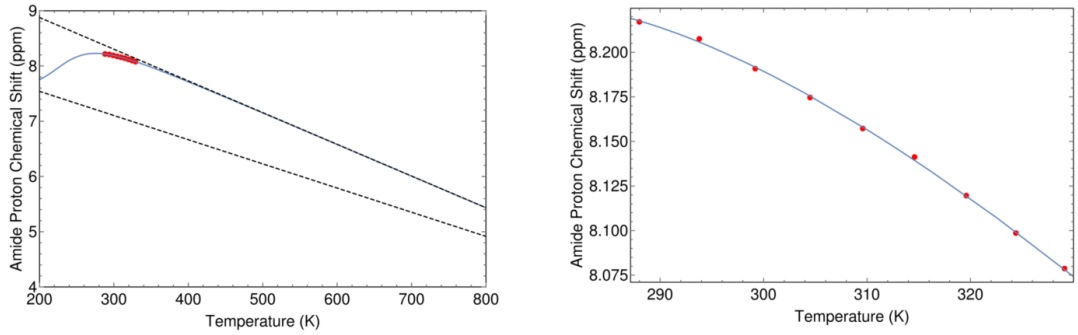


Figure I.155: The minimum RMSE fit of L78I Adnectin residue 46 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

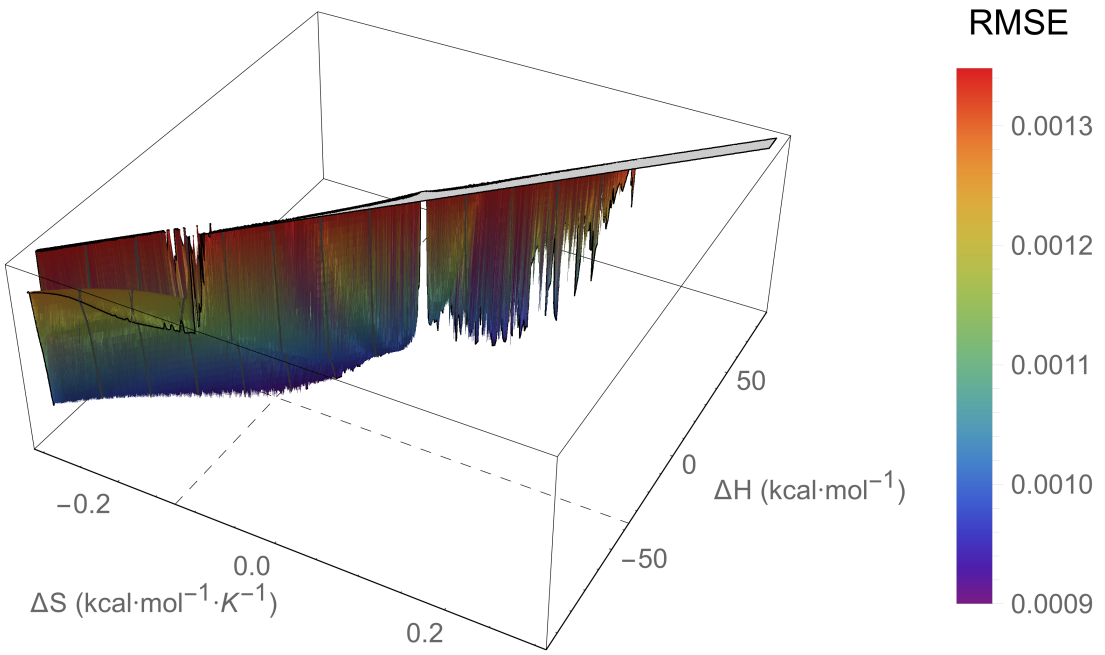


Figure I.156: An RMSE landscape from fitting L78I Adnectin residue 54 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

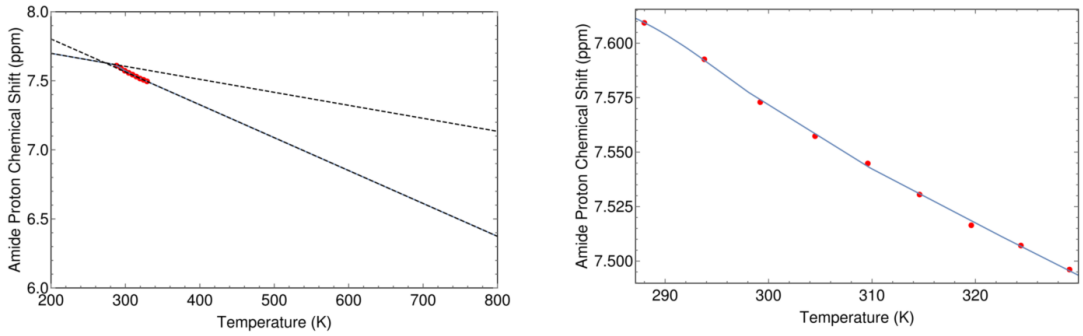


Figure I.157: The minimum RMSE fit of L78I Adnectin residue 54 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

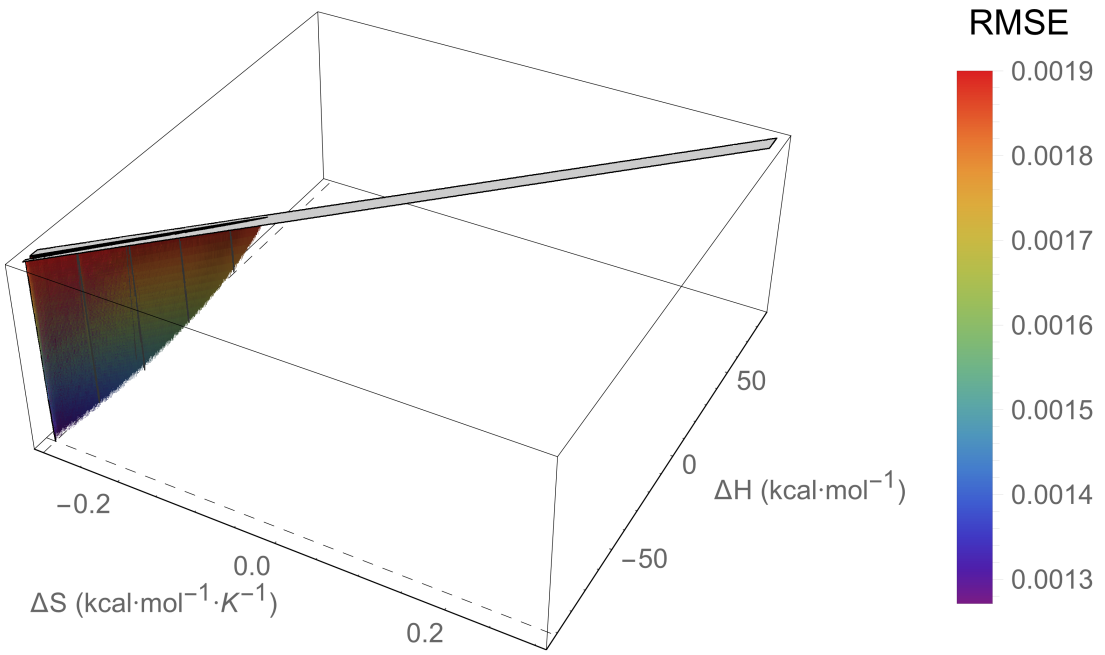


Figure I.158: An RMSE landscape from fitting L78I Adnectin residue 60 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

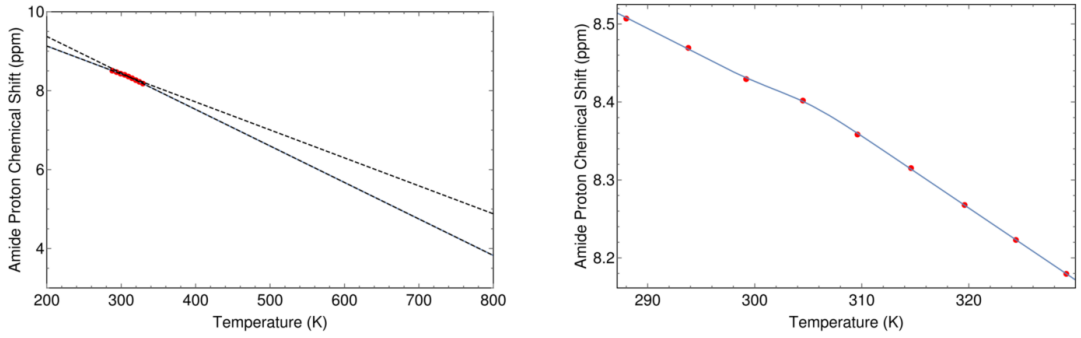


Figure I.159: The minimum RMSE fit of L78I Adnectin residue 60 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

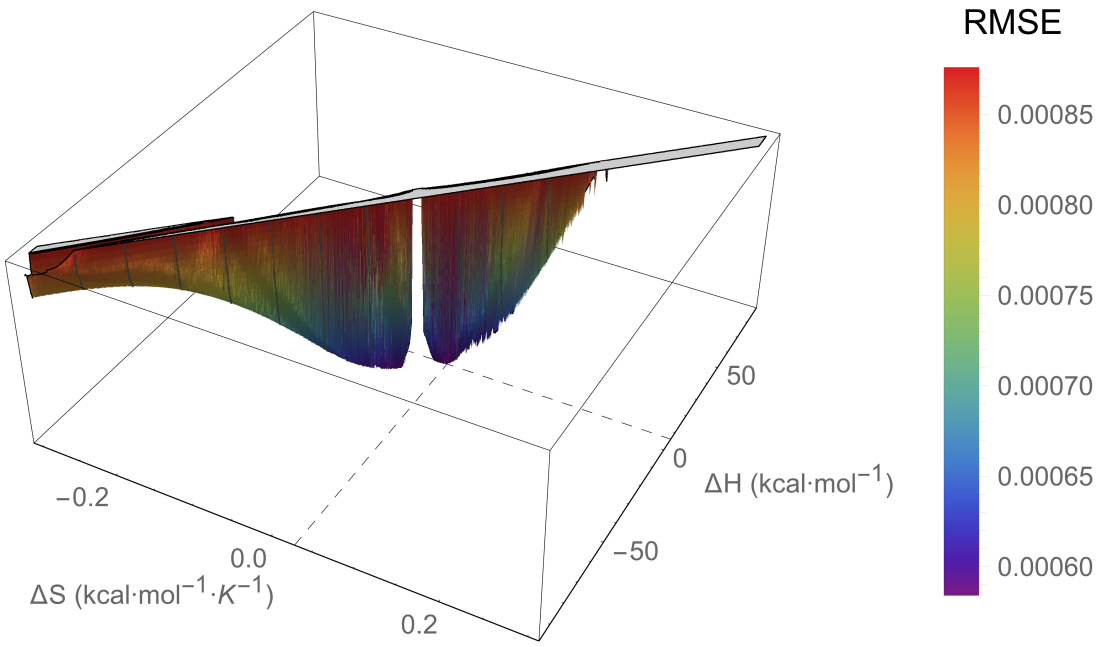


Figure I.160: An RMSE landscape from fitting L78I Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

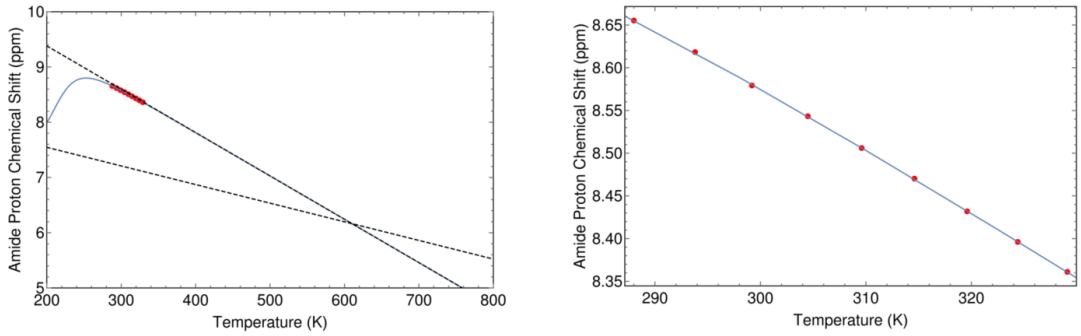


Figure I.161: The minimum RMSE fit of L78I Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

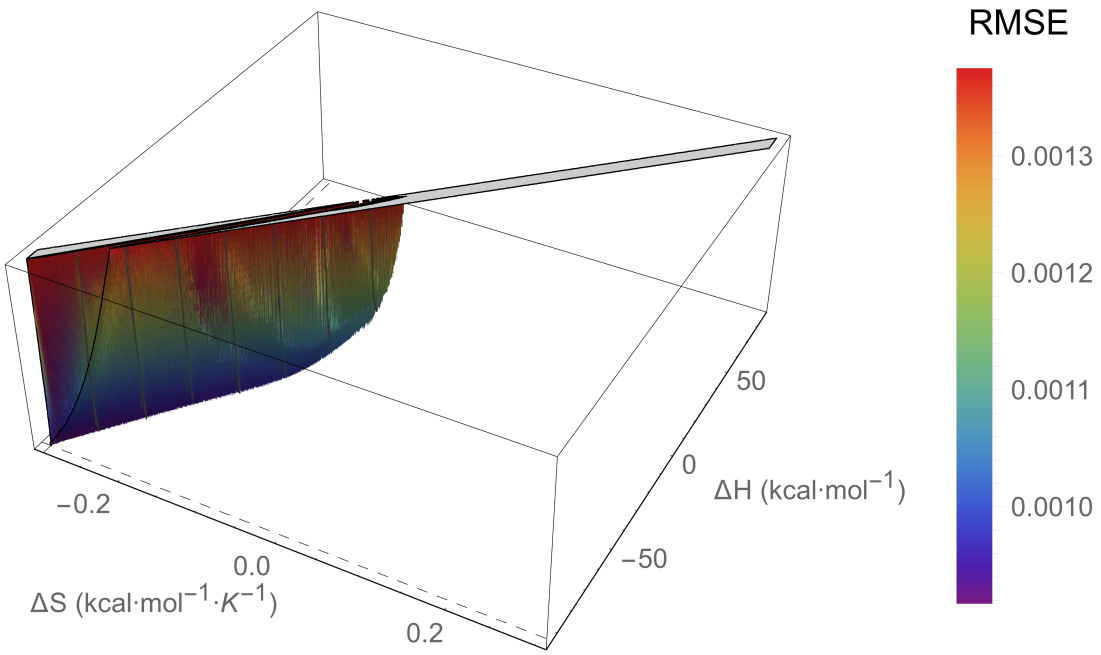


Figure I.162: An RMSE landscape from fitting L78I Adnectin residue 62 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE_{min}) are indicated in gray.

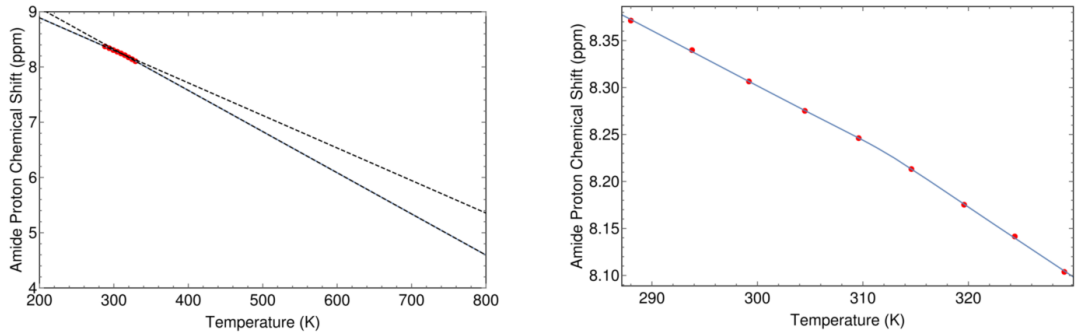


Figure I.163: The minimum RMSE fit of L78I Adnectin residue 62 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

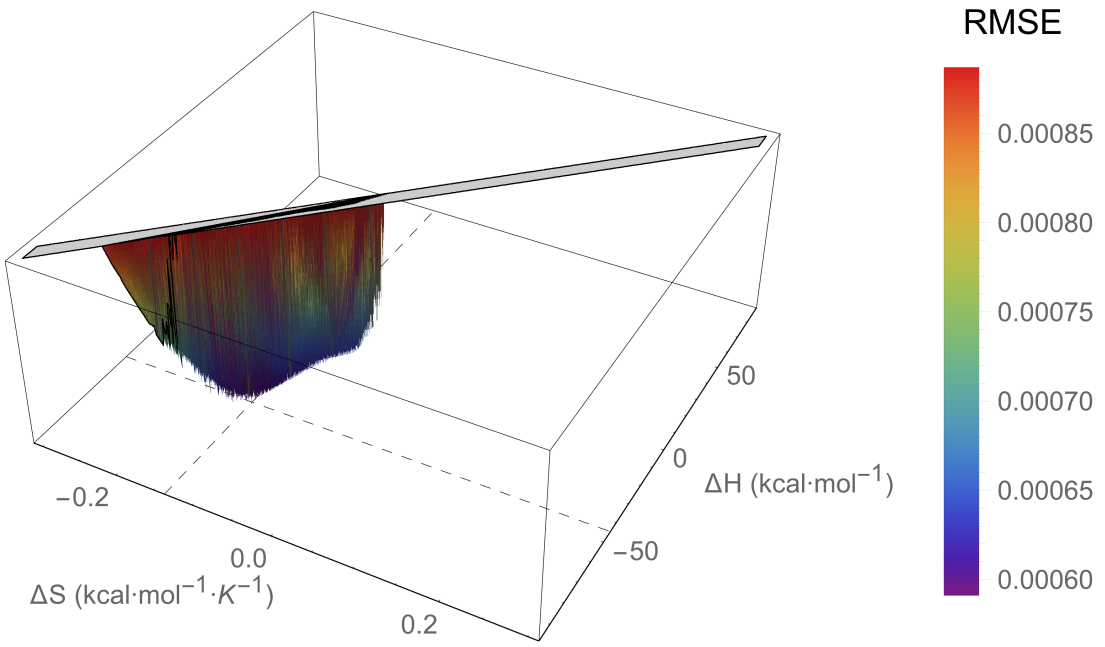


Figure I.164: An RMSE landscape from fitting L78I Adnectin residue 67 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

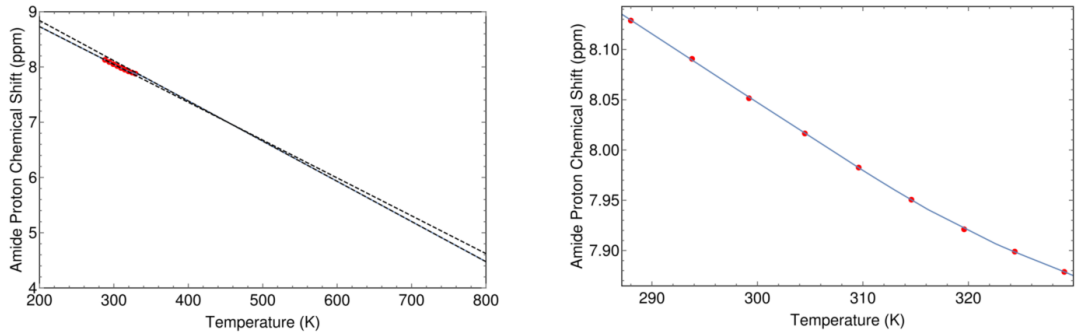


Figure I.165: The minimum RMSE fit of L78I Adnectin residue 67 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

I.9 V75R Adnectin Curvalyzer Results

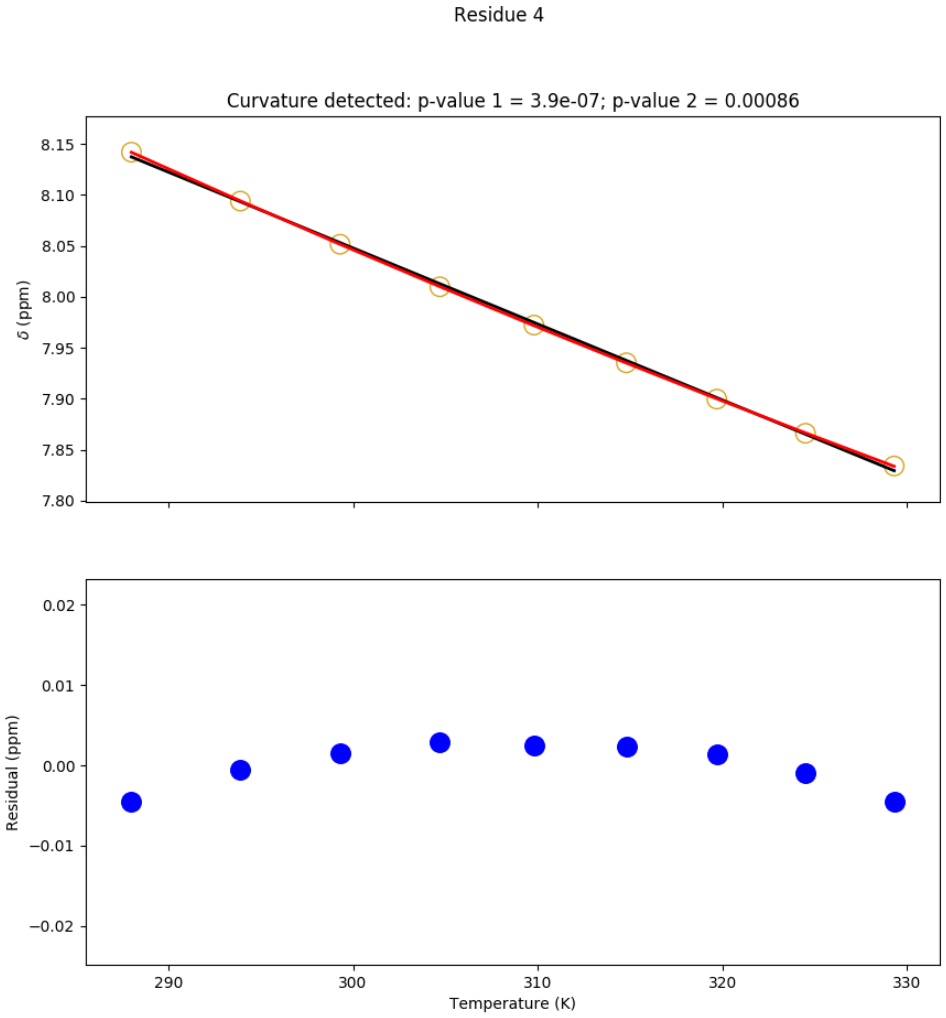


Figure I.166: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 4. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

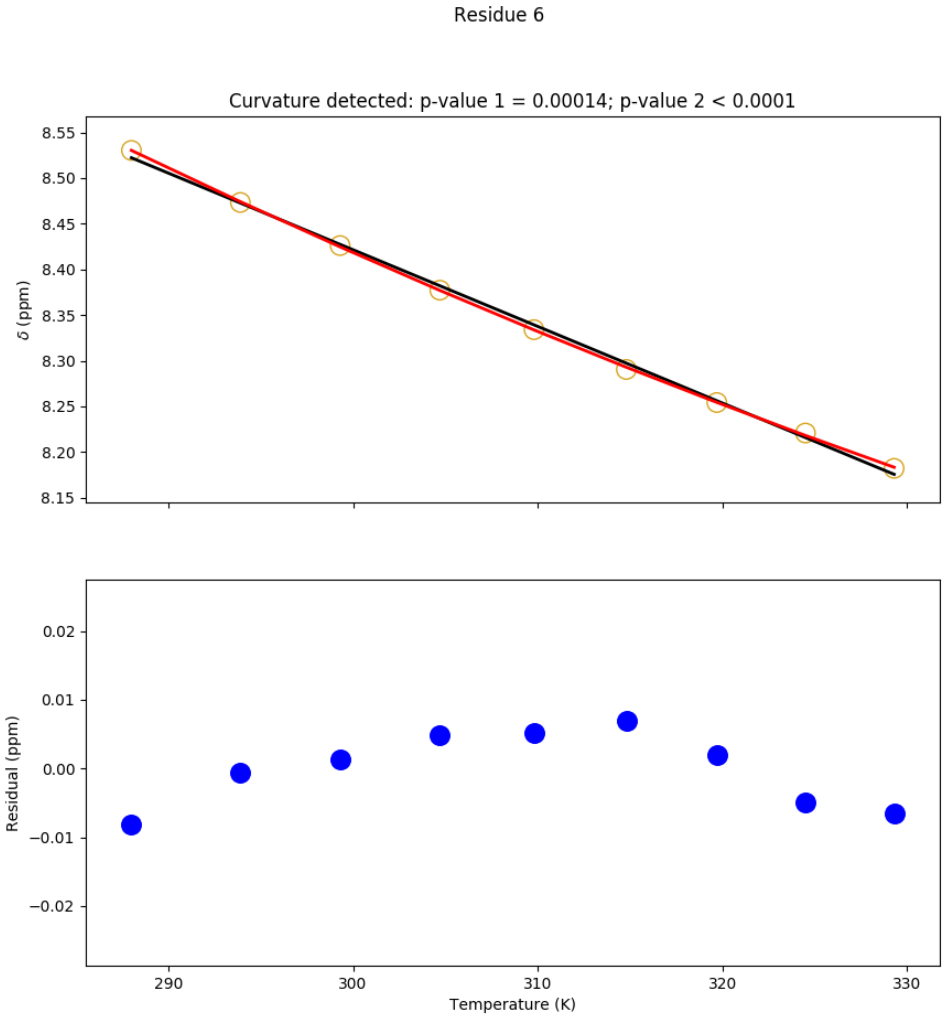


Figure I.167: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 6. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

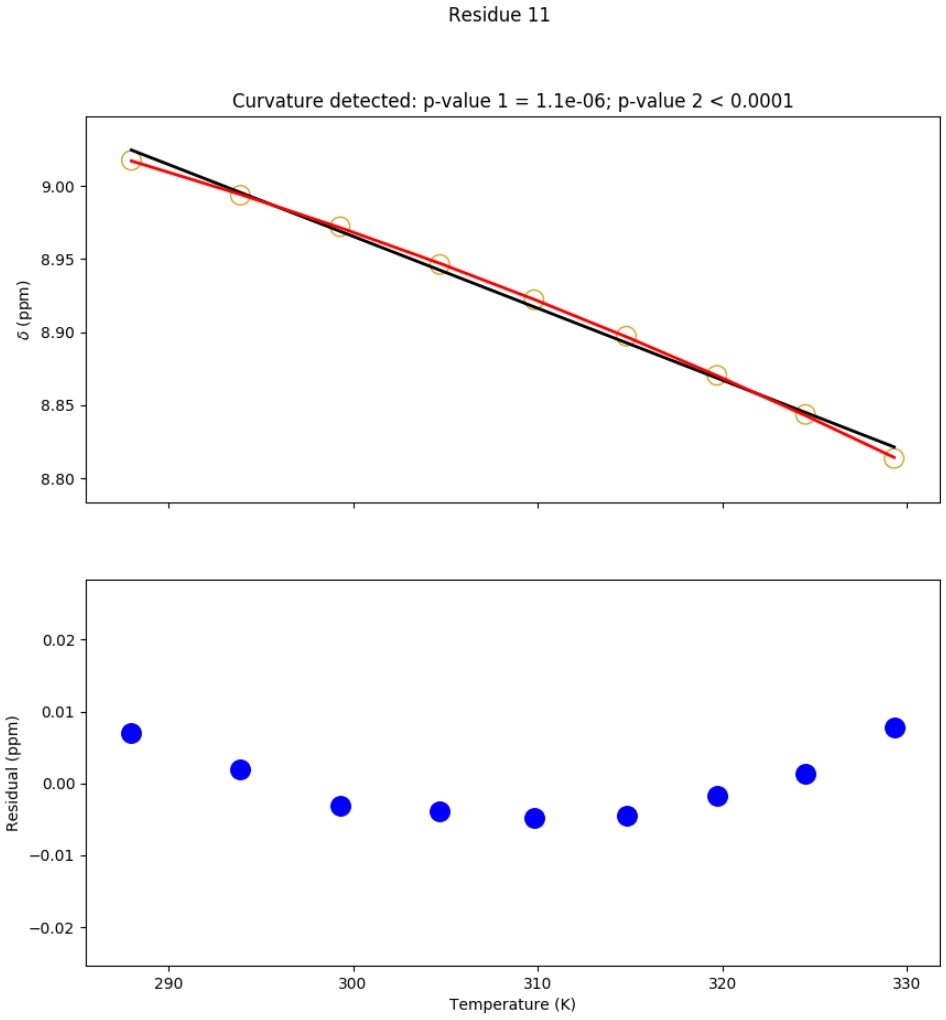


Figure I.168: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 11. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

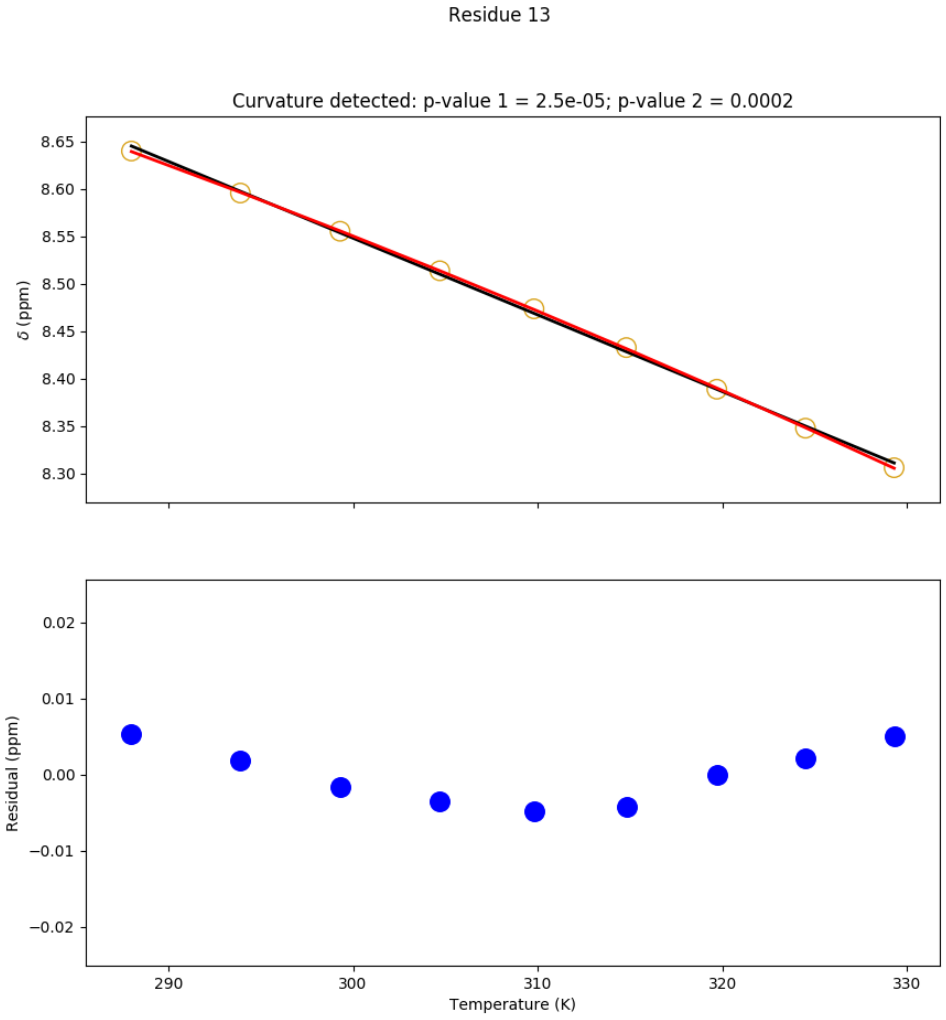


Figure I.169: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 13. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

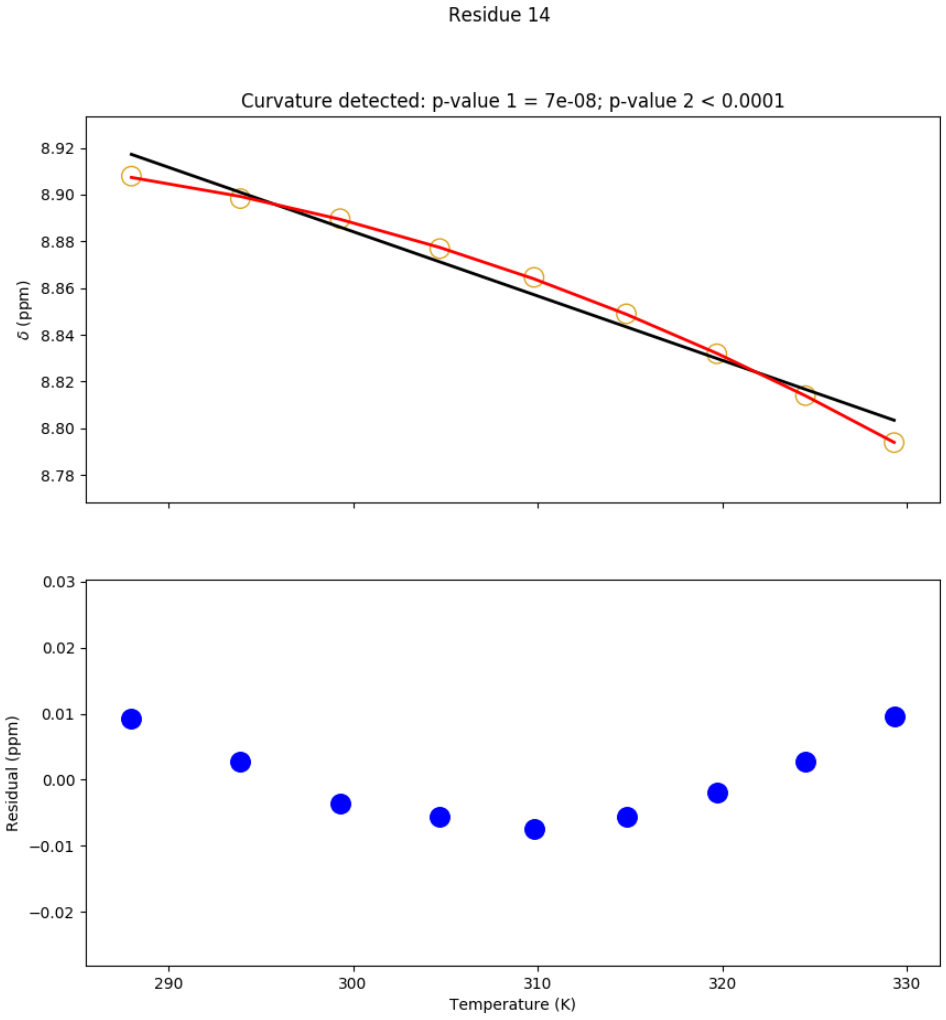


Figure I.170: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 14. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

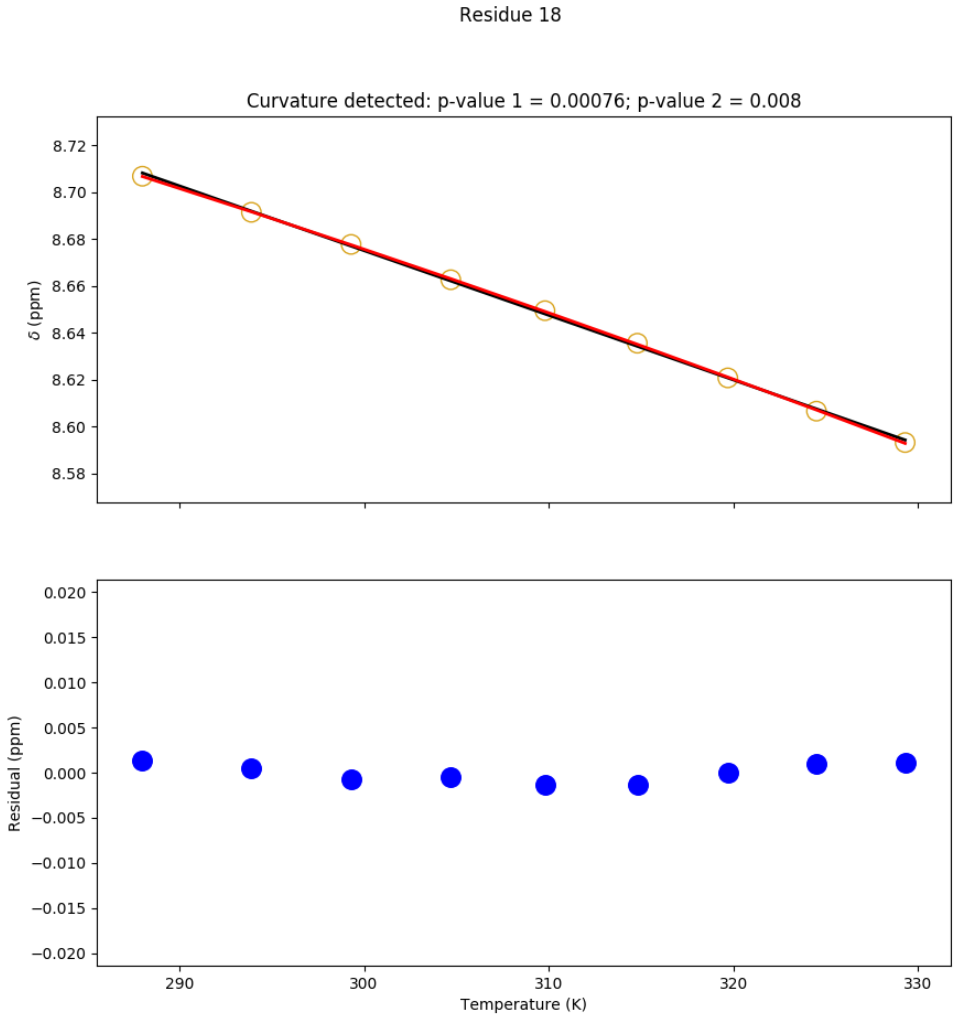


Figure I.171: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 18. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

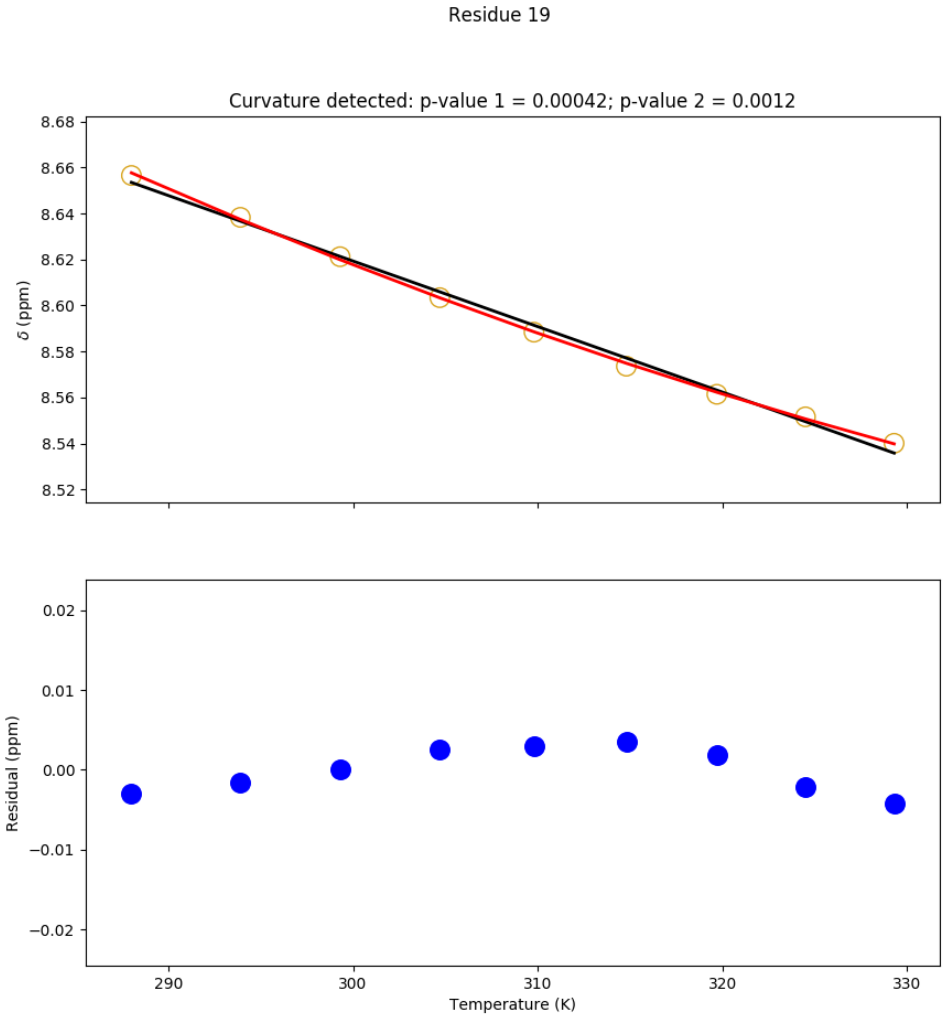


Figure I.172: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 19. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

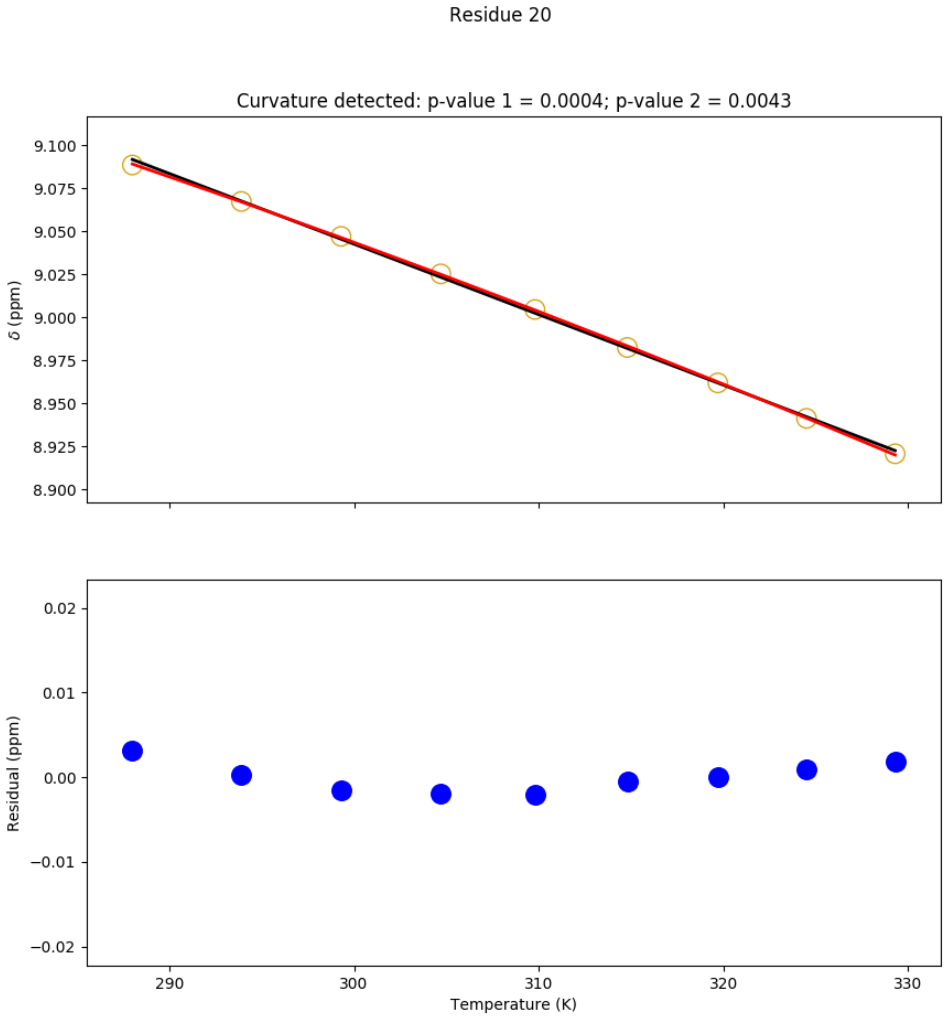


Figure I.173: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 20. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

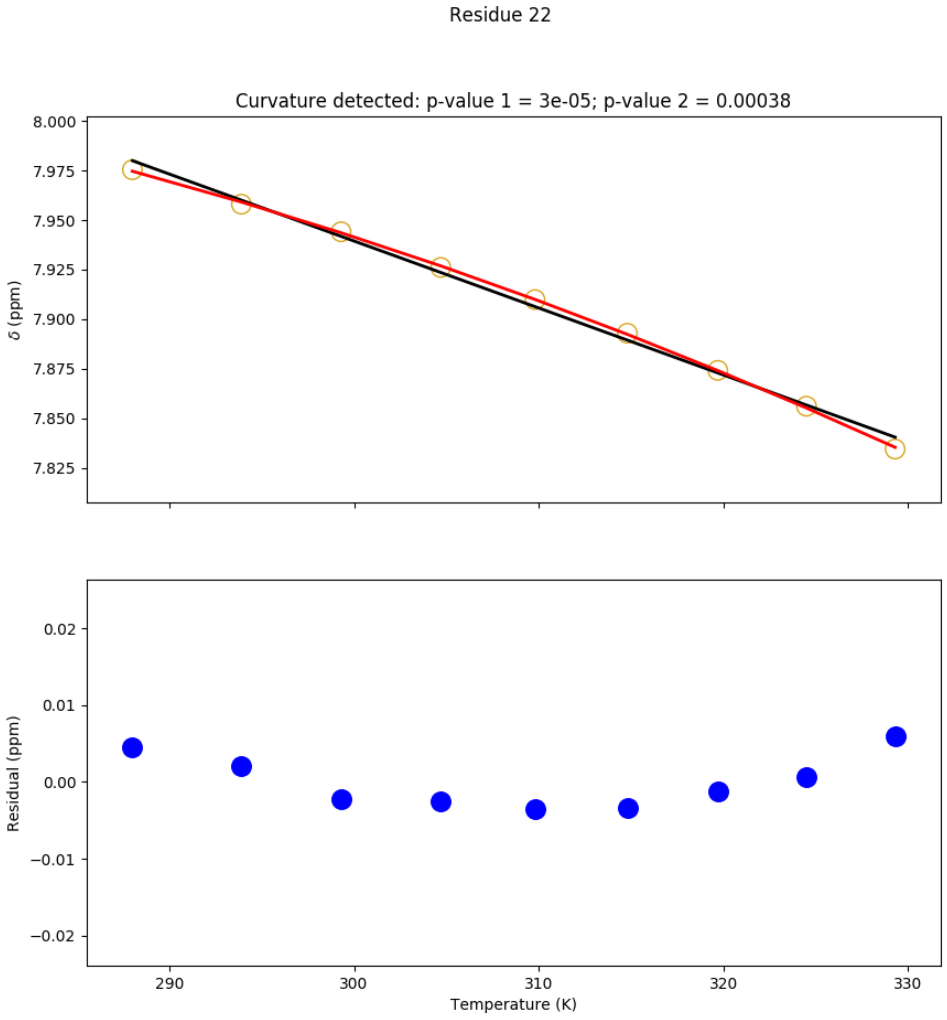


Figure I.174: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 22. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

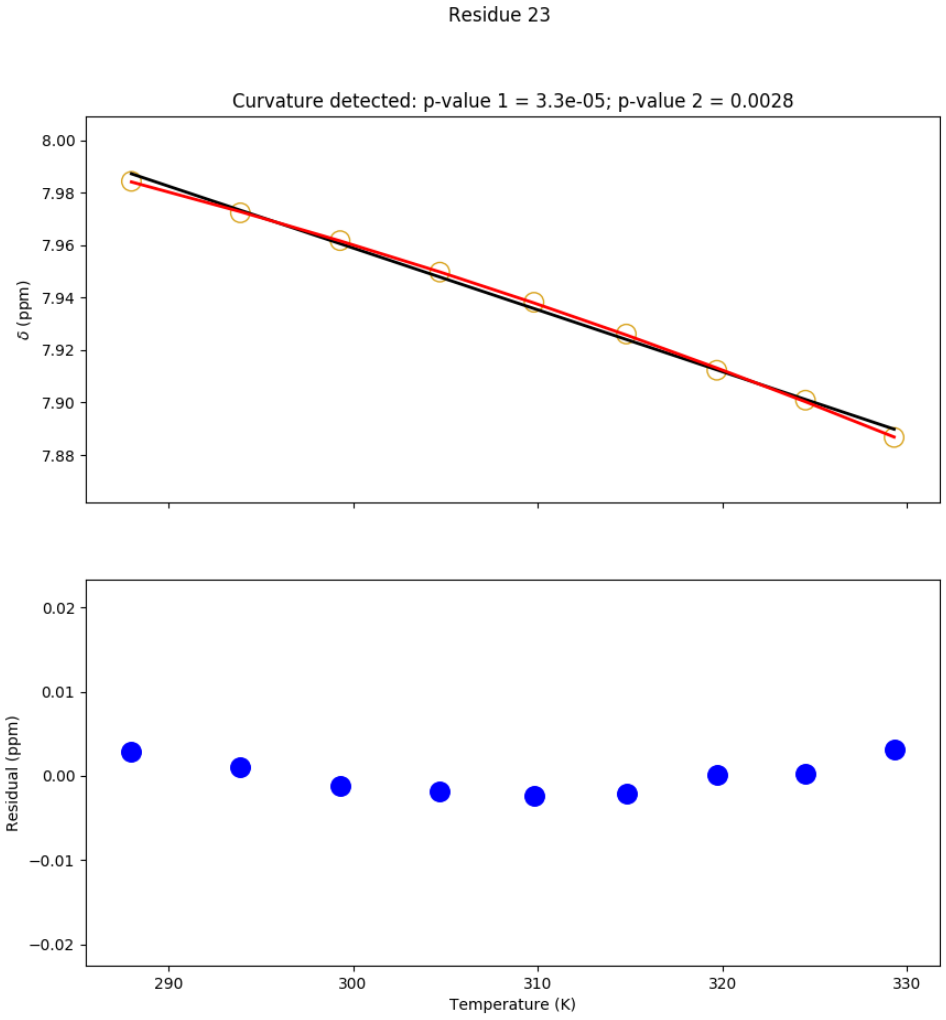


Figure I.175: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 23. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

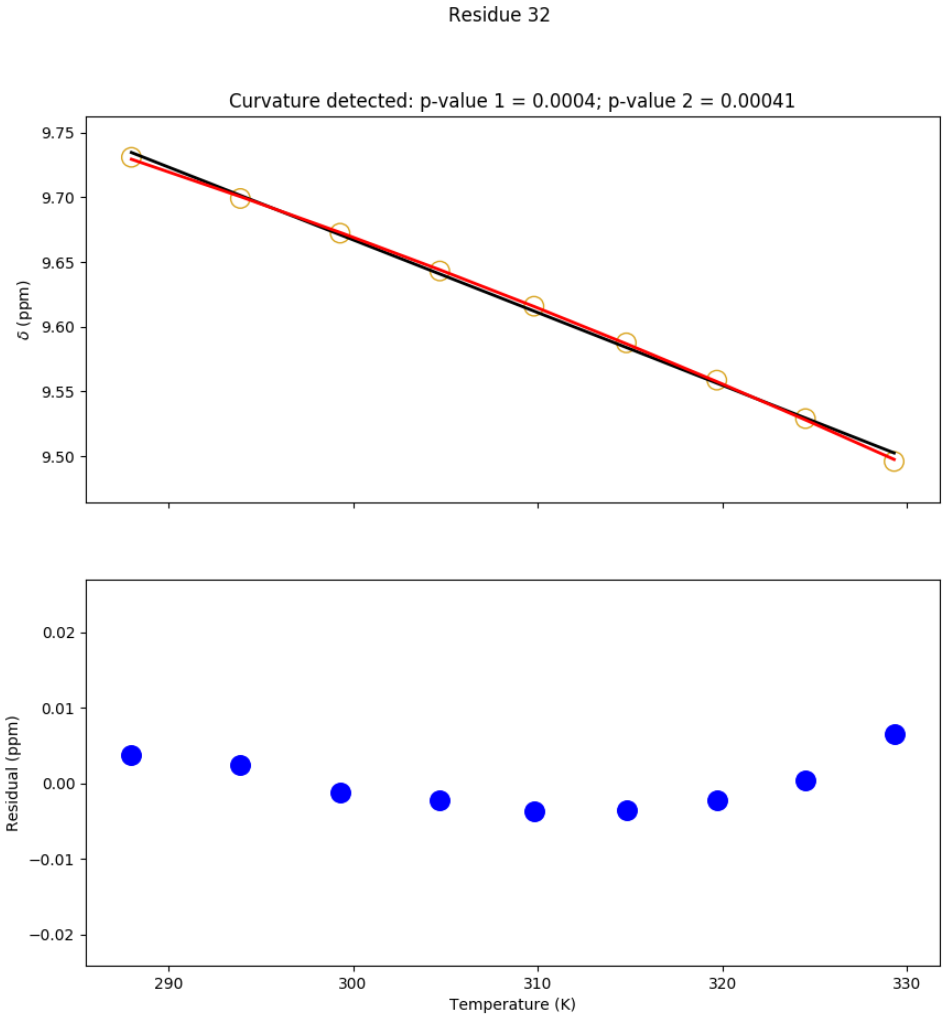


Figure I.176: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 32. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

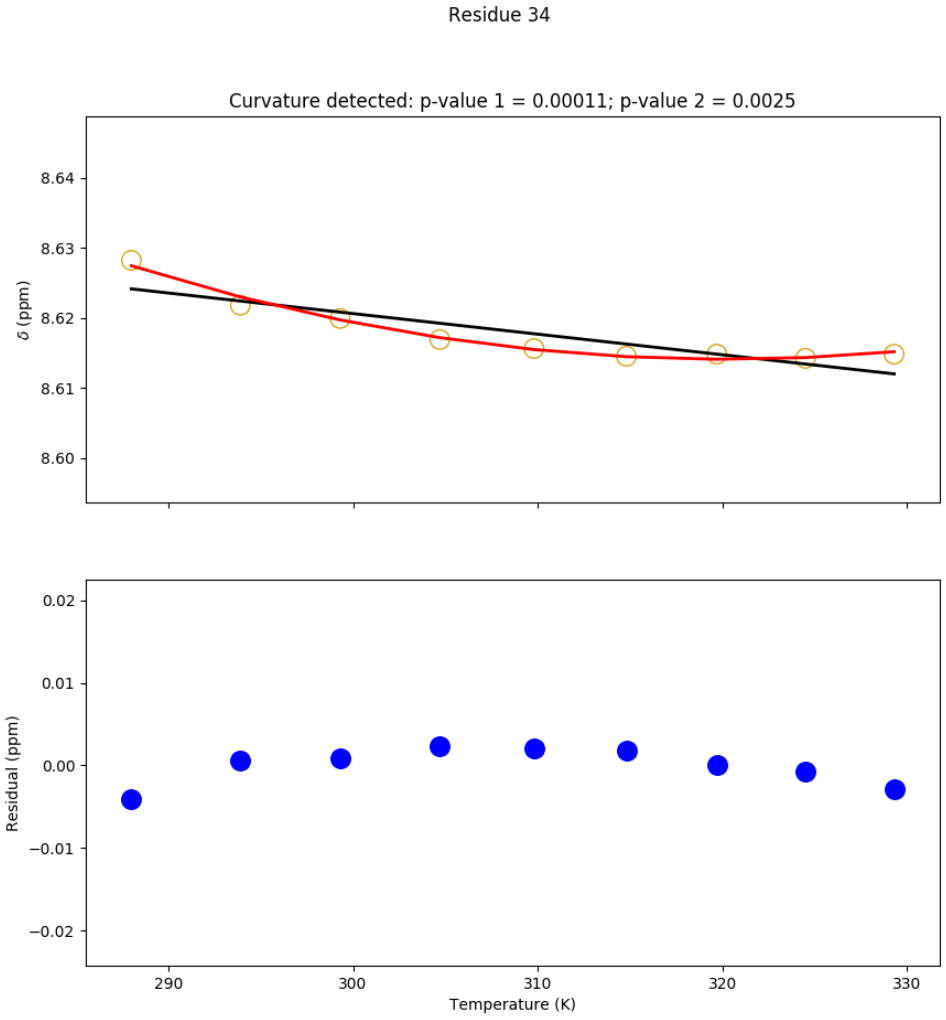


Figure I.177: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 34. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

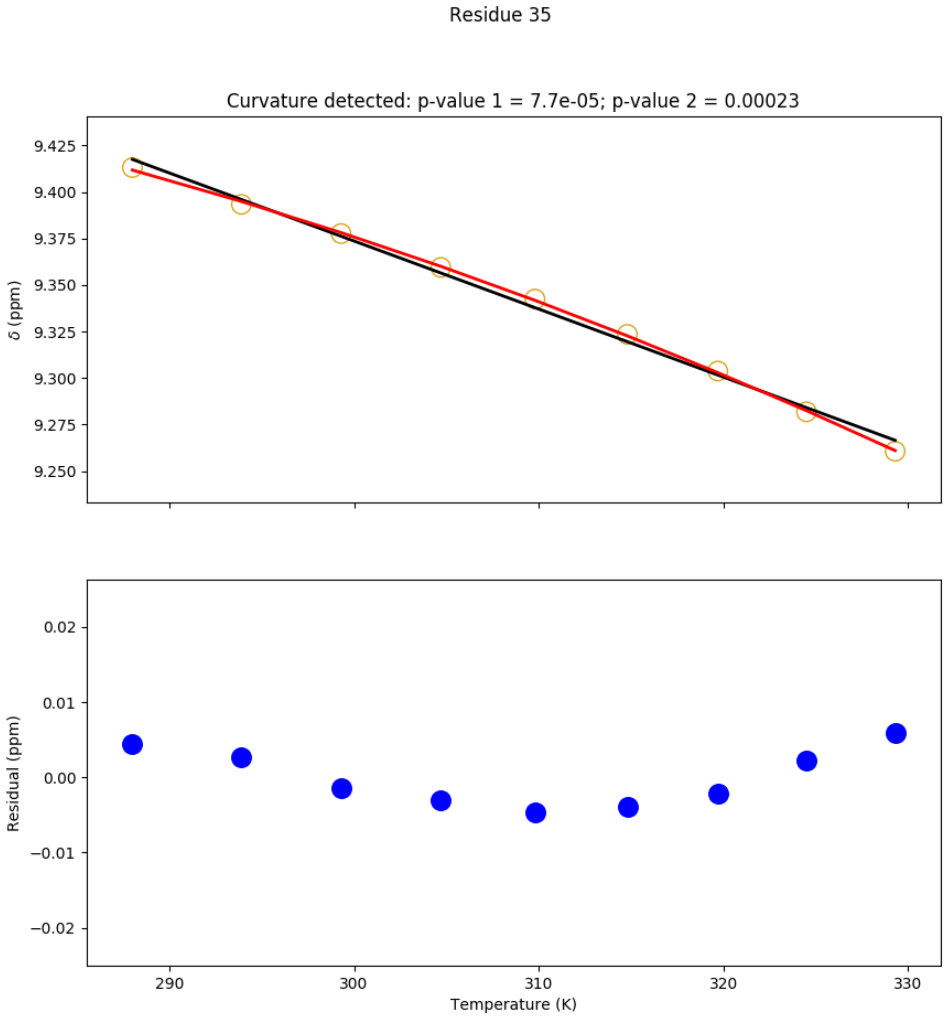


Figure I.178: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 35. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

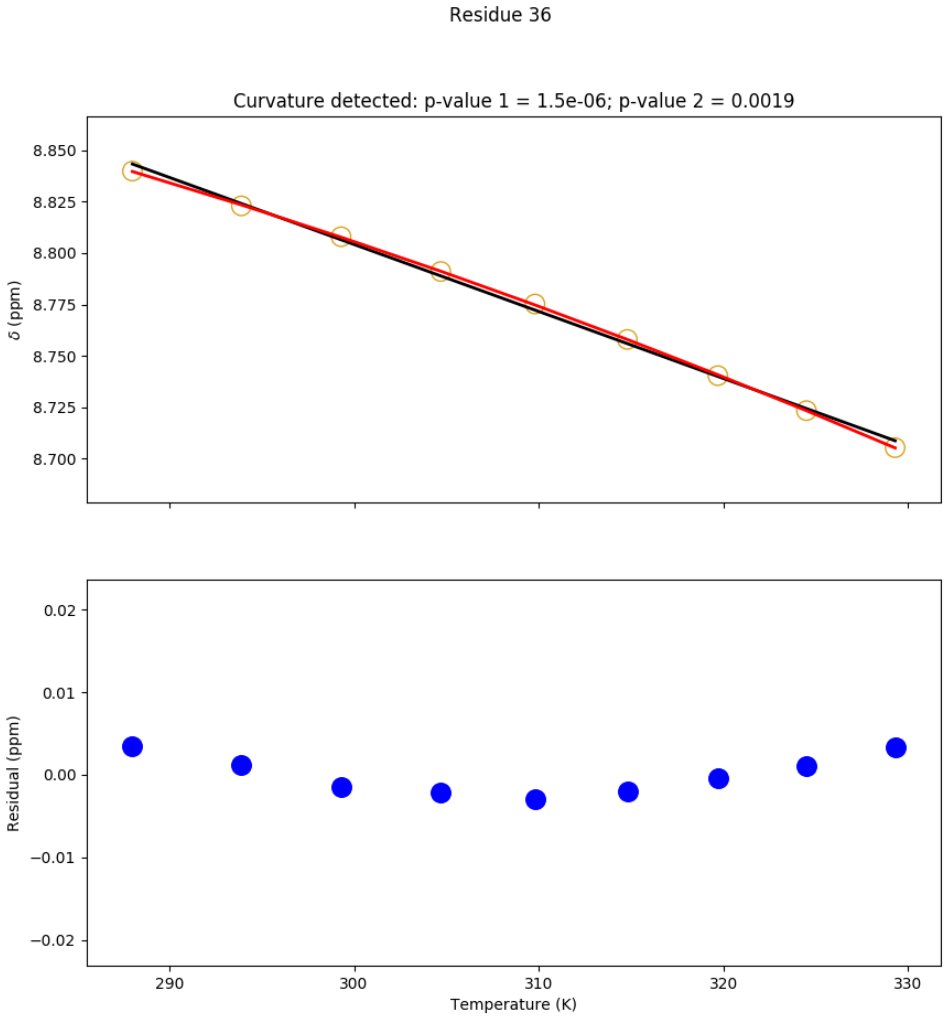


Figure I.179: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 36. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

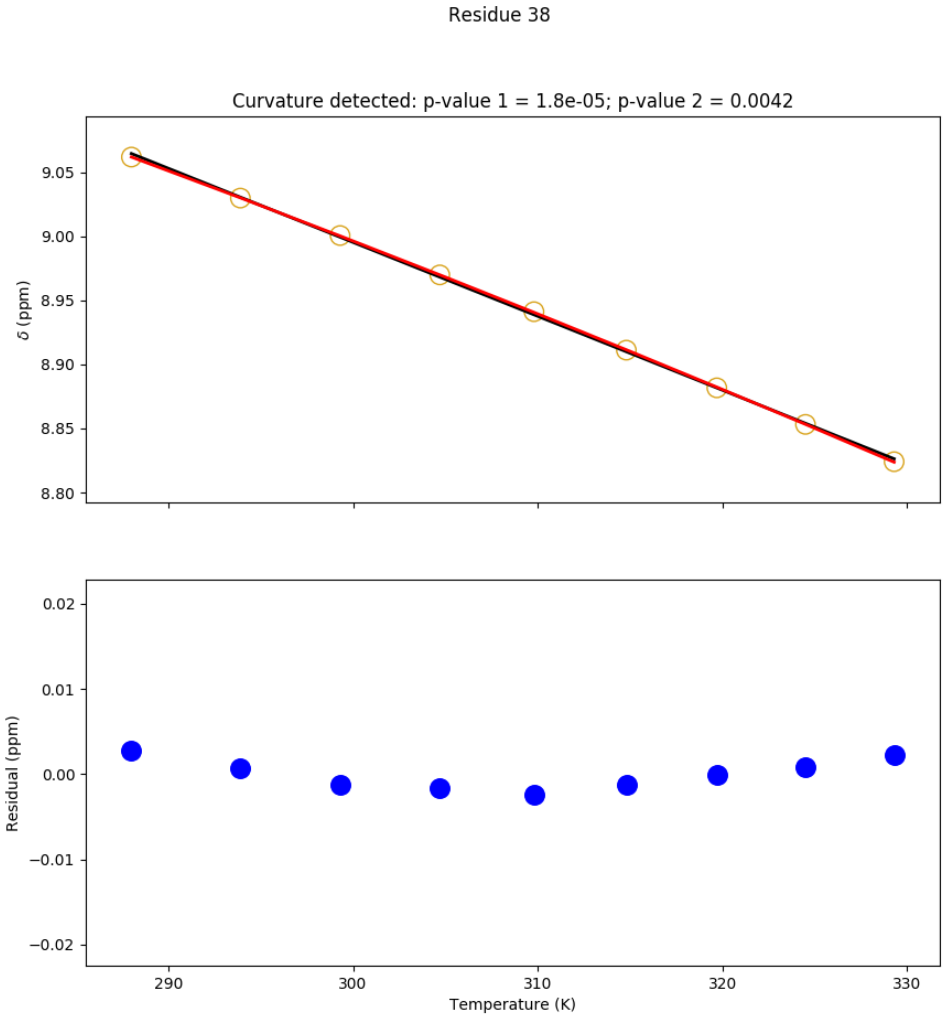


Figure I.180: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 38. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

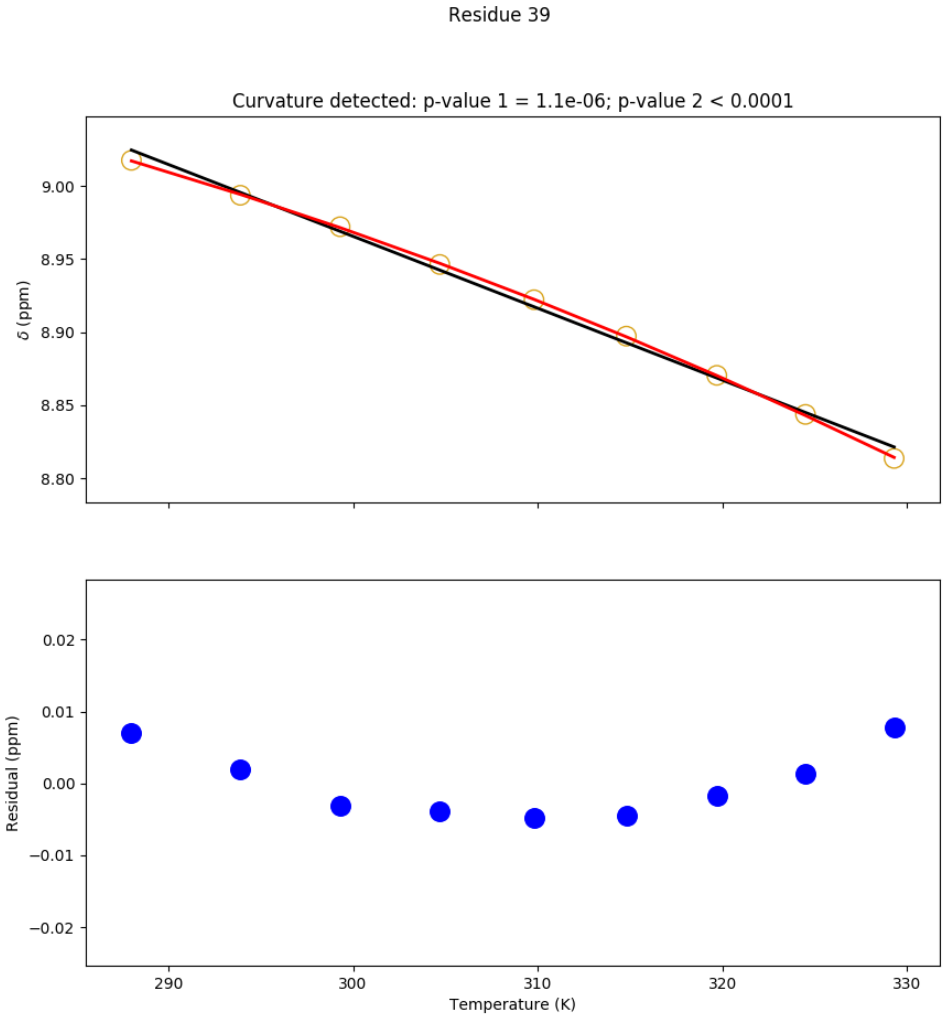


Figure I.181: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 39. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

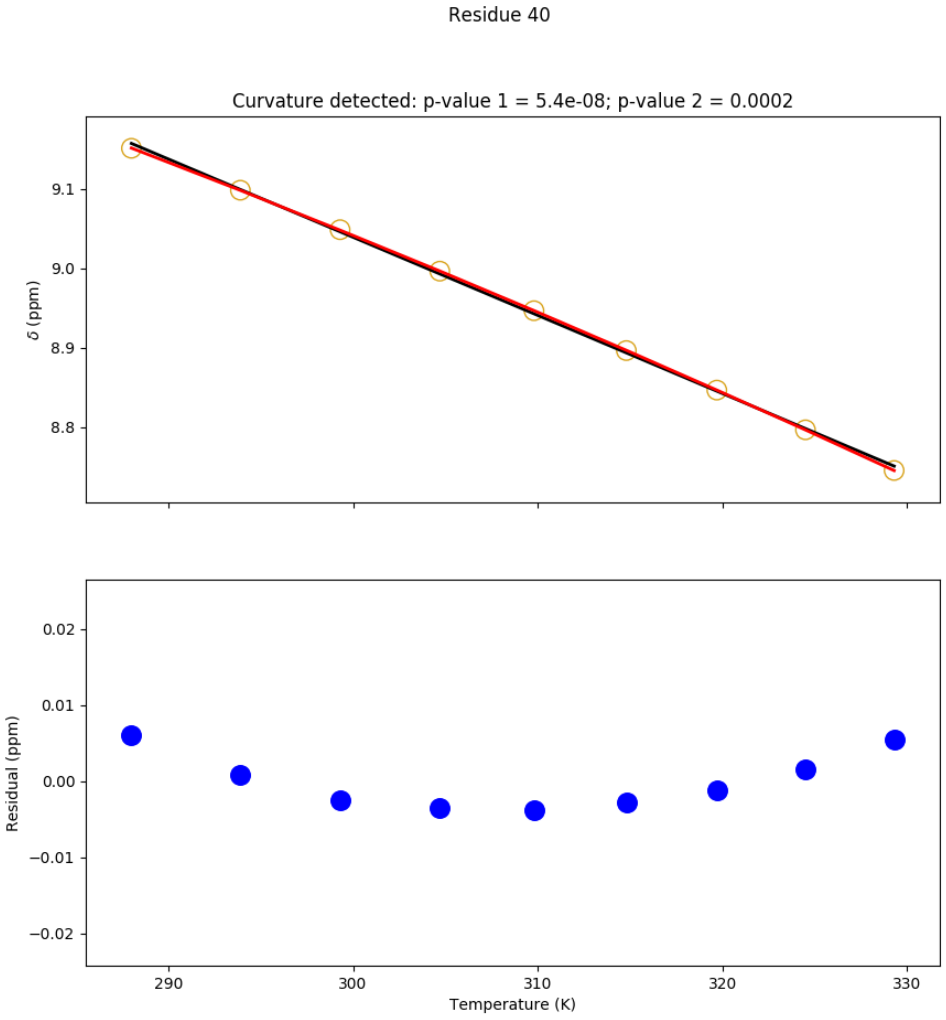


Figure I.182: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 40. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

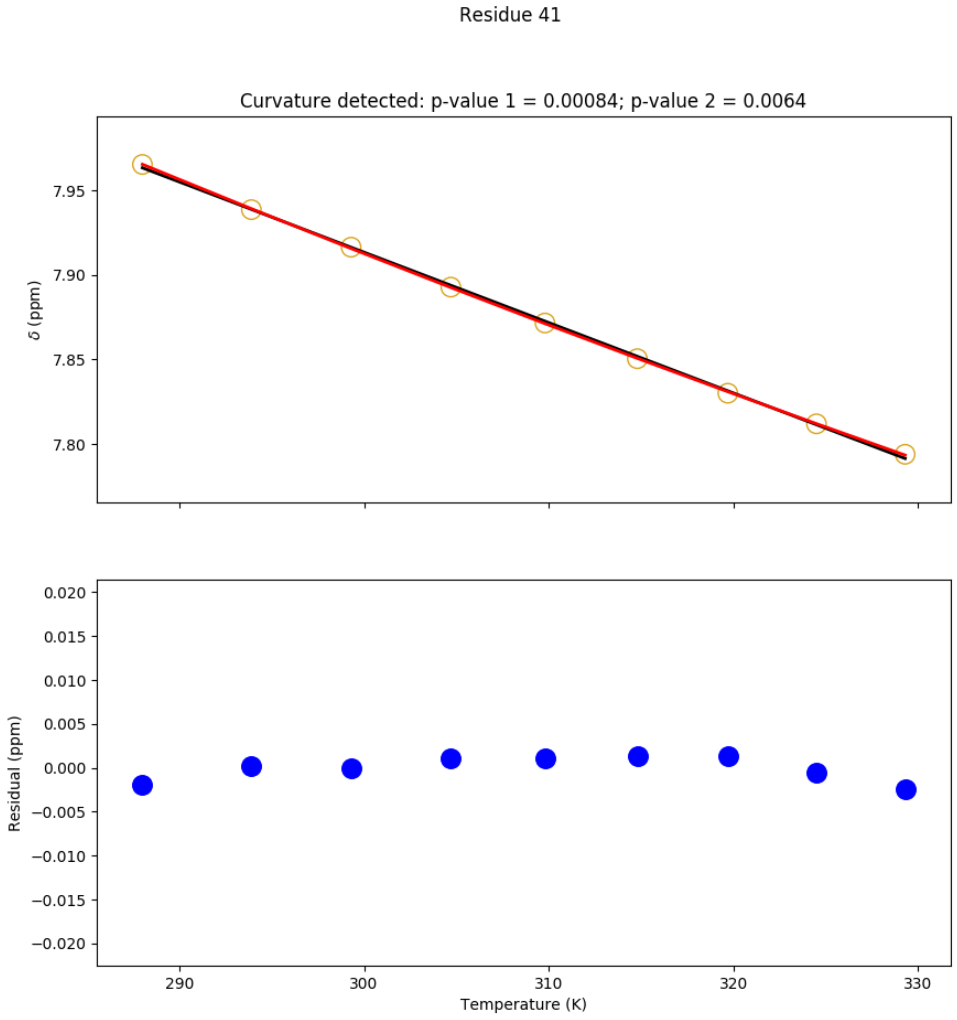


Figure I.183: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 41. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

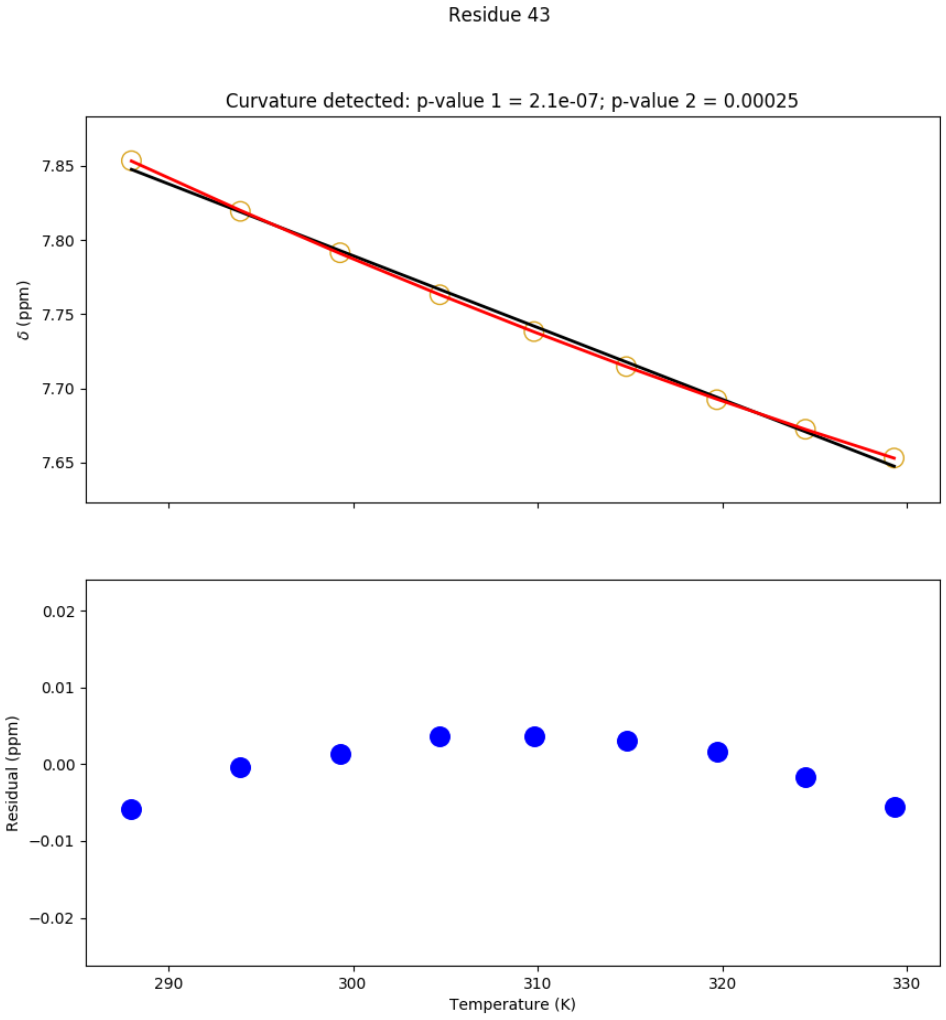


Figure I.184: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 43. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

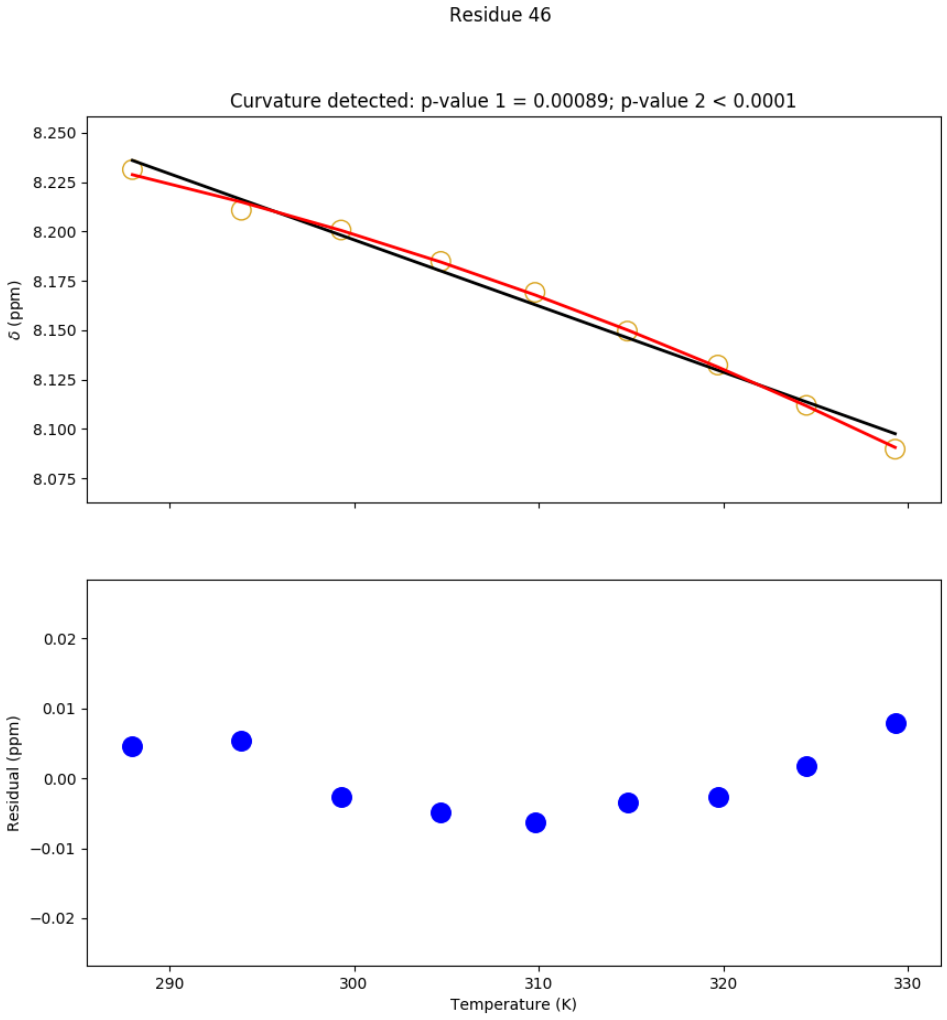


Figure I.185: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 46. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

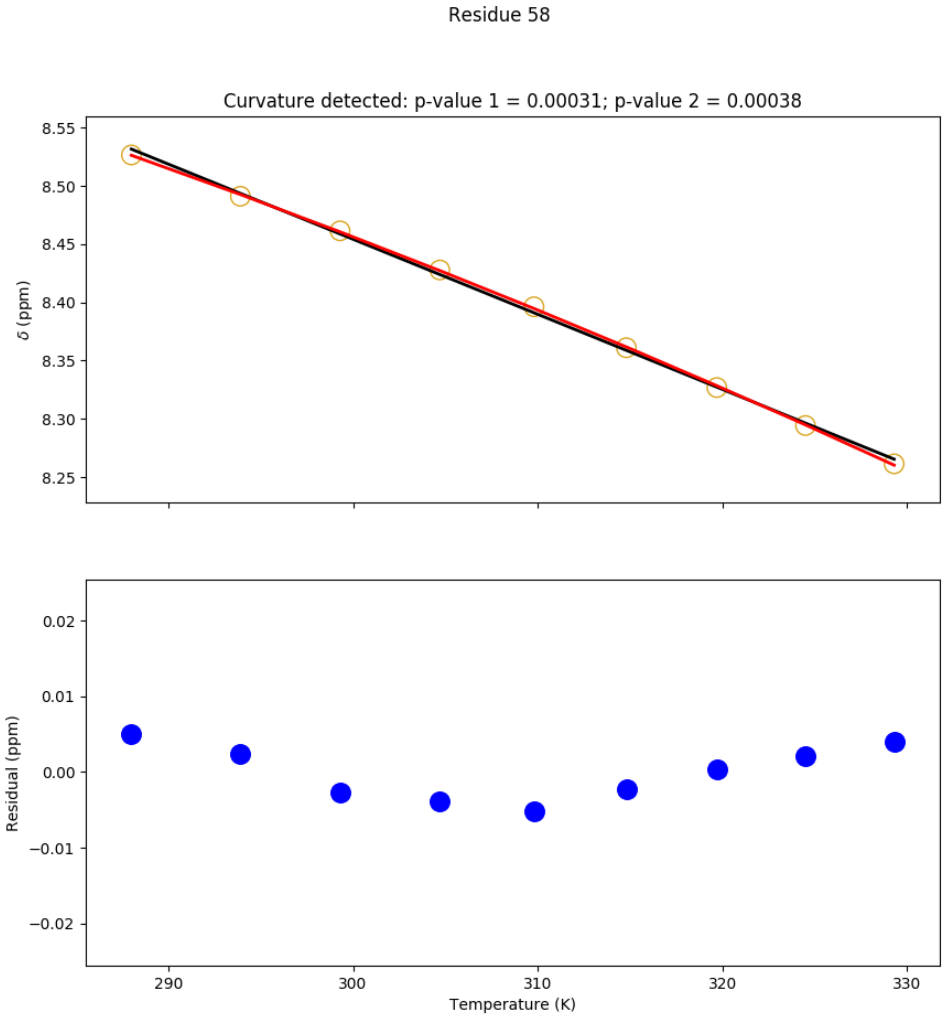


Figure I.186: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 58. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

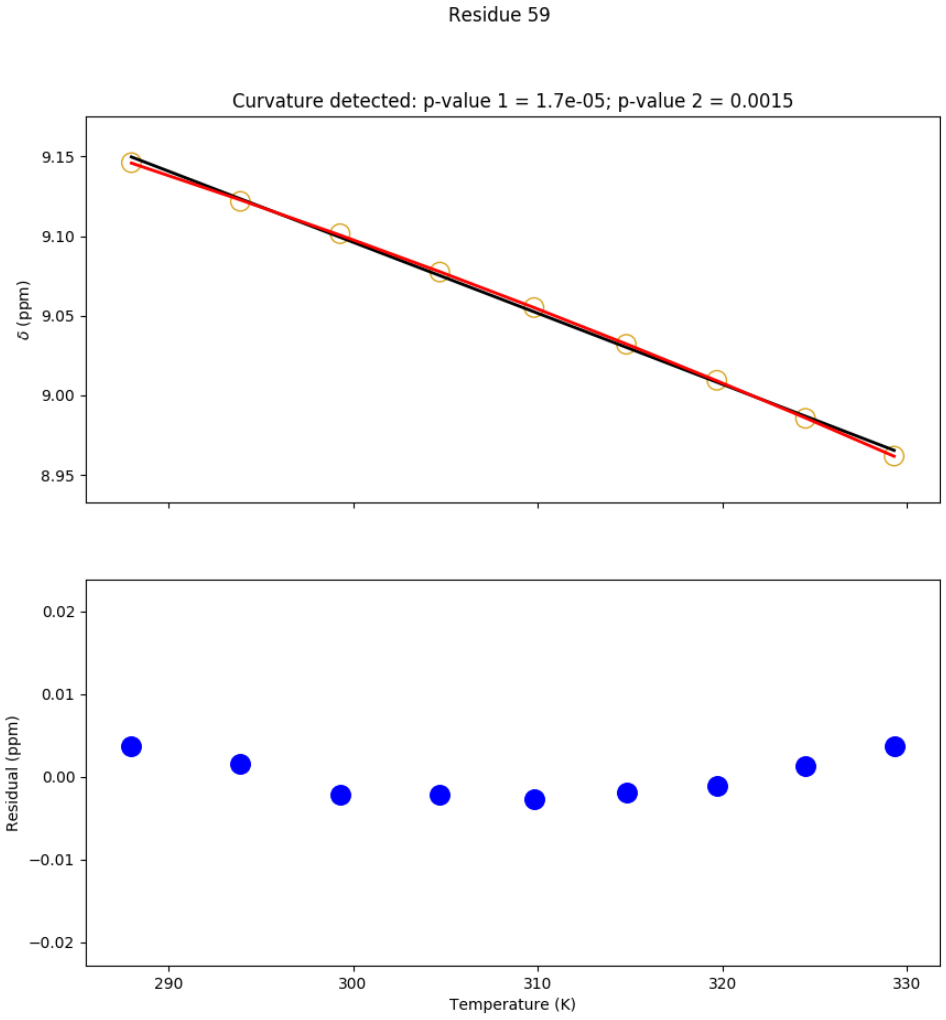


Figure I.187: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 59. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

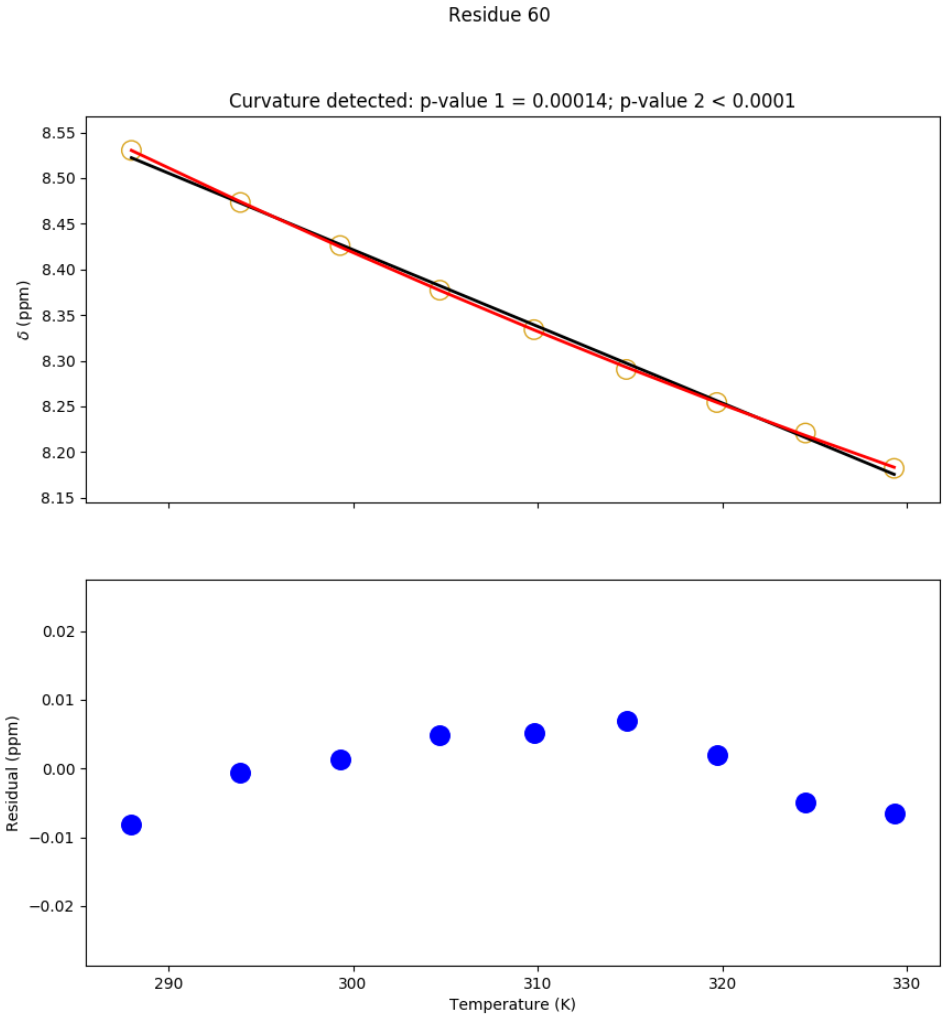


Figure I.188: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 60. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

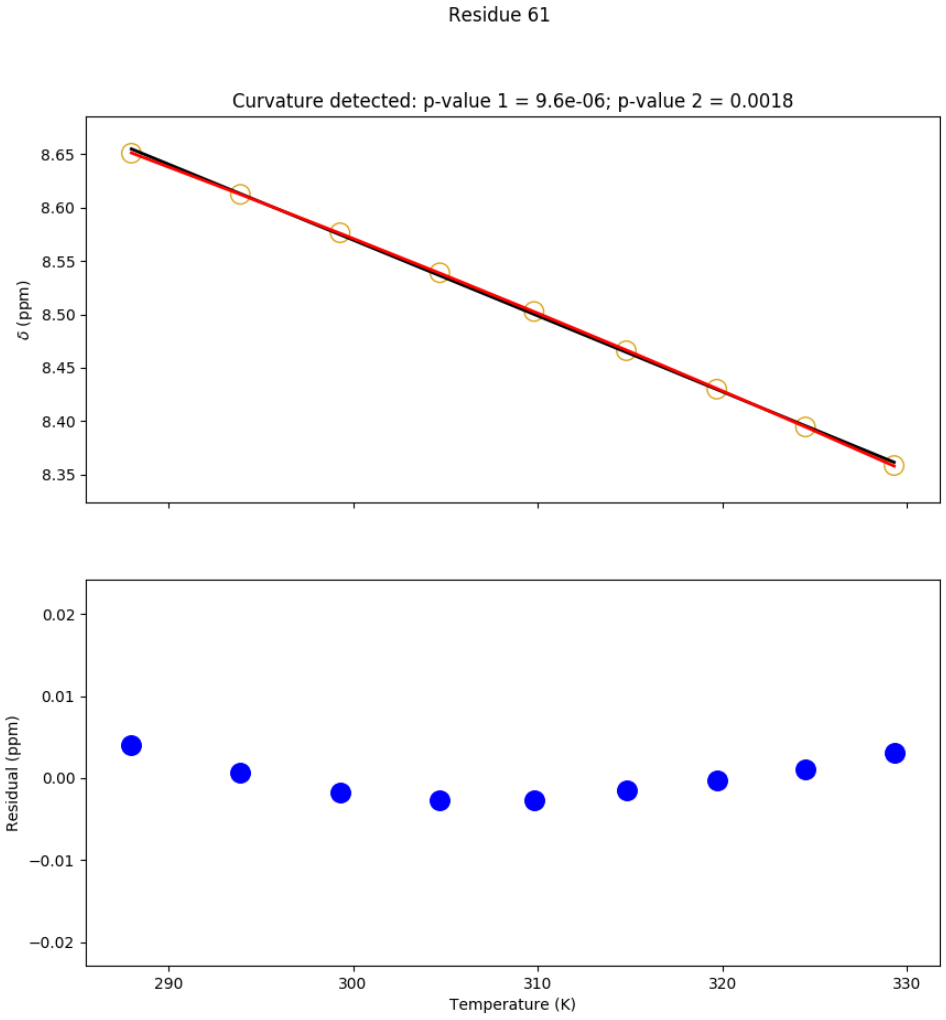


Figure I.189: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 61. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

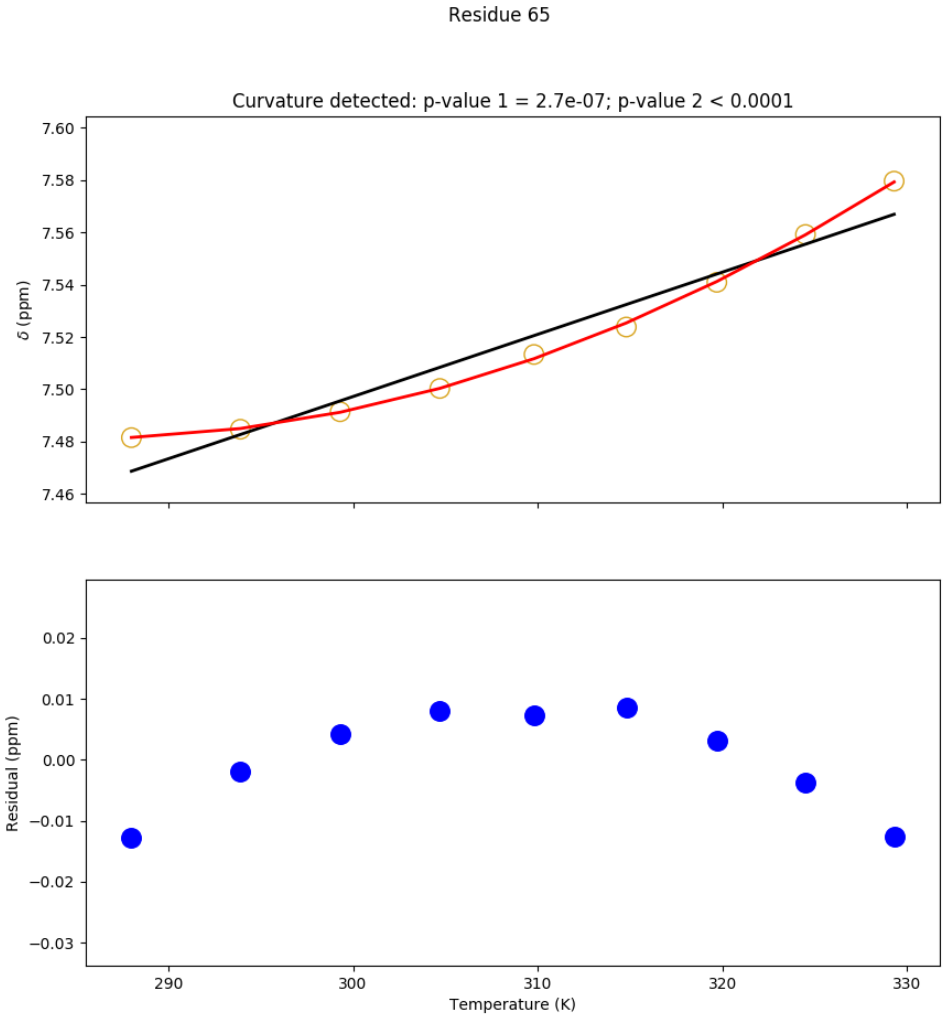


Figure I.190: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 65. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

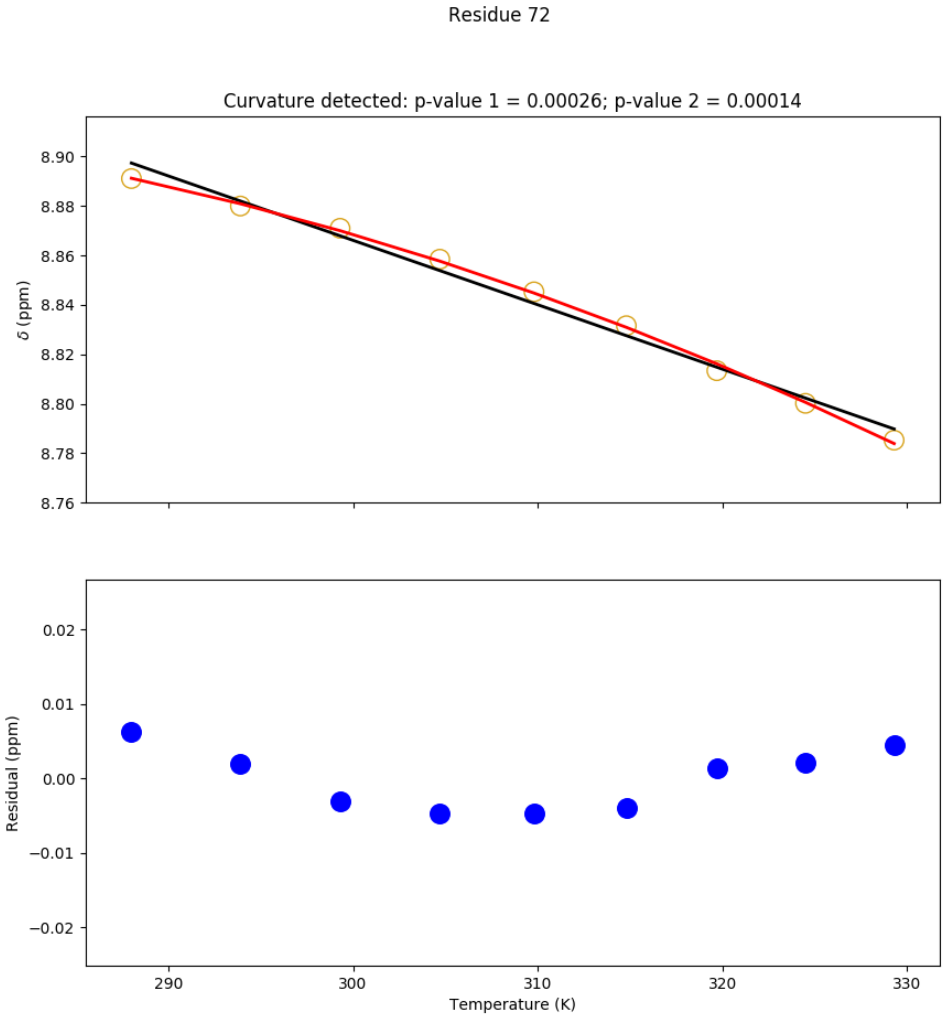


Figure I.191: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 72. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

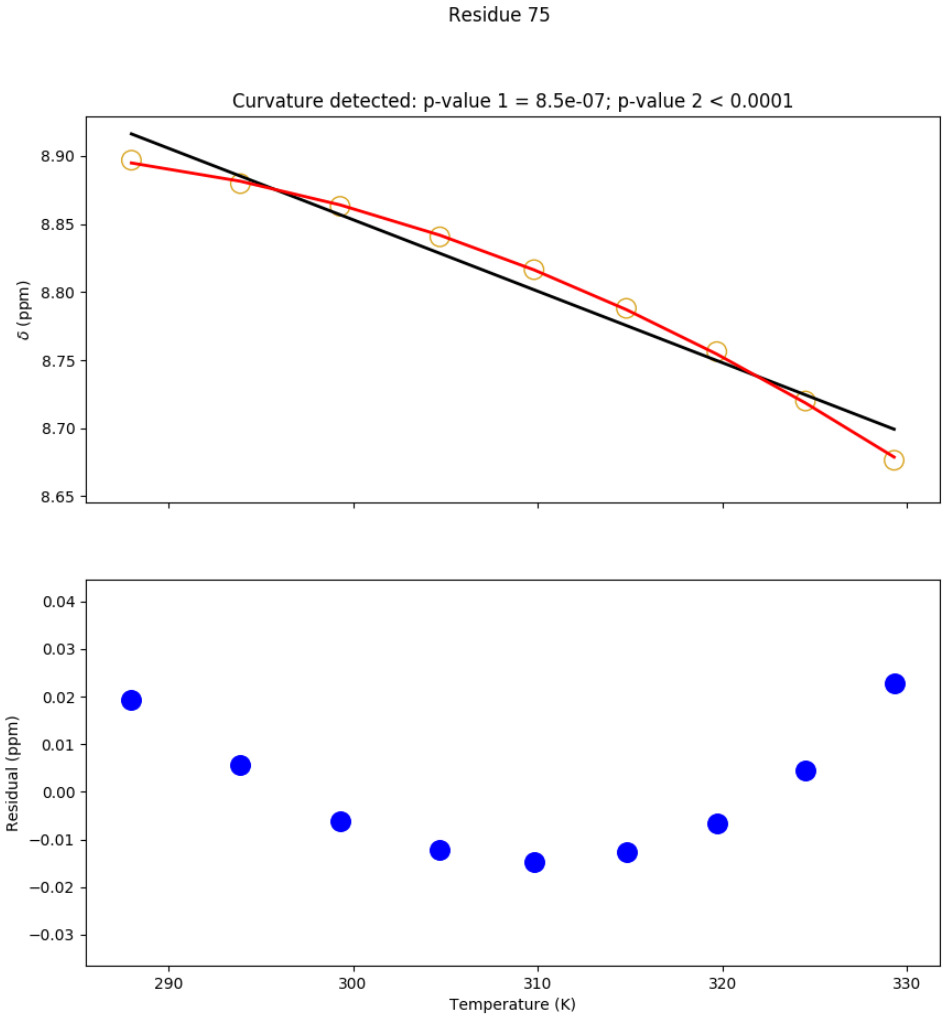


Figure I.192: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 75. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

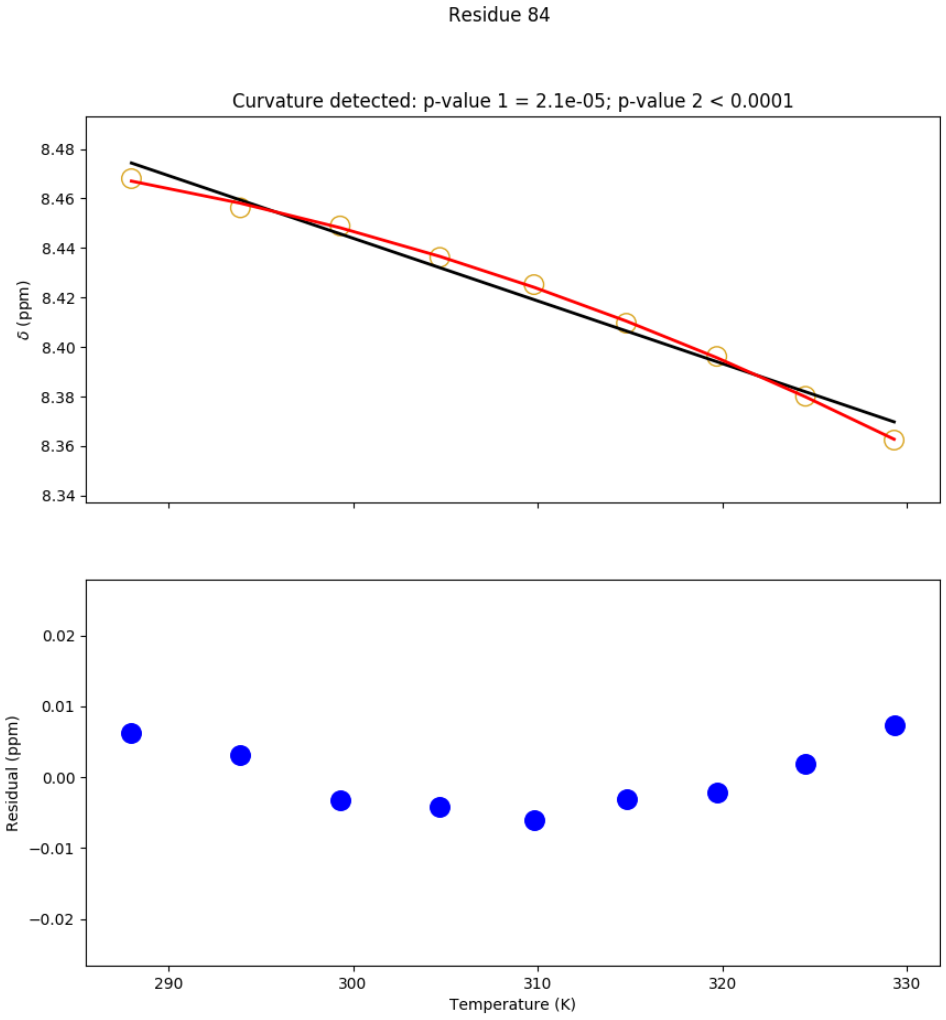


Figure I.193: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 84. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

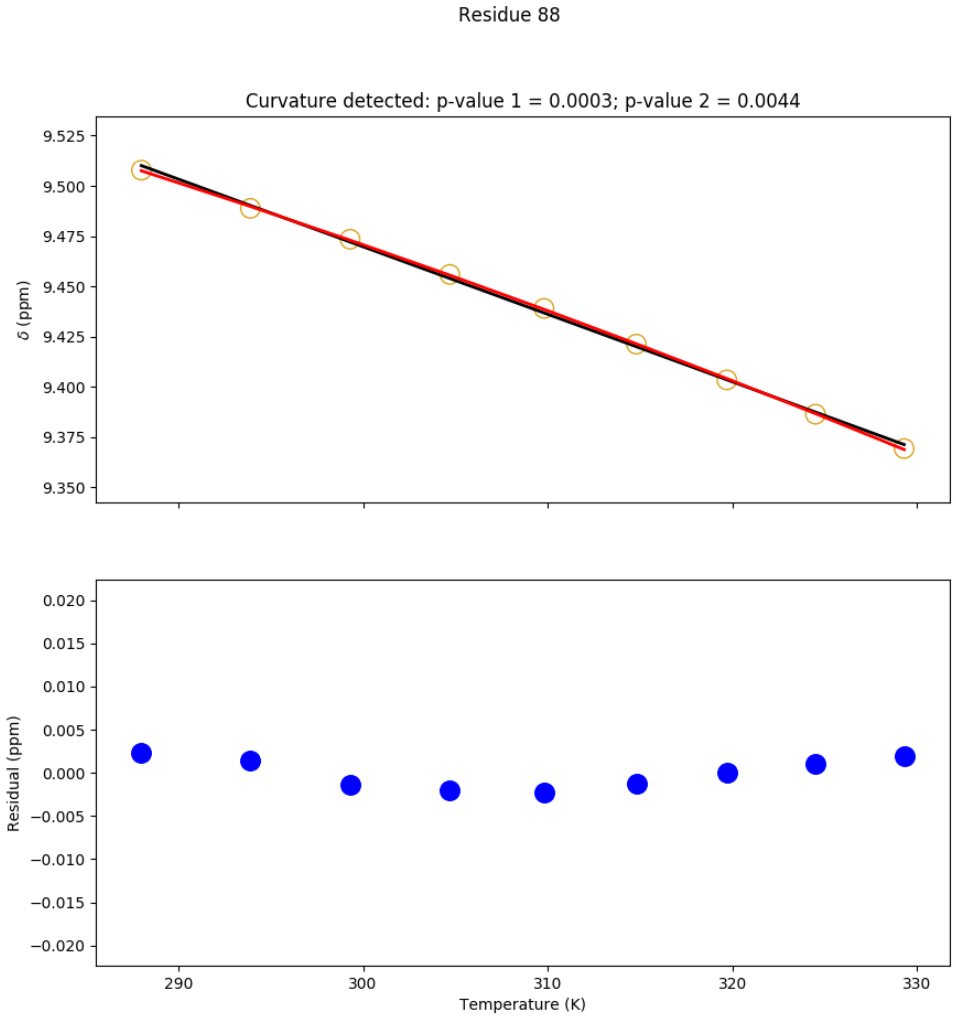


Figure I.194: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 88. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

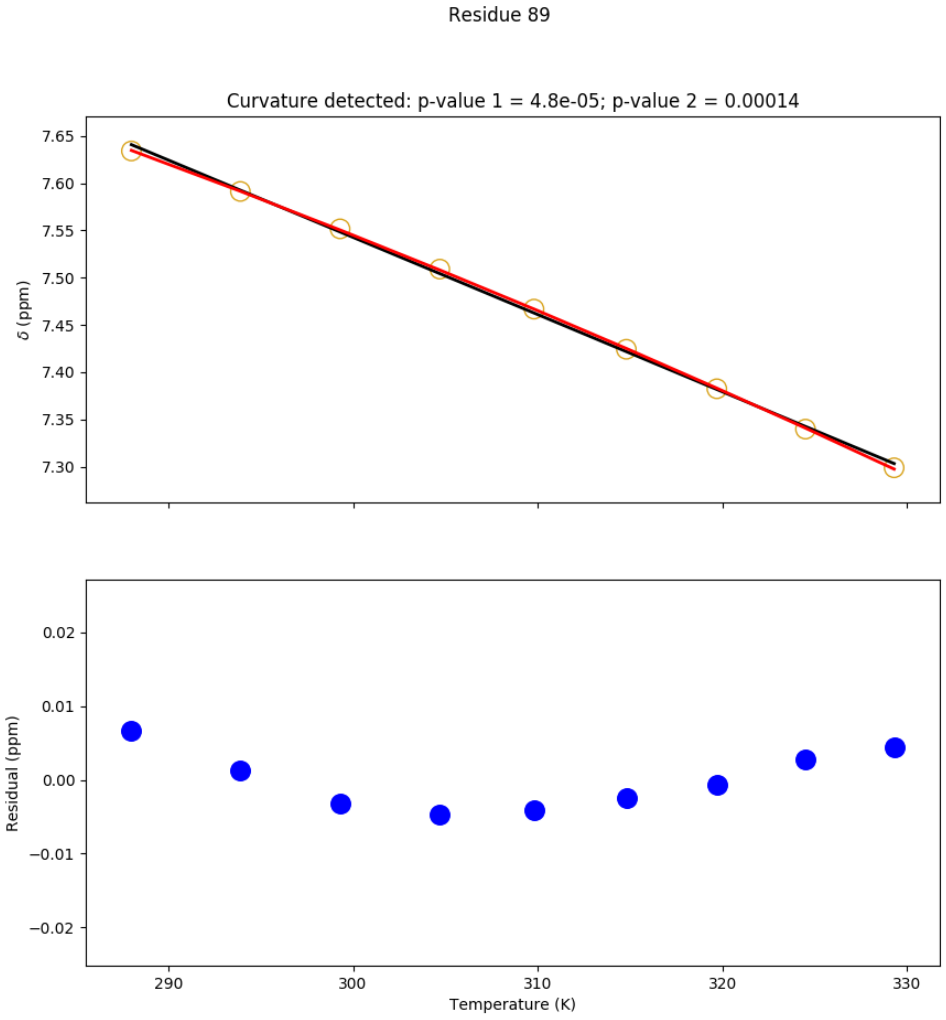


Figure I.195: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 89. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

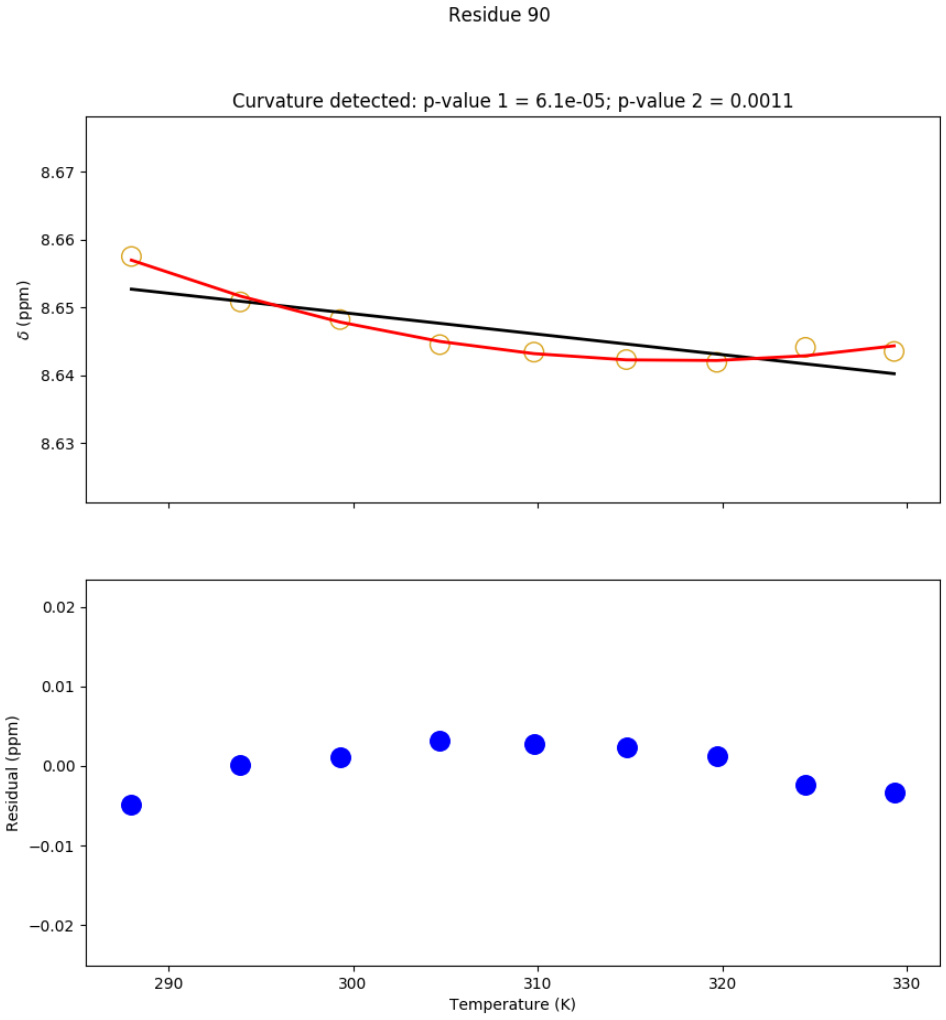


Figure I.196: Temperature dependence of the chemical shift of the V75R Adnectin amide proton from residue 90. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

I.10 V75R Adnectin Curvature Modelling

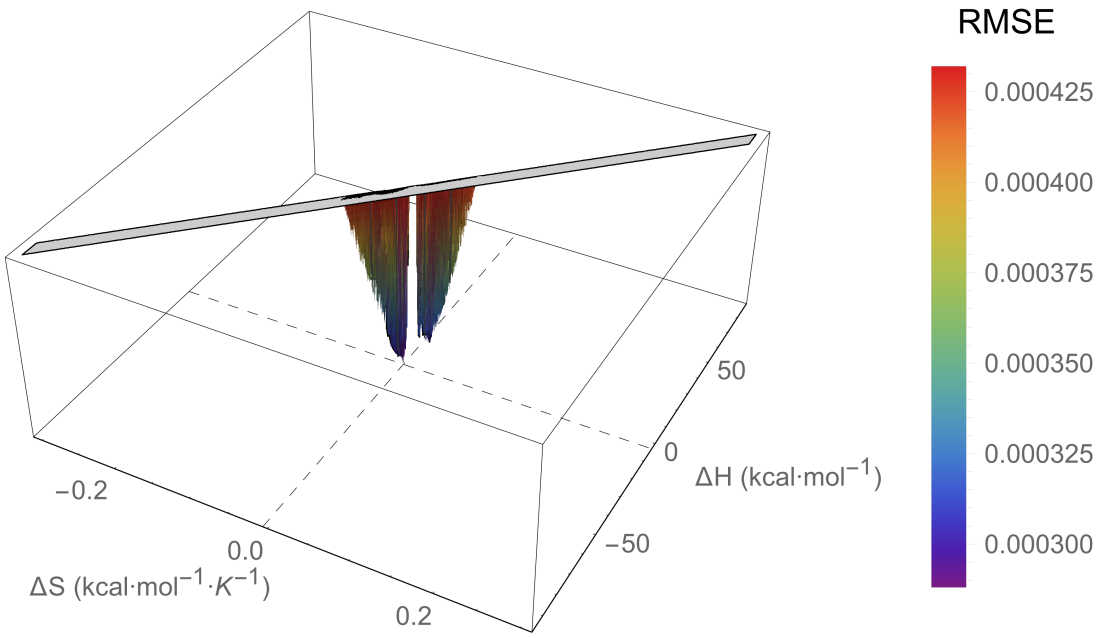


Figure I.197: An RMSE landscape from fitting V75R Adnectin residue 4 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

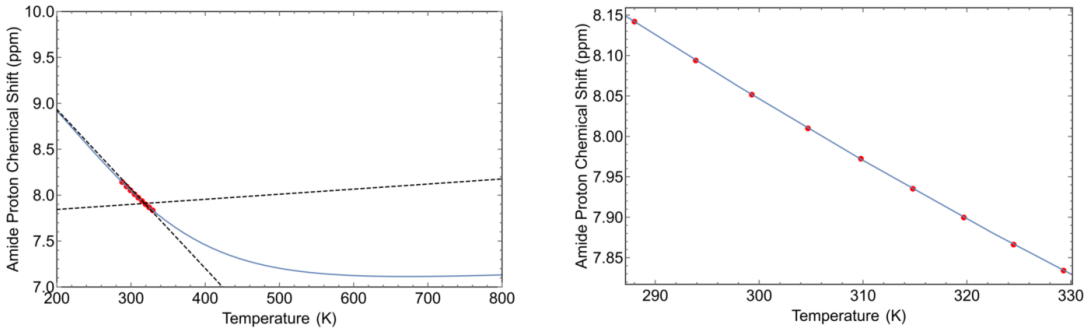


Figure I.198: The minimum RMSE fit of V75R Adnectin residue 4 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

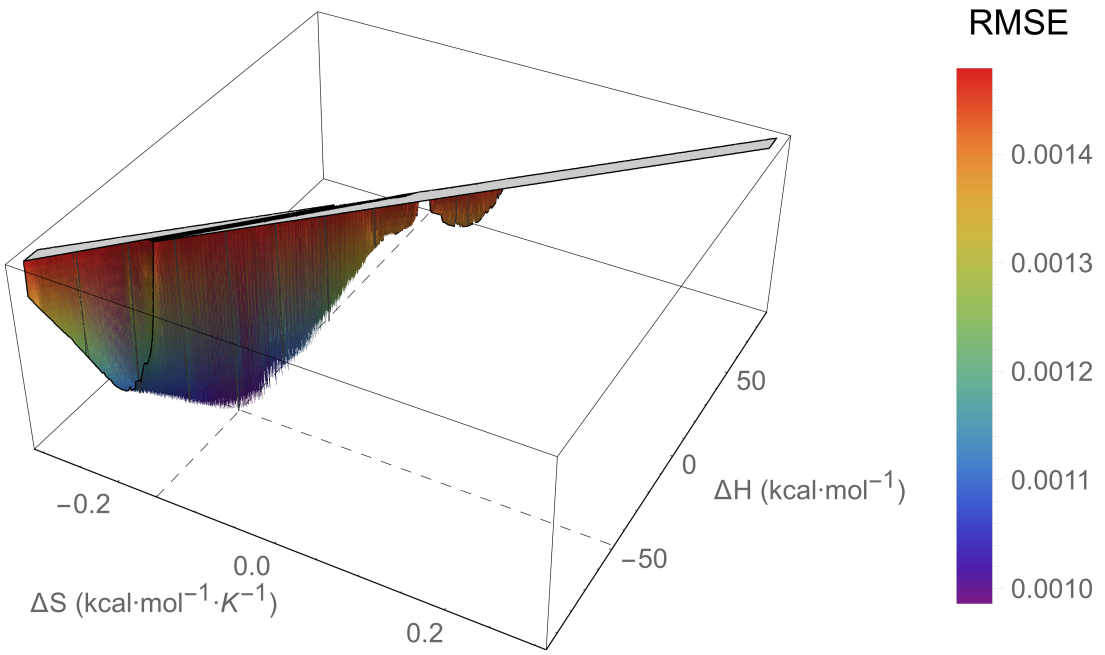


Figure I.199: An RMSE landscape from fitting V75R Adnectin residue 6 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

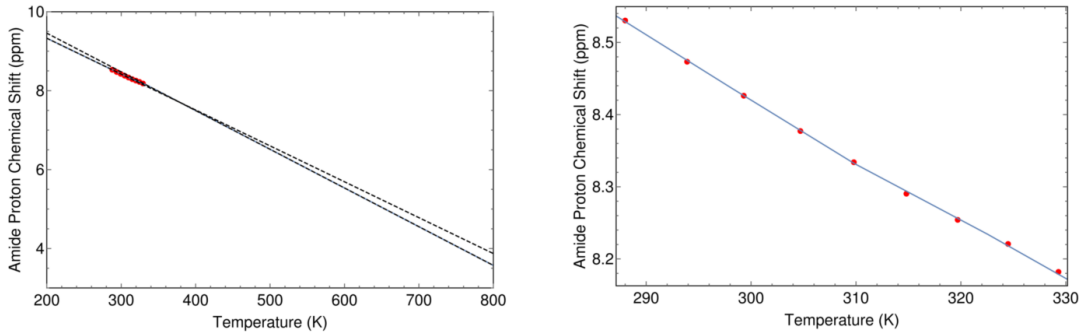


Figure I.200: The minimum RMSE fit of V75R Adnectin residue 6 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

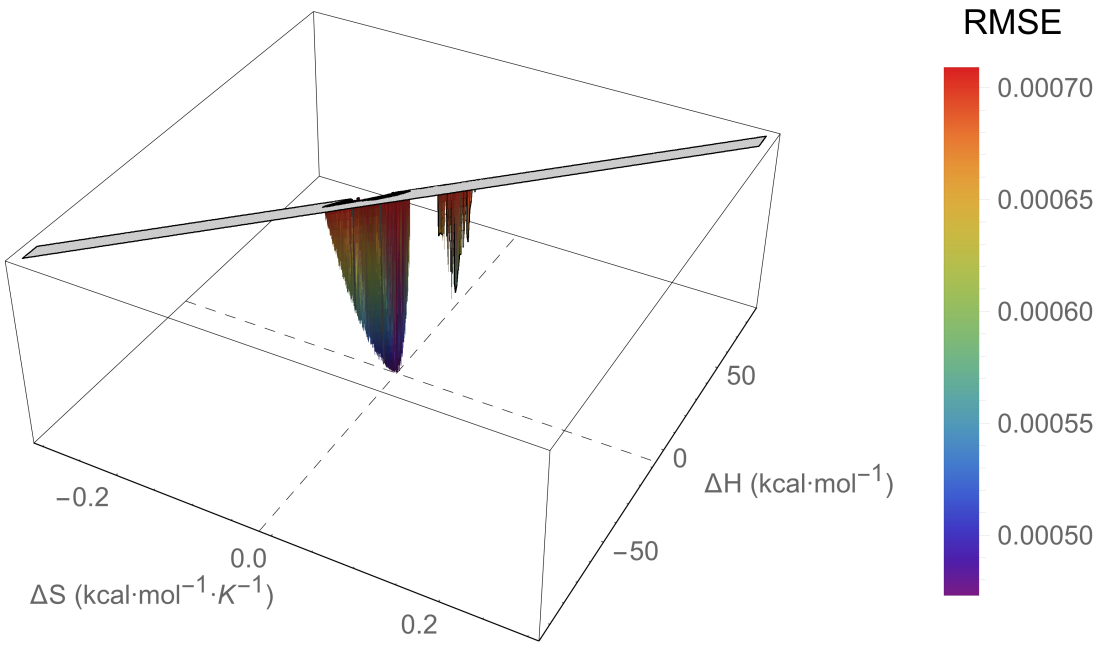


Figure I.201: An RMSE landscape from fitting V75R Adnectin residue 11 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

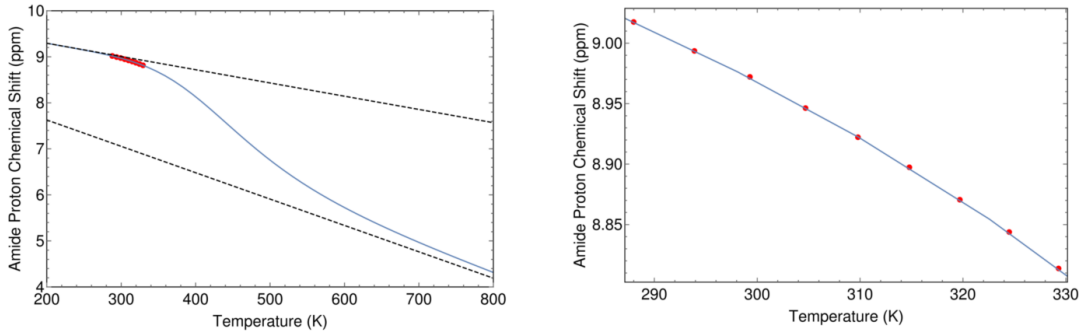


Figure I.202: The minimum RMSE fit of V75R Adnectin residue 11 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

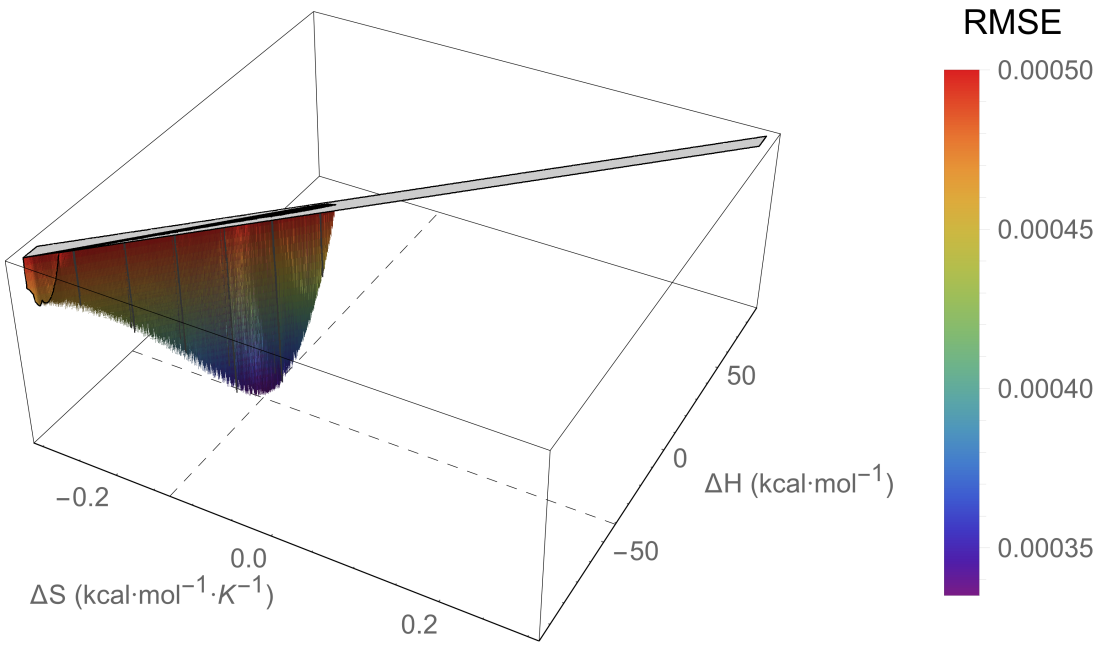


Figure I.203: An RMSE landscape from fitting V75R Adnectin residue 13 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

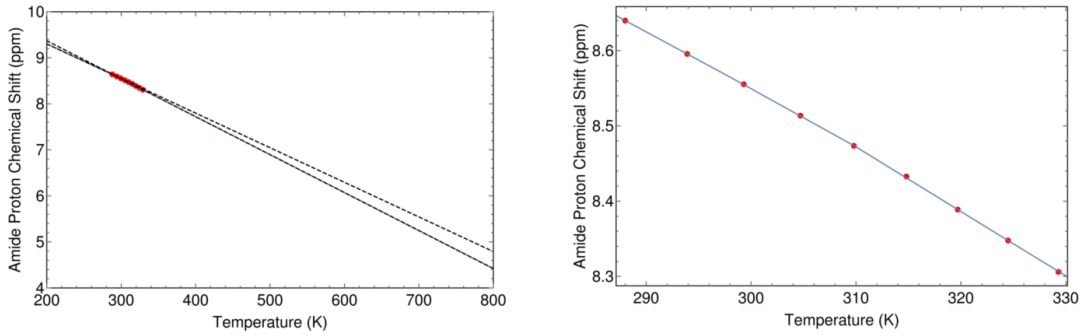


Figure I.204: The minimum RMSE fit of V75R Adnectin residue 13 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

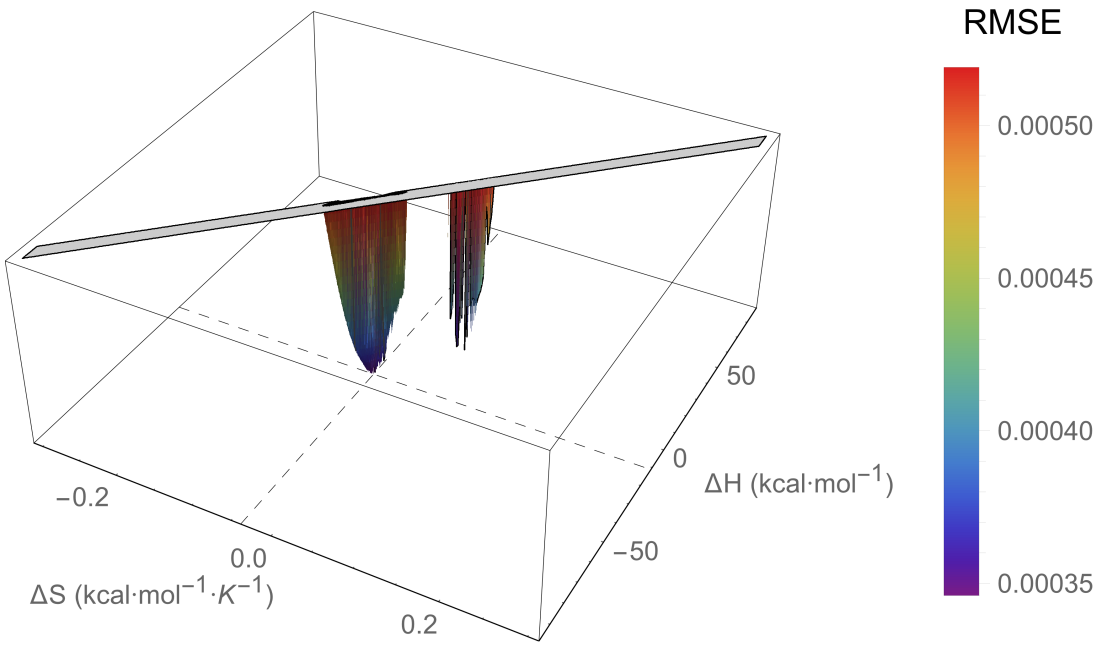


Figure I.205: An RMSE landscape from fitting V75R Adnectin residue 14 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

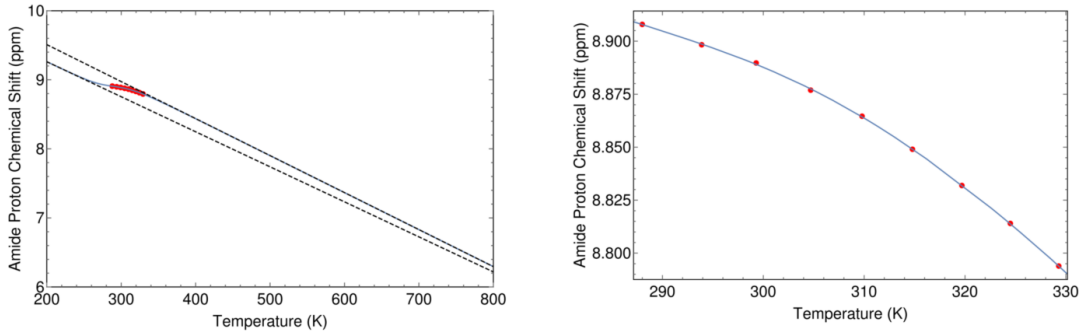


Figure I.206: The minimum RMSE fit of V75R Adnectin residue 14 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

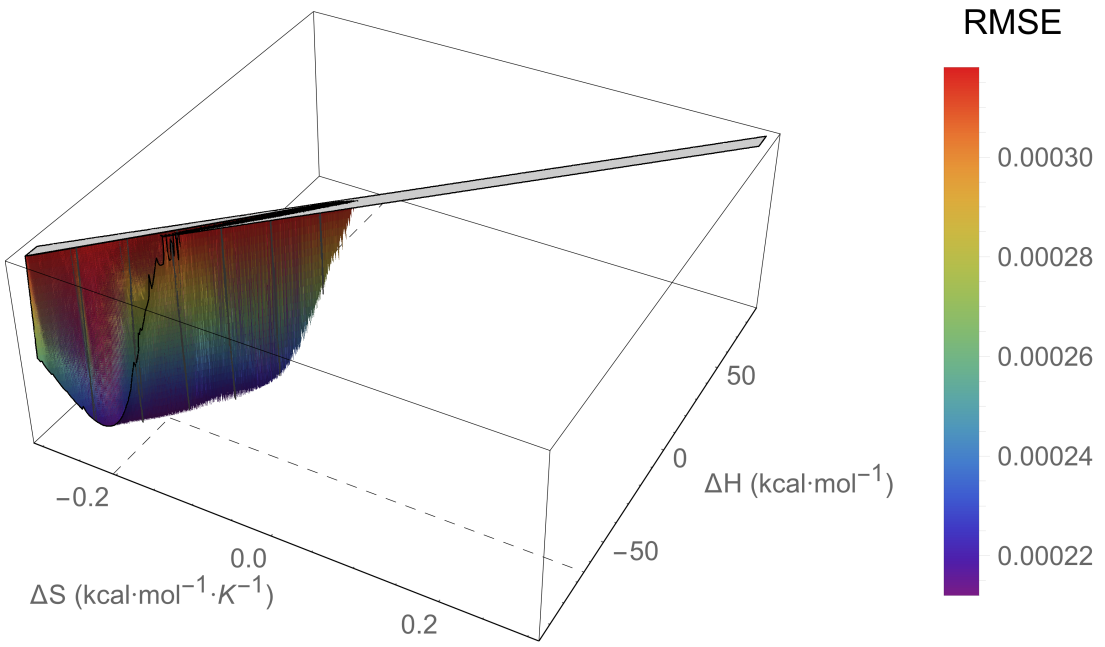


Figure I.207: An RMSE landscape from fitting V75R Adnectin residue 18 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

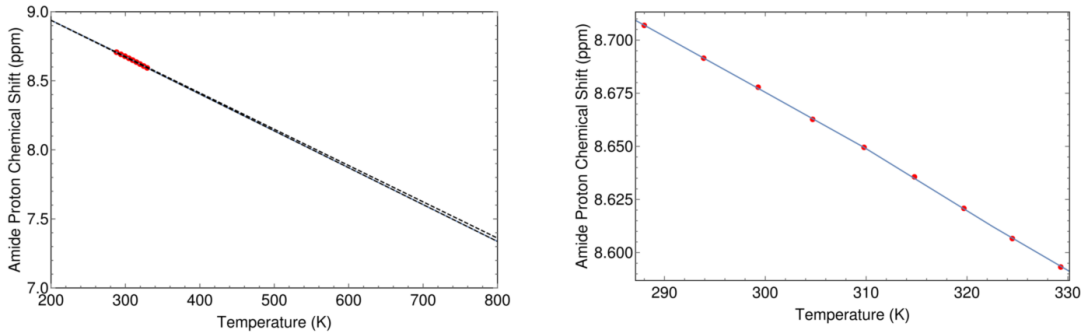


Figure I.208: The minimum RMSE fit of V75R Adnectin residue 18 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

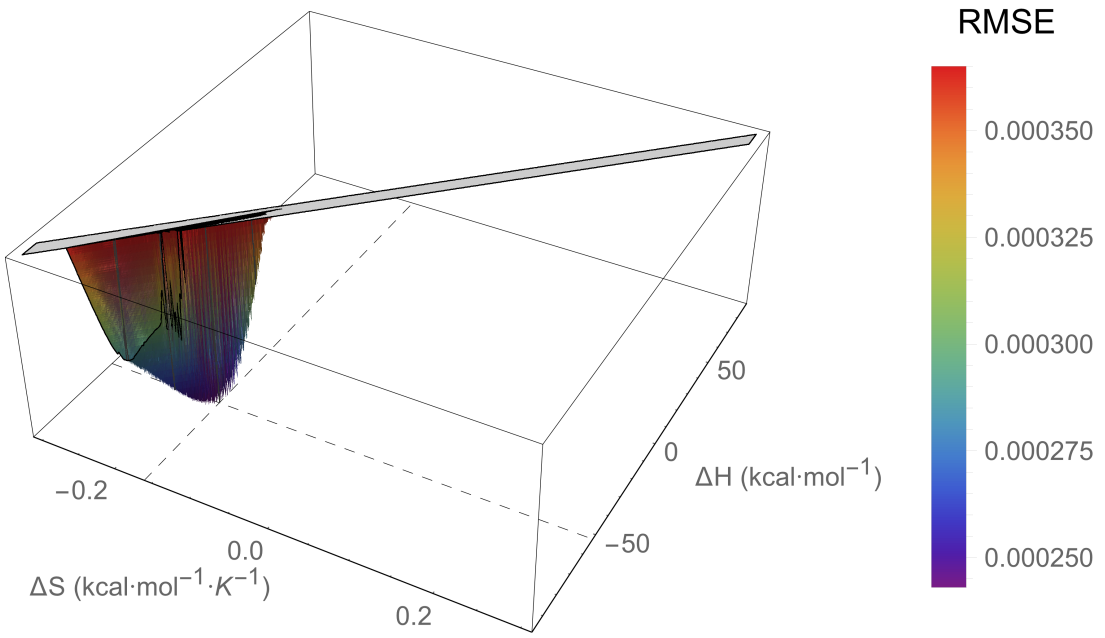


Figure I.209: An RMSE landscape from fitting V75R Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

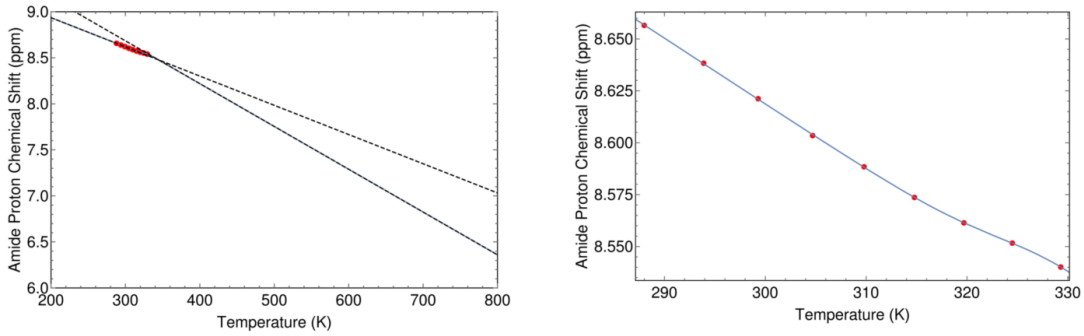


Figure I.210: The minimum RMSE fit of V75R Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

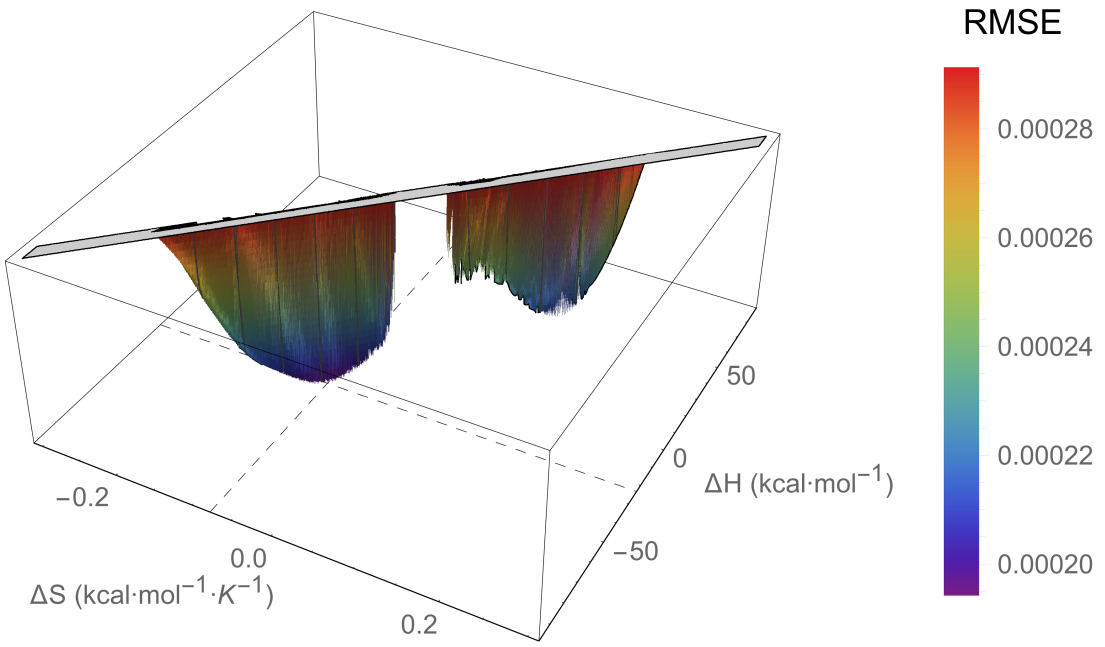


Figure I.211: An RMSE landscape from fitting V75R Adnectin residue 20 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

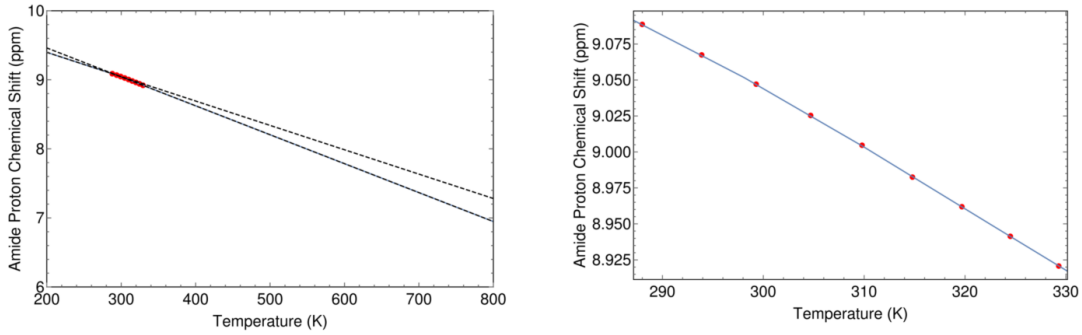


Figure I.212: The minimum RMSE fit of V75R Adnectin residue 20 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

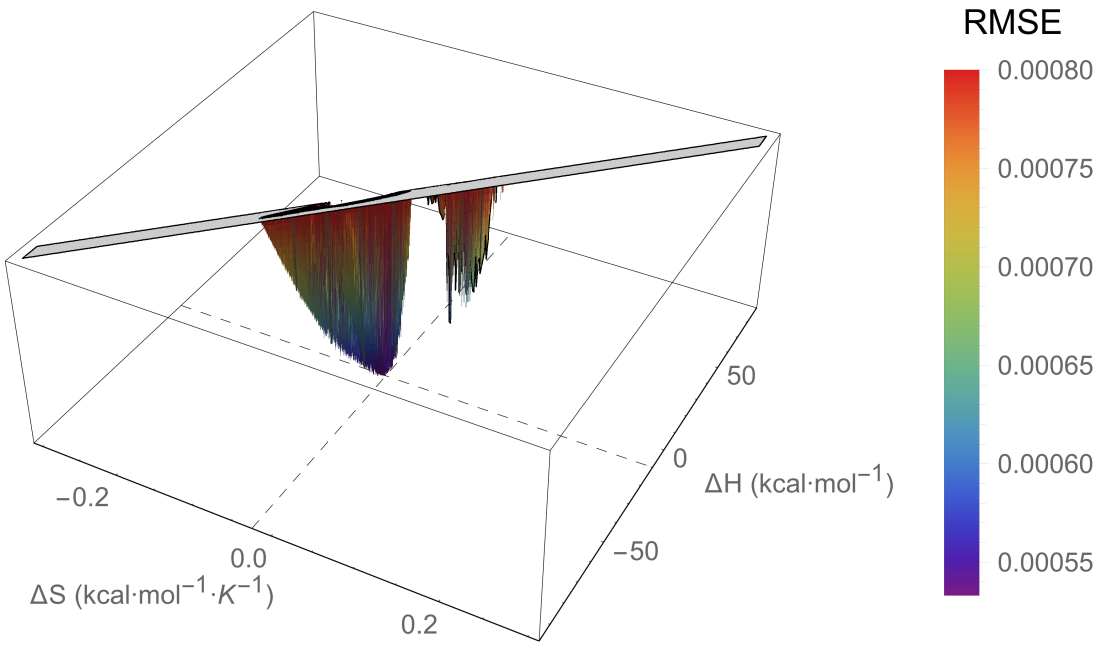


Figure I.213: An RMSE landscape from fitting V75R Adnectin residue 22 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

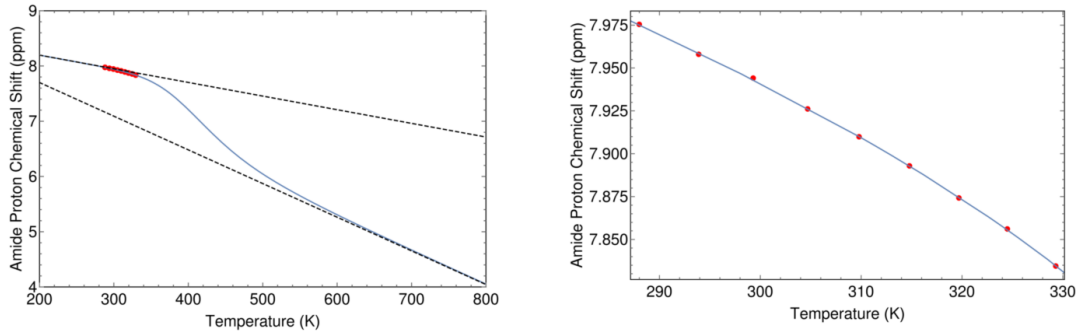


Figure I.214: The minimum RMSE fit of V75R Adnectin residue 22 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

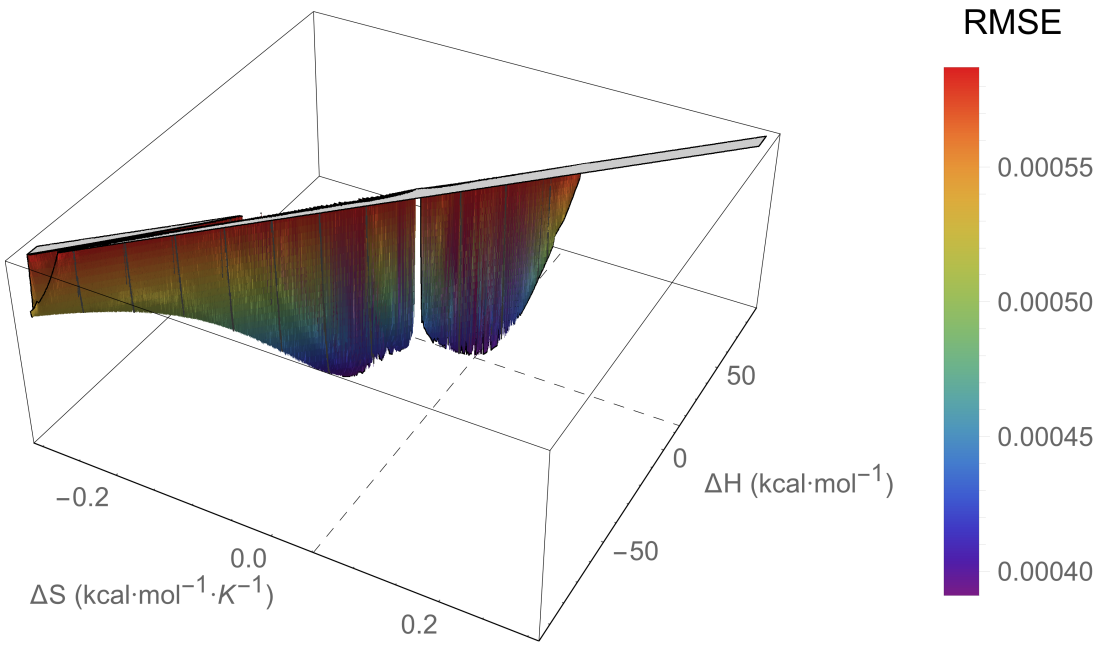


Figure I.215: An RMSE landscape from fitting V75R Adnectin residue 23 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

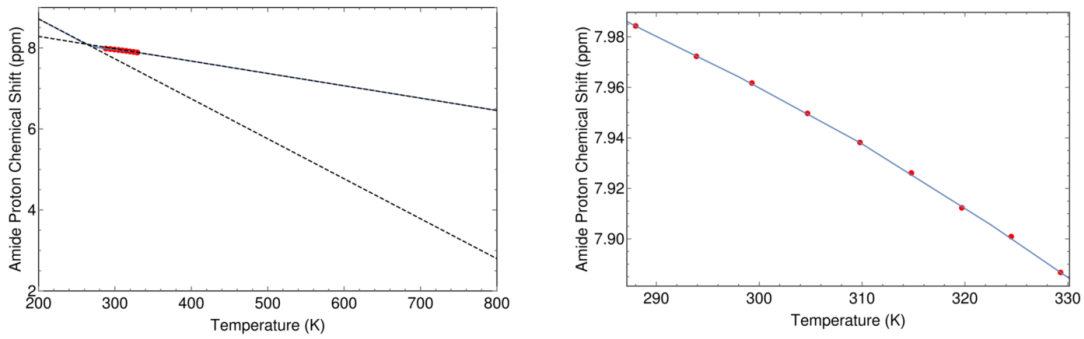


Figure I.216: The minimum RMSE fit of V75R Adnectin residue 23 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

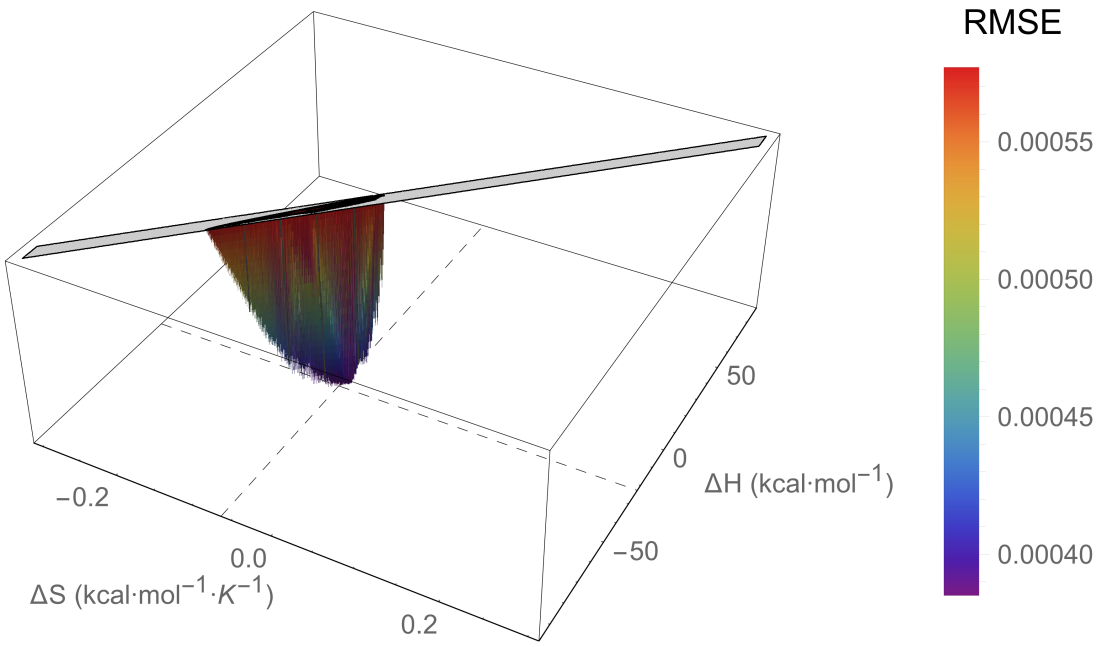


Figure I.217: An RMSE landscape from fitting V75R Adnectin residue 32 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

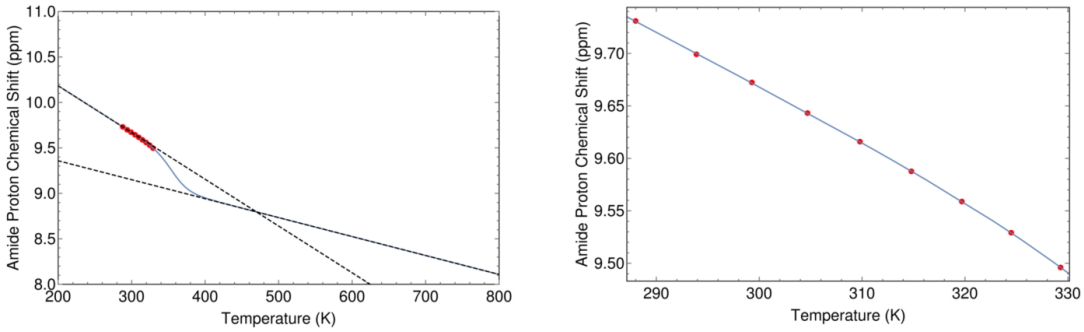


Figure I.218: The minimum RMSE fit of V75R Adnectin residue 32 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

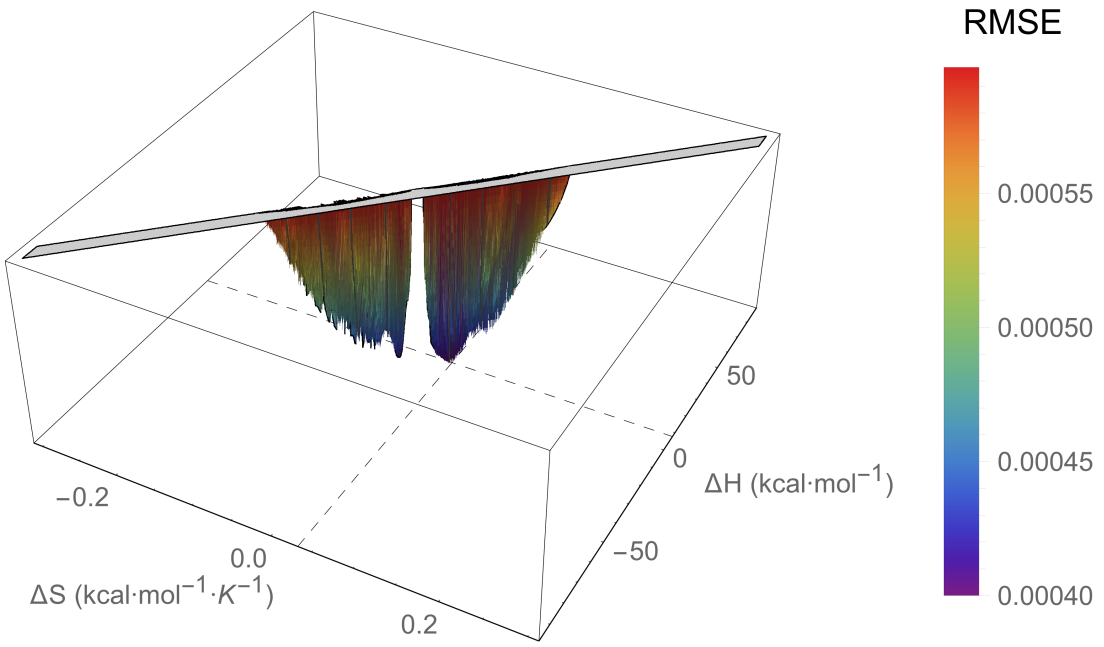


Figure I.219: An RMSE landscape from fitting V75R Adnectin residue 34 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

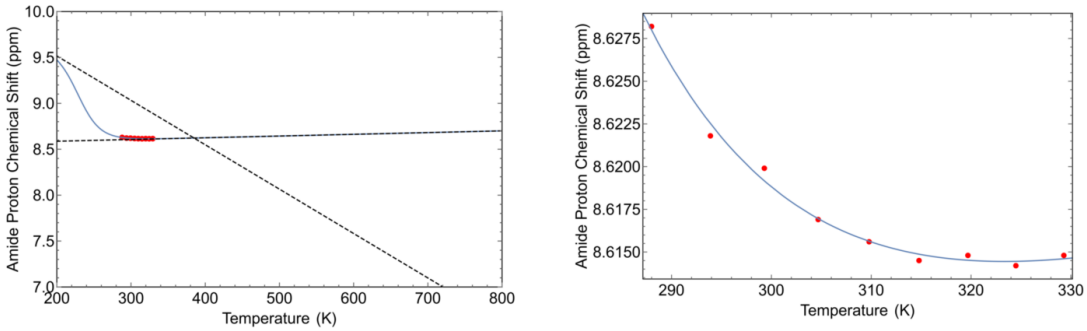


Figure I.220: The minimum RMSE fit of V75R Adnectin residue 34 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

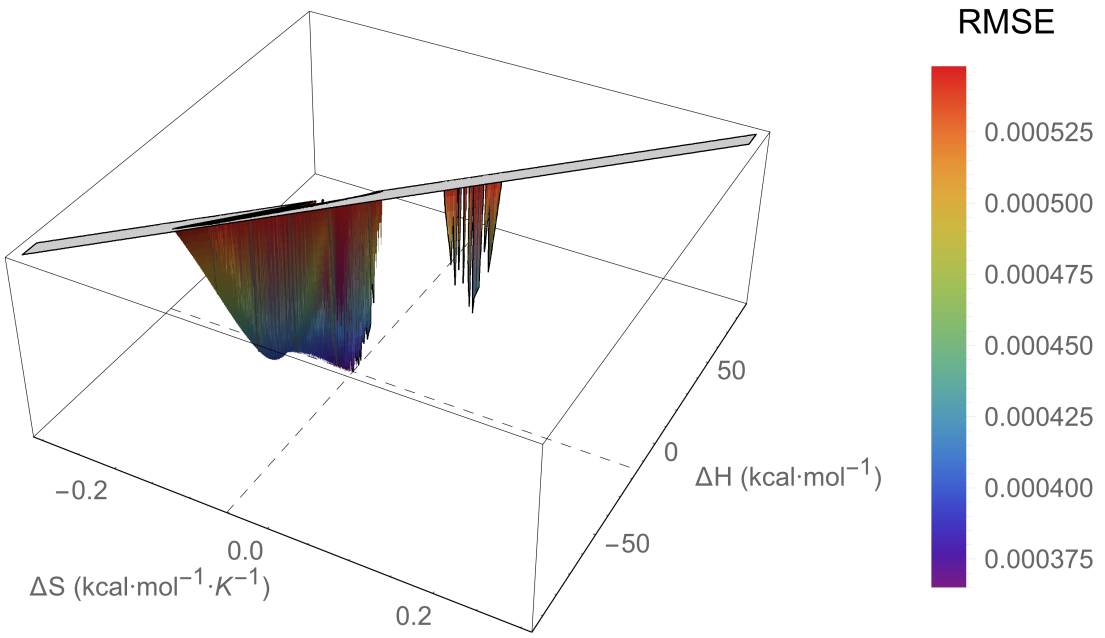


Figure I.221: An RMSE landscape from fitting V75R Adnectin residue 35 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

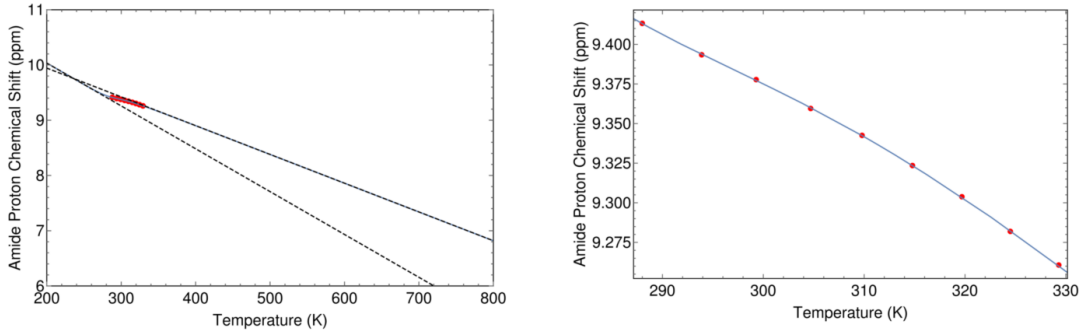


Figure I.222: The minimum RMSE fit of V75R Adnectin residue 35 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

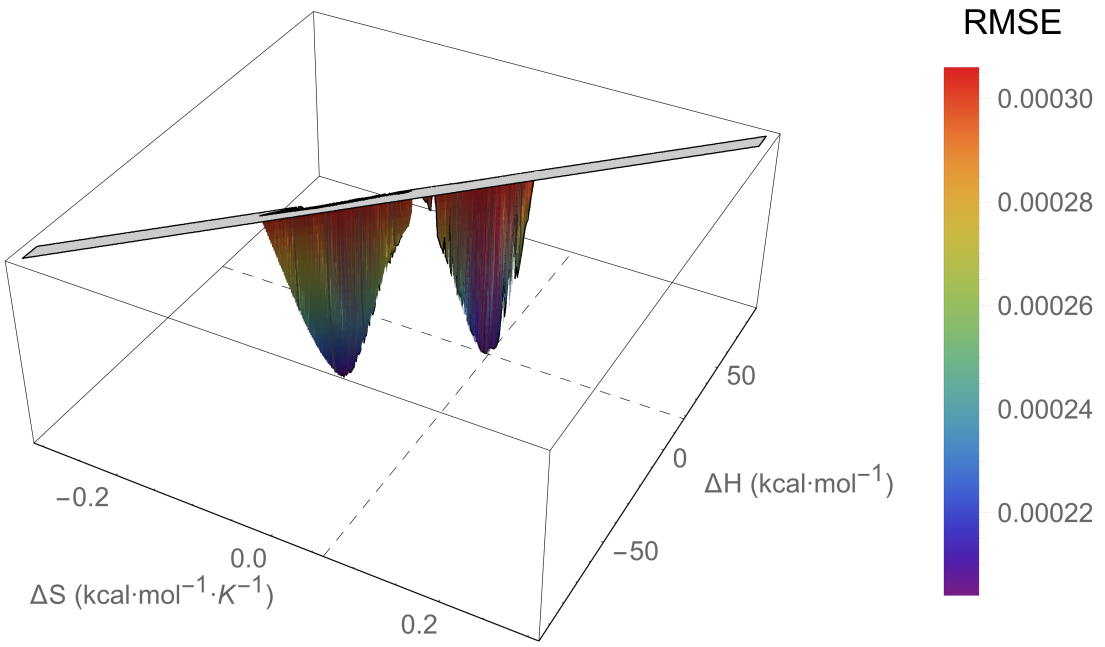


Figure I.223: An RMSE landscape from fitting V75R Adnectin residue 36 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

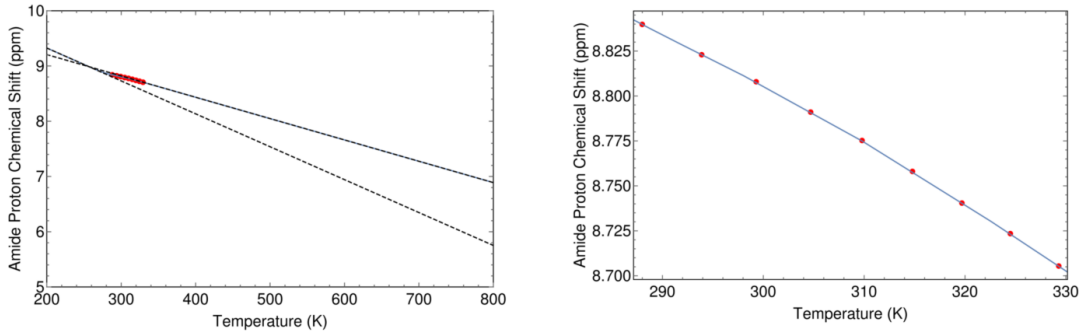


Figure I.224: The minimum RMSE fit of V75R Adnectin residue 36 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

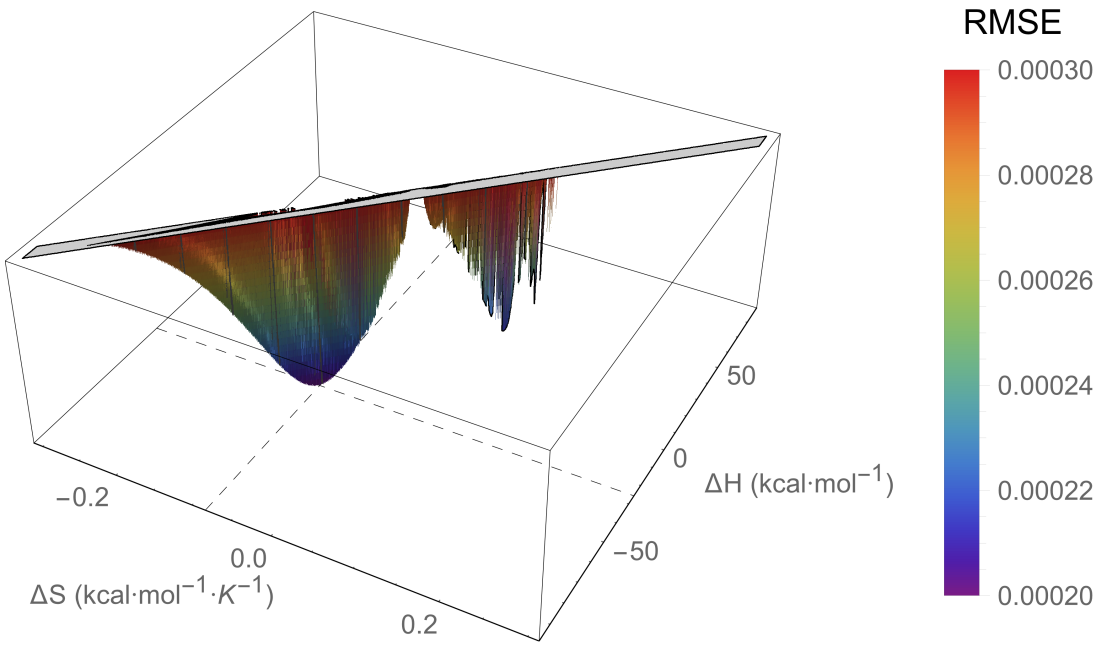


Figure I.225: An RMSE landscape from fitting V75R Adnectin residue 38 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

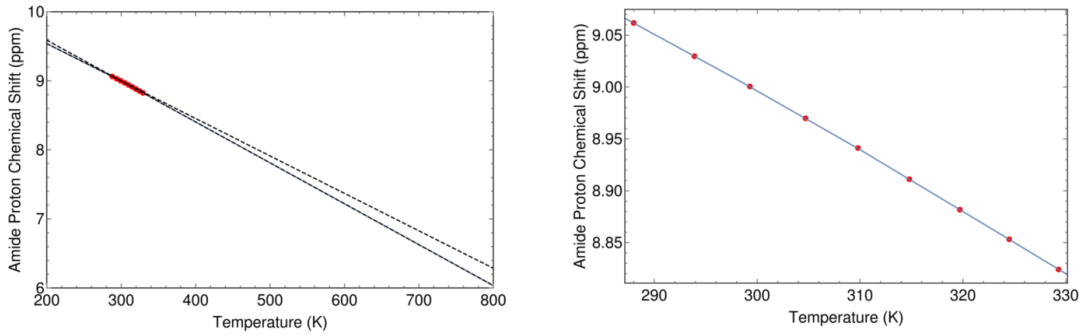


Figure I.226: The minimum RMSE fit of V75R Adnectin residue 38 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

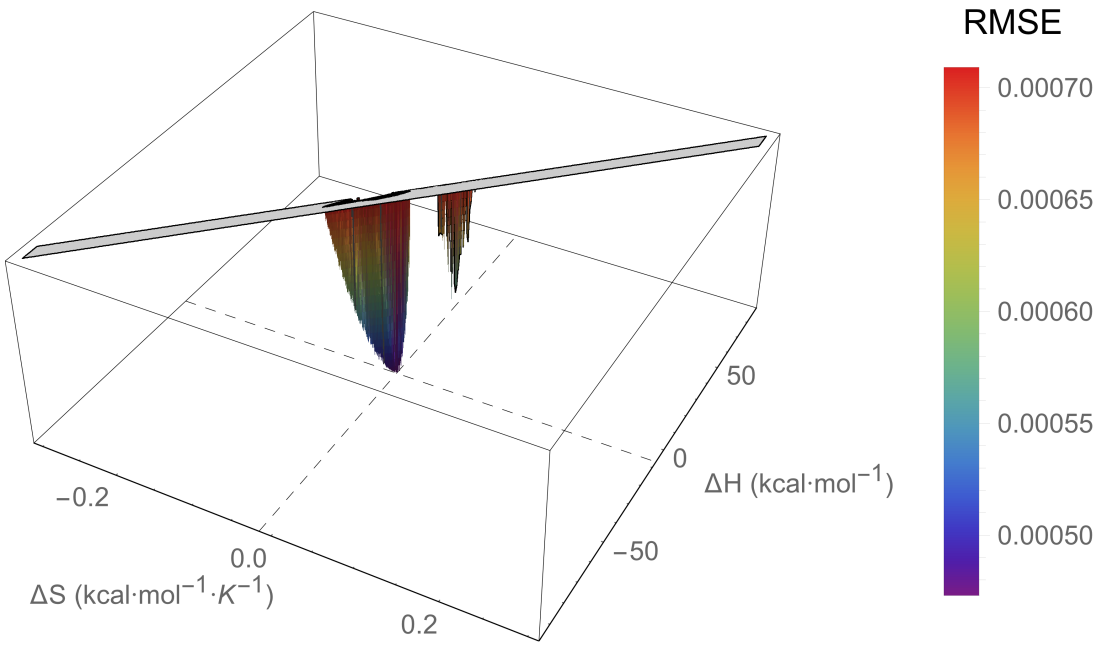


Figure I.227: An RMSE landscape from fitting V75R Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

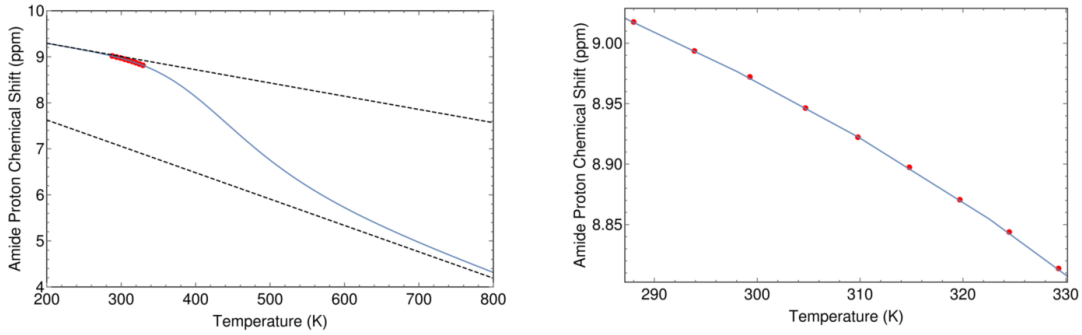


Figure I.228: The minimum RMSE fit of V75R Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

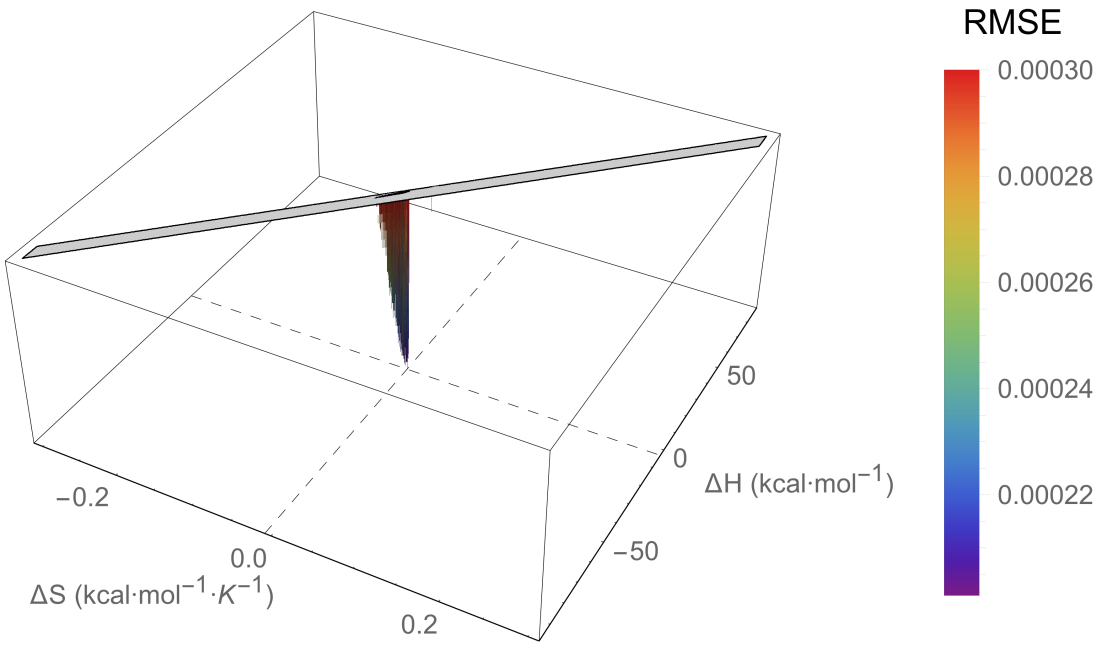


Figure I.229: An RMSE landscape from fitting V75R Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

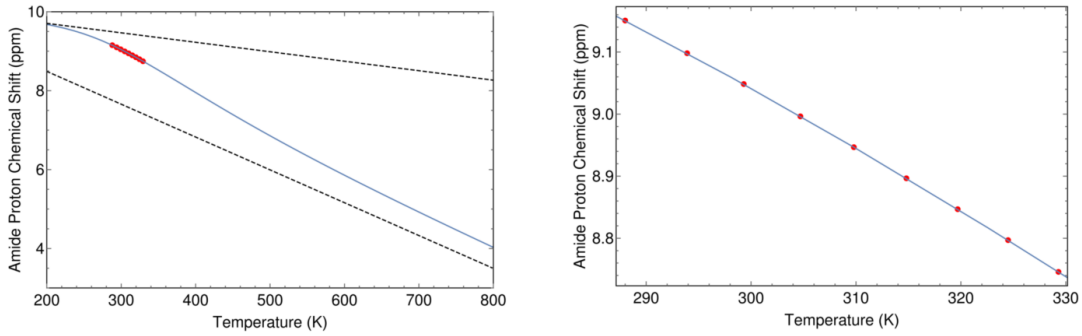


Figure I.230: The minimum RMSE fit of V75R Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

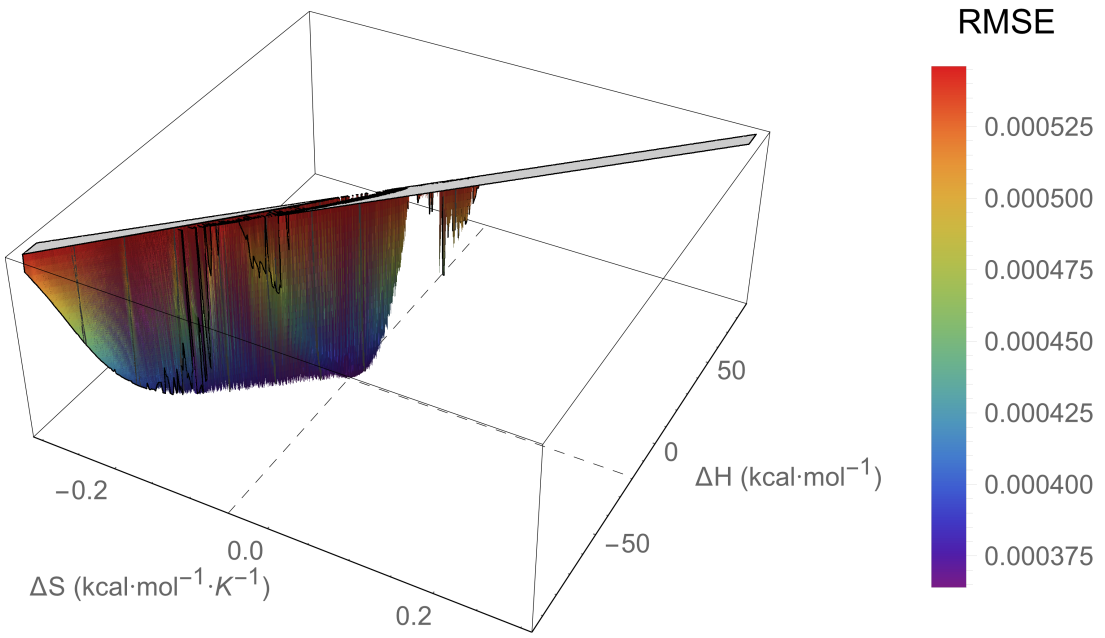


Figure I.231: An RMSE landscape from fitting V75R Adnectin residue 41 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

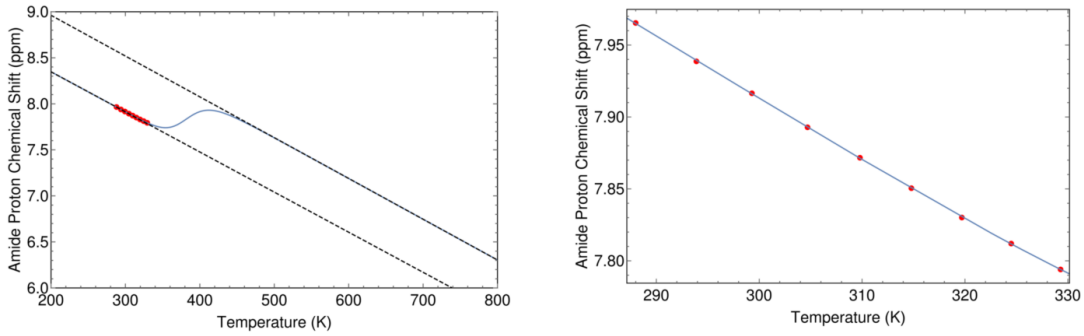


Figure I.232: The minimum RMSE fit of V75R Adnectin residue 41 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

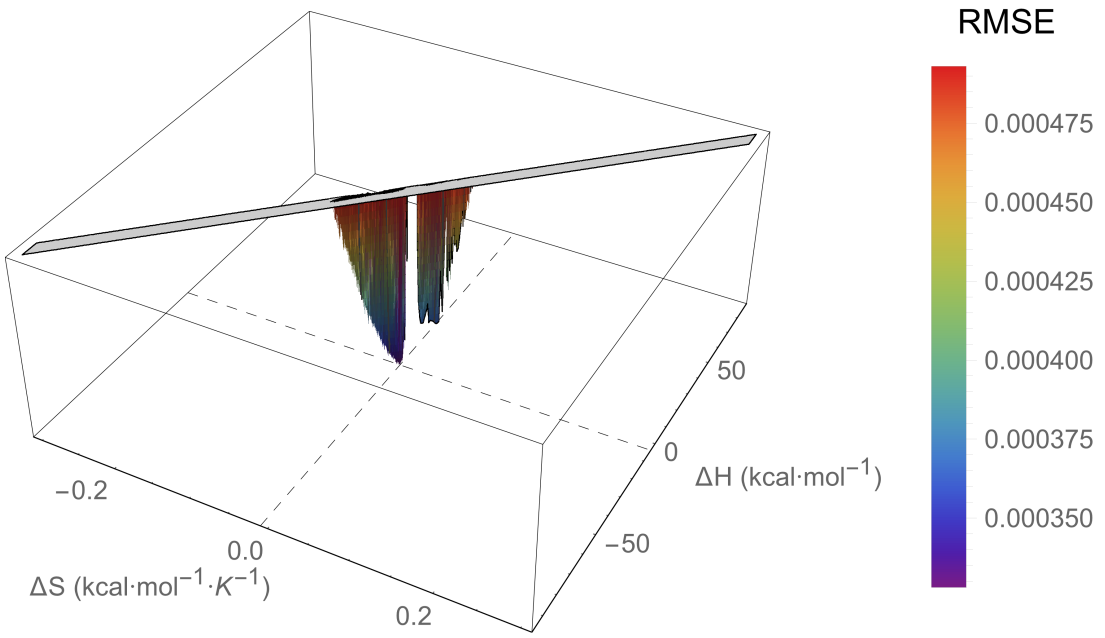


Figure I.233: An RMSE landscape from fitting V75R Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

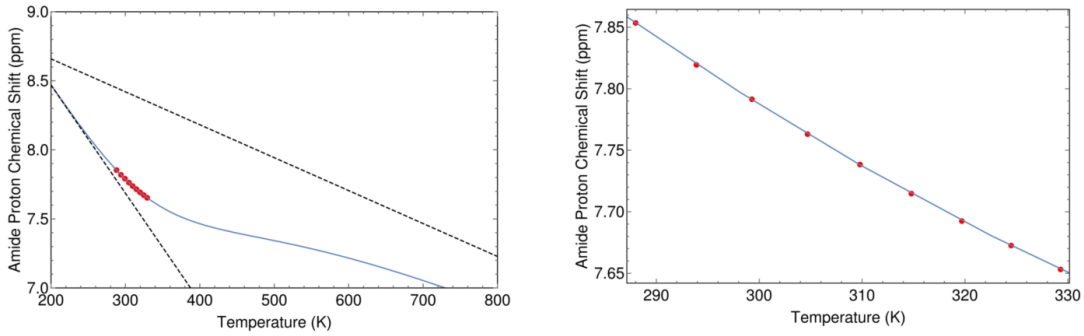


Figure I.234: The minimum RMSE fit of V75R Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

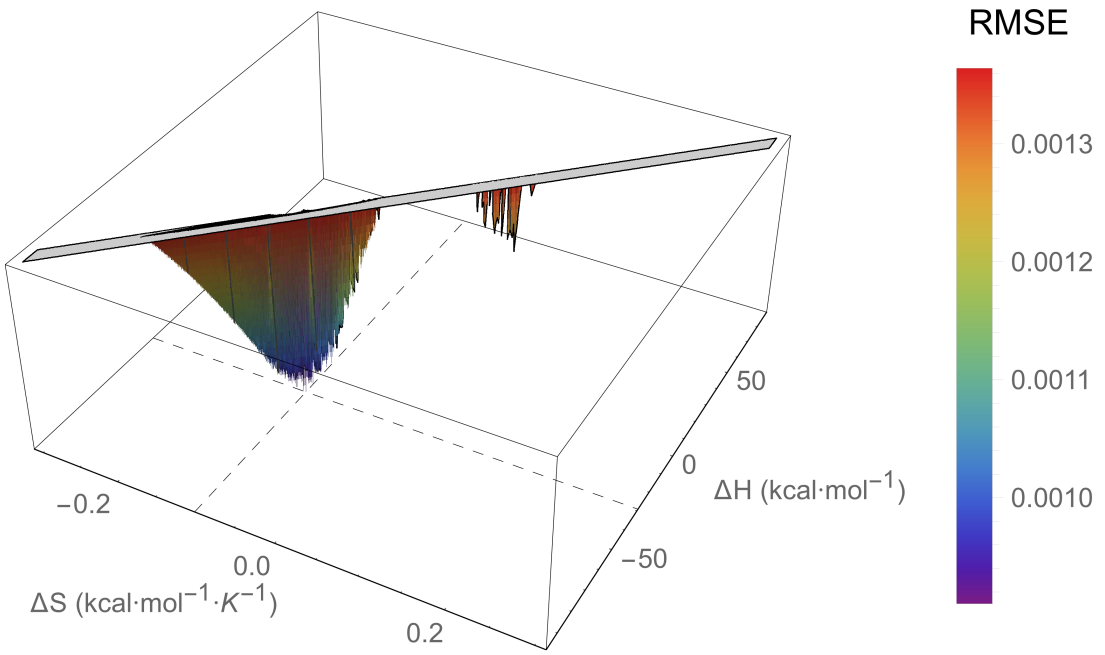


Figure I.235: An RMSE landscape from fitting V75R Adnectin residue 46 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

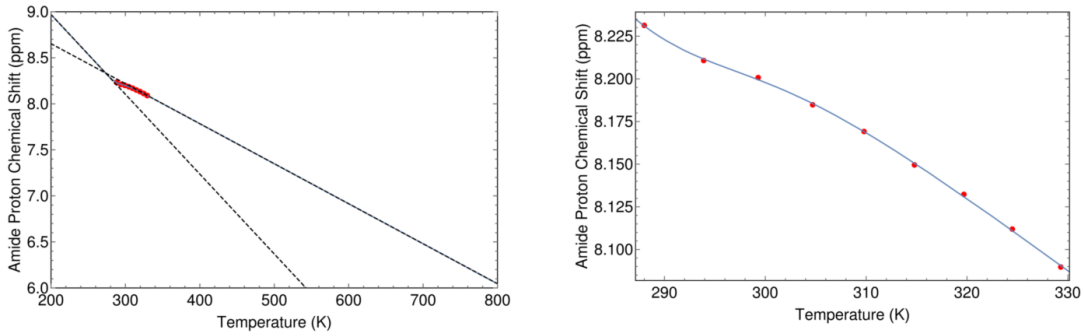


Figure I.236: The minimum RMSE fit of V75R Adnectin residue 46 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

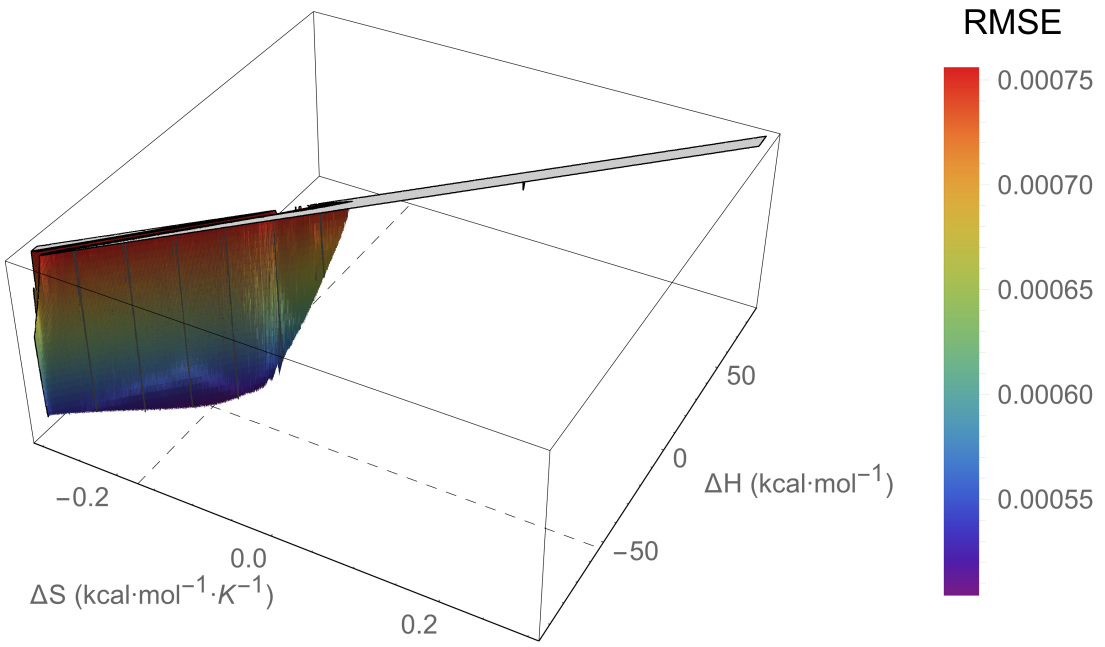


Figure I.237: An RMSE landscape from fitting V75R Adnectin residue 58 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

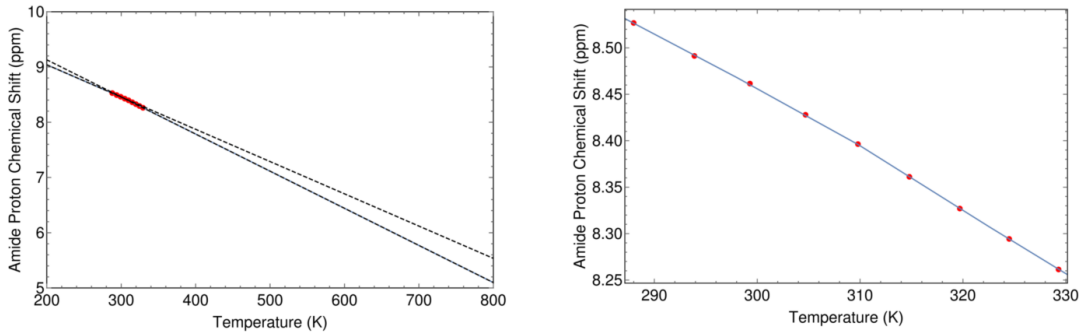


Figure I.238: The minimum RMSE fit of V75R Adnectin residue 58 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

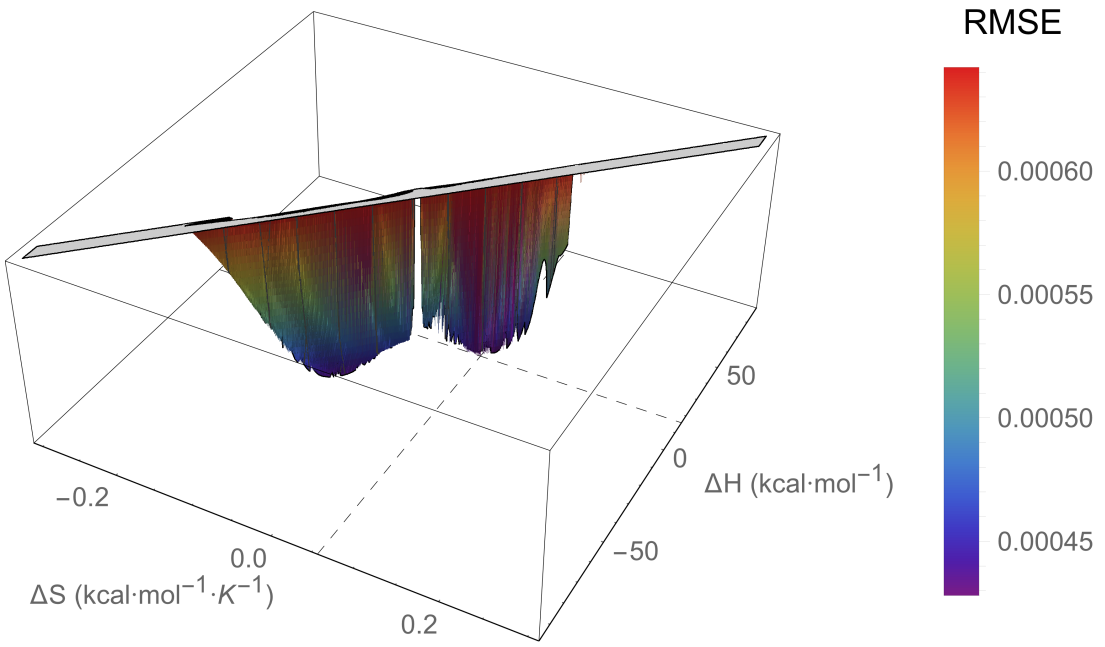


Figure I.239: An RMSE landscape from fitting V75R Adnectin residue 59 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

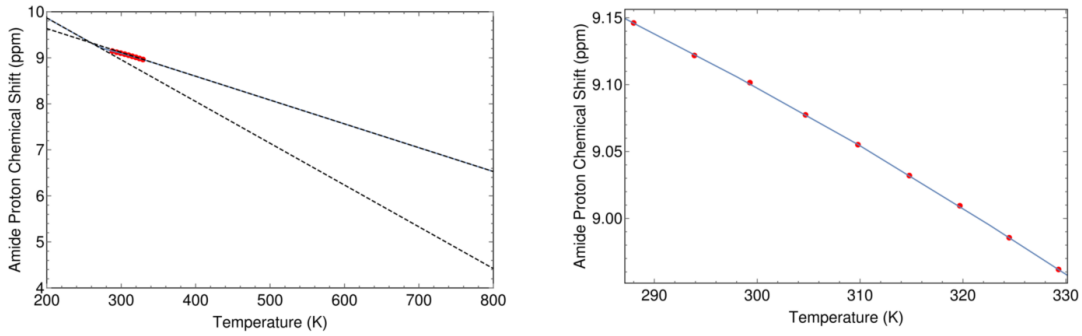


Figure I.240: The minimum RMSE fit of V75R Adnectin residue 59 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

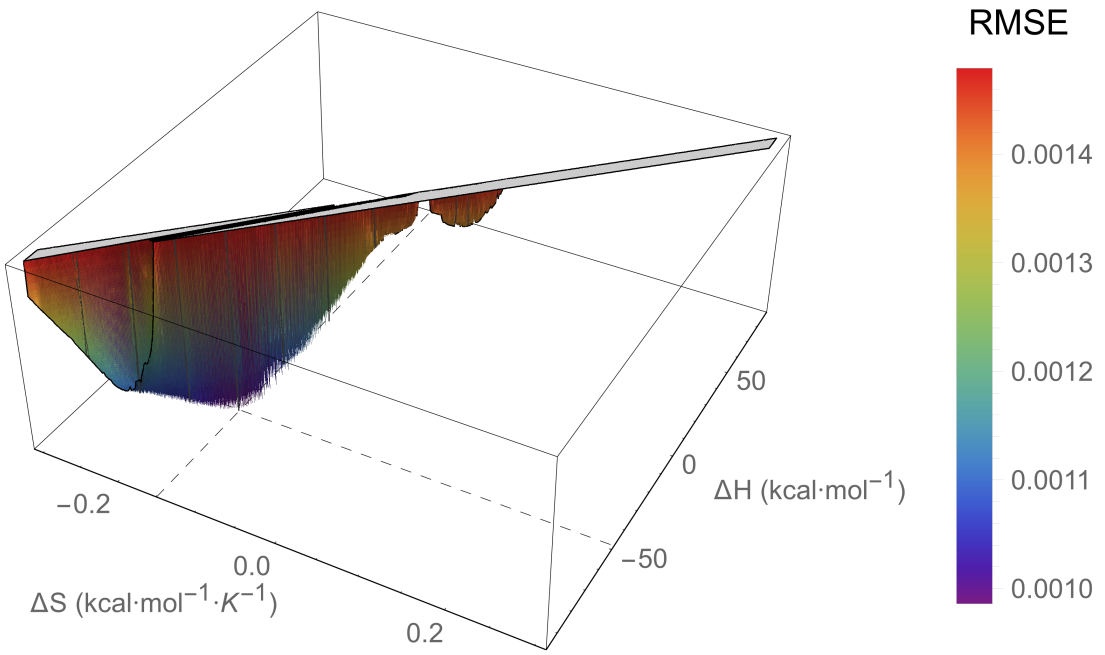


Figure I.241: An RMSE landscape from fitting V75R Adnectin residue 60 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

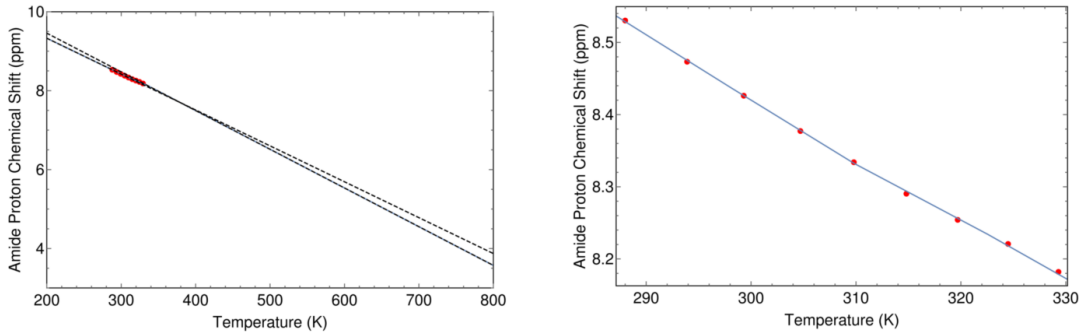


Figure I.242: The minimum RMSE fit of V75R Adnectin residue 60 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

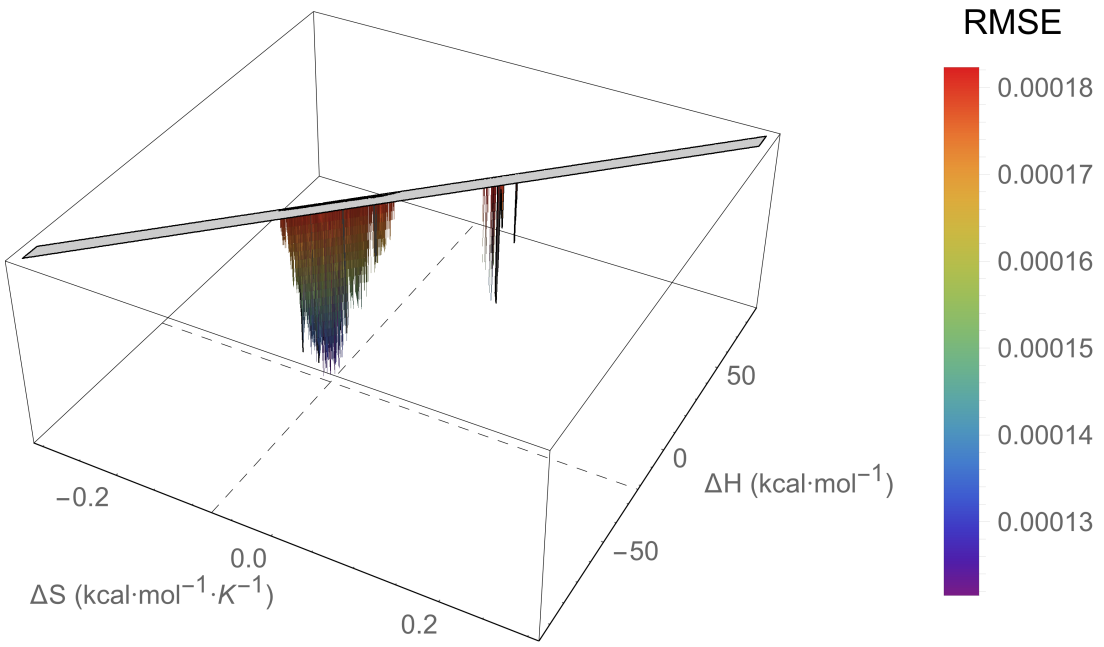


Figure I.243: An RMSE landscape from fitting V75R Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

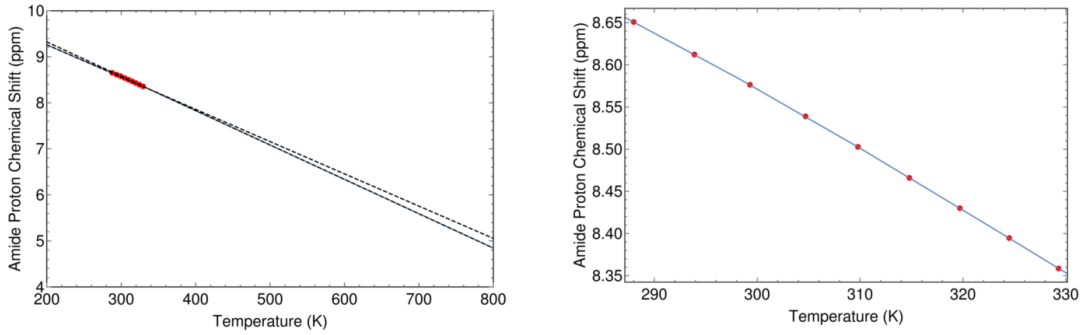


Figure I.244: The minimum RMSE fit of V75R Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

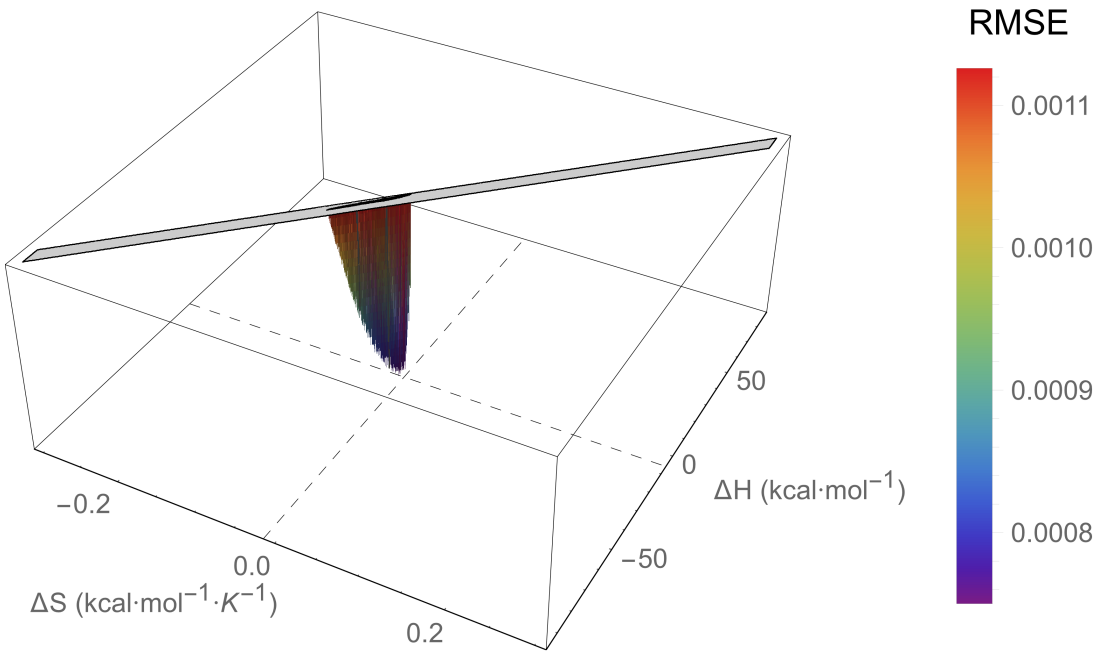


Figure I.245: An RMSE landscape from fitting V75R Adnectin residue 65 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

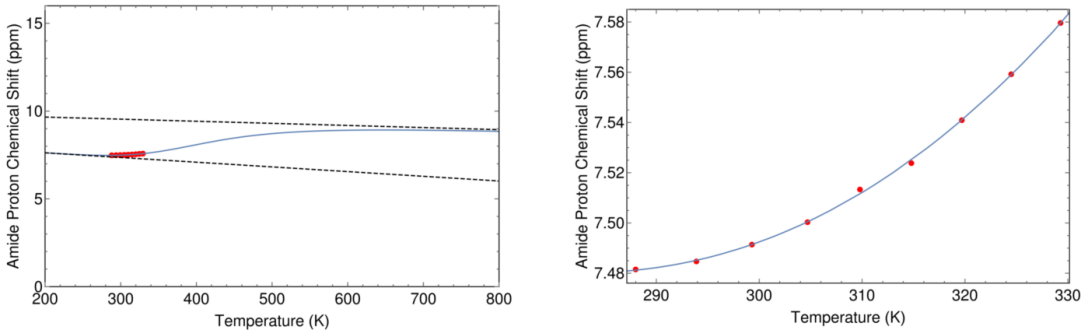


Figure I.246: The minimum RMSE fit of V75R Adnectin residue 65 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

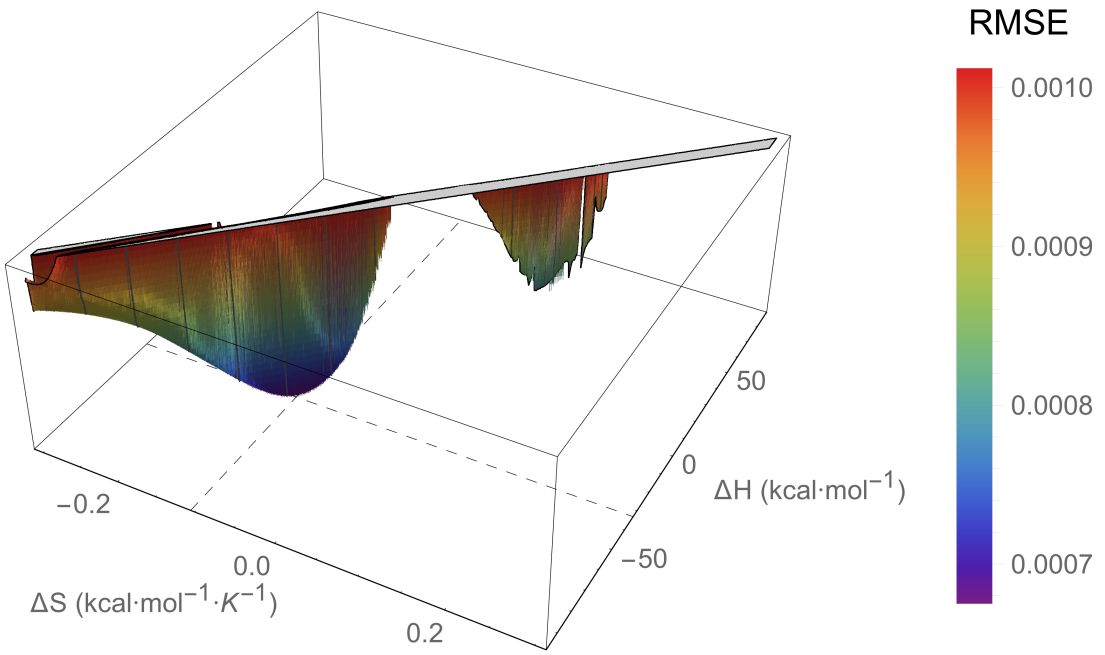


Figure I.247: An RMSE landscape from fitting V75R Adnectin residue 72 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

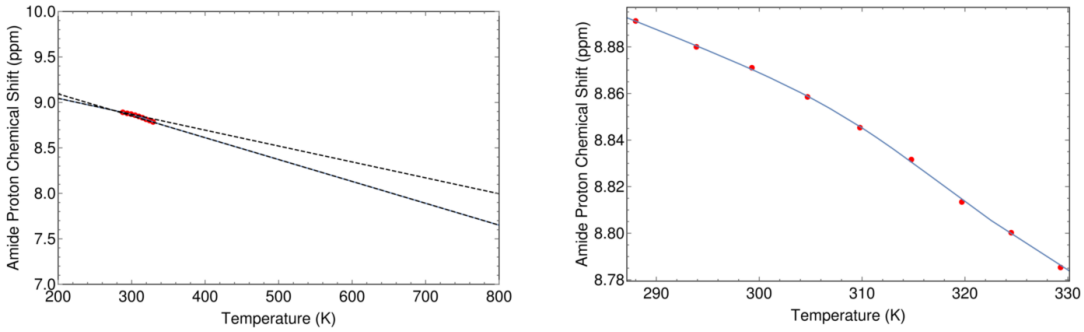


Figure I.248: The minimum RMSE fit of V75R Adnectin residue 72 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

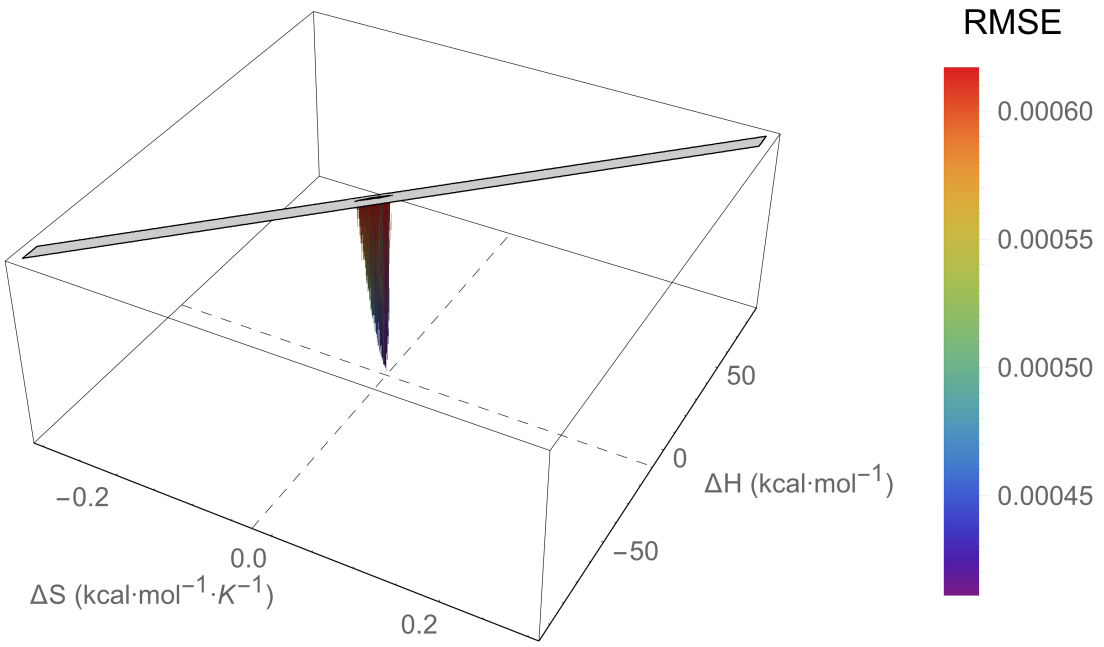


Figure I.249: An RMSE landscape from fitting V75R Adnectin residue 75 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

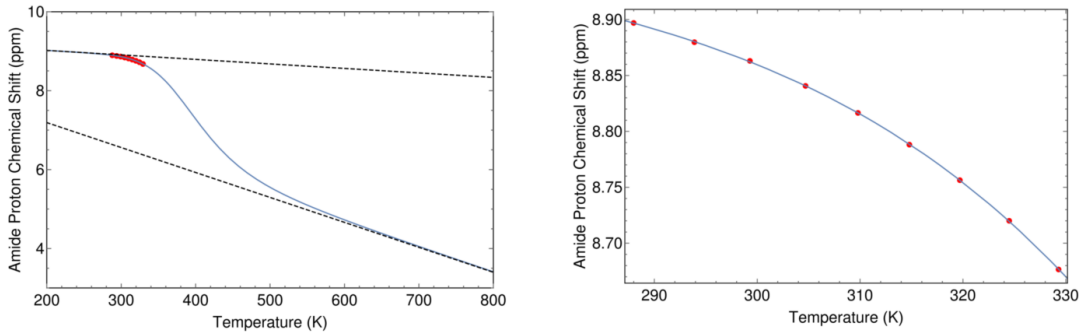


Figure I.250: The minimum RMSE fit of V75R Adnectin residue 75 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

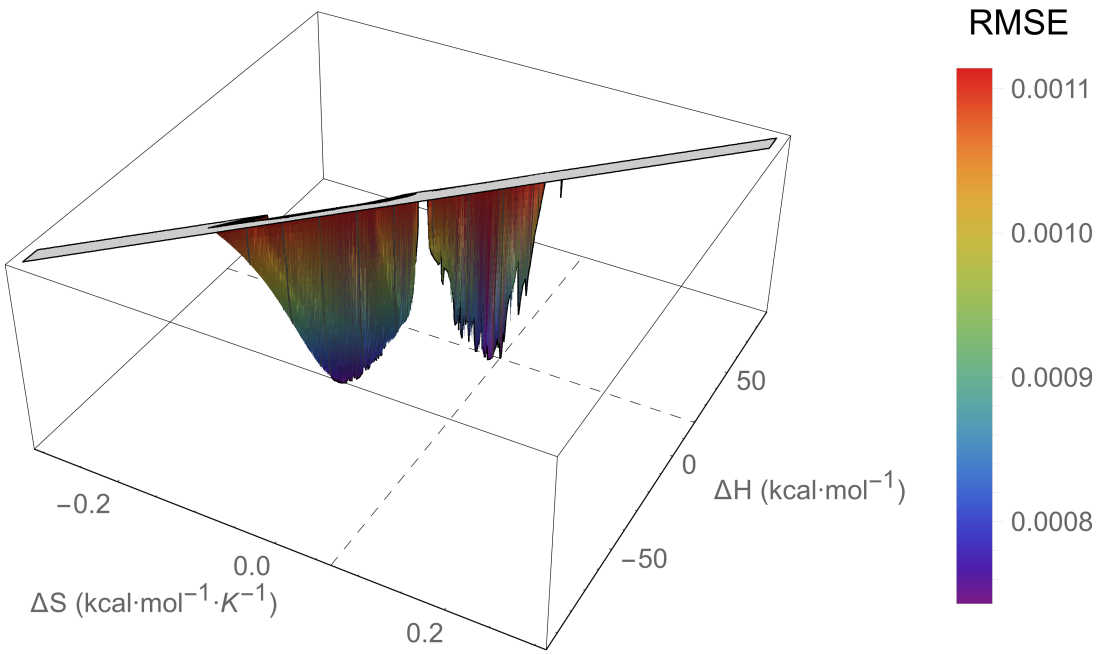


Figure I.251: An RMSE landscape from fitting V75R Adnectin residue 84 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

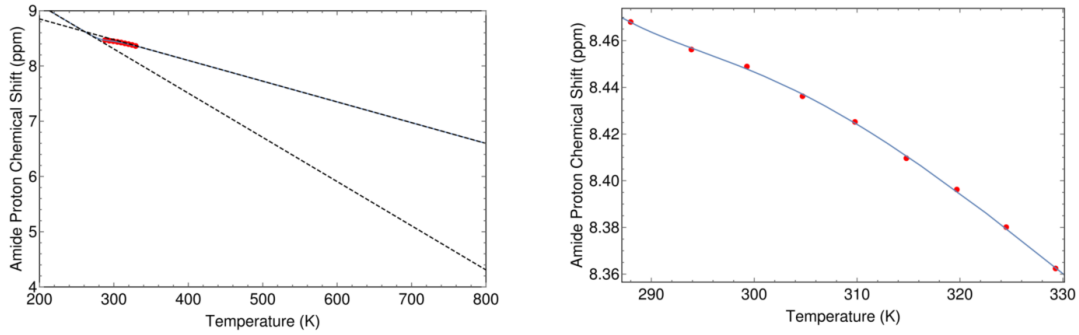


Figure I.252: The minimum RMSE fit of V75R Adnectin residue 84 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

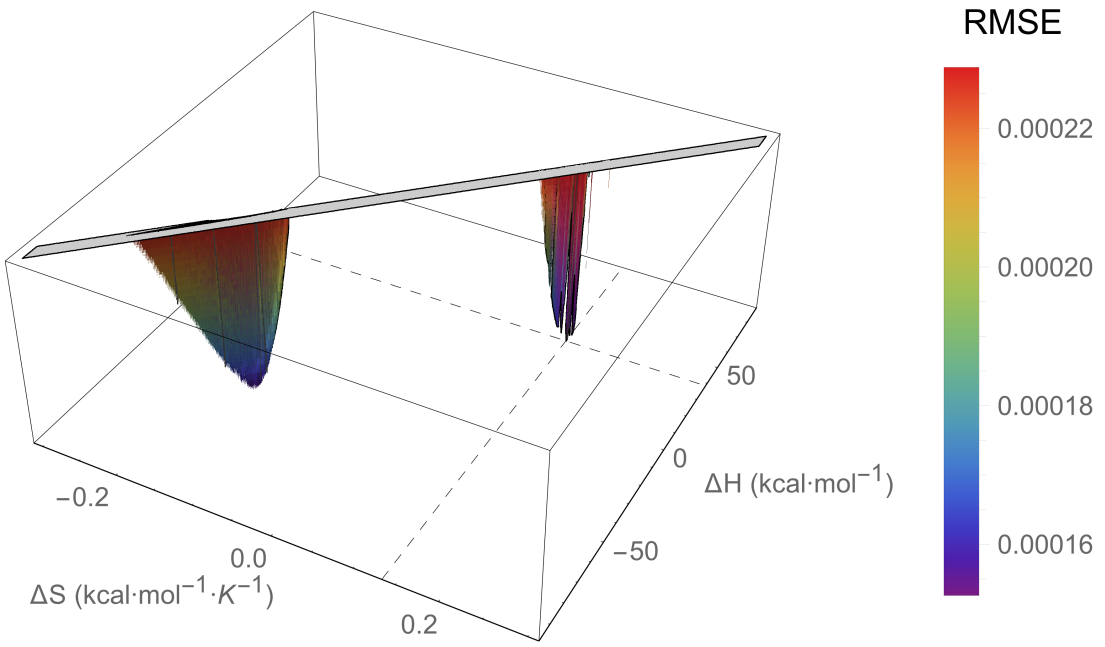


Figure I.253: An RMSE landscape from fitting V75R Adnectin residue 88 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

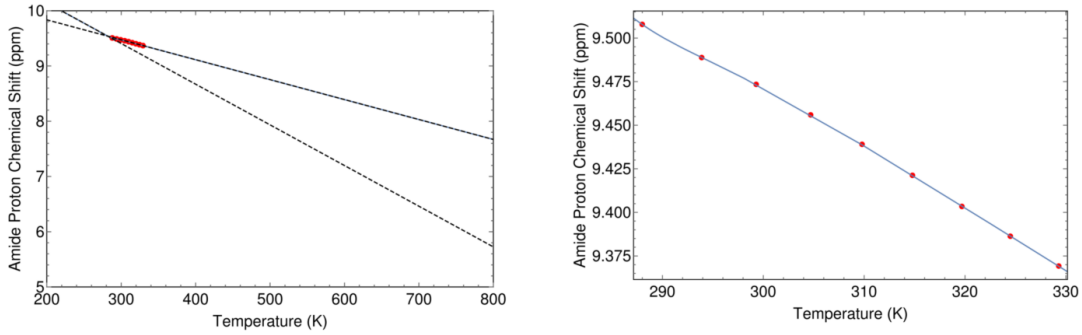


Figure I.254: The minimum RMSE fit of V75R Adnectin residue 88 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

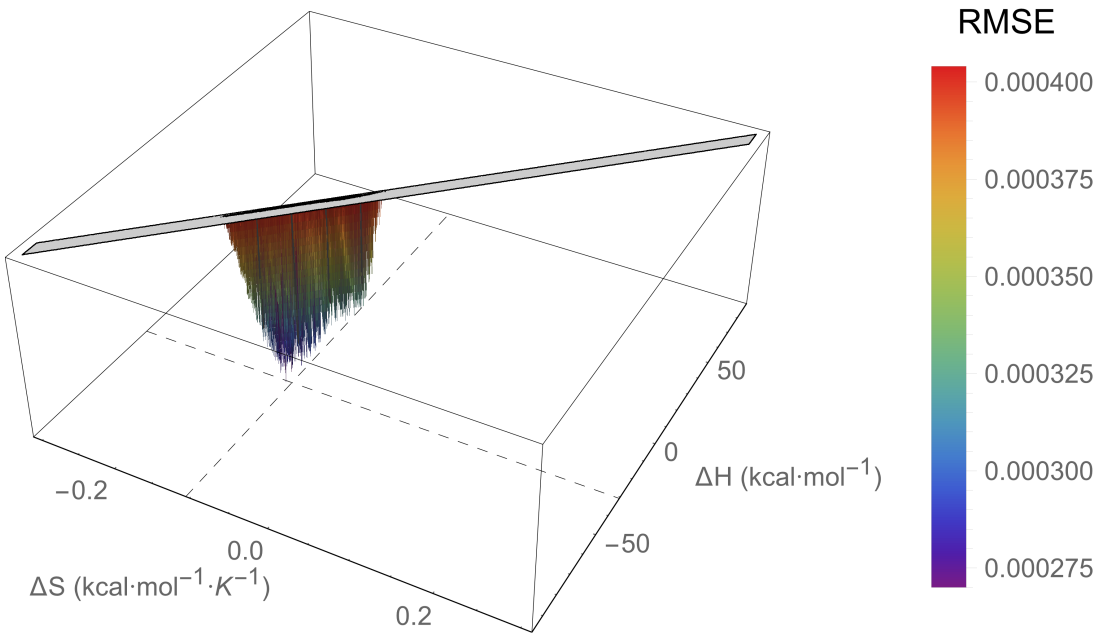


Figure I.255: An RMSE landscape from fitting V75R Adnectin residue 89 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

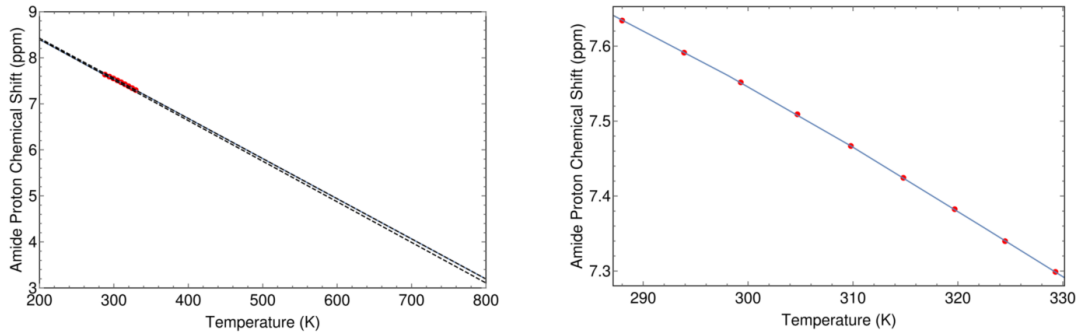


Figure I.256: The minimum RMSE fit of V75R Adnectin residue 89 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

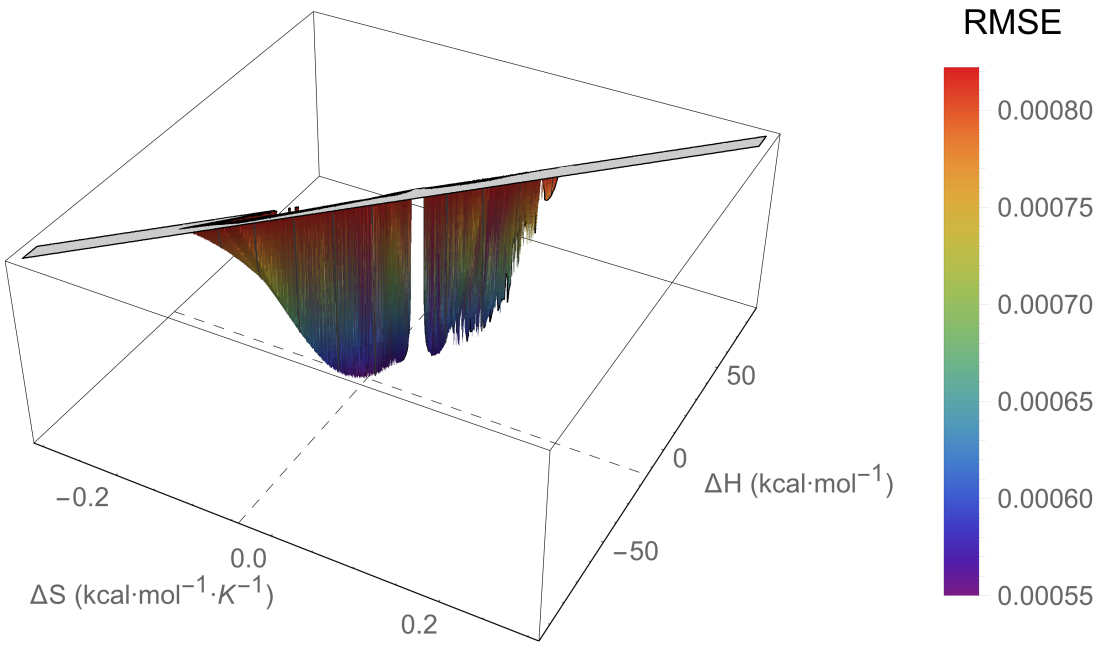


Figure I.257: An RMSE landscape from fitting V75R Adnectin residue 90 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

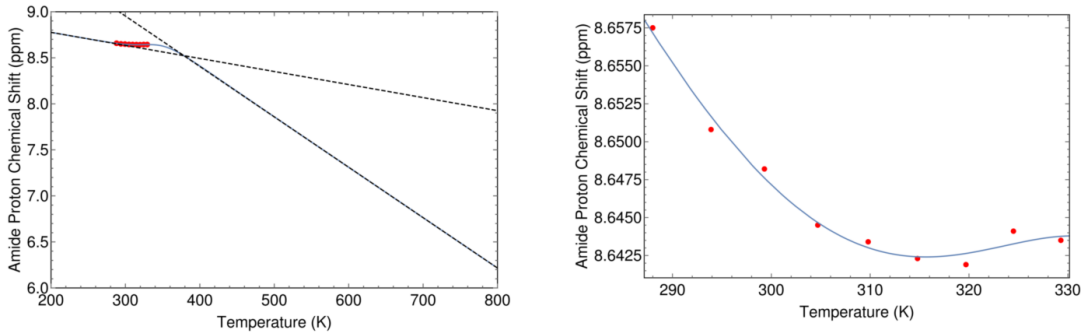


Figure I.258: The minimum RMSE fit of V75R Adnectin residue 90 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (black dashed lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

I.11 L18V-Y88F Adnectin Curvalyzer Results

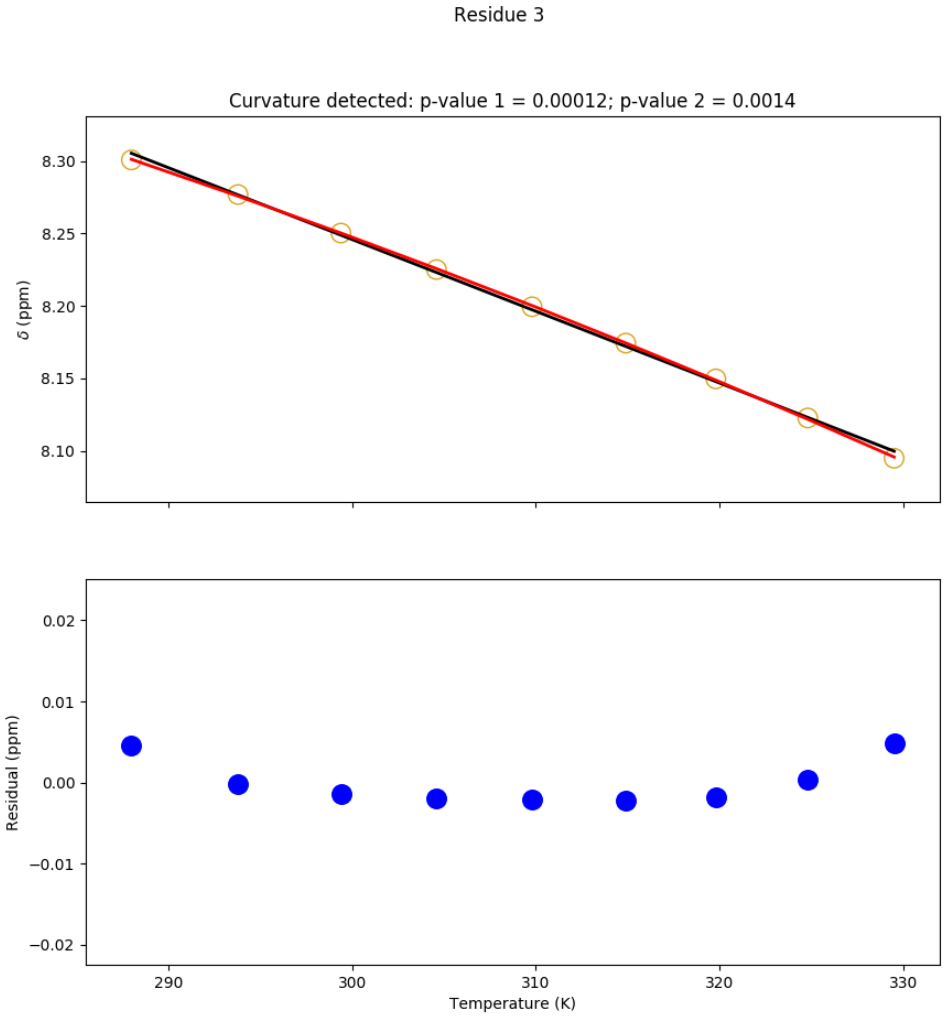


Figure I.259: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 3. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

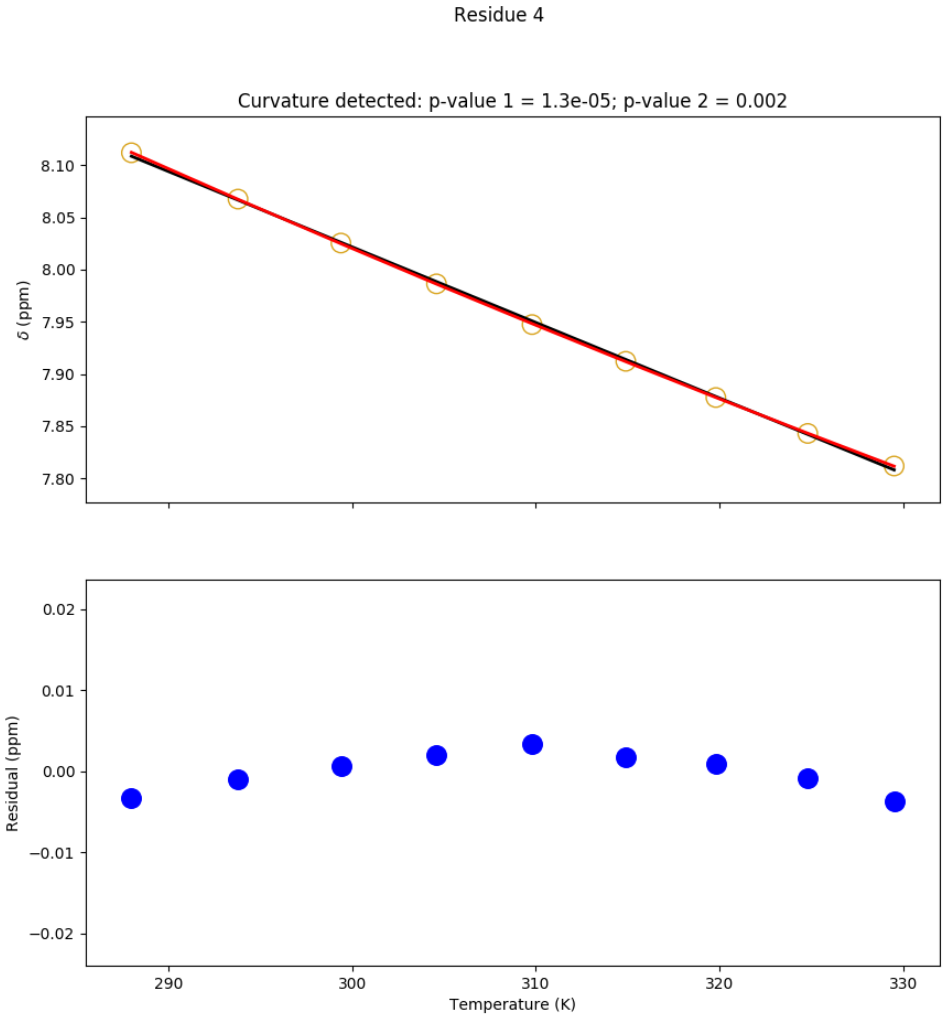


Figure I.260: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 4. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

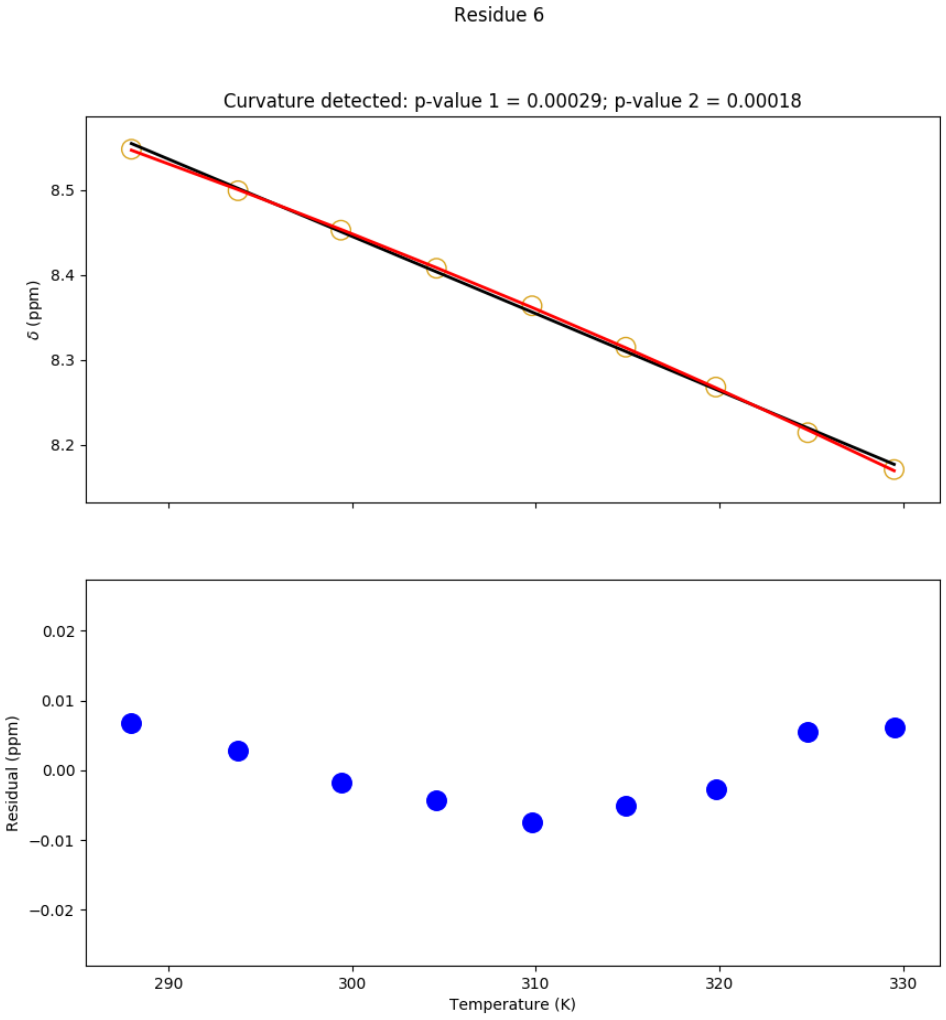


Figure I.261: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 6. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

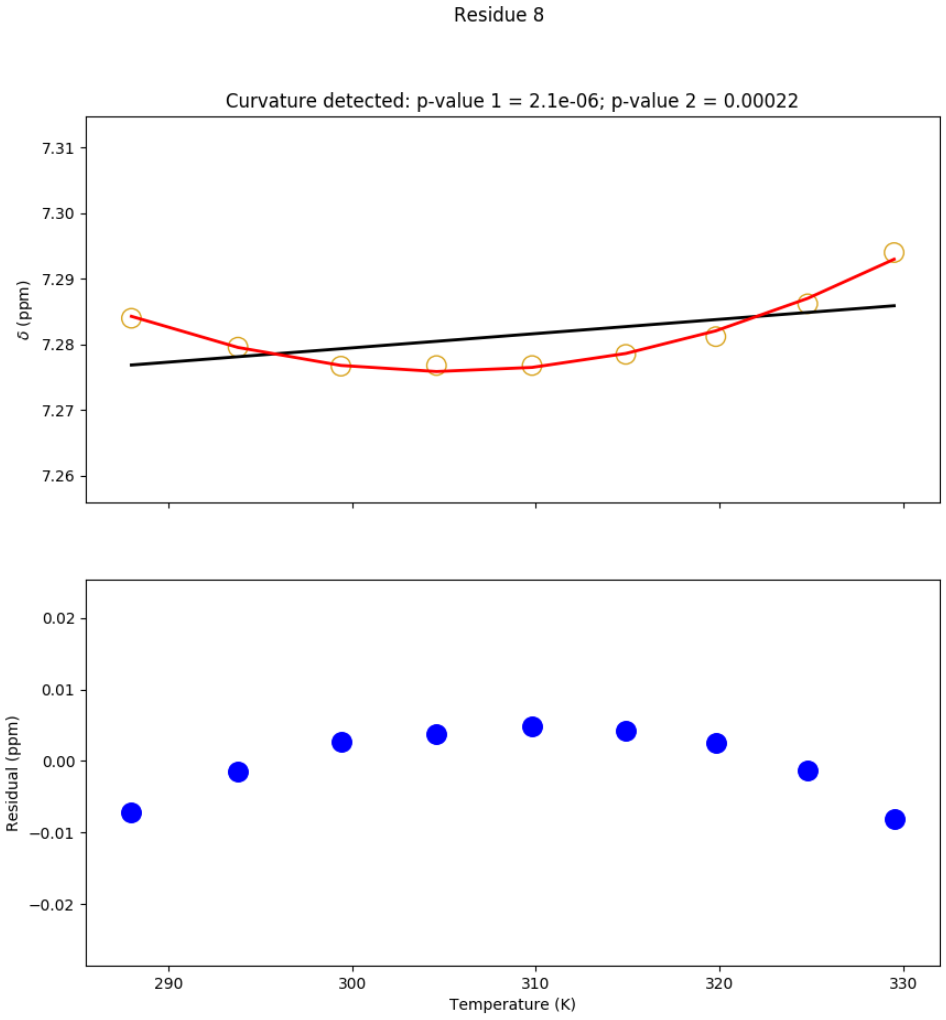


Figure I.262: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 8. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

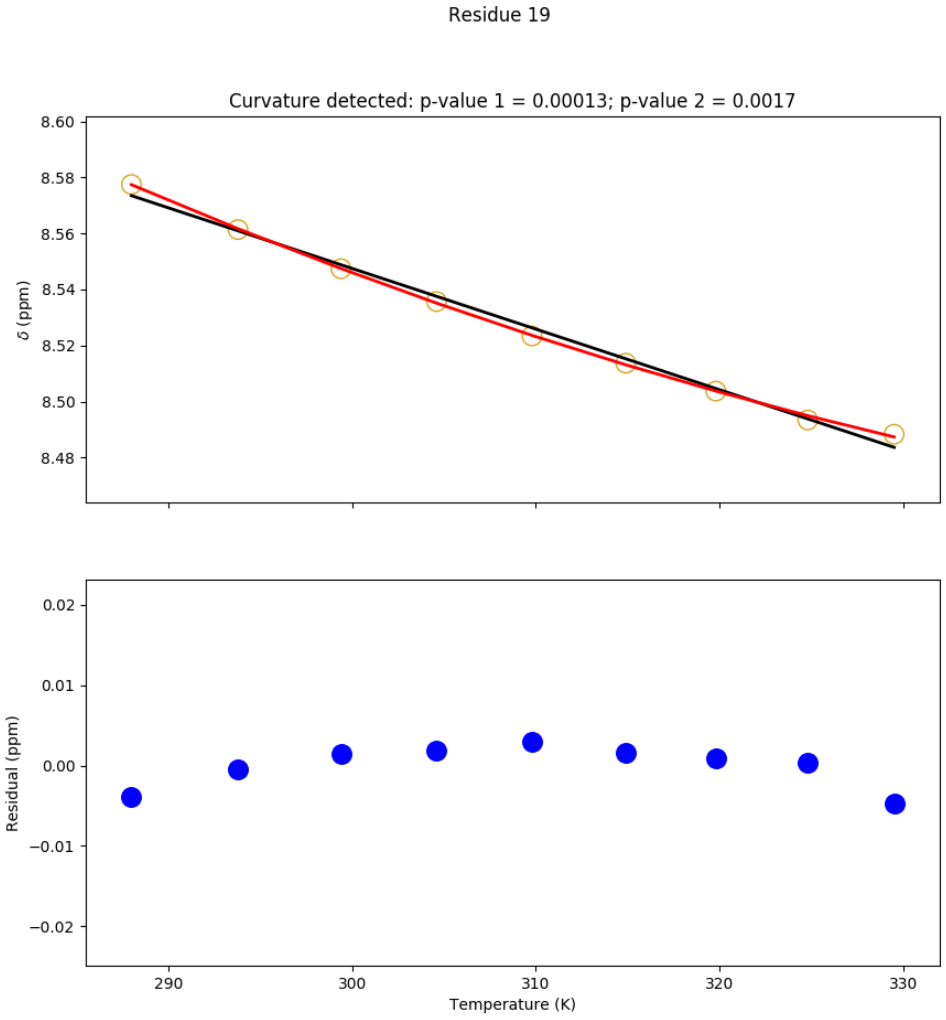


Figure I.263: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 19. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

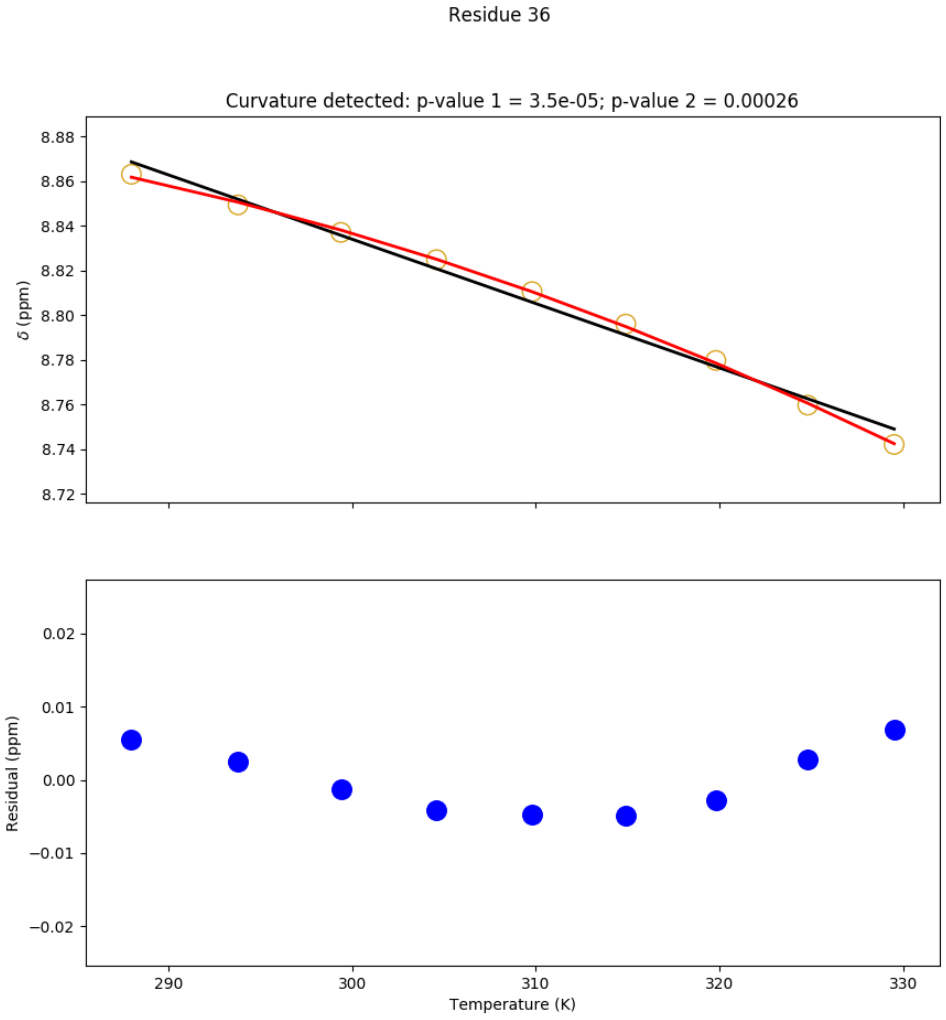


Figure I.264: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 36. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

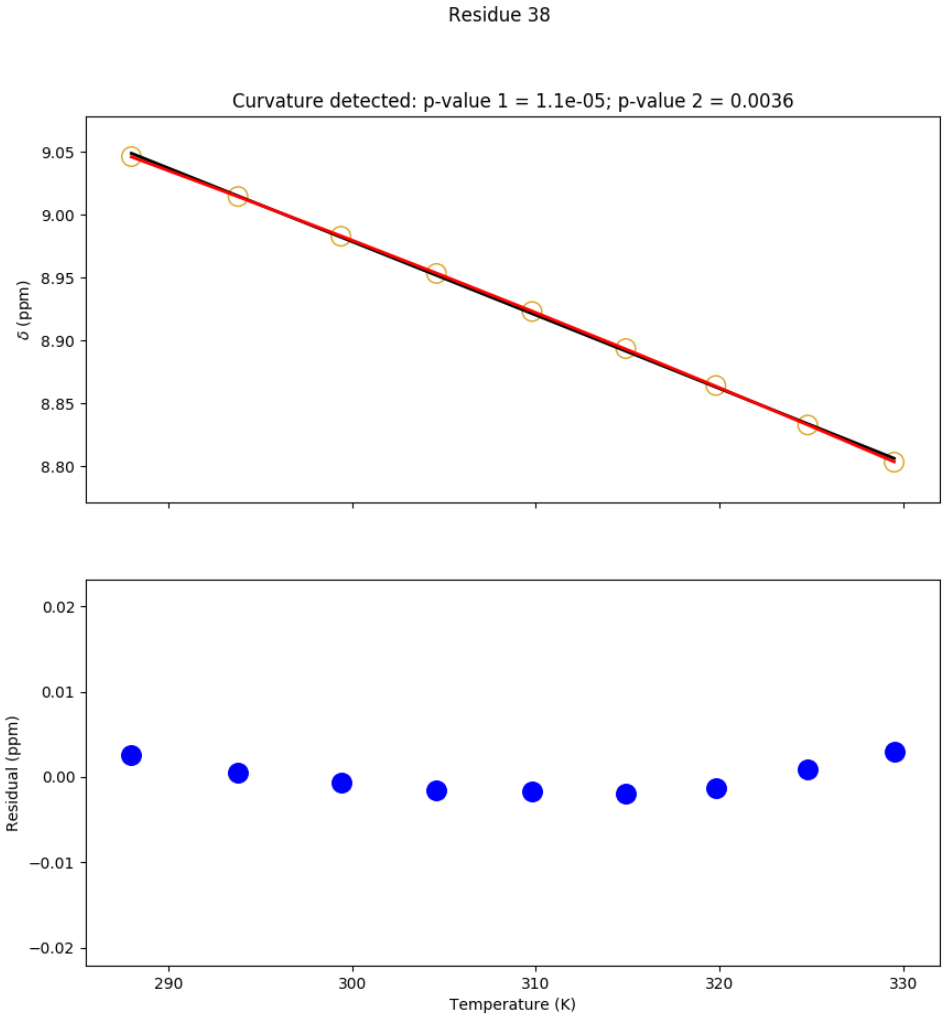


Figure I.265: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 38. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

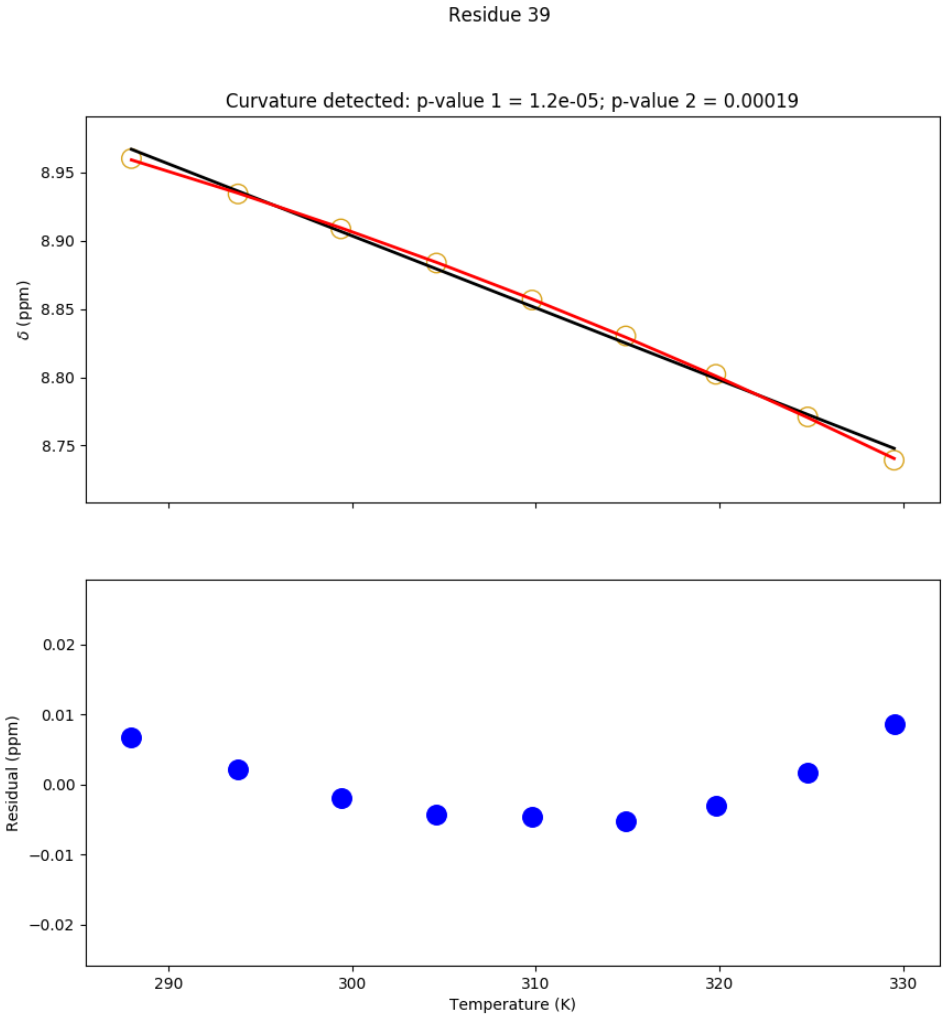


Figure I.266: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 39. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

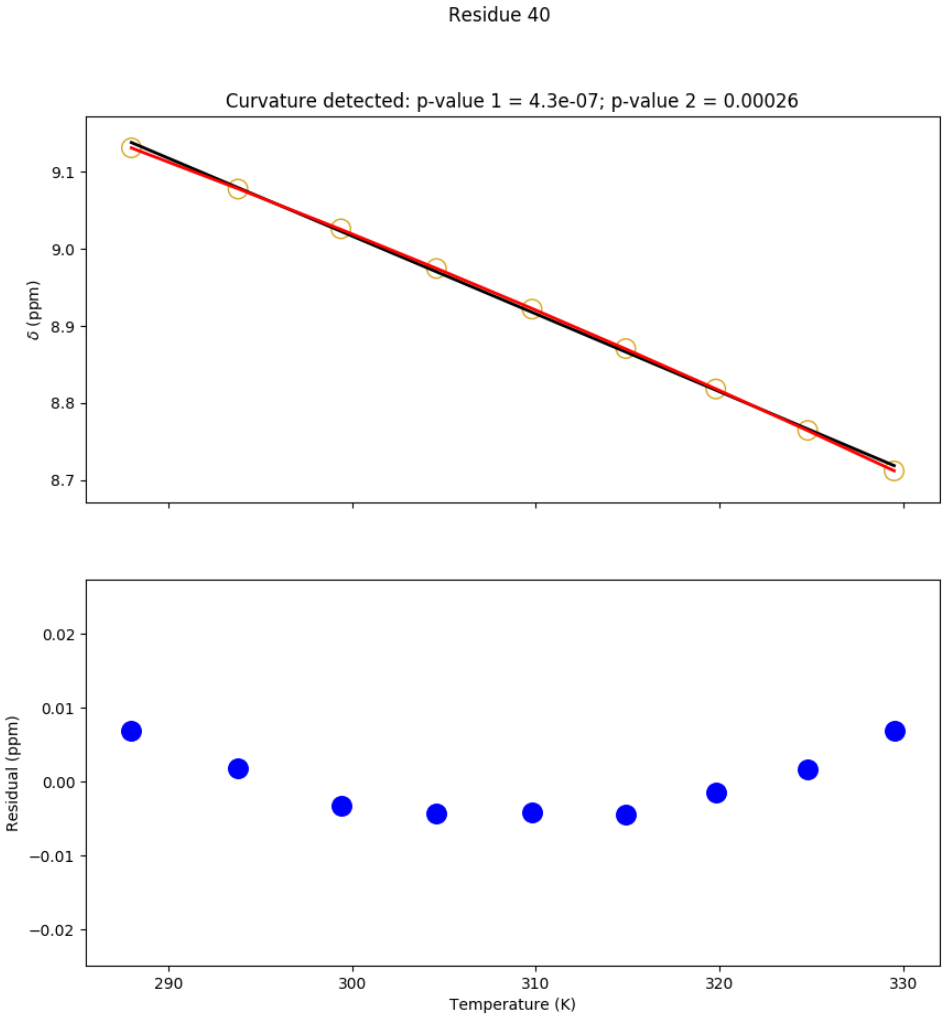


Figure I.267: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 40. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

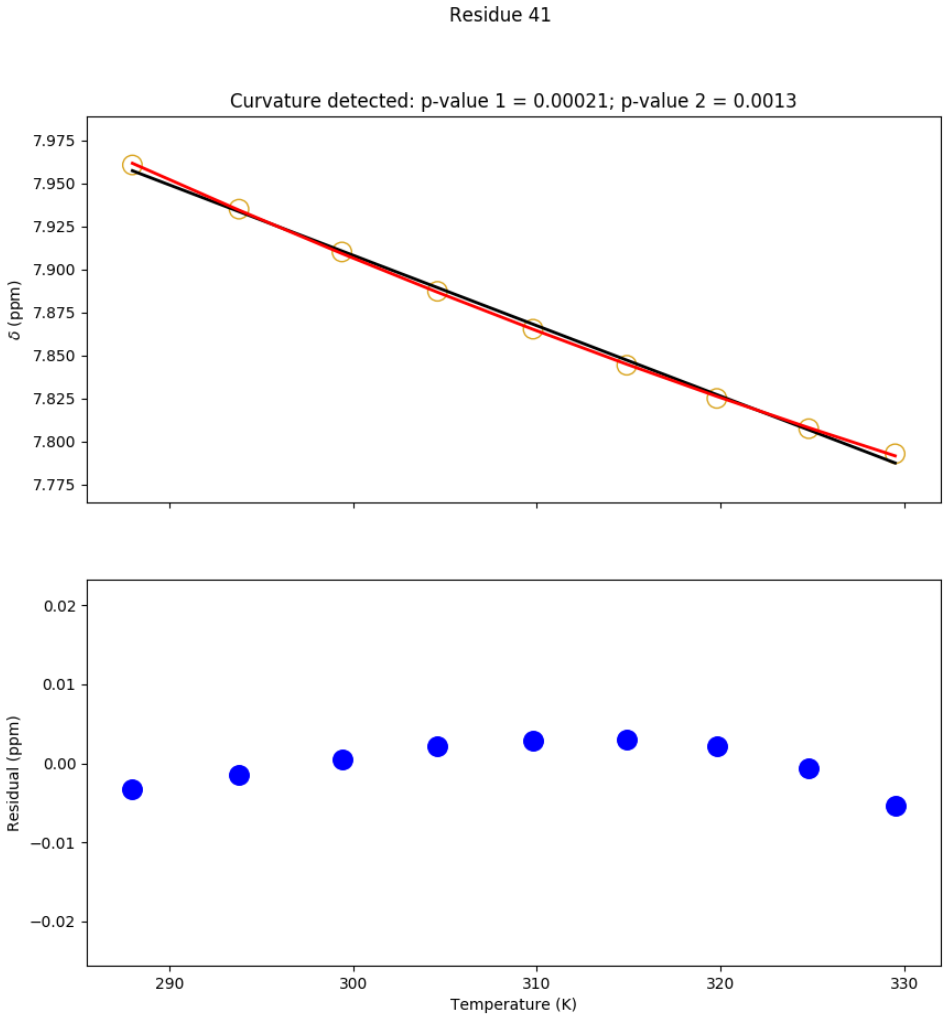


Figure I.268: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 41. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

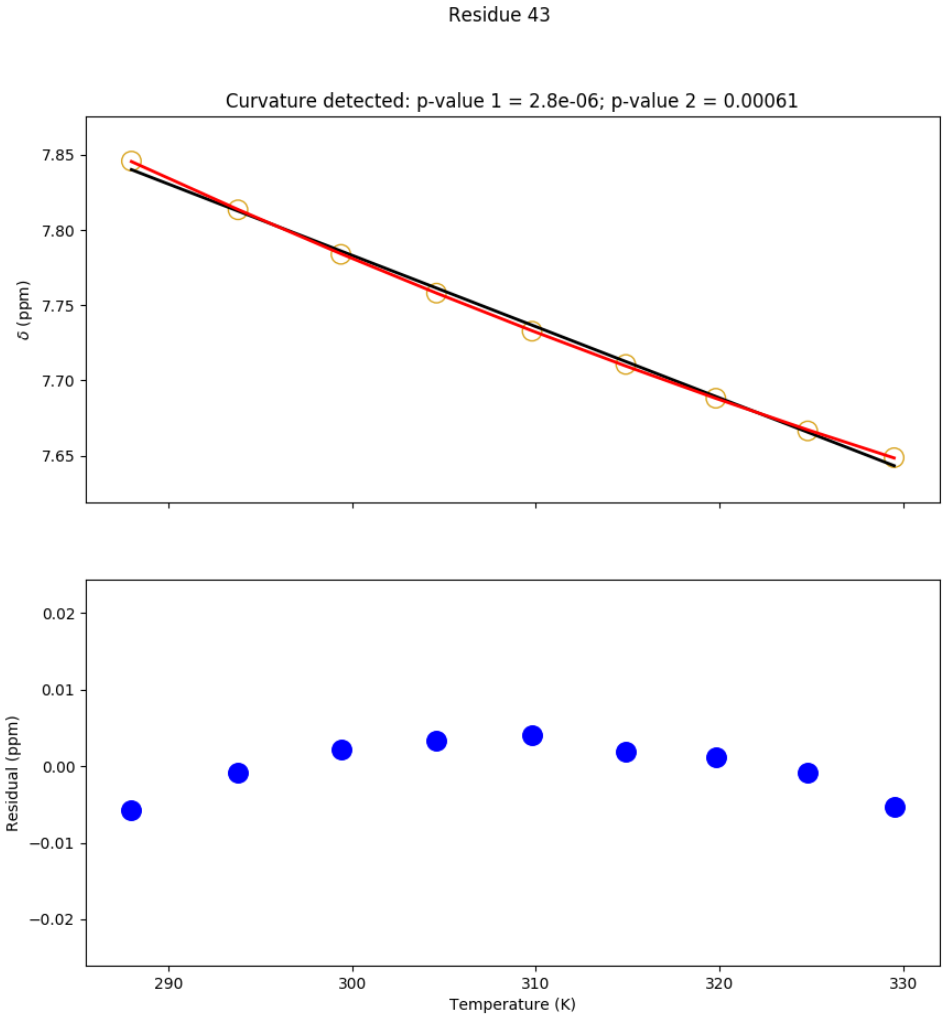


Figure I.269: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 43. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

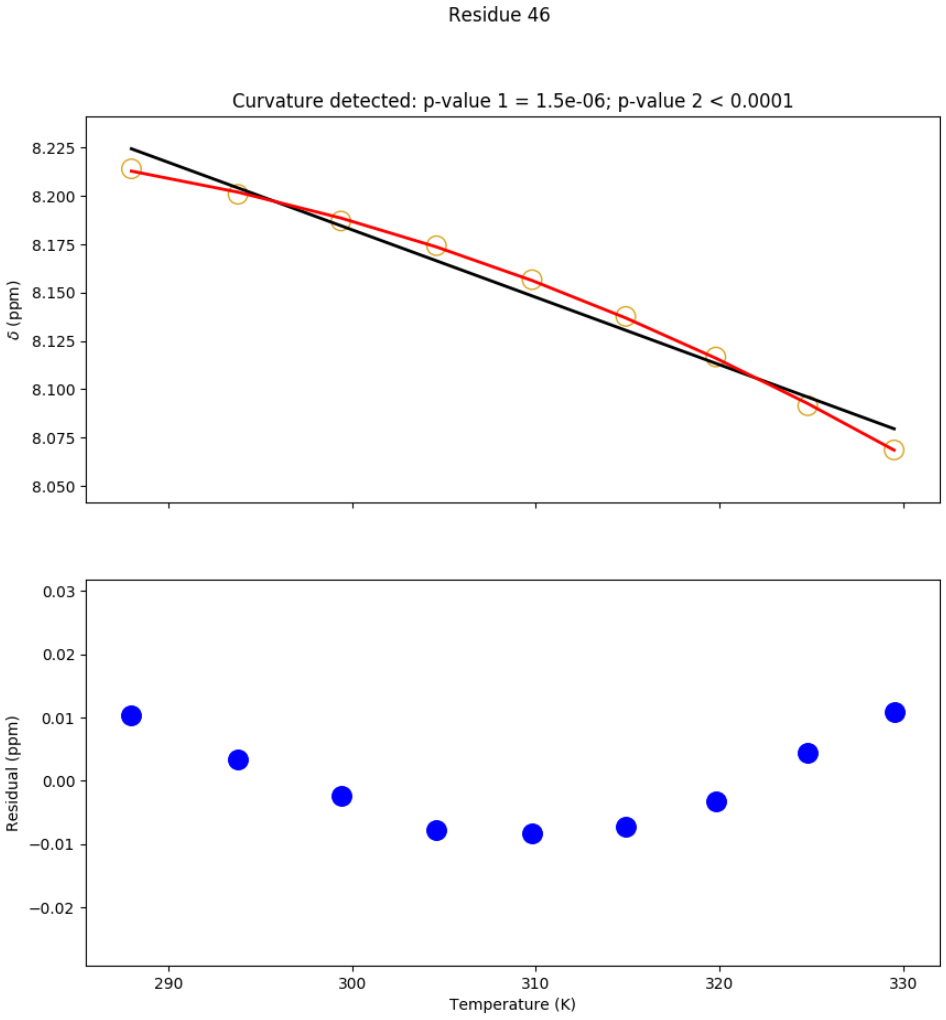


Figure I.270: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 46. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

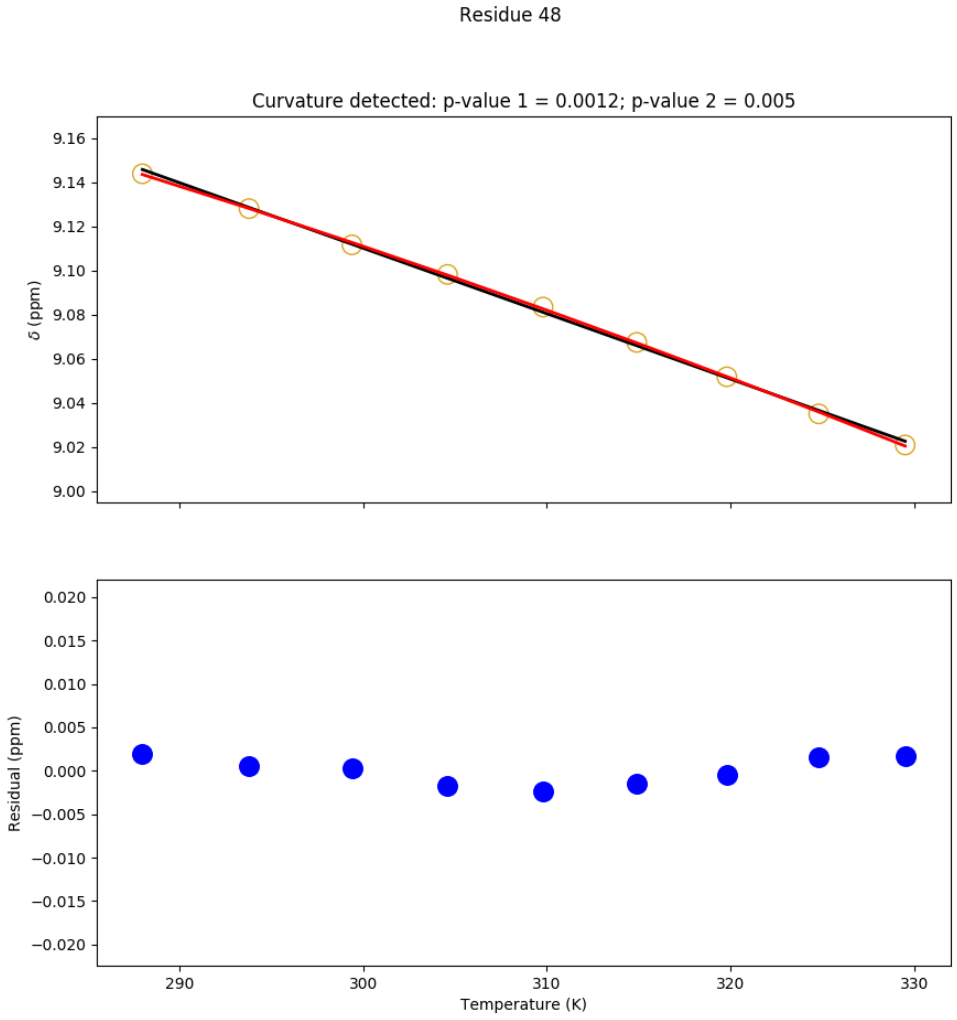


Figure I.271: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 48. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

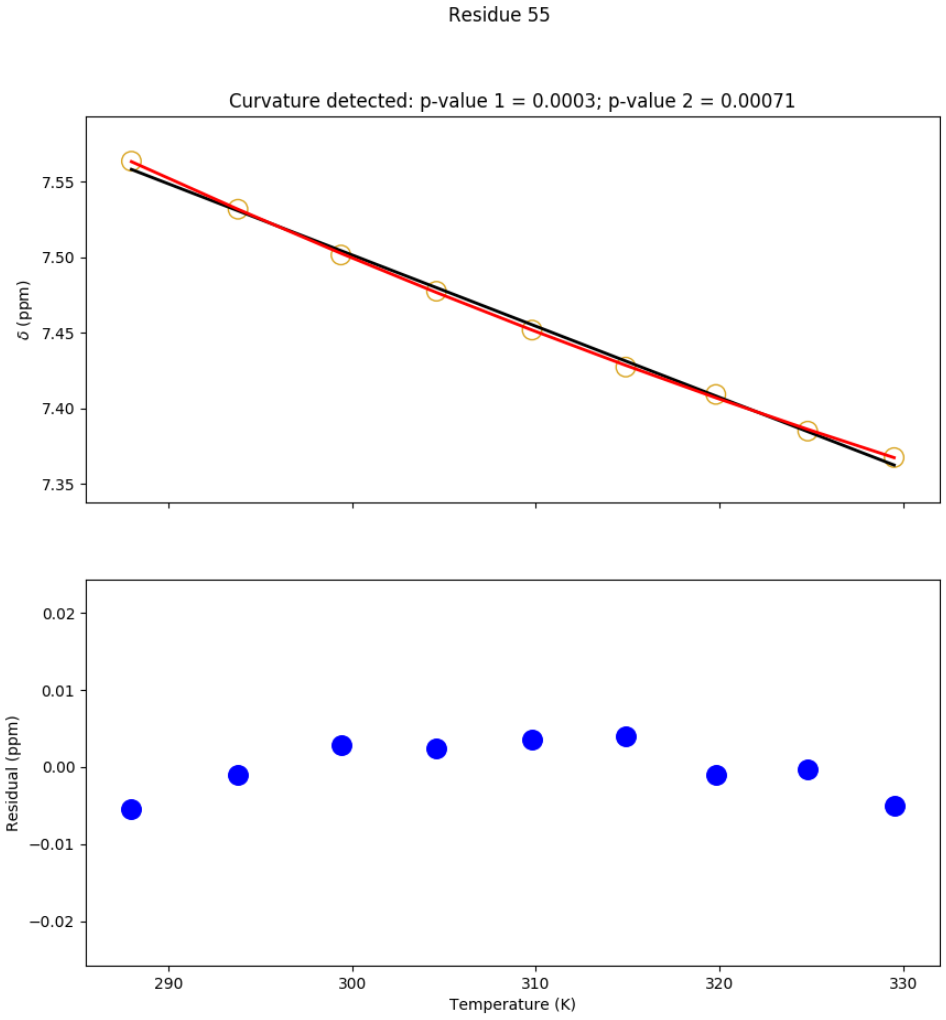


Figure I.272: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 55. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

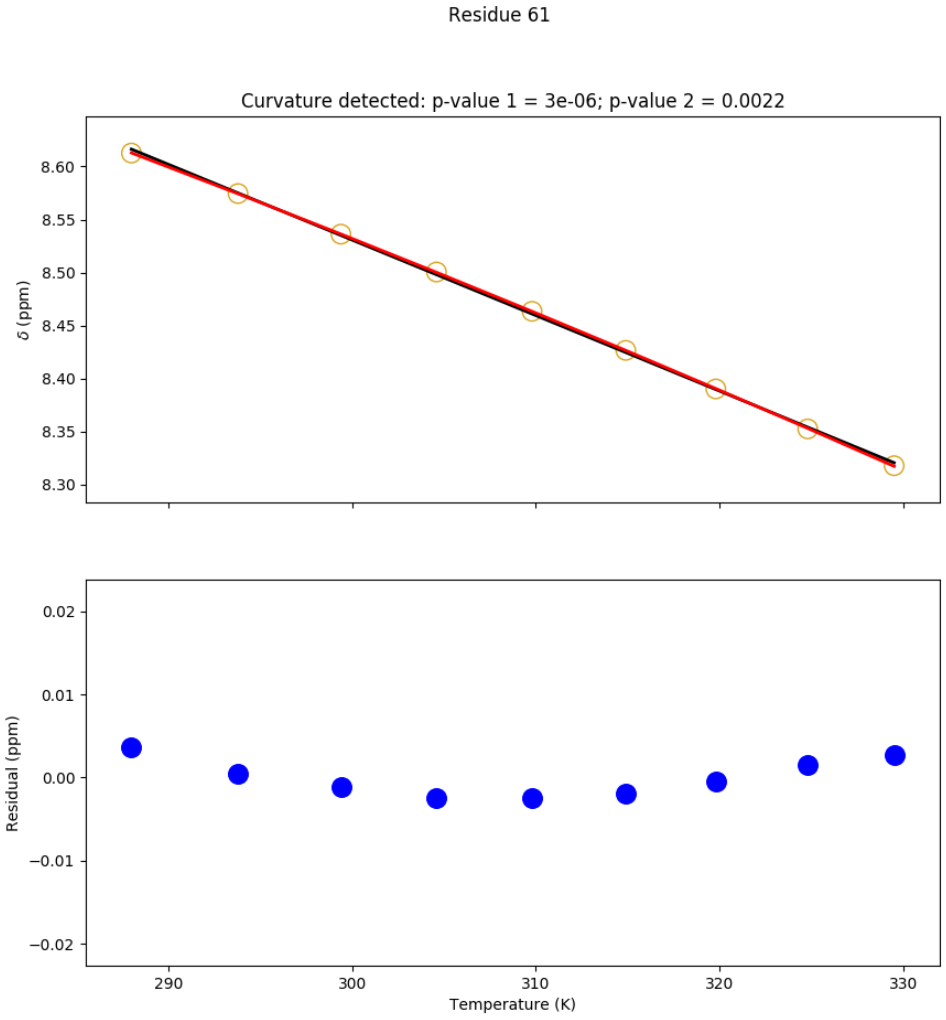


Figure I.273: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 61. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

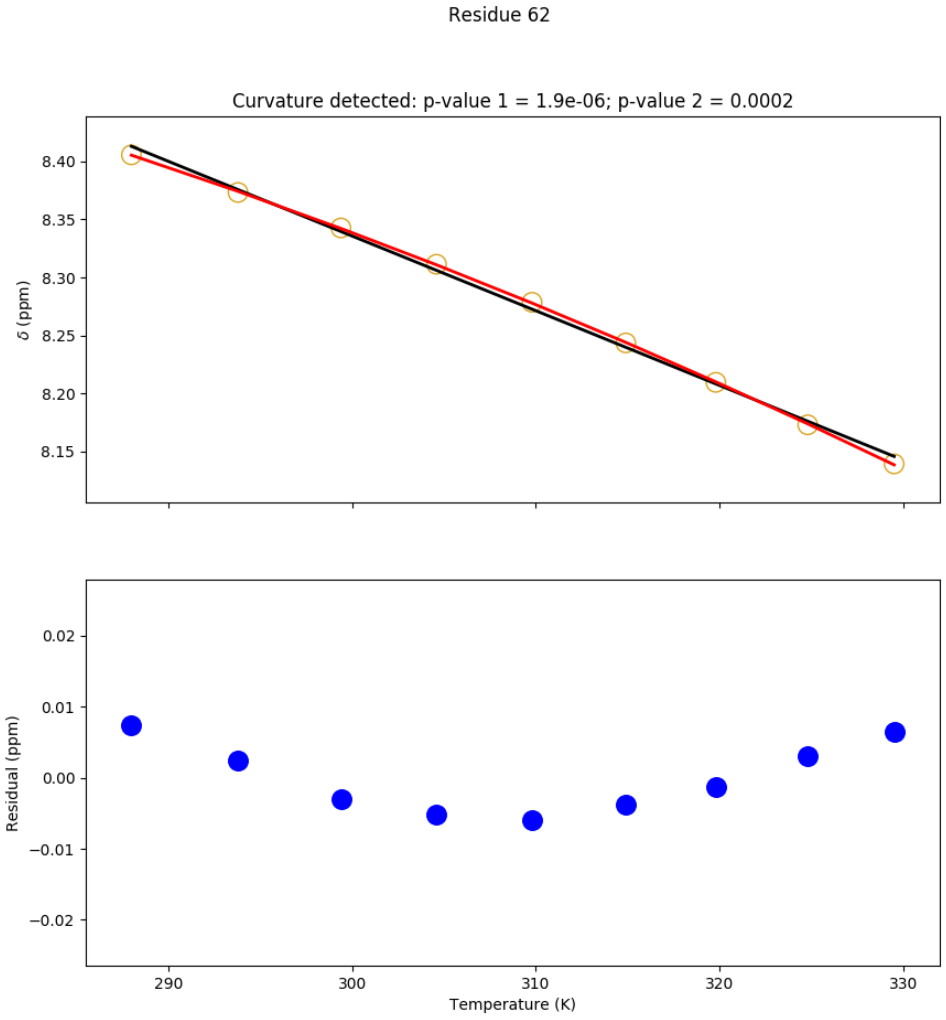


Figure I.274: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 62. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

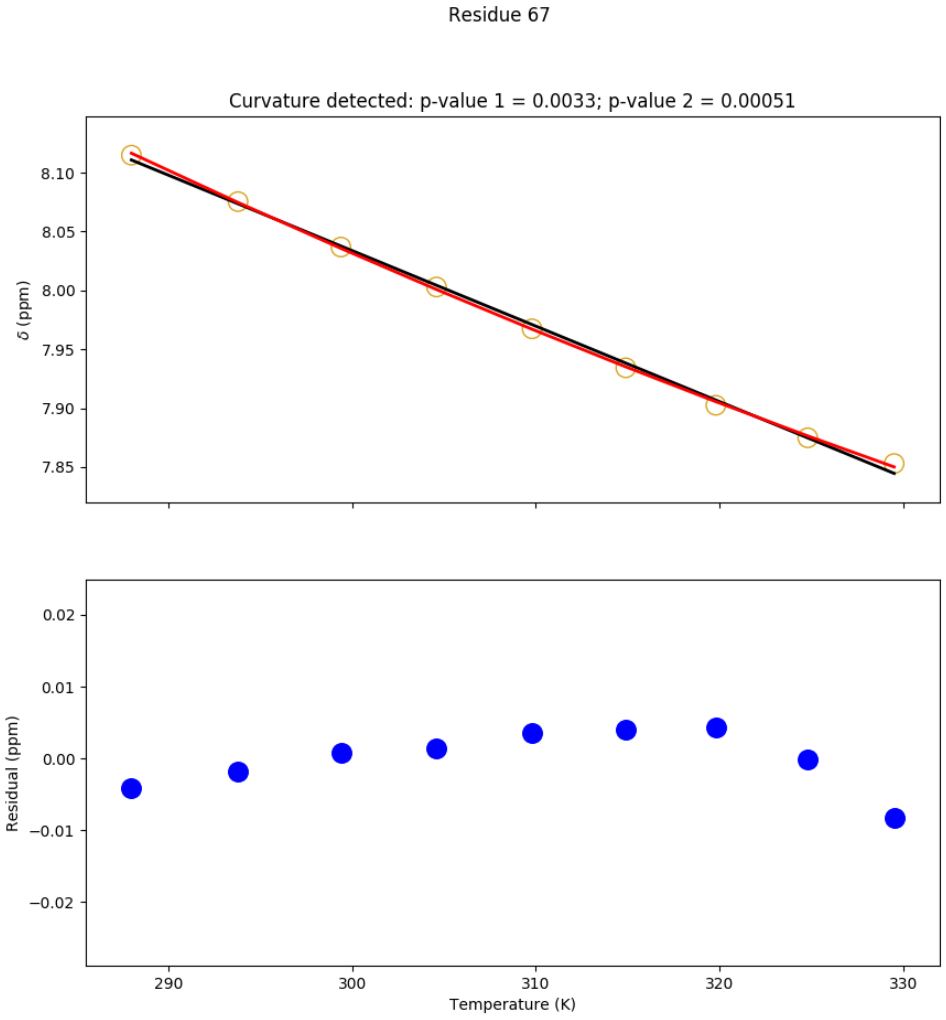


Figure I.275: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 67. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

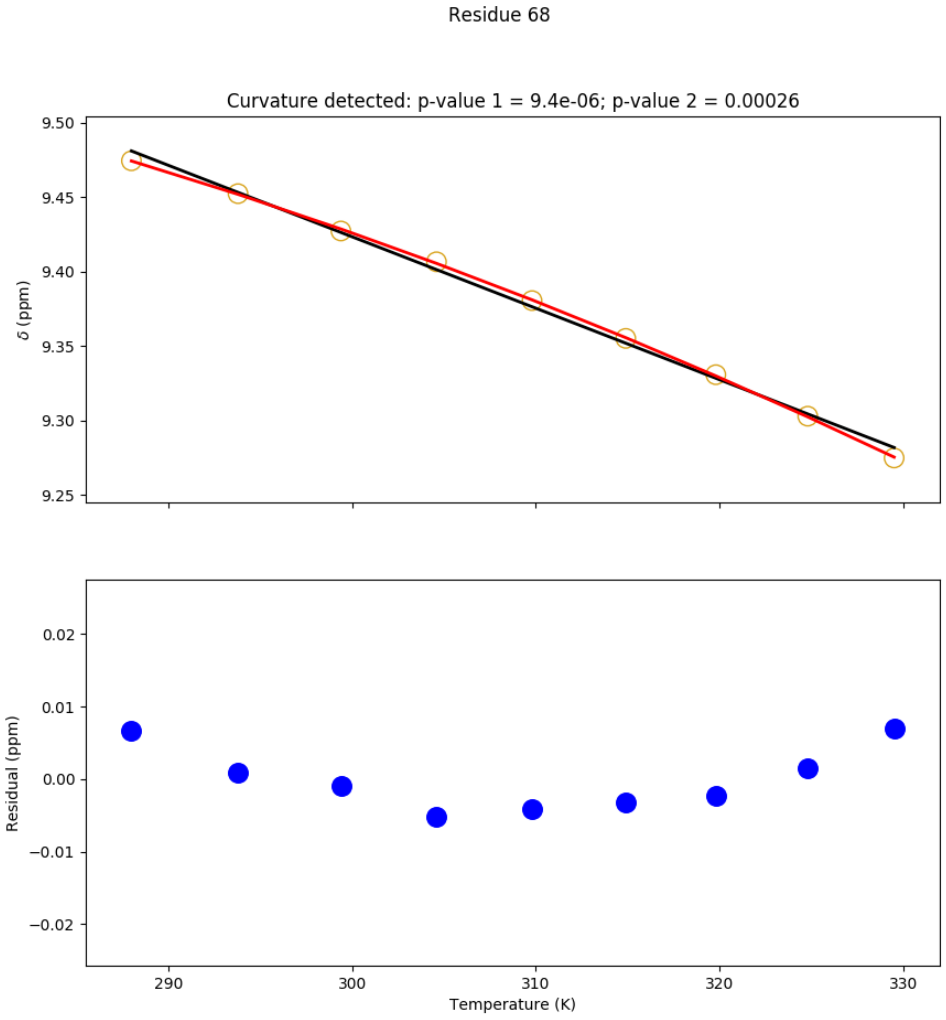


Figure I.276: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 68. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

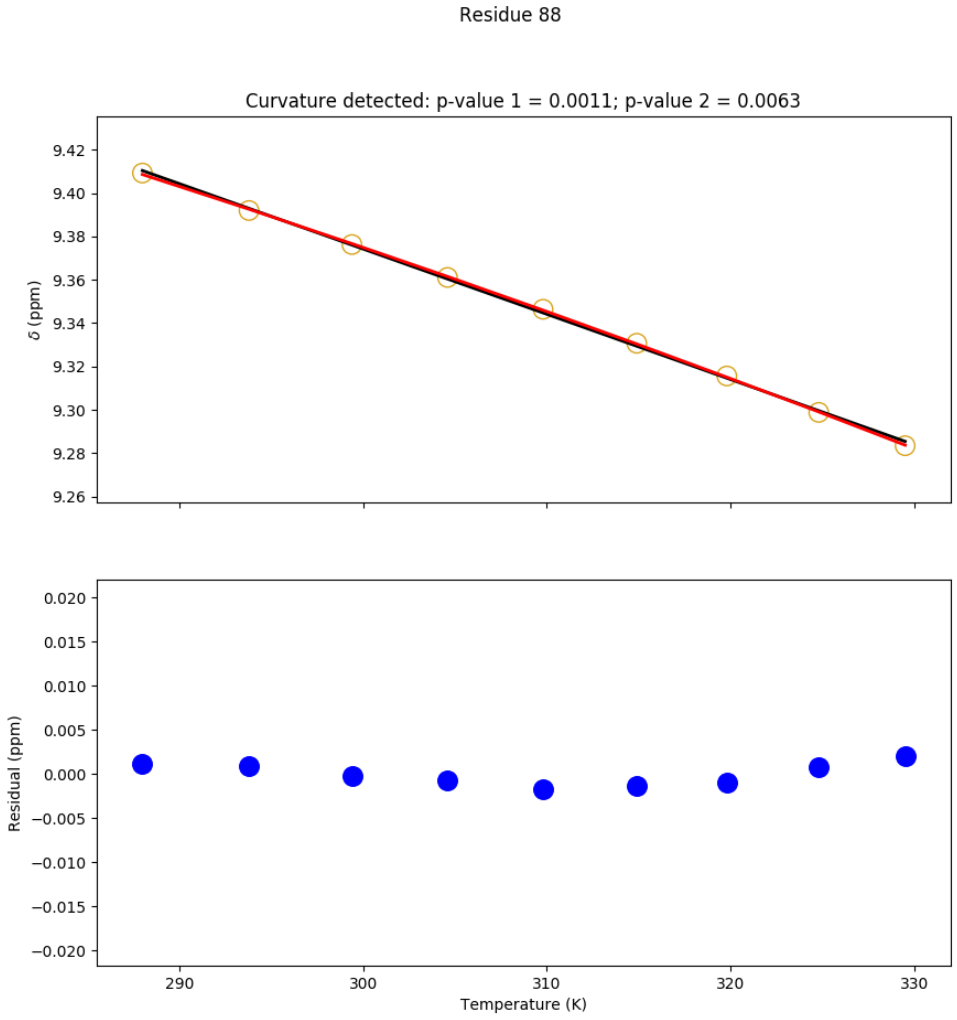


Figure I.277: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 88. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

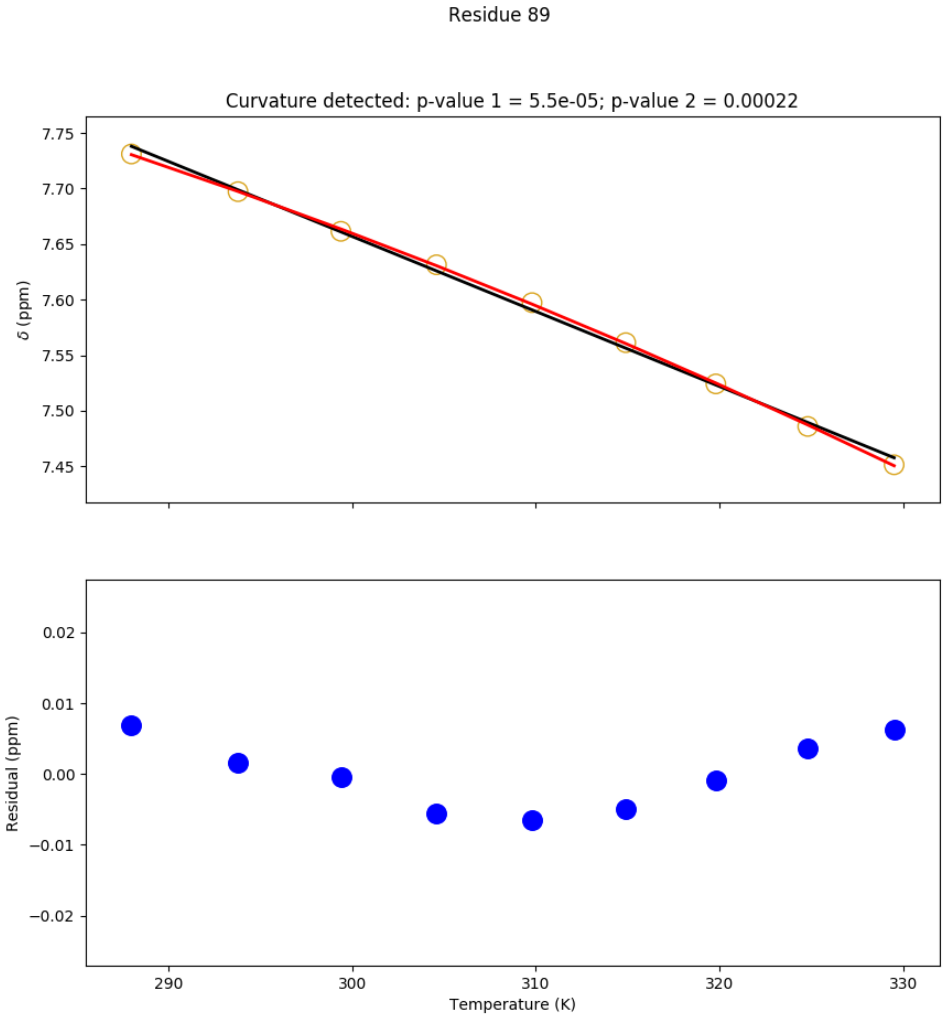


Figure I.278: Temperature dependence of the chemical shift of the L18V-Y88F Adnectin amide proton from residue 89. Top: amide proton chemical shifts (yellow), linear fit (black), and quadratic fit (red). Bottom: residuals (blue; linear fit minus chemical shift). The null hypothesis that the linear model is correct is tested to produce p-value 1. The null hypothesis that the observed curvature is the result of measurement errors is tested to produce p-value 2.

I.12 L18V-Y88F Adnectin Curvature Modelling

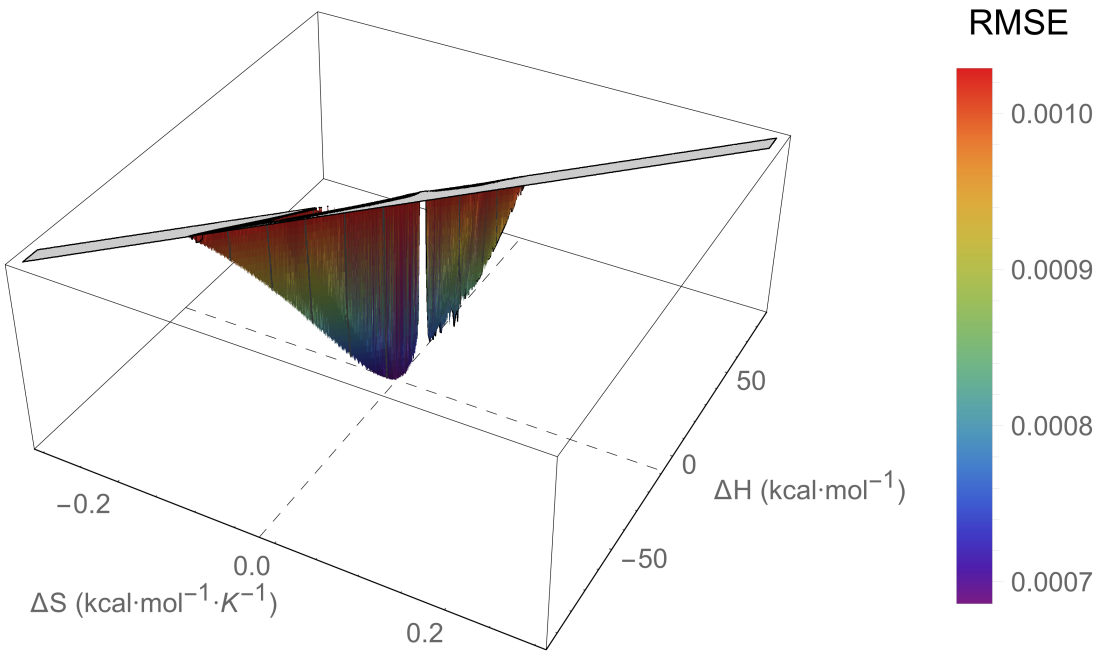


Figure I.279: An RMSE landscape from fitting L18V-Y88F Adnectin residue 3 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

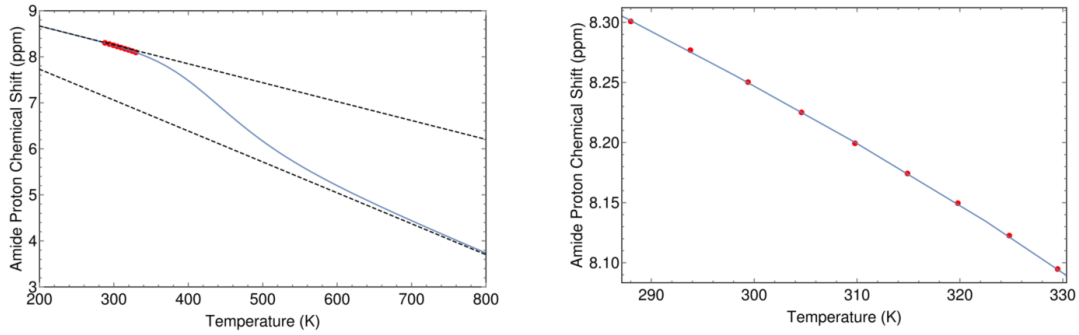


Figure I.280: The minimum RMSE fit of L18V-Y88F Adnectin residue 3 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

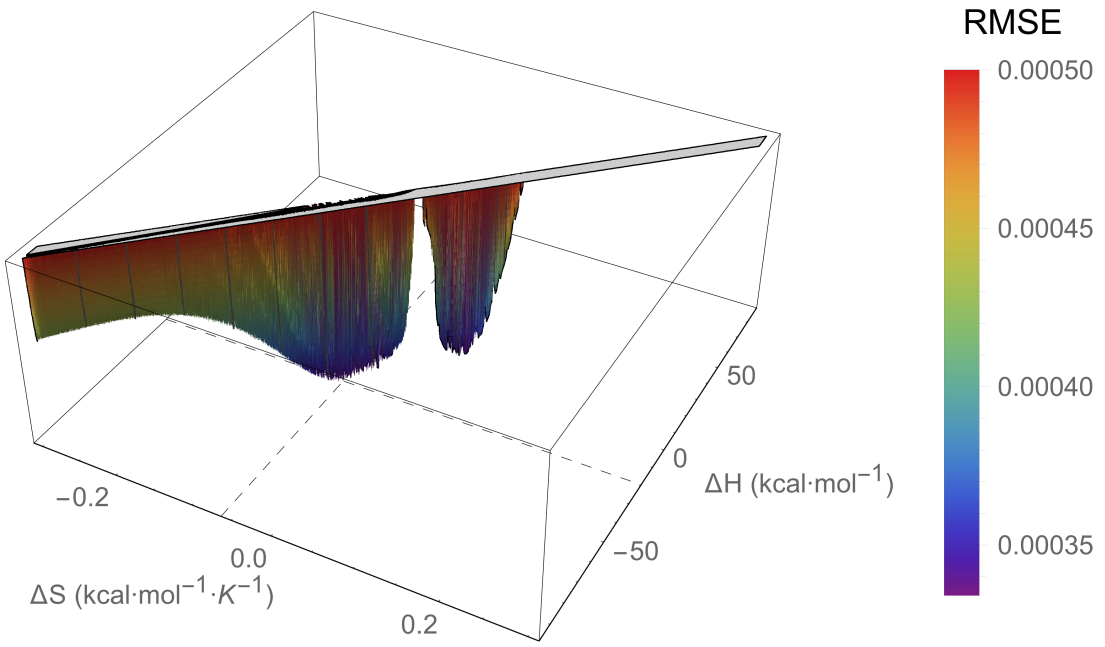


Figure I.281: An RMSE landscape from fitting L18V-Y88F Adnectin residue 4 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

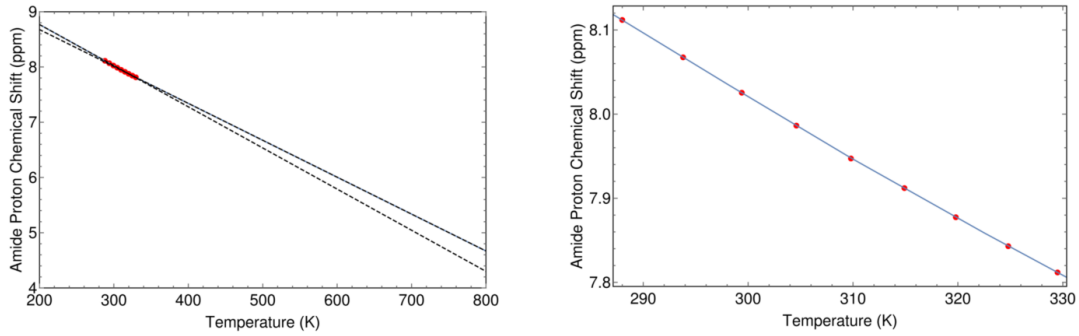


Figure I.282: The minimum RMSE fit of L18V-Y88F Adnectin residue 4 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

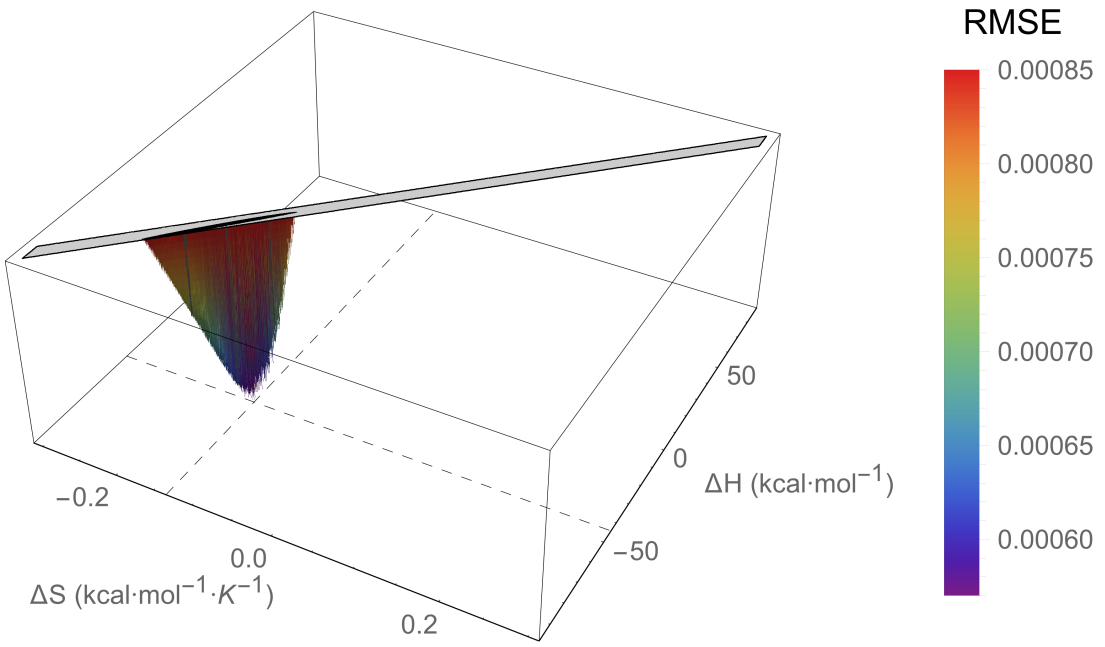


Figure I.283: An RMSE landscape from fitting L18V-Y88F Adnectin residue 6 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

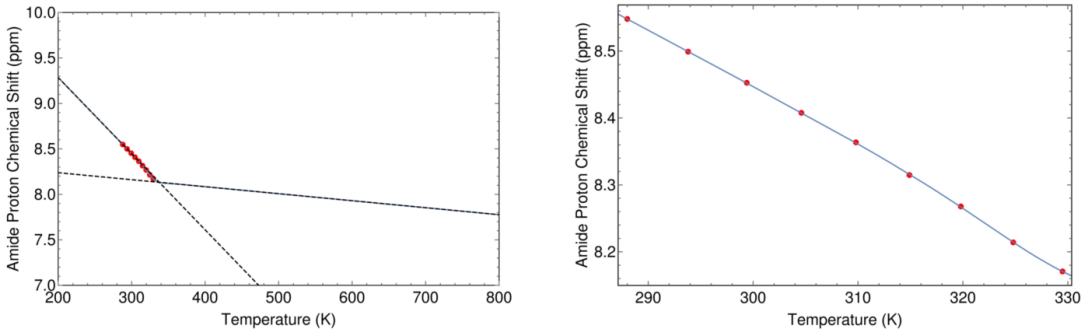


Figure I.284: The minimum RMSE fit of L18V-Y88F Adnectin residue 6 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

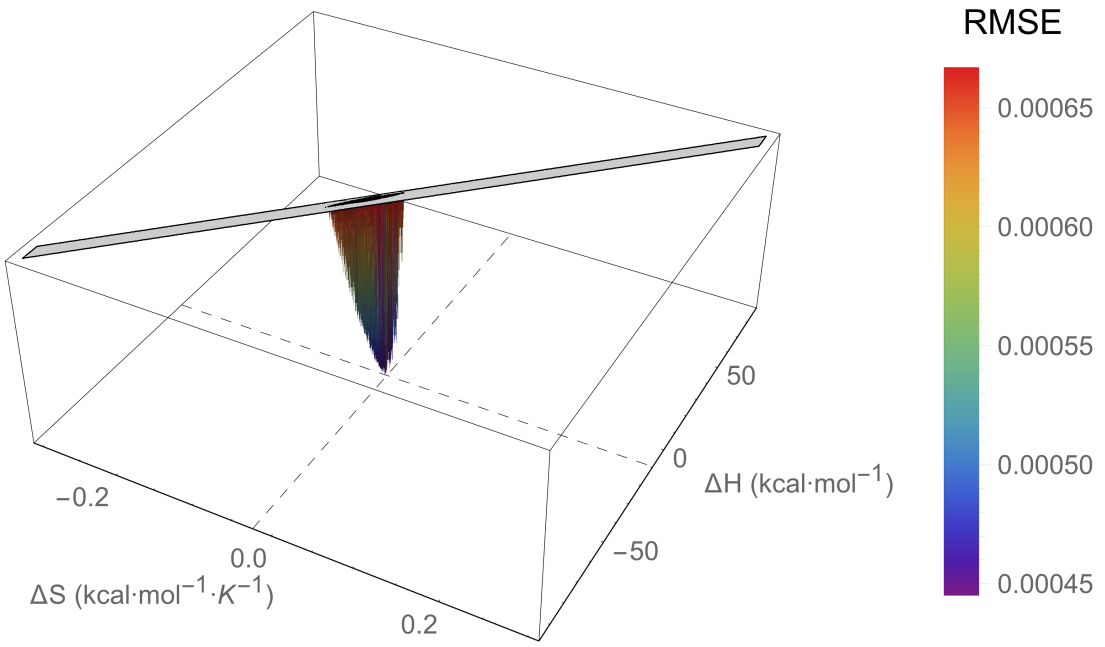


Figure I.285: An RMSE landscape from fitting L18V-Y88F Adnectin residue 8 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

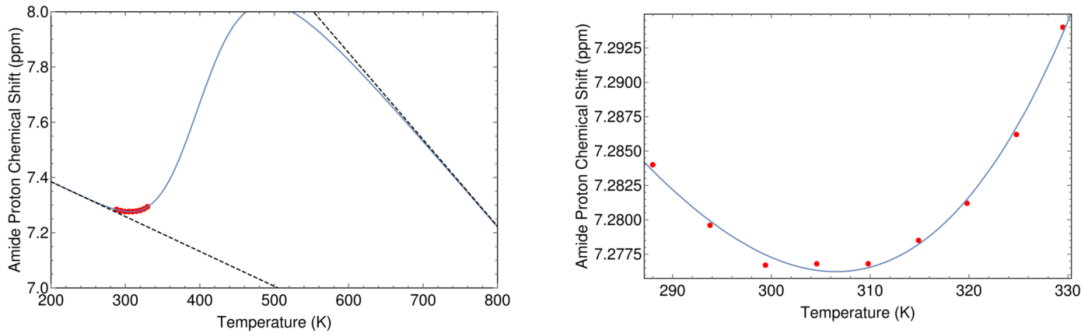


Figure I.286: The minimum RMSE fit of L18V-Y88F Adnectin residue 8 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

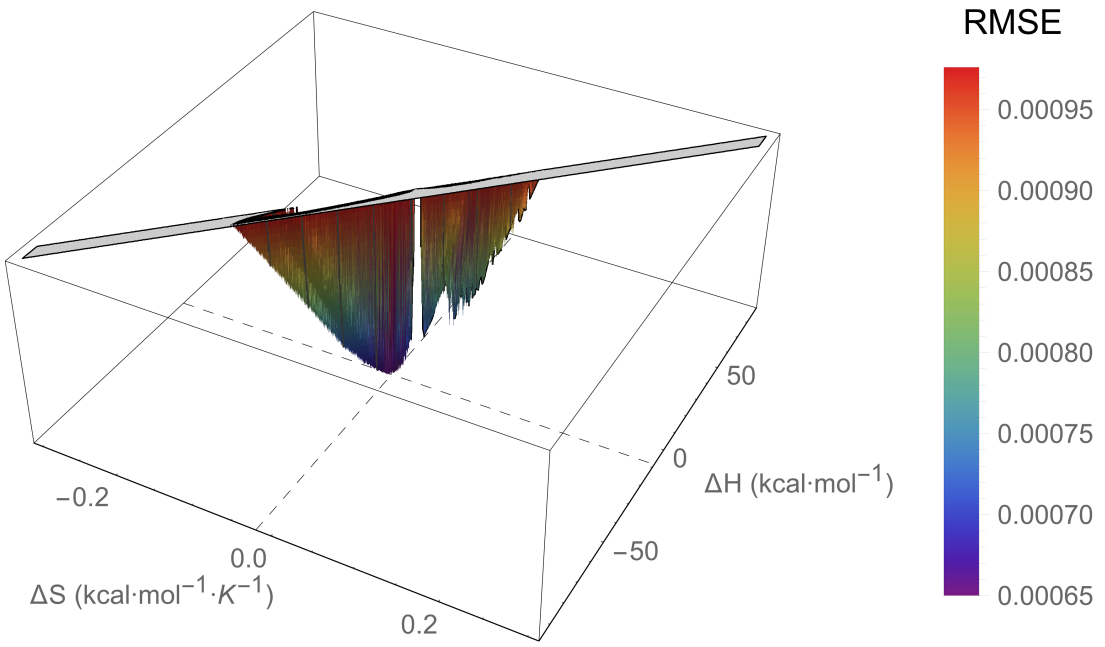


Figure I.287: An RMSE landscape from fitting L18V-Y88F Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

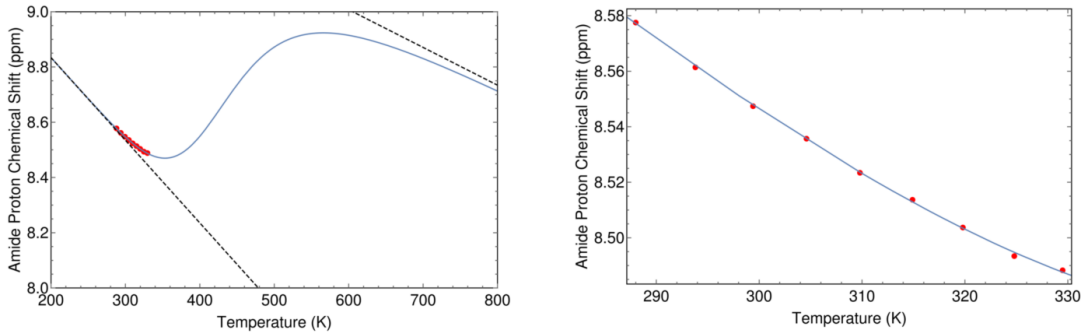


Figure I.288: The minimum RMSE fit of L18V-Y88F Adnectin residue 19 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

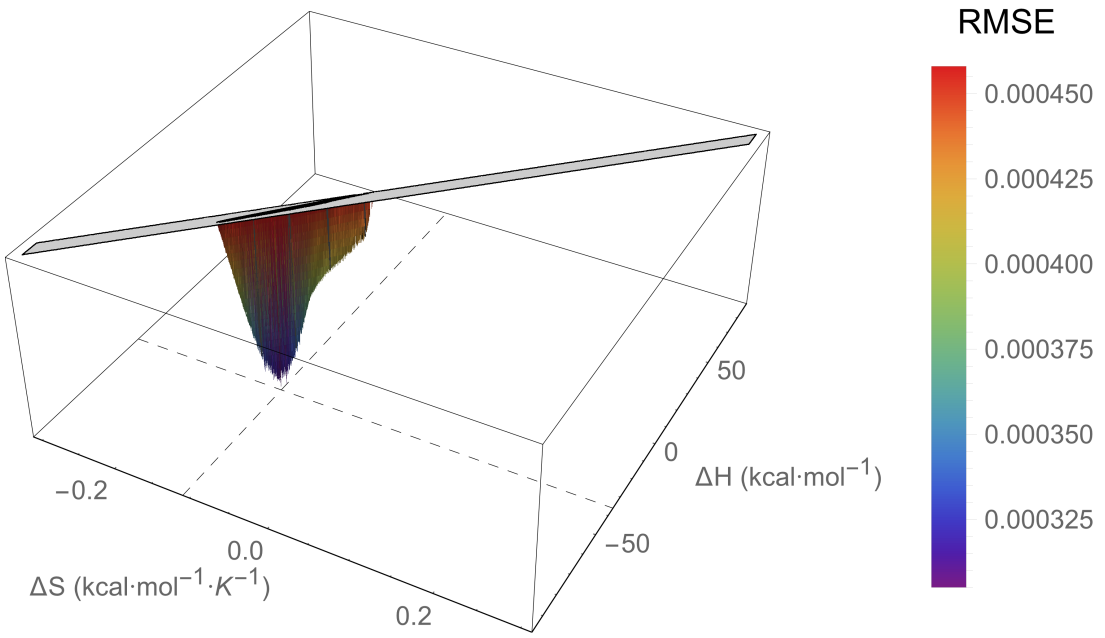


Figure I.289: An RMSE landscape from fitting L18V-Y88F Adnectin residue 36 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

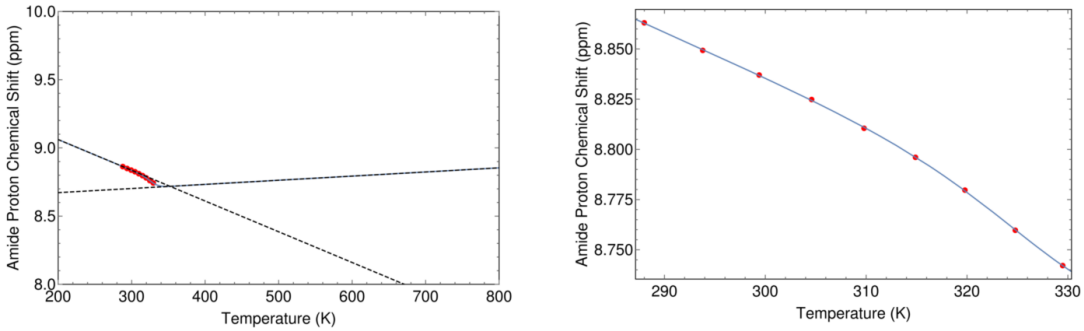


Figure I.290: The minimum RMSE fit of L18V-Y88F Adnectin residue 36 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

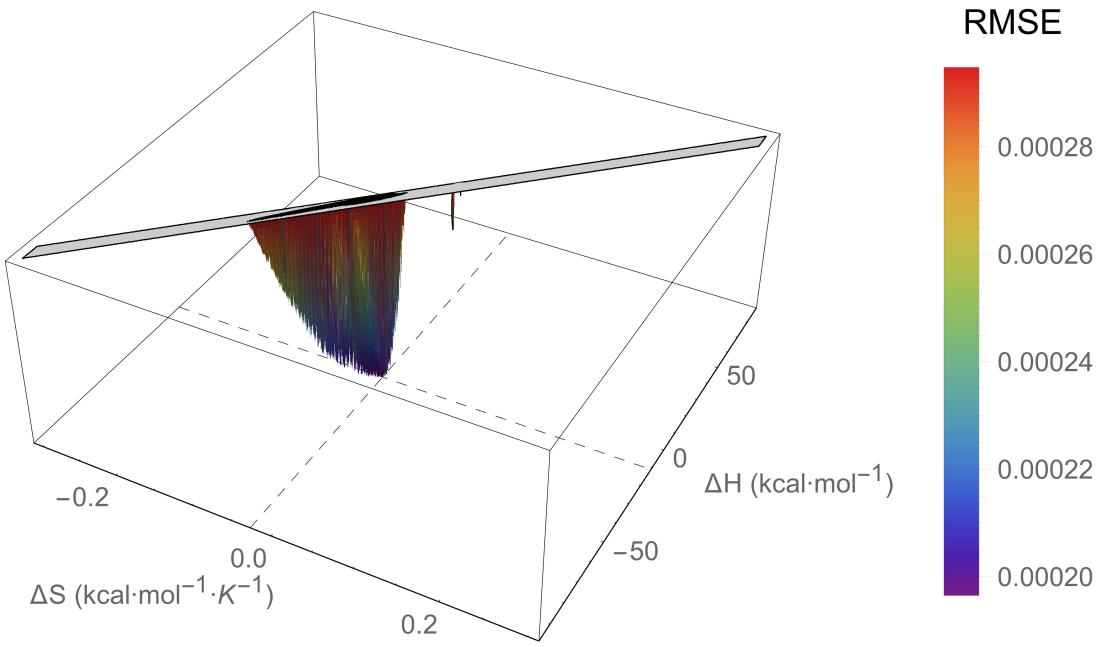


Figure I.291: An RMSE landscape from fitting L18V-Y88F Adnectin residue 38 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

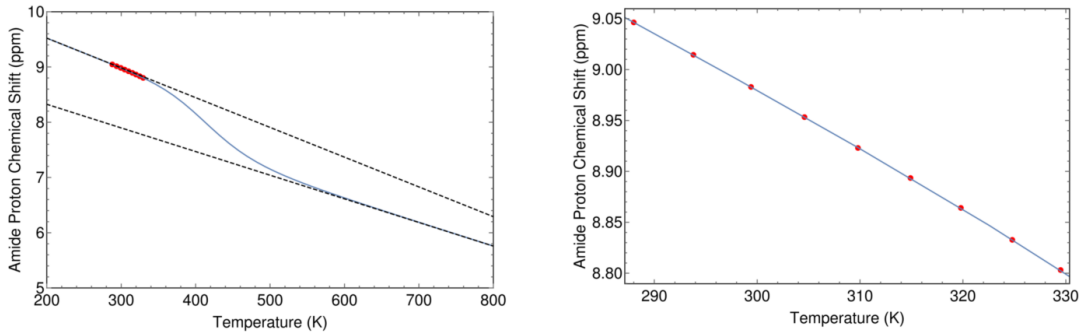


Figure I.292: The minimum RMSE fit of L18V-Y88F Adnectin residue 38 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

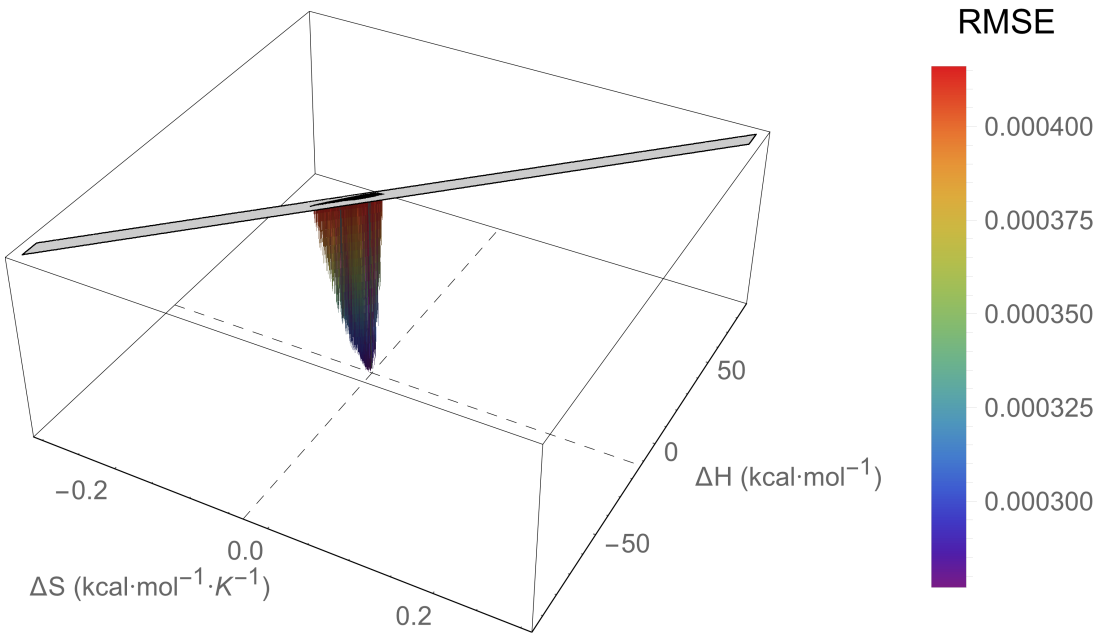


Figure I.293: An RMSE landscape from fitting L18V-Y88F Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE$_{\text{min}}$) are indicated in gray.

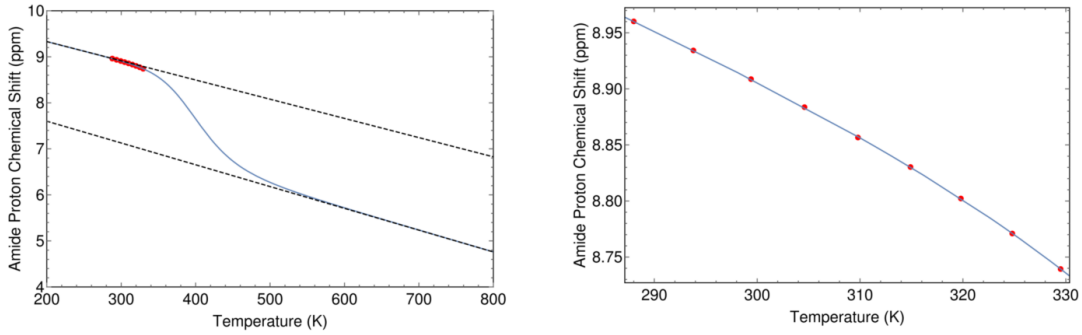


Figure I.294: The minimum RMSE fit of L18V-Y88F Adnectin residue 39 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

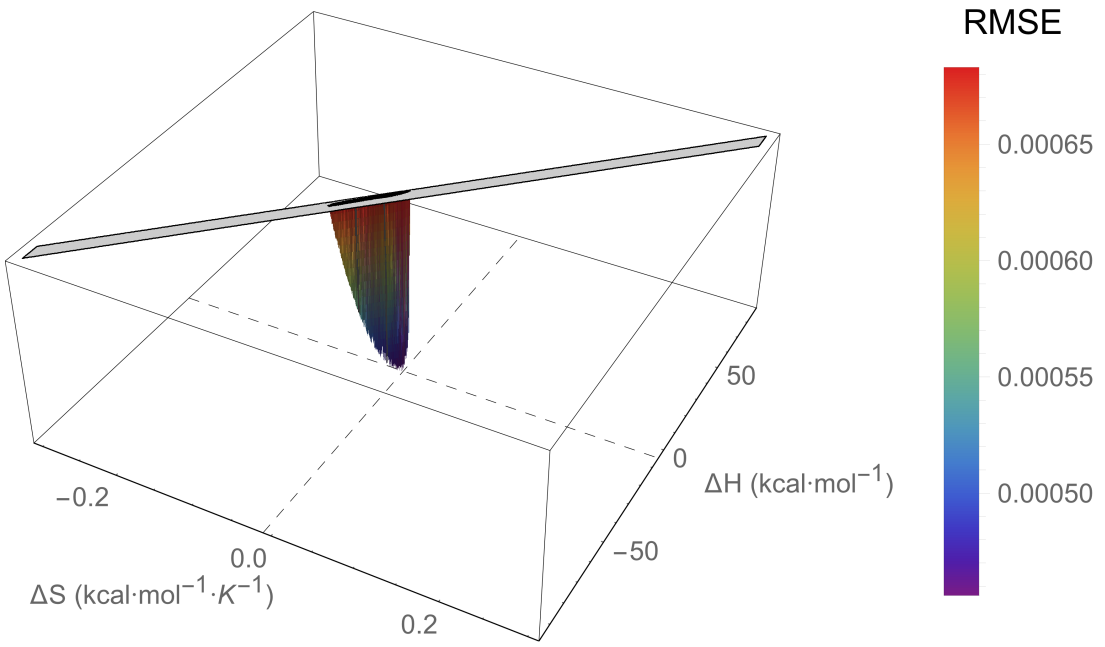


Figure I.295: An RMSE landscape from fitting L18V-Y88F Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

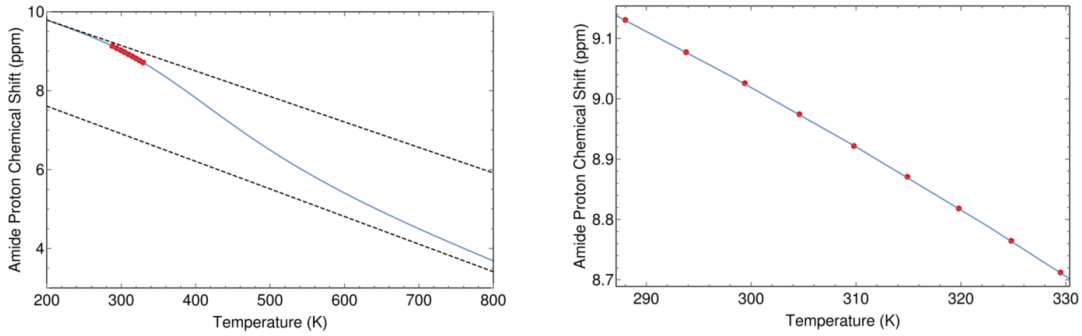


Figure I.296: The minimum RMSE fit of L18V-Y88F Adnectin residue 40 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

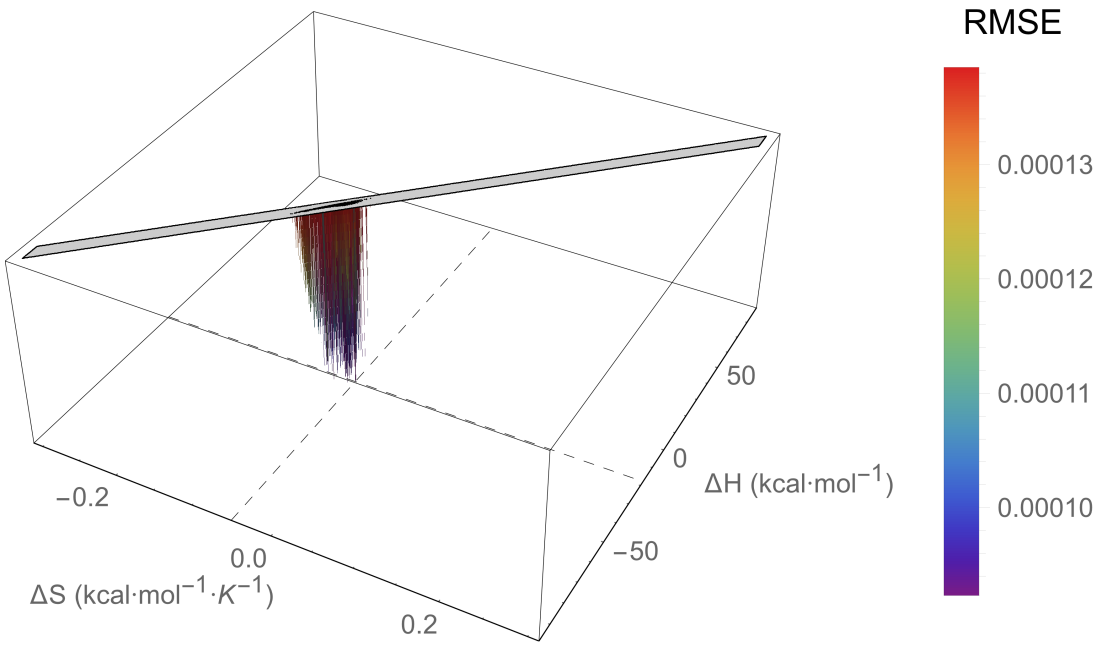


Figure I.297: An RMSE landscape from fitting L18V-Y88F Adnectin residue 41 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

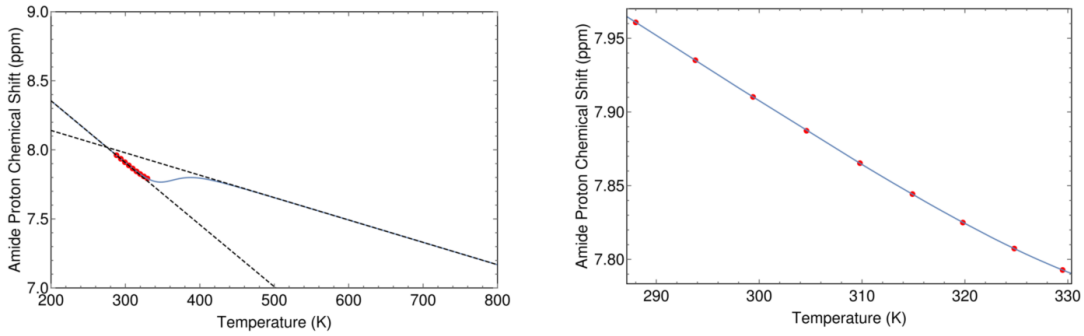


Figure I.298: The minimum RMSE fit of L18V-Y88F Adnectin residue 41 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

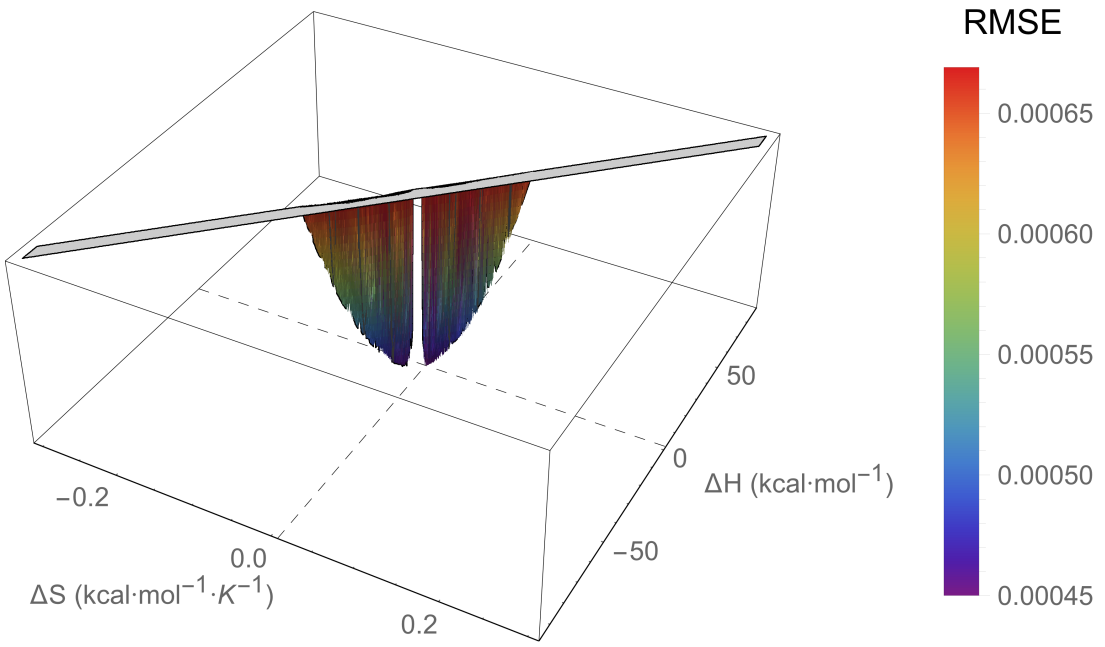


Figure I.299: An RMSE landscape from fitting L18V-Y88F Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

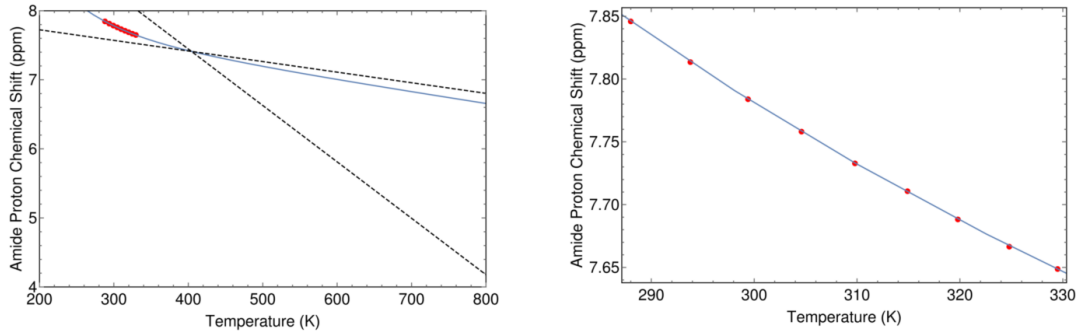


Figure I.300: The minimum RMSE fit of L18V-Y88F Adnectin residue 43 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

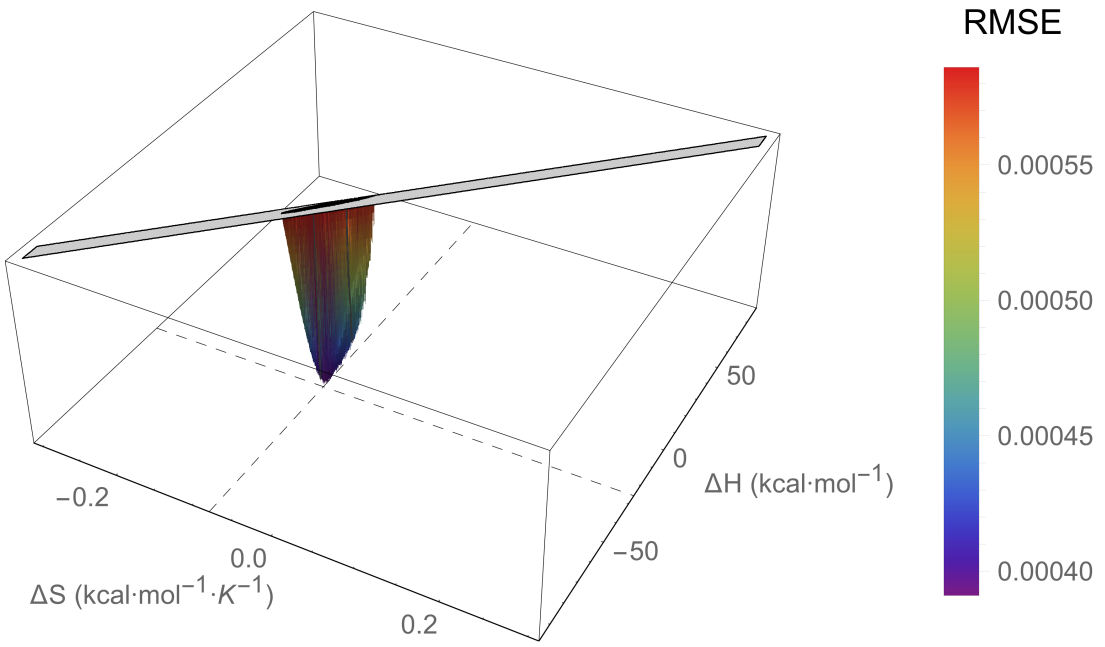


Figure I.301: An RMSE landscape from fitting L18V-Y88F Adnectin residue 46 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

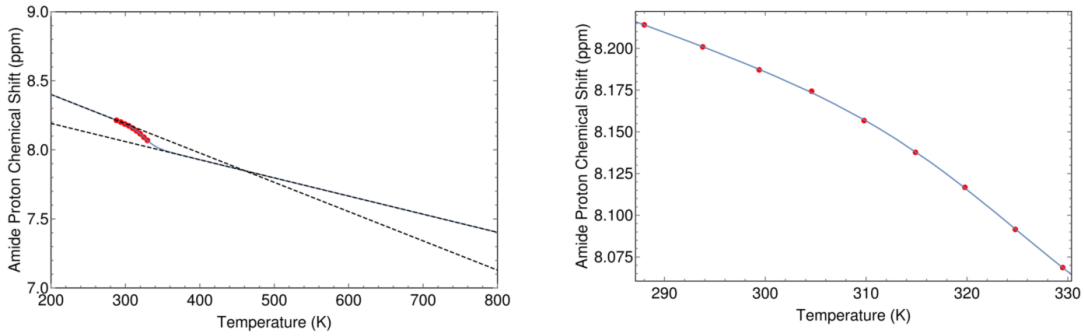


Figure I.302: The minimum RMSE fit of L18V-Y88F Adnectin residue 46 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

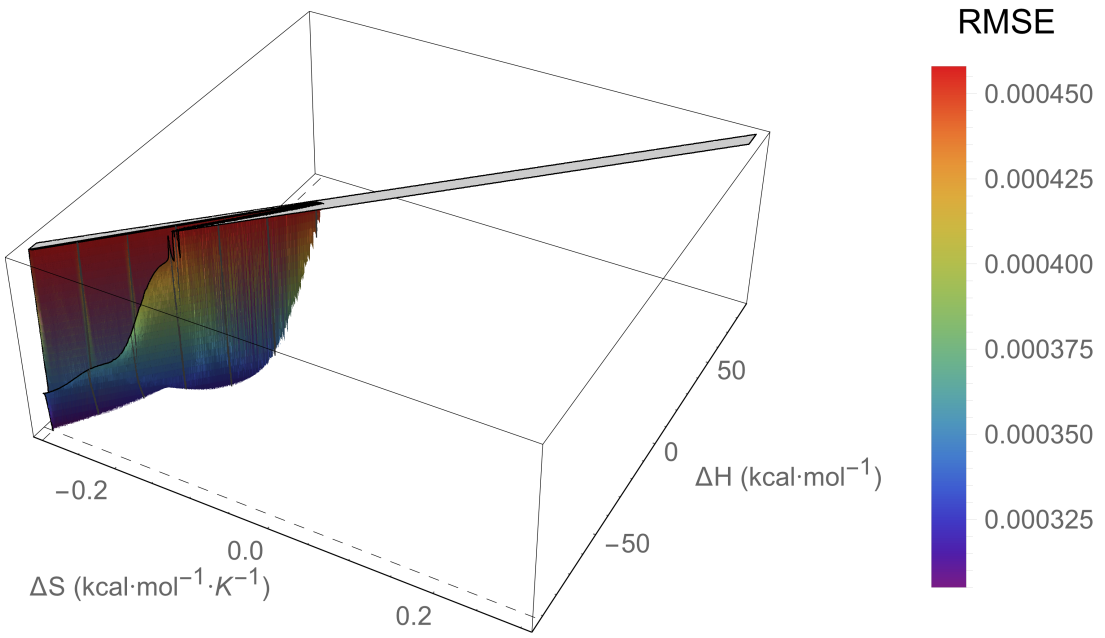


Figure I.303: An RMSE landscape from fitting L18V-Y88F Adnectin residue 48 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE_{min}) are indicated in gray.

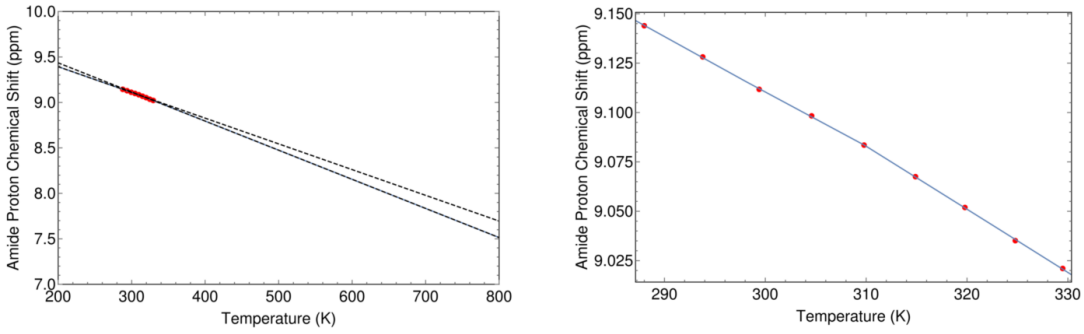


Figure I.304: The minimum RMSE fit of L18V-Y88F Adnectin residue 48 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

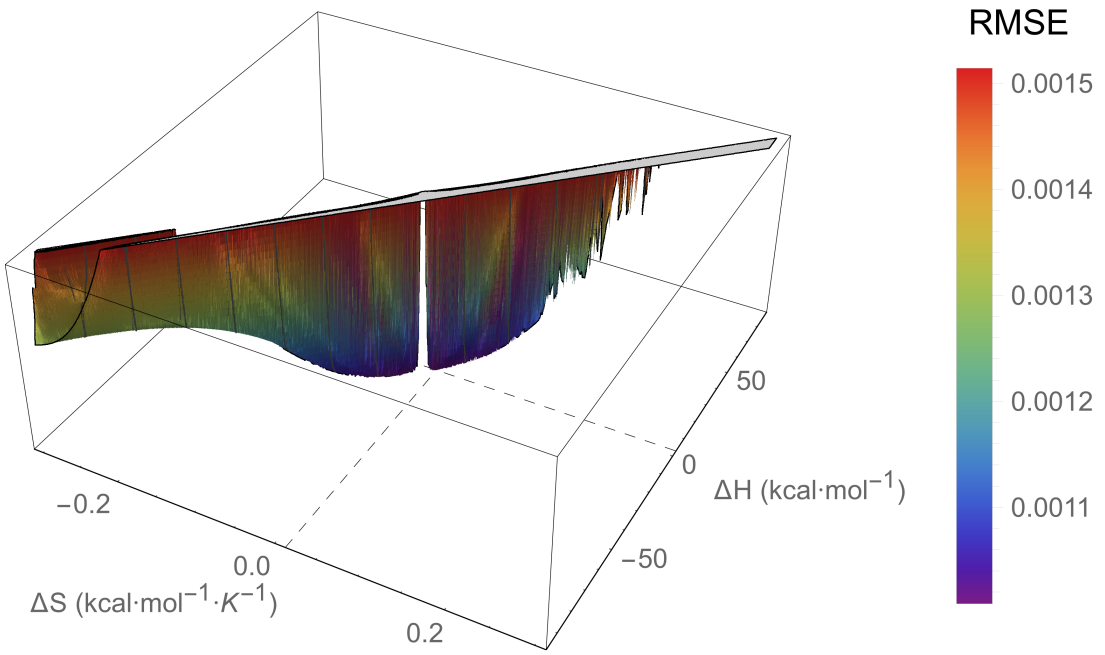


Figure I.305: An RMSE landscape from fitting L18V-Y88F Adnectin residue 55 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

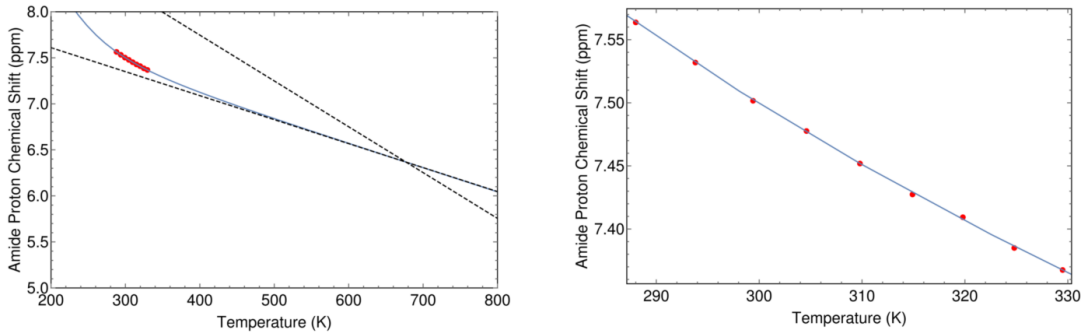


Figure I.306: The minimum RMSE fit of L18V-Y88F Adnectin residue 55 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

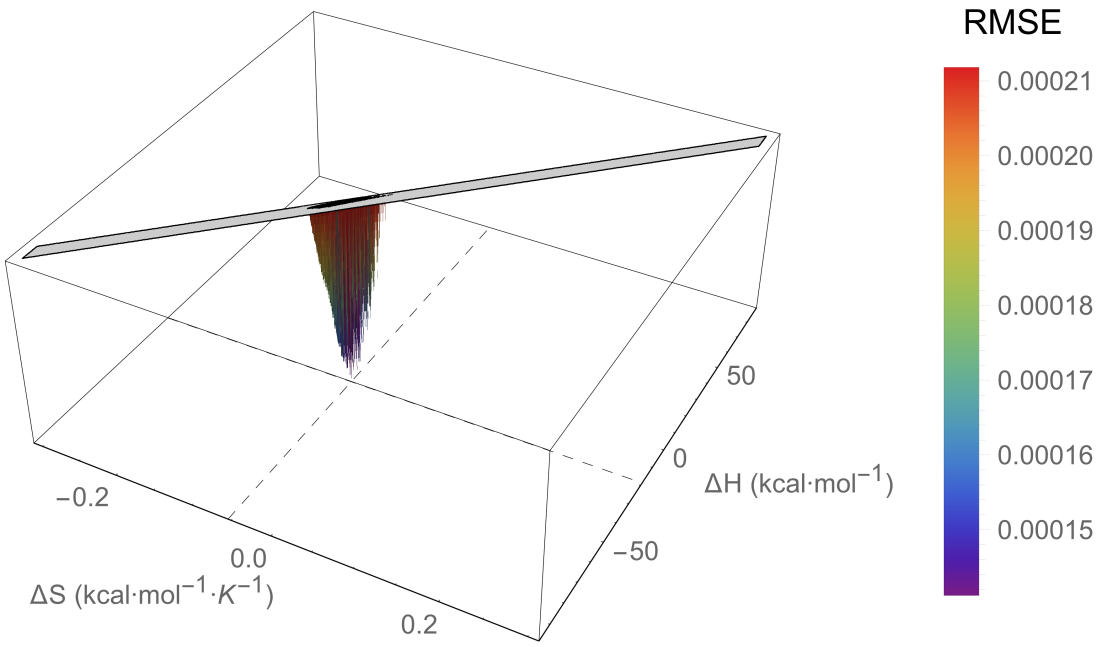


Figure I.307: An RMSE landscape from fitting L18V-Y88F Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

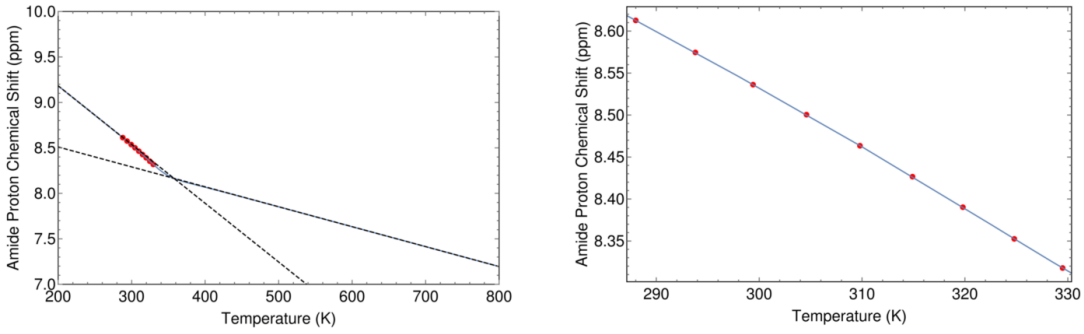


Figure I.308: The minimum RMSE fit of L18V-Y88F Adnectin residue 61 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

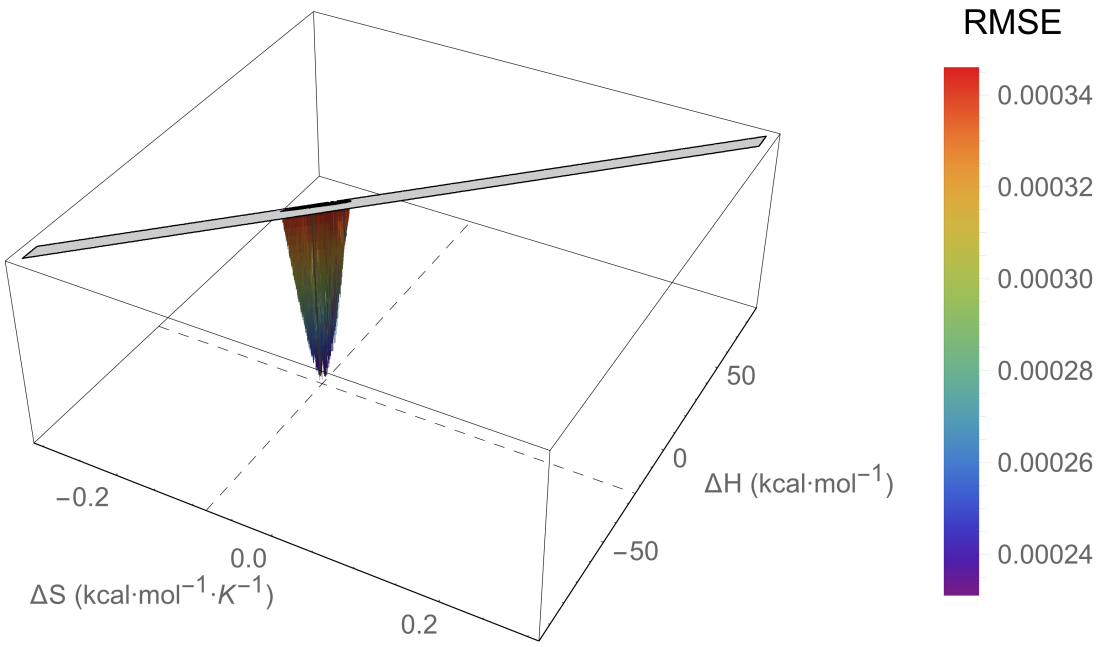


Figure I.309: An RMSE landscape from fitting L18V-Y88F Adnectin residue 62 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

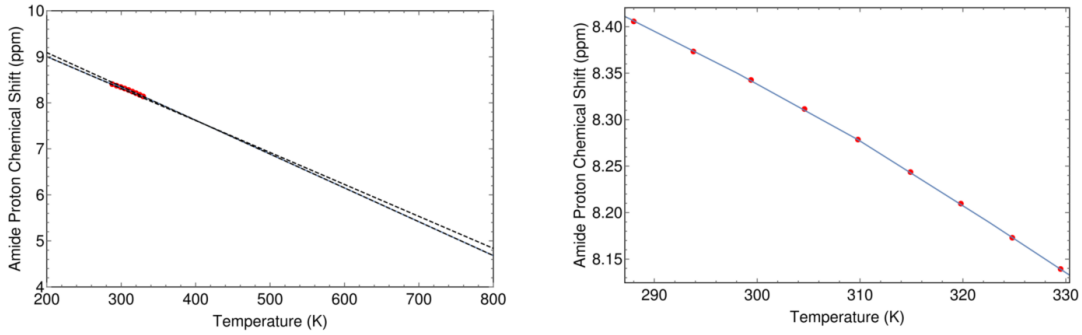


Figure I.310: The minimum RMSE fit of L18V-Y88F Adnectin residue 62 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

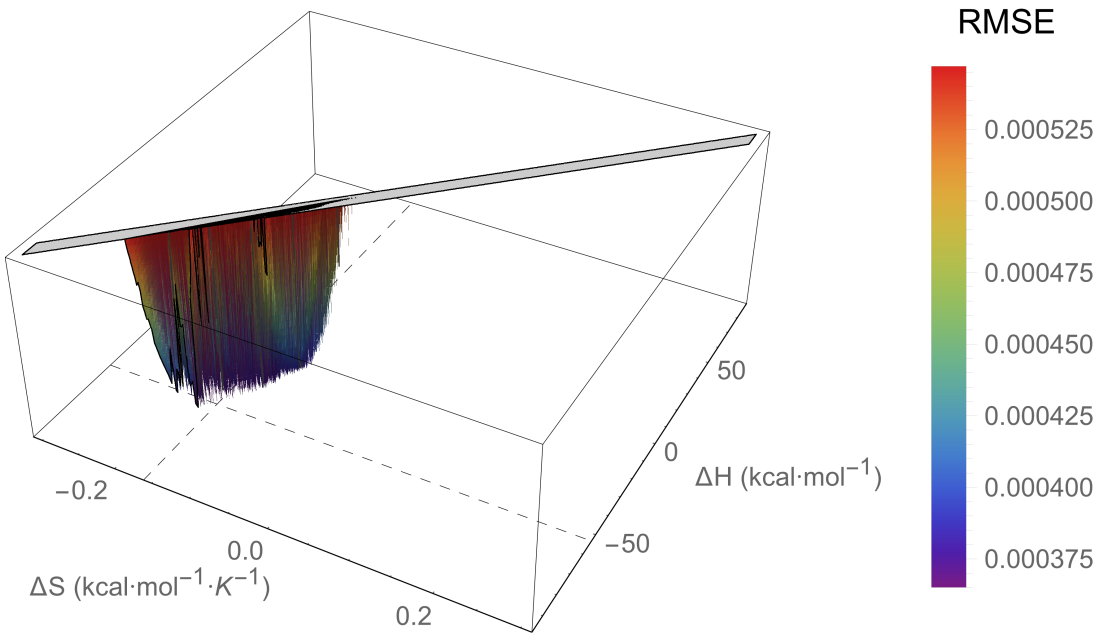


Figure I.311: An RMSE landscape from fitting L18V-Y88F Adnectin residue 67 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

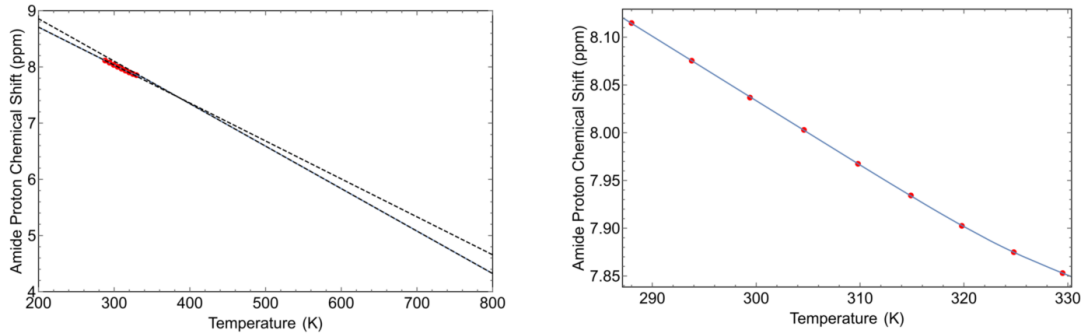


Figure I.312: The minimum RMSE fit of L18V-Y88F Adnectin residue 67 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

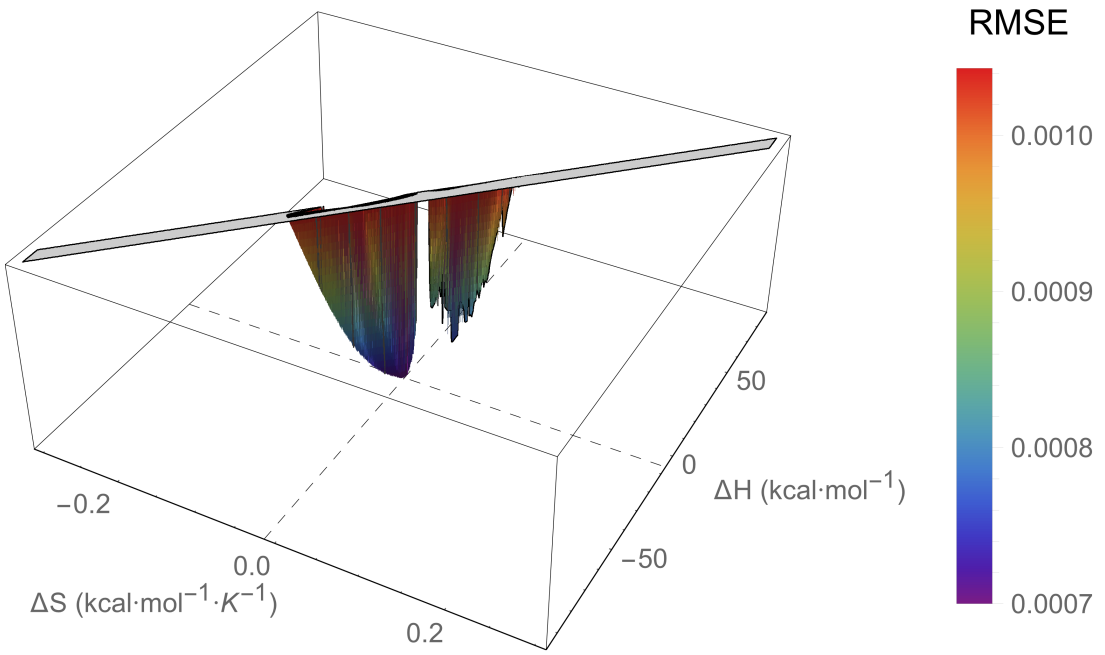


Figure I.313: An RMSE landscape from fitting L18V-Y88F Adnectin residue 68 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give $0 > \Delta G \geq -7.5$ kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above $1.5 \cdot \text{RMSE}_{\text{min}}$) are indicated in gray.

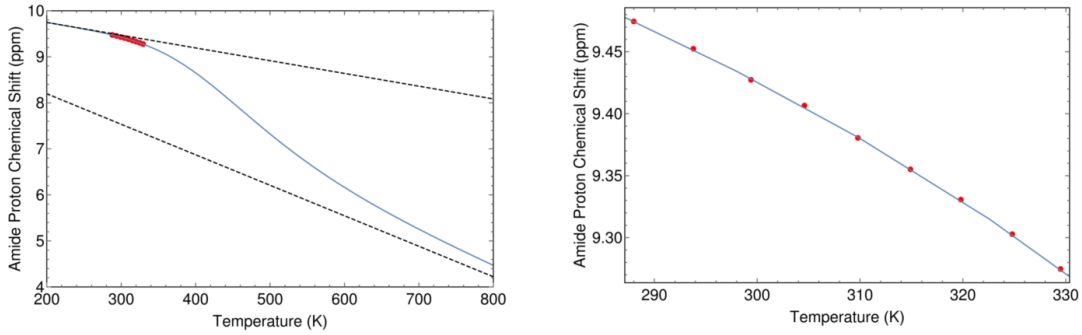


Figure I.314: The minimum RMSE fit of L18V-Y88F Adnectin residue 68 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

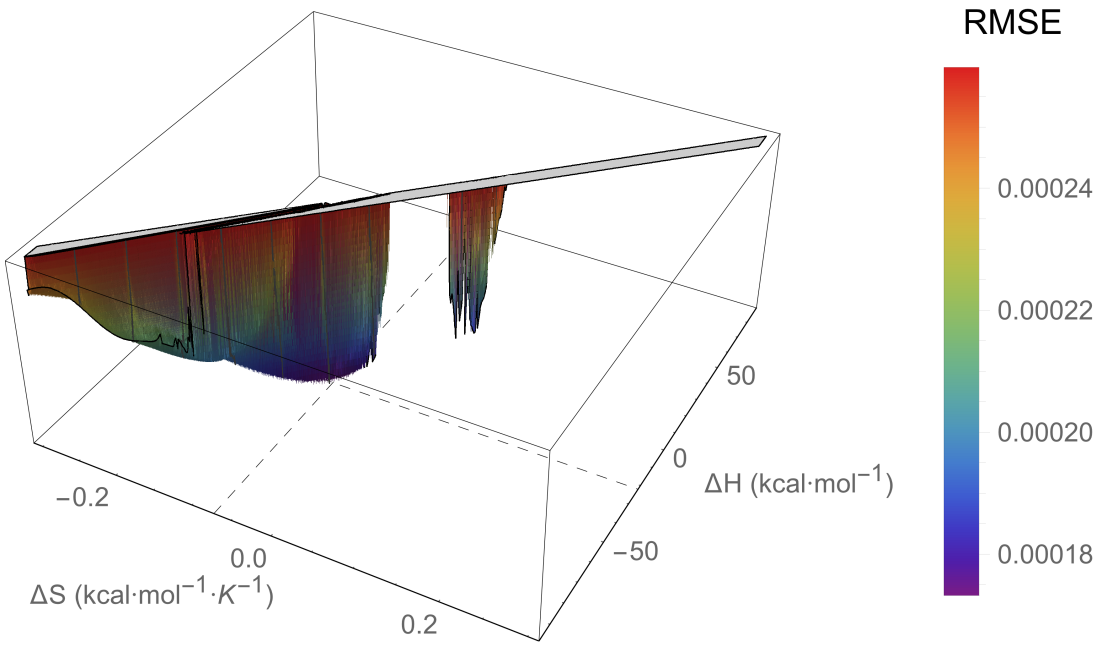


Figure I.315: An RMSE landscape from fitting L18V-Y88F Adnectin residue 88 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE$_{\text{min}}$) are indicated in gray.

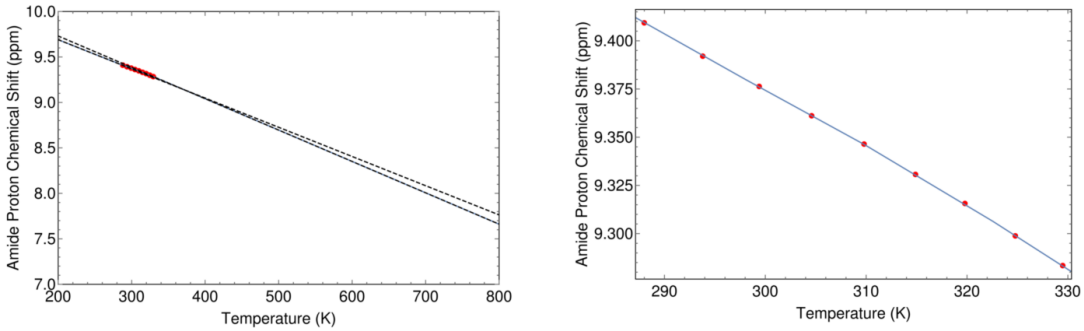


Figure I.316: The minimum RMSE fit of L18V-Y88F Adnectin residue 88 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

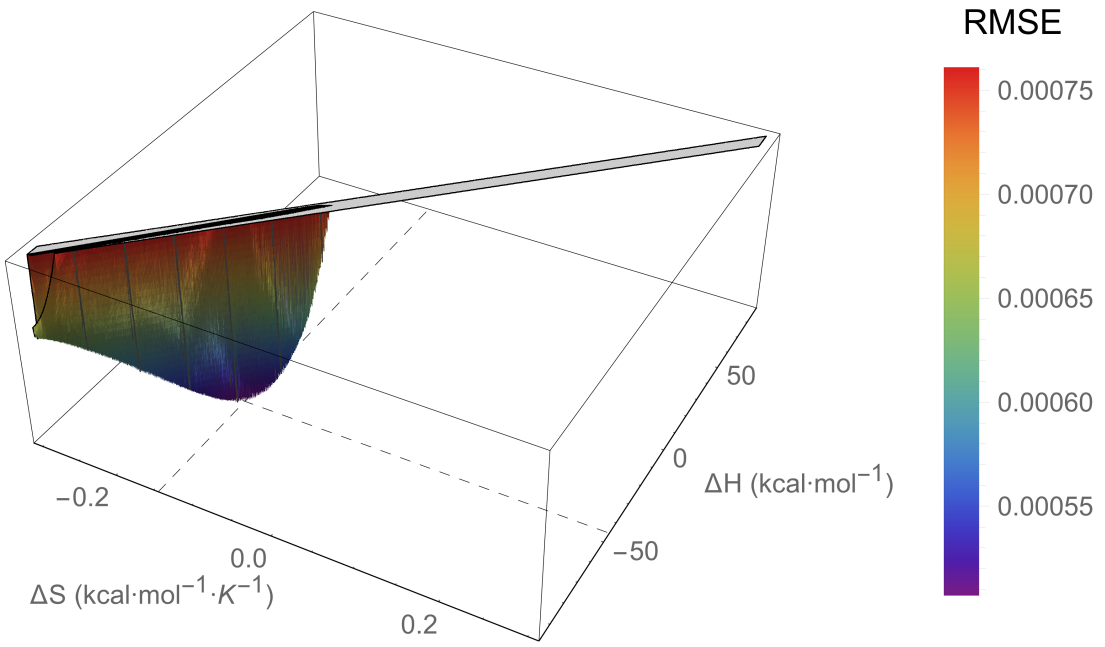


Figure I.317: An RMSE landscape from fitting L18V-Y88F Adnectin residue 89 curvature to the $\Delta C_p = 0$ model. ΔH and ΔS are swept over combinations of values that give 0>ΔG≥-7.5 kcal/mol at 288 K, and at each point the remaining model parameters are determined by linear regression. RMSE values off the scale (above 1.5·RMSE_{min}) are indicated in gray.

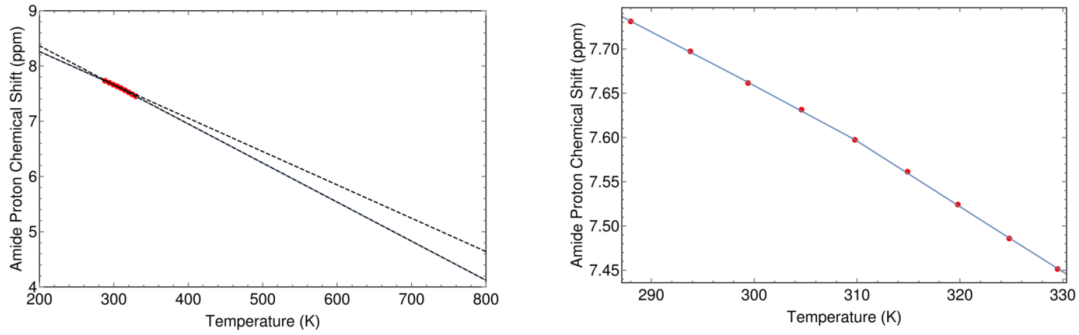


Figure I.318: The minimum RMSE fit of L18V-Y88F Adnectin residue 89 curvature to the $\Delta C_p = 0$ model. Shown are chemical shifts (red), the $\Delta C_p = 0$ model fit (blue), and the linear temperature dependences of states A and B (dashed black lines). Left: an unrealistic temperature range showing more of the sigmoidal transition; Right: the experimental temperature range.

Appendix J

Amide Nitrogen Temperature Coefficients: Supplemental Information

J.1 Adnectin Amide Nitrogen Temperature Coefficients

Table J.1: **Adnectin amide nitrogen linear temperature coefficients.**

Residue #[2]	2° Struct.[3]	Temperature Coefficients (ppb/K)[1]				
		pWT[4]	Parent[5]	L78I[6]	V75R[6]	L18V/ Y88F[6]
1		-20.0	-4.0	-3.1	-4.9	-4.5
2		-29.9	N/S	-19.7	-14.8	-18.1
3		N/S	N/S	-13.5	-14.6	-15.1
4		-40.7	-34.5	-33.7	-36.6	-33.2
5						
6	E (β)	N/S	N/S	-15.4	-19.8	-18.8
7	E (β)	5.0	N/S	-19.6	N/S	N/S
8	E (β)	4.9	15.8	17.5	4.1	14.0
9	E (β)	-10.5	-4.8	-7.6	-3.0	N/A
10	E (β)	10.6	N/S	N/S	N/S	N/S
11	E (β)	N/S	1.3	0.5	-35.0	N/S
12	E (β)	-6.0	-1.3	N/S	-6.7	N/A
13	E (β)	2.4	0.1	-0.4	-1.7	N/A
14	E (β)	N/S	-7.3	-8.6	-4.8	N/A
15	T (turn)					
16	T (turn)	18.1	22.7	23.2	17.4	21.2
17	E (β)	-1.2	-0.8	-0.7	-0.6	-3.1
18	E (β)	12.1	15.9	16.0	13.6	32.3

Table J.1: **Adnectin amide nitrogen linear temperature coefficients.**

Residue #[2]	2° Struct.[3]	Temperature Coefficients (ppb/K)[1]				
		pWT[4]	Parent[5]	L78I[6]	V75R[6]	L18V/ Y88F[6]
19	E (β)	-7.2	-9.5	-9.5	-11.5	4.0
20	E (β)	19.0	10.7	12.2	15.9	14.9
21	E (β)	11.0	24.1	1.4	5.1	12.2
22	E (β)	3.6	-5.2	-6.6	-6.6	-8.3
23	E (β)	N/S	7.3	6.2	5.1	6.3
24		0.5	-8.6	-7.1	-22.1	-9.5
25						
26	S (bend)	-13.0	N/A	N/A	N/A	N/A
27	S (bend)	1.4	N/A	N/A	N/A	N/A
28		-21.1	N/A	N/A	N/A	N/A
29		-41.4	N/A	N/A	N/S	N/A
30	S (bend)	N/S	N/A	N/A	N/S	N/A
31	E (β)	-14.4	N/A	N/A	-30.6	N/A
32	E (β)	-0.8	N/S	N/S	5.5	N/A
33	E (β)	N/S	N/A	N/A	-18.8	N/A
34	E (β)	17.8	-11.0	-1.5	-3.1	-7.8
35	E (β)	18.9	24.8	25.1	15.5	25.6
36	E (β)	21.4	24.1	23.6	15.9	24.8
37	E (β)	-12.9	-11.6	-12.2	-10.8	-13.6
38	E (β)	-3.8	N/S	-5.0	-3.6	-6.2
39	T (turn)	-37.1	-38.7	-39.2	-35.0	-41.1
40	T (turn)	-32.1	-30.6	-30.6	-31.3	-31.2
41		-9.5	-8.7	-8.3	-9.5	-8.5
42		-3.3	-5.8	N/A	-5.5	-5.3
43	S (bend)	-14.7	-14.4	-14.4	-14.9	-15.2
44						
45		-10.4	-5.1	-9.2	-6.2	N/S
46	E (β)	N/S	N/S	-14.3	-11.2	-14.9
47	E (β)	N/S	-5.8	N/A	N/S	-5.3
48	E (β)	-22.4	-10.8	-10.6	-7.9	-9.1
49	E (β)	-19.3	21.5	22.1	12.1	23.0
50	E (β)	21.9	N/S	N/S	-11.0	N/A
51	E (β)					
52	T (turn)	-12.4	N/A	N/A	N/A	N/A
53	T (turn)	-10.4	N/A	N/A	N/A	N/A
54		N/S	-32.9	-32.7	-31.1	-33.6
55		1.4	2.7	0.1	2.0	1.9
56	E (β)	14.9	12.4	13.5	8.7	13.7
57	E (β)	N/S	N/S	2.8	N/S	-6.3
58	E (β)	-8.6	N/S	N/S	-6.9	-8.6
59	E (β)	-20.6	-25.6	-26.0	-30.0	-18.5
60		-13.7	-17.9	-15.8	-19.8	-15.9
61	S (bend)	-0.9	-0.6	-0.5	-1.0	-1.9
62		N/A	-6.4	2.1	6.2	1.4

Table J.1: **Adnectin amide nitrogen linear temperature coefficients.**

Residue #[2]	2° Struct.[3]	Temperature Coefficients (ppb/K)[1]				
		pWT[4]	Parent[5]	L78I[6]	V75R[6]	L18V/ Y88F[6]
63		N/S	N/S	N/S	-16.3	N/A
64	T (turn)					
65	T (turn)	-27.2	-27.5	-25.2	-24.3	-32.0
66		-28.1	-25.0	-25.3	-24.7	-24.1
67		-7.9	-8.4	-8.3	-8.1	-5.6
68	E (β)	N/S	-21.4	-20.8	-20.3	-21.7
69	E (β)	5.3	8.2	8.4	7.3	2.8
70	E (β)	-20.0	-18.0	-18.7	-13.6	-27.2
71	E (β)	-12.0	-1.3	-1.4	0.2	-3.1
72	E (β)	5.7	8.7	8.2	3.4	7.8
73	E (β)	N/S	N/S	N/S	-2.1	N/A
74	E (β)	17.5	N/A	N/A	9.1	N/A
75	E (β)	5.9	N/A	N/A	-2.3	N/A
76		N/S	N/A	N/A	37.2	N/A
77		-19.3	N/A	N/A	N/A	N/A
78		19.1	N/A	N/A	N/A	N/A
79	S (bend)	1.6	N/A	N/A	N/A	N/A
80	S (bend)	18.6	N/A	N/A	N/A	N/A
81	S (bend)	9.1	N/A	N/A	N/A	N/A
82			N/A	N/A	N/A	N/A
83 (NA)		-33.6				
84 (NA)	B (β)	-37.2				
85 (NA)		-42.6				
86 (NA)		N/S				
87 (83)						
88 (84)	E (β)	14.2	N/A	N/A	-23.5	N/A
89 (85)	E (β)	-8.1	-6.4	-9.8	-6.7	-6.3
90 (86)	E (β)	7.0	8.9	9.5	9.2	9.8
91 (87)	E (β)	16.3	N/S	22.9	N/S	16.9
92 (88)	E (β)	-12.1	-8.7	-8.7	-13.2	-15.6
93 (89)		-22.6	N/S	-21.9	-21.0	-44.8
94 (90)		-4.8	-0.9	-0.9	-3.6	-13.8

[1]Proline residues indicated by gray background; N/A: no assignment; N/S: no solution found by the ShiftTrack algorithm (Chapter 7)
[2]Pseudo-wild type ^{10}Fn3 numbering (Adnectin numbering in brackets where different)
[3]2° structure from DSSP (1FNF); blue indicates intra-β-sheet hydrogen bonds (as determined by submission of the ^{10}Fn3 domain from the 1FNF PDB structure to the VADAR protein structural analysis web server [258])
[4]Pseudo-wild type ^{10}Fn3: D80E mutation (in the FG loop) relative to wild type
[5]Parent Adnectin: differs from pWT ^{10}Fn3 in the BC, DE, and FG loops (Chapter 1)
[6]Mutation relative to the 'Parent' Adnectin amino acid sequence

J.2 DFT Chemical Shift Calculations

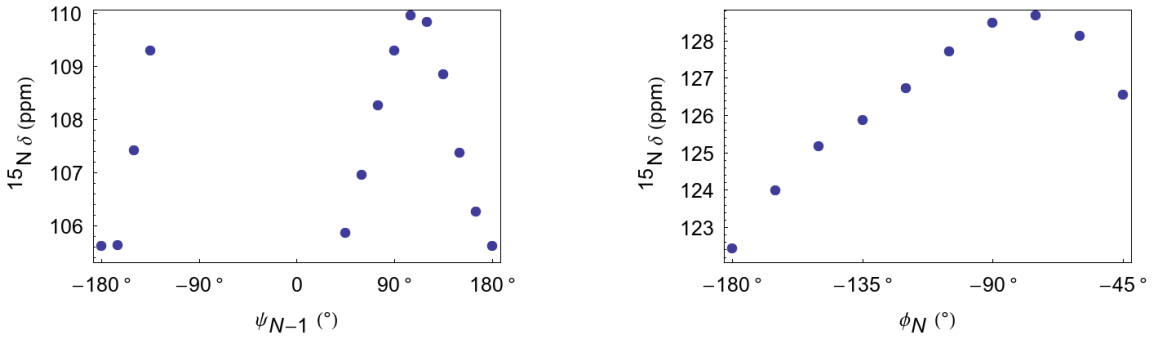
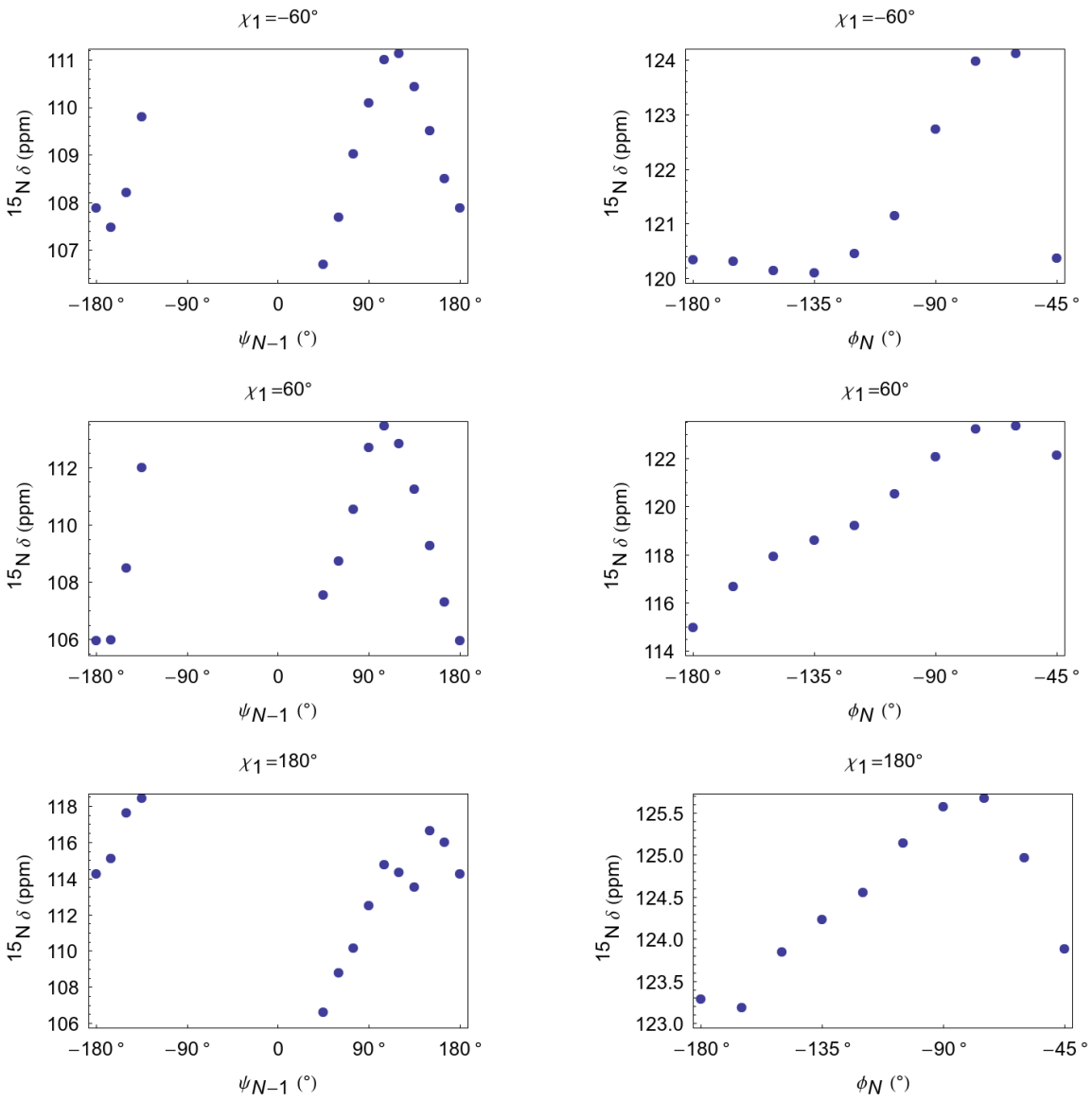


Figure J.1: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=alanine) and ϕ_N (AA_N=alanine). DFT calculations as described in Section 11.3.



DFT calculations as described in Section 11.3.

Figure J.2: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=cysteine) and ϕ_N (AA_N=cysteine).

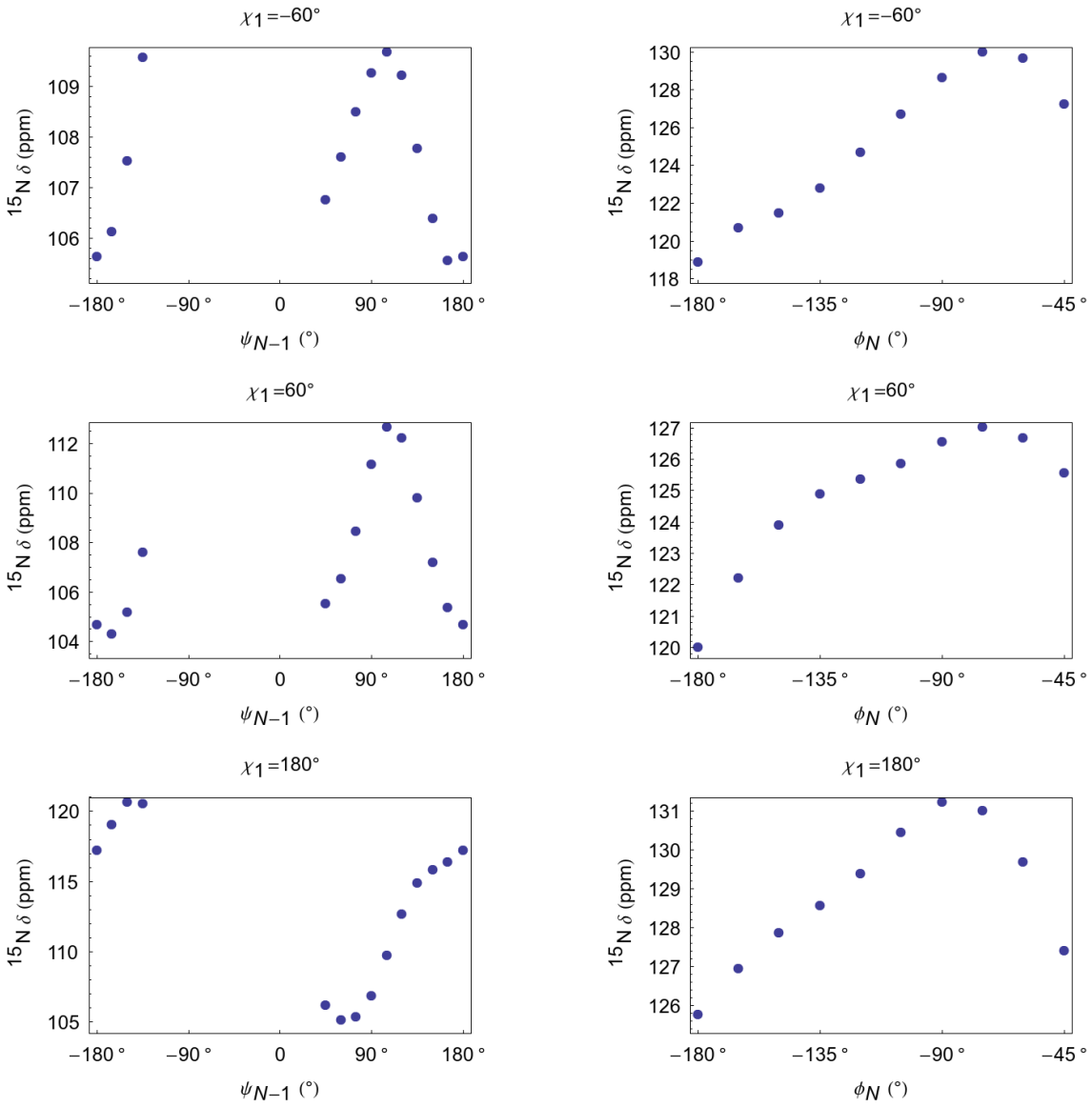


Figure J.3: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=aspartate) and ϕ_N (AA_N=aspartate). DFT calculations as described in Section 11.3.

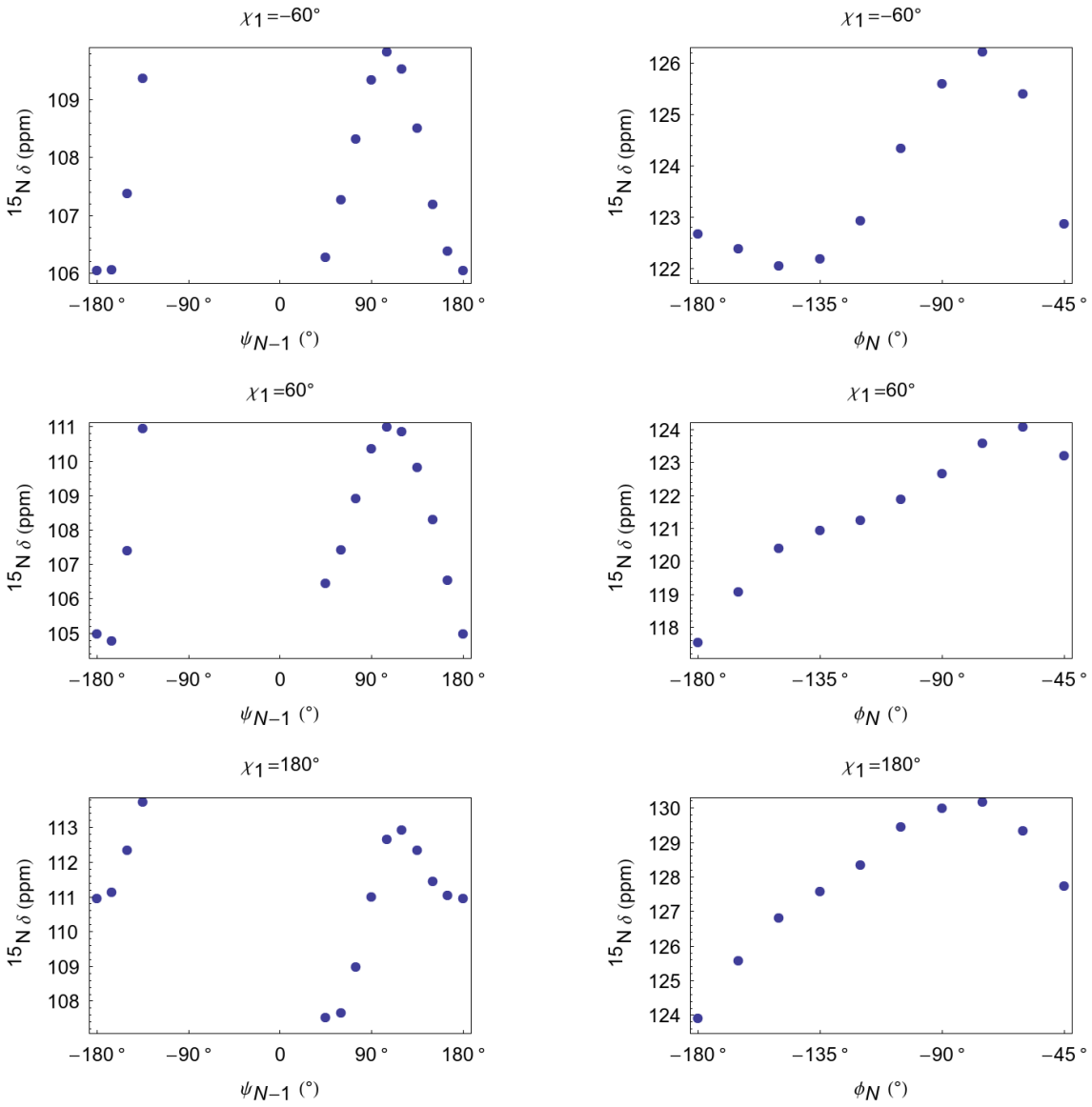


Figure J.4: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=glutamate) and ϕ_N (AA_N=glutamate). DFT calculations as described in Section 11.3.

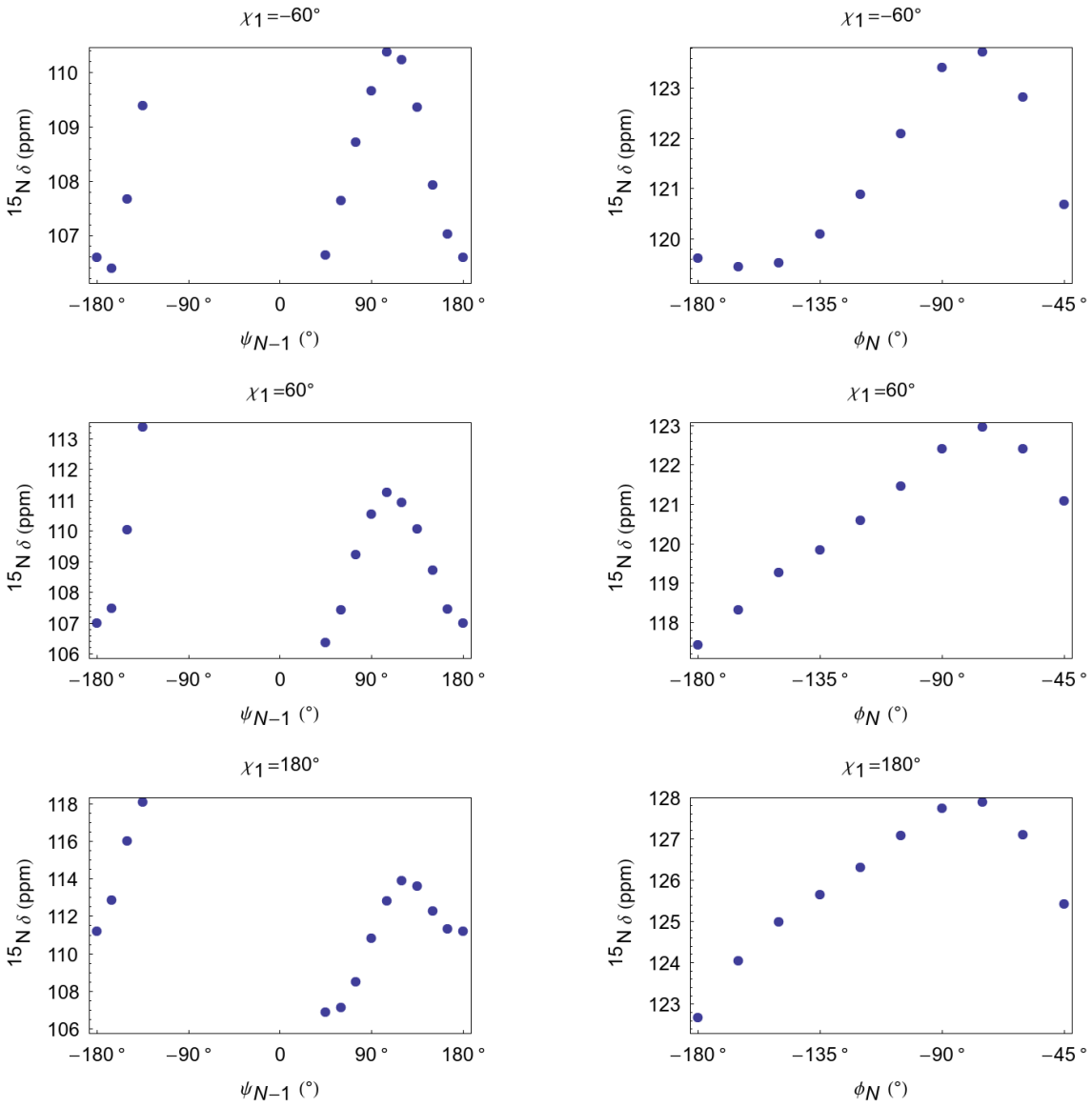


Figure J.5: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=phenylalanine) and ϕ_N (AA_N=phenylalanine). DFT calculations as described in Section 11.3.

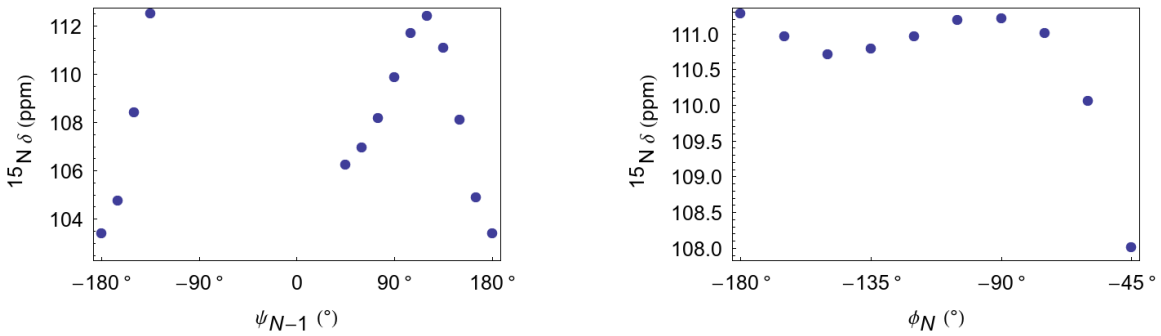


Figure J.6: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=glycine) and ϕ_N (AA_N=glycine). DFT calculations as described in Section 11.3.

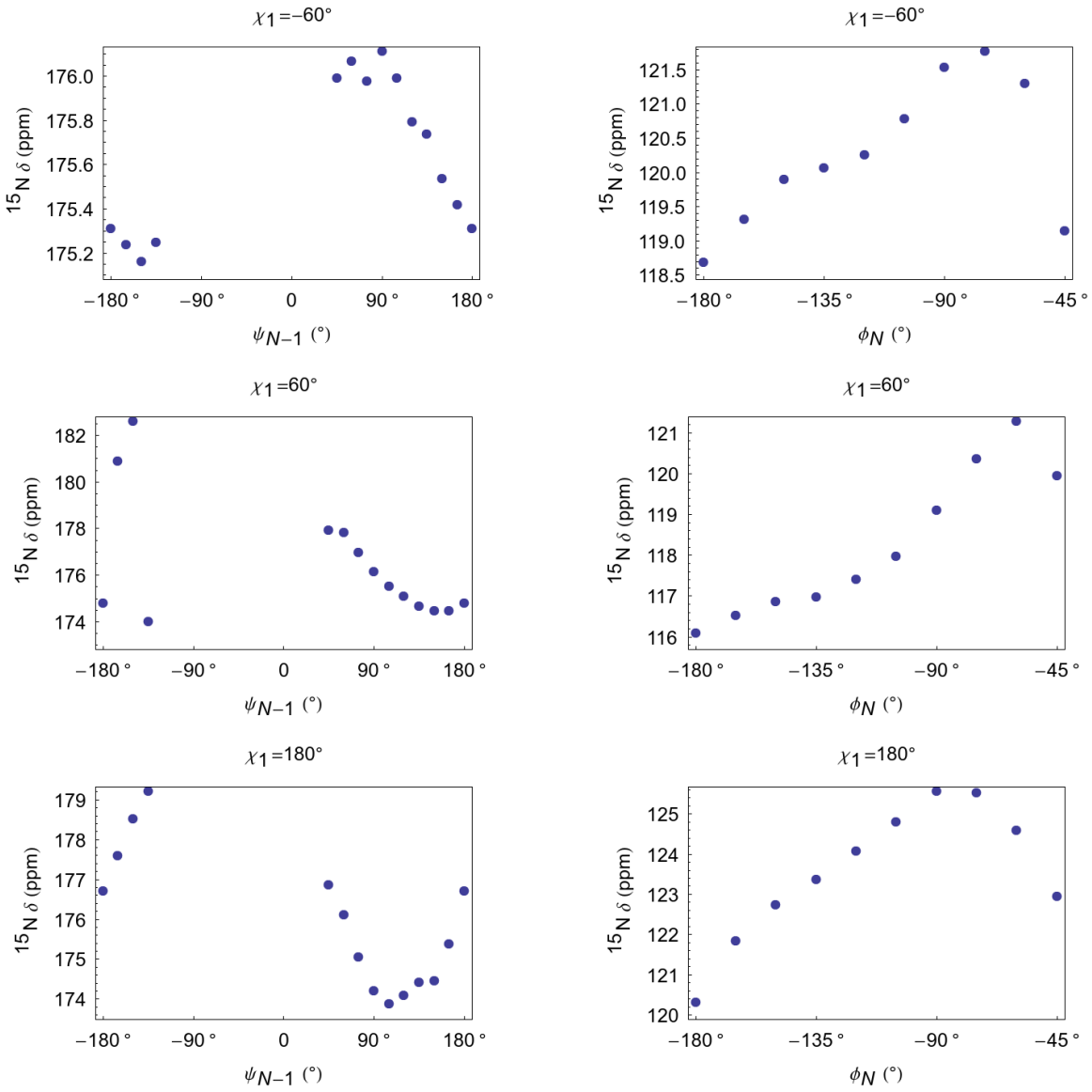


Figure J.7: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=histidine) and ϕ_N (AA_N=histidine). DFT calculations as described in Section 11.3.

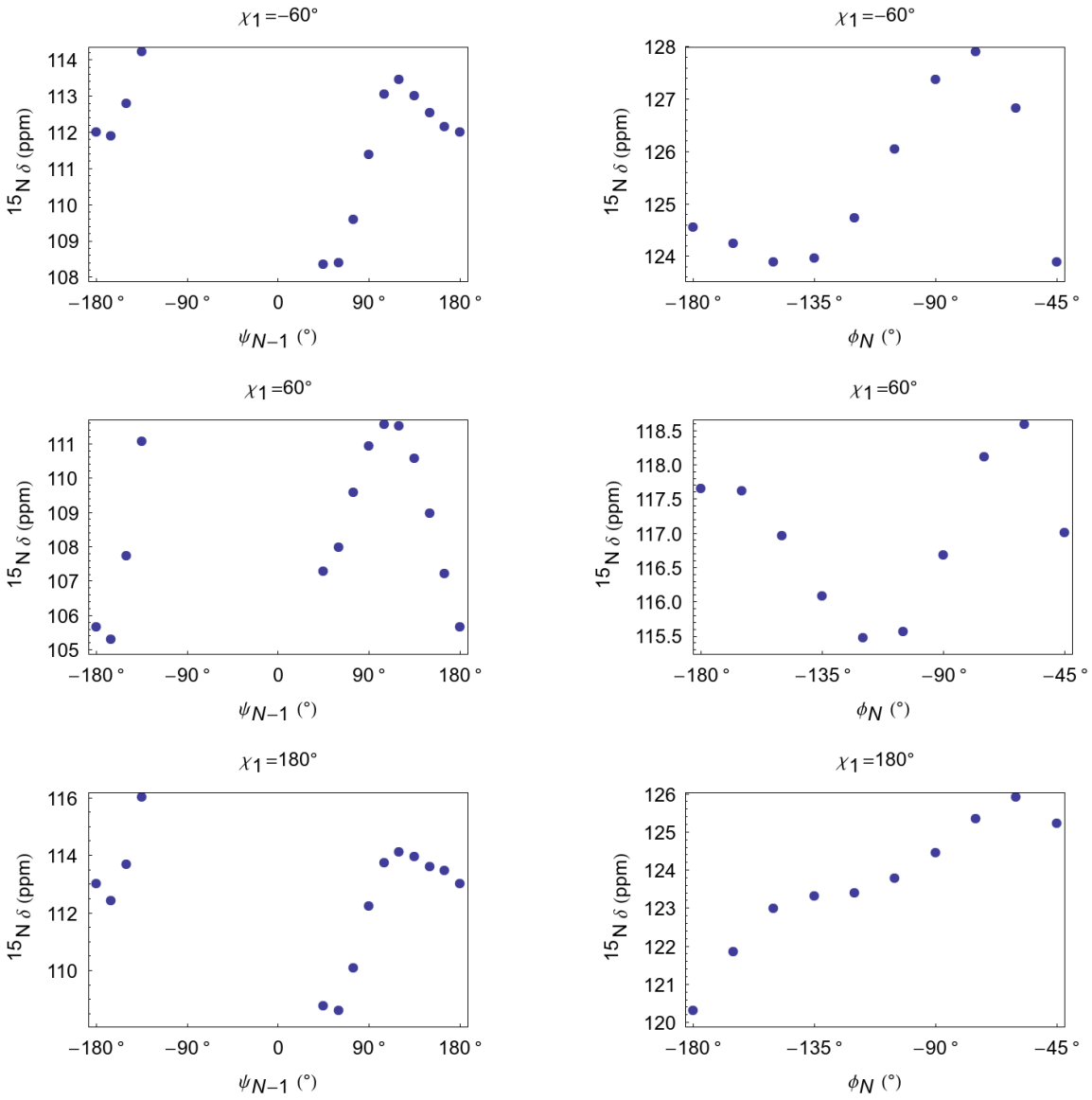


Figure J.8: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=isoleucine) and ϕ_N (AA_N=isoleucine). DFT calculations as described in Section 11.3.

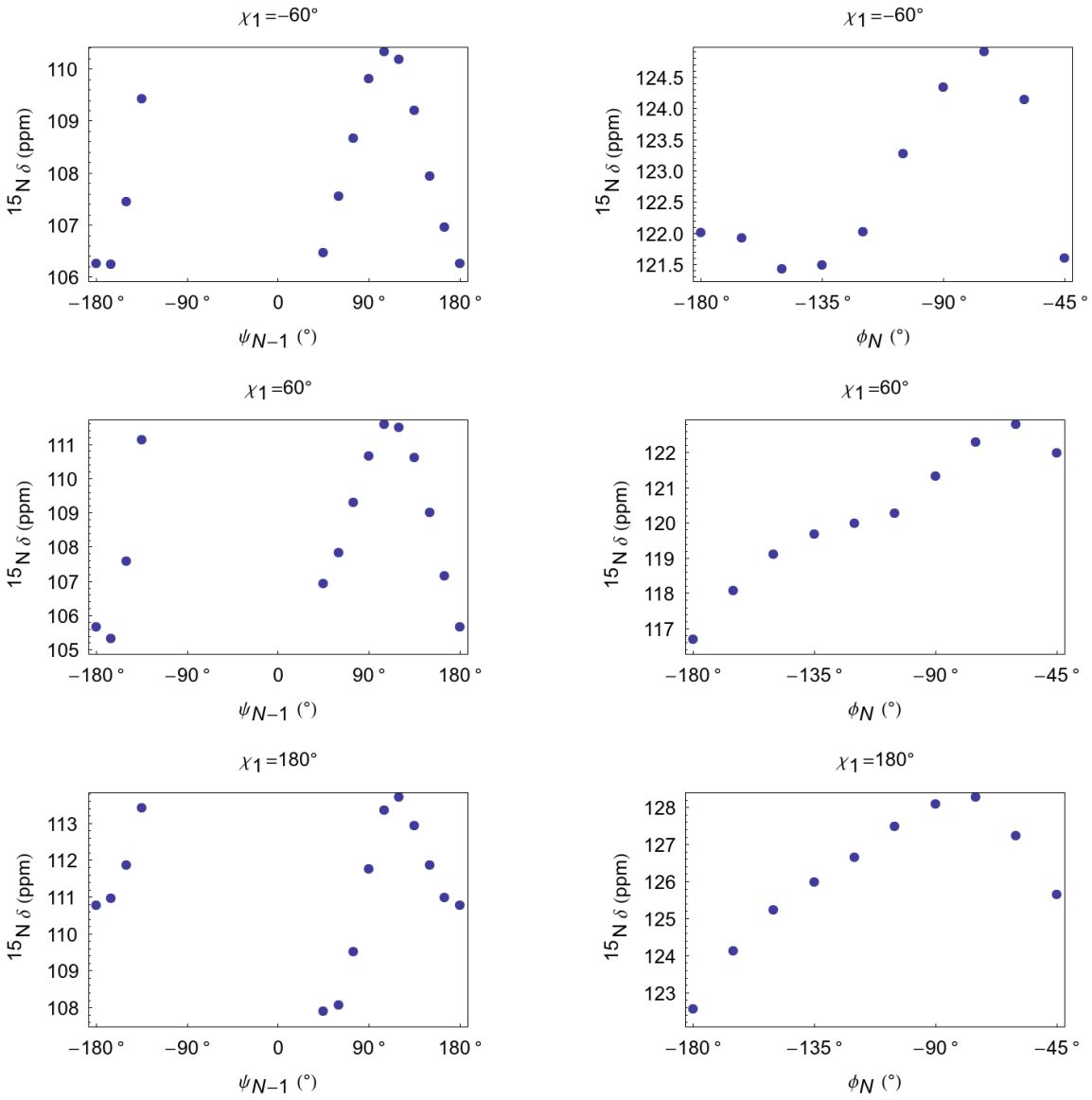


Figure J.9: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=lysine) and ϕ_N (AA_N=lysine). DFT calculations as described in Section 11.3.

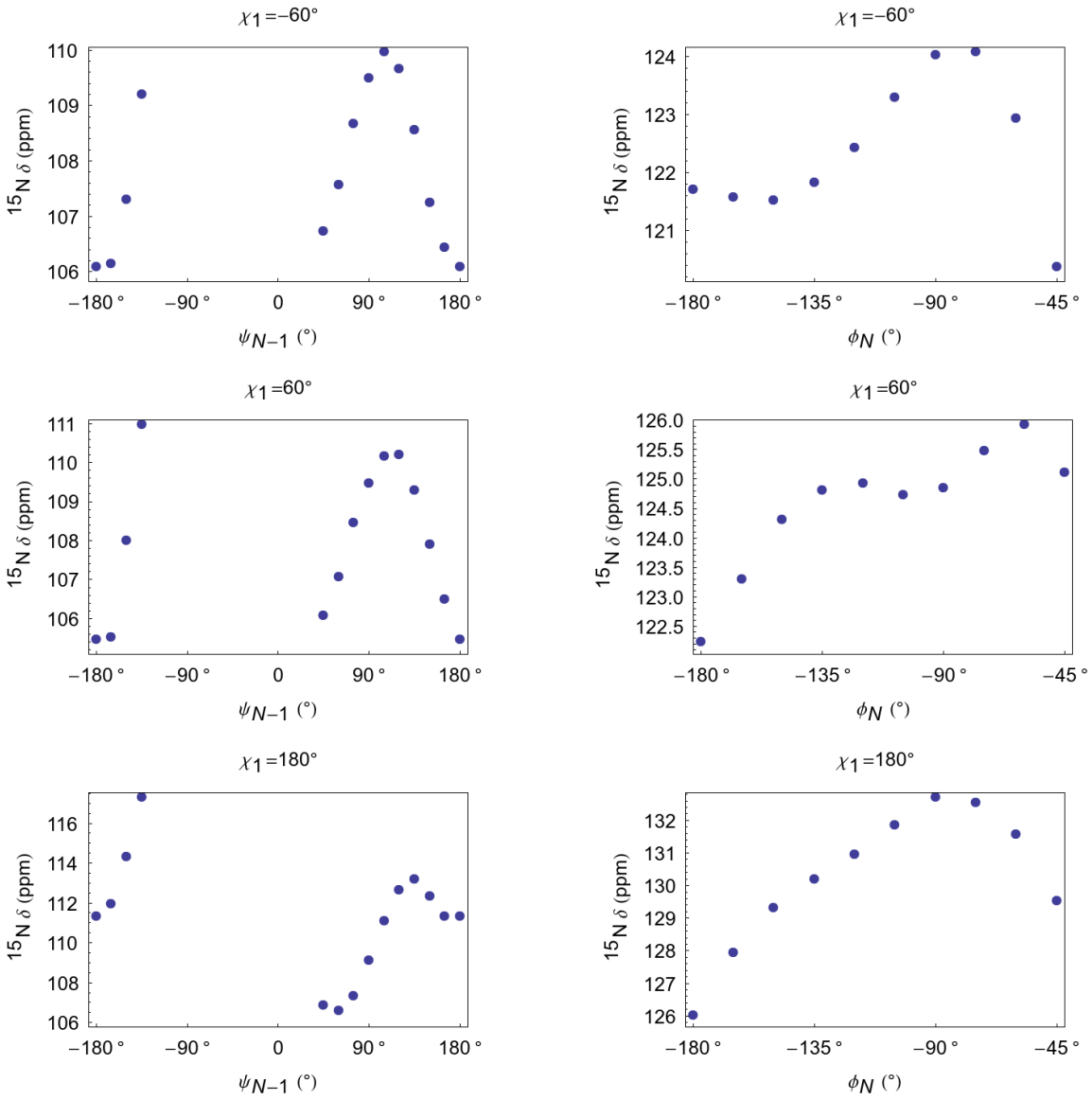


Figure J.10: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=leucine) and ϕ_N (AA_N=leucine). DFT calculations as described in Section 11.3.

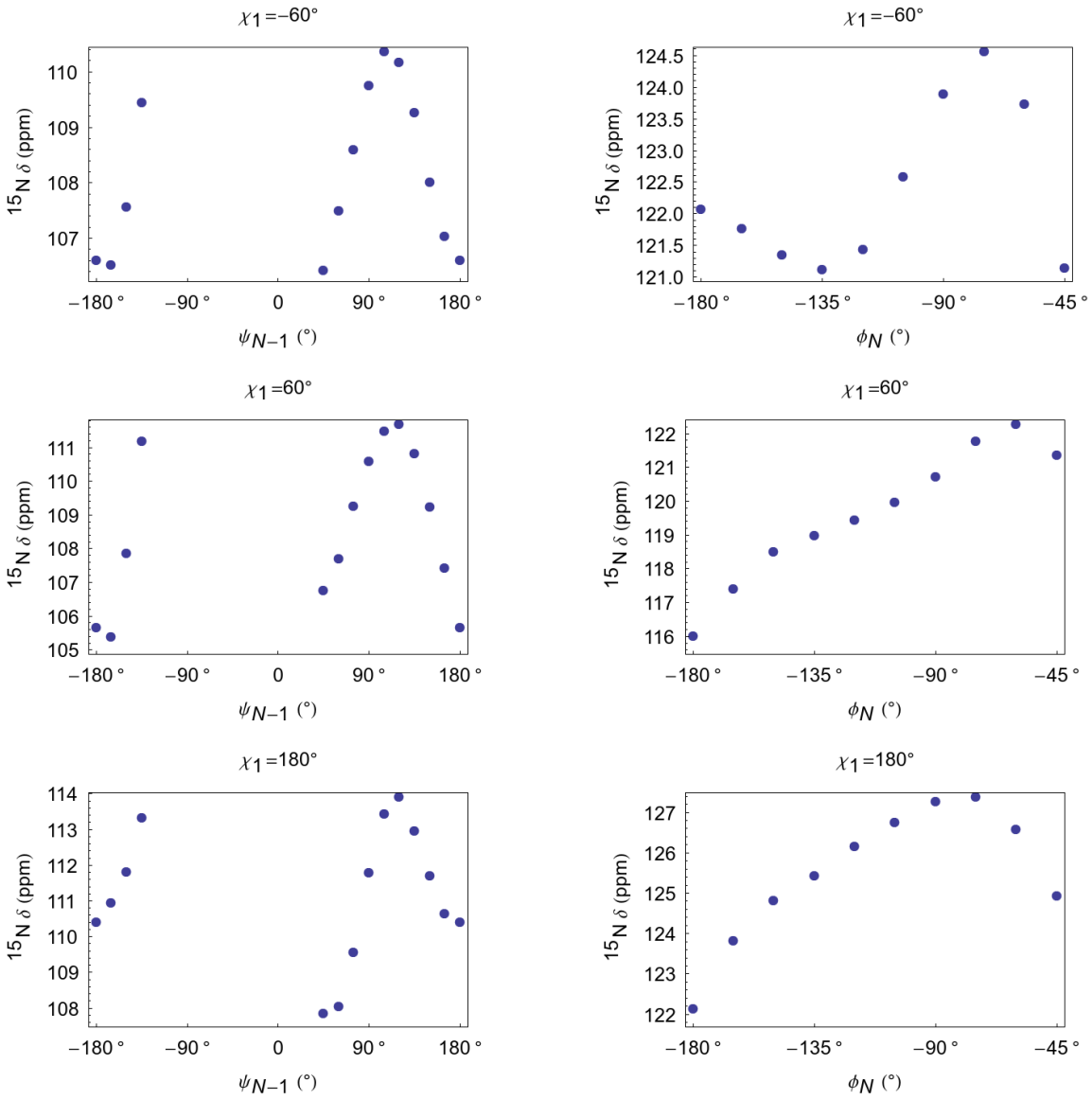


Figure J.11: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=methionine) and ϕ_N (AA_N=methionine). DFT calculations as described in Section 11.3.

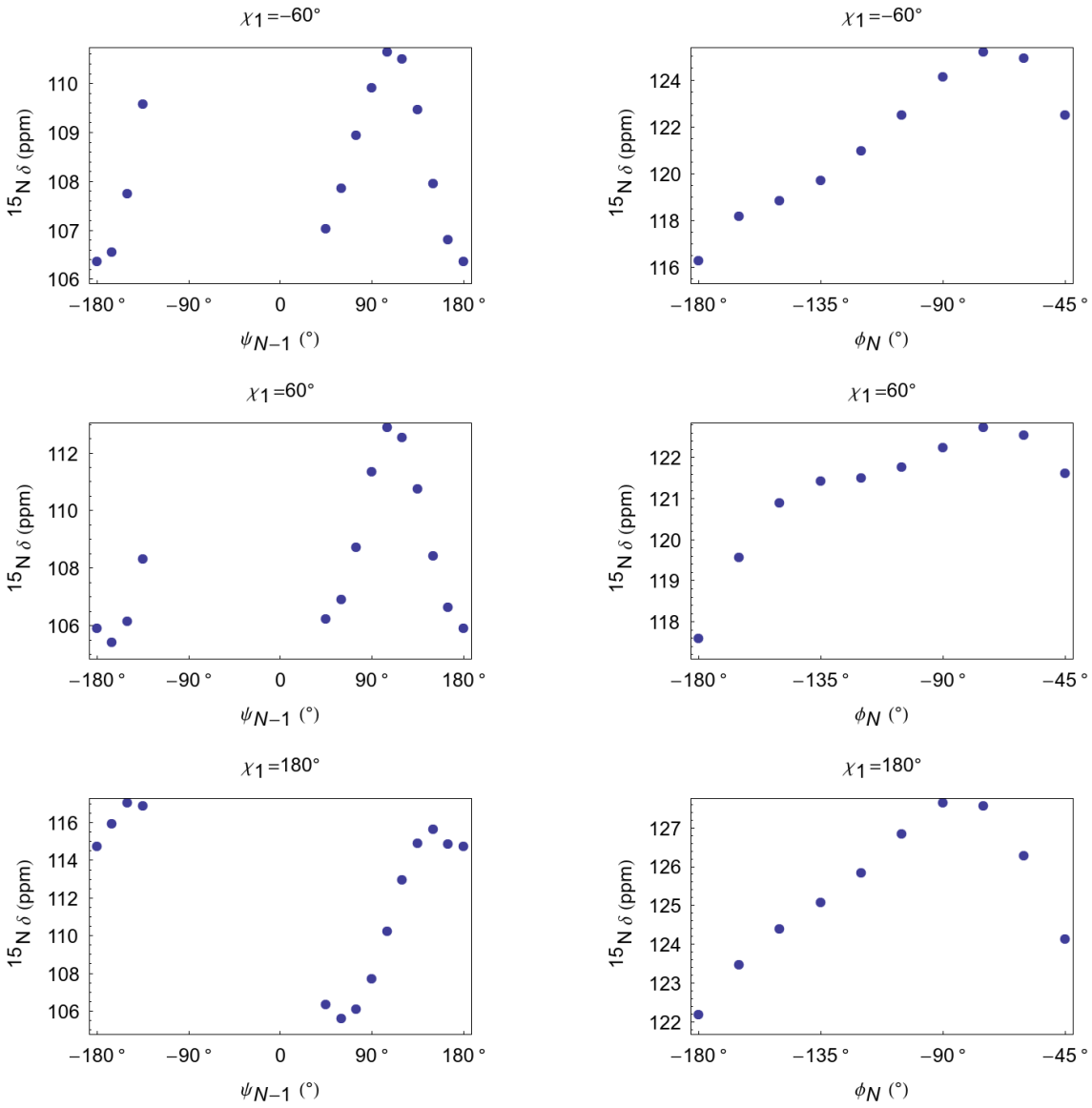


Figure J.12: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=asparagine) and ϕ_N (AA_N=asparagine). DFT calculations as described in Section 11.3.

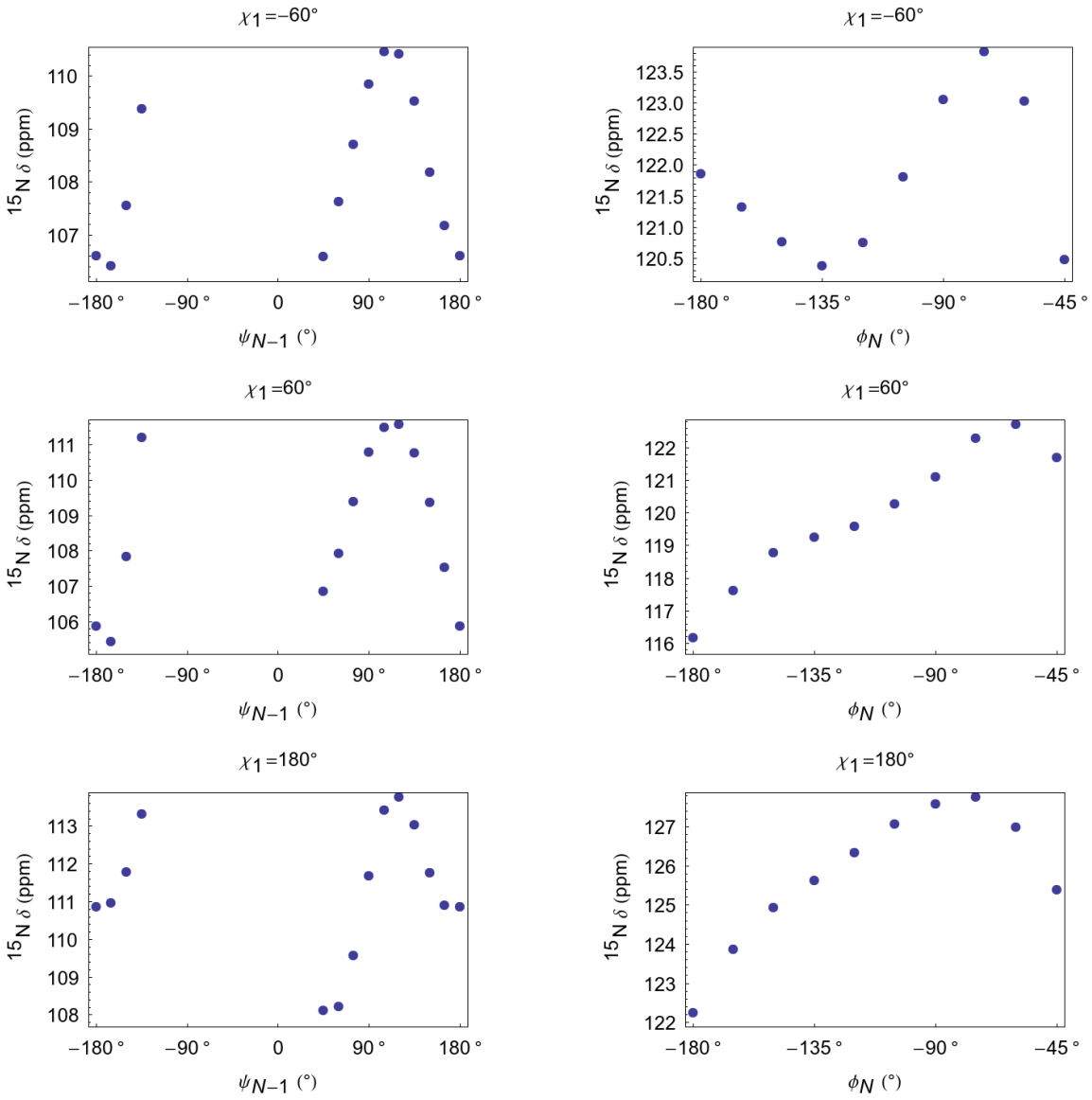


Figure J.13: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=glutamine) and ϕ_N (AA_N=glutamine). DFT calculations as described in Section 11.3.

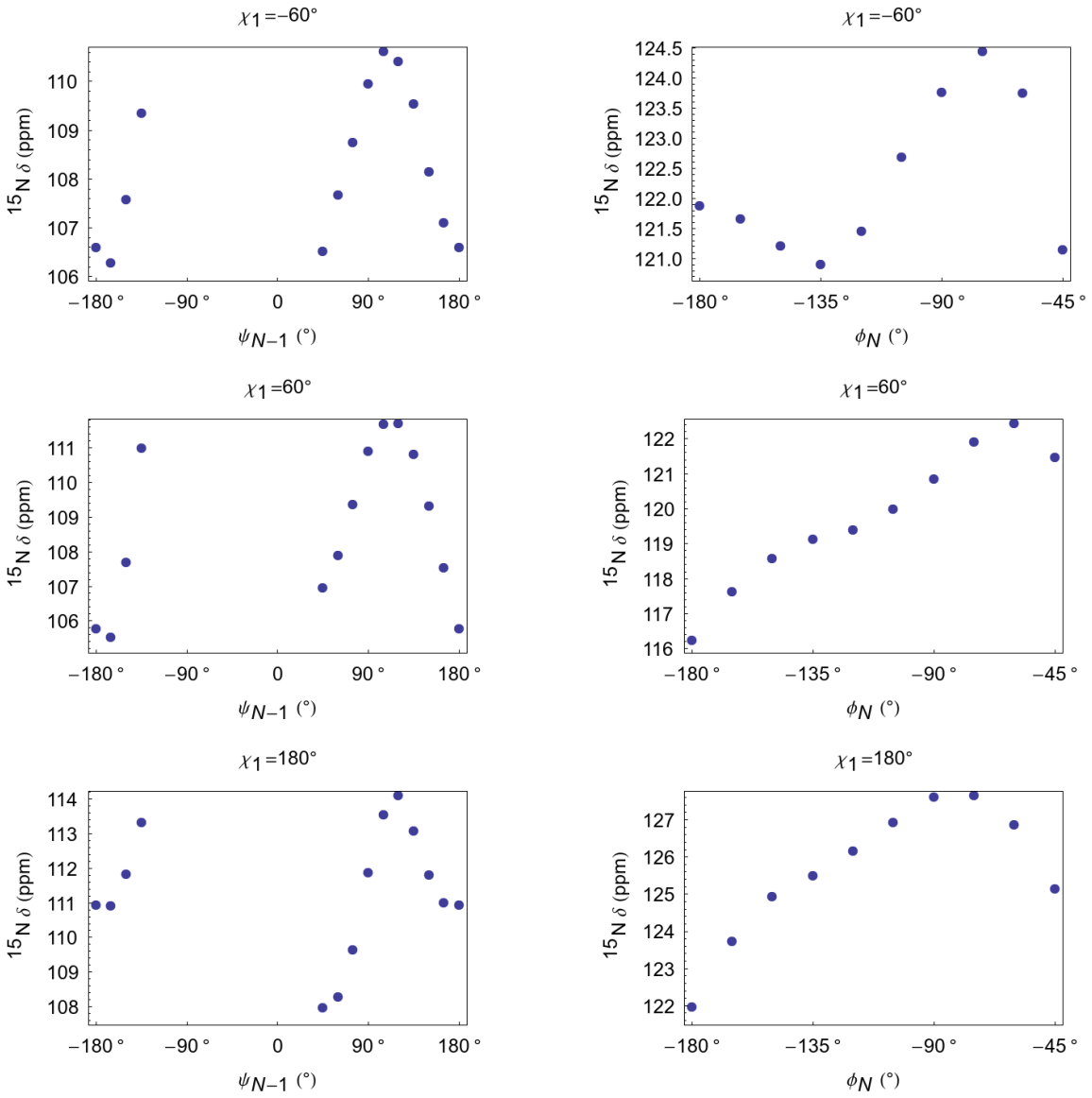


Figure J.14: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=arginine) and ϕ_N (AA_N=arginine). DFT calculations as described in Section 11.3.

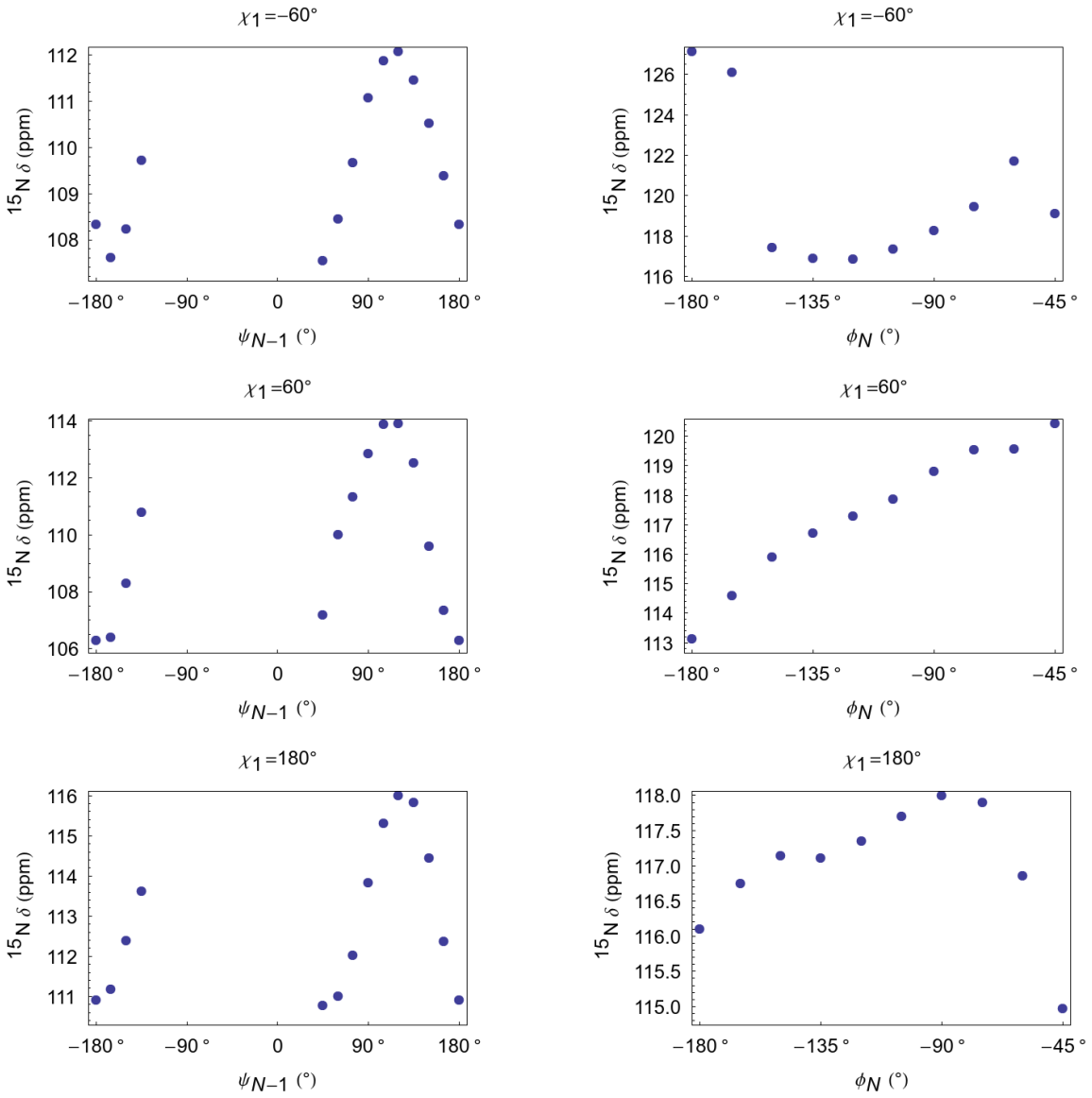


Figure J.15: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=serine) and ϕ_N (AA_N=serine). DFT calculations as described in Section 11.3.

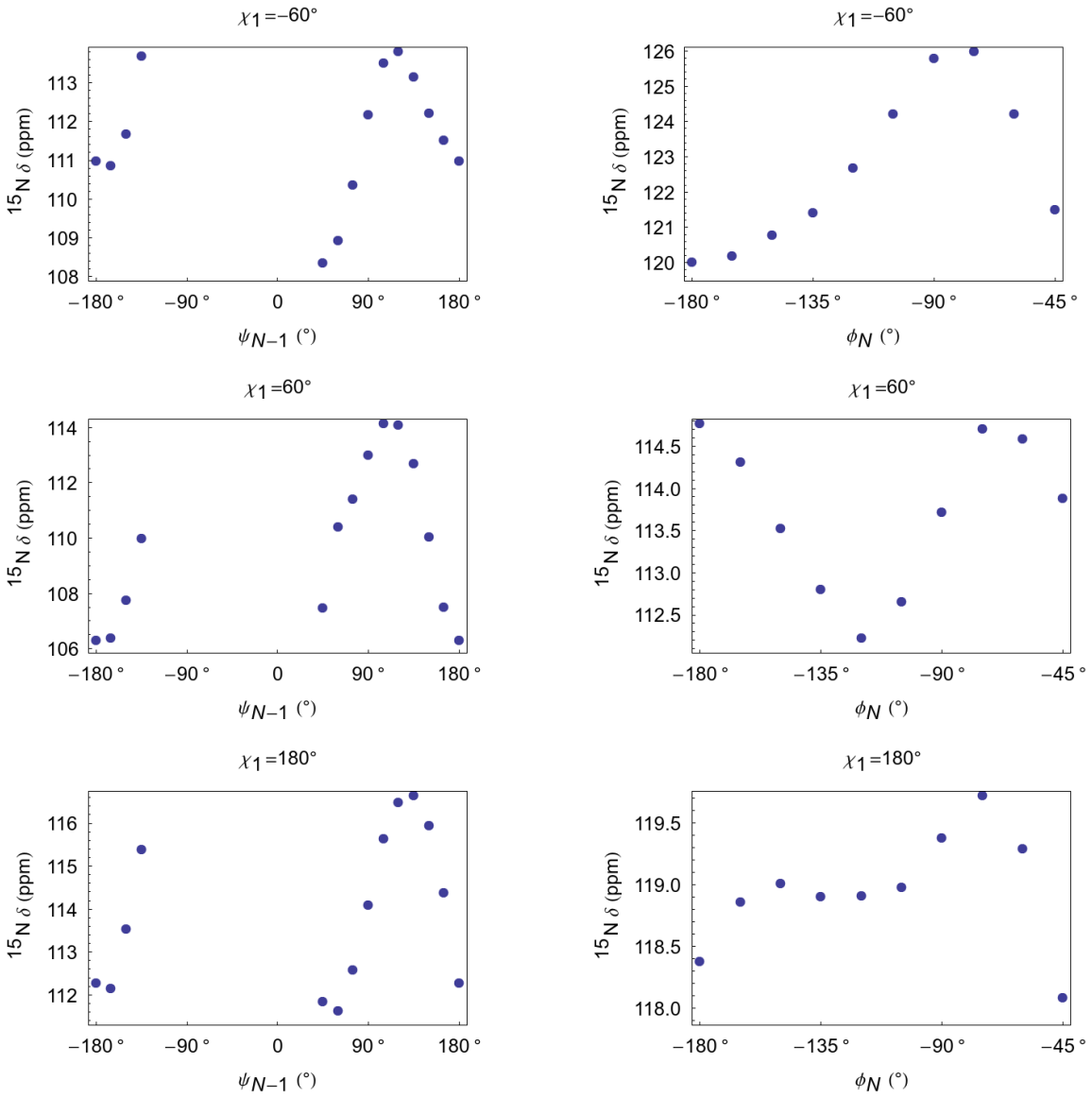


Figure J.16: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=threonine) and ϕ_N (AA_N=threonine). DFT calculations as described in Section 11.3.

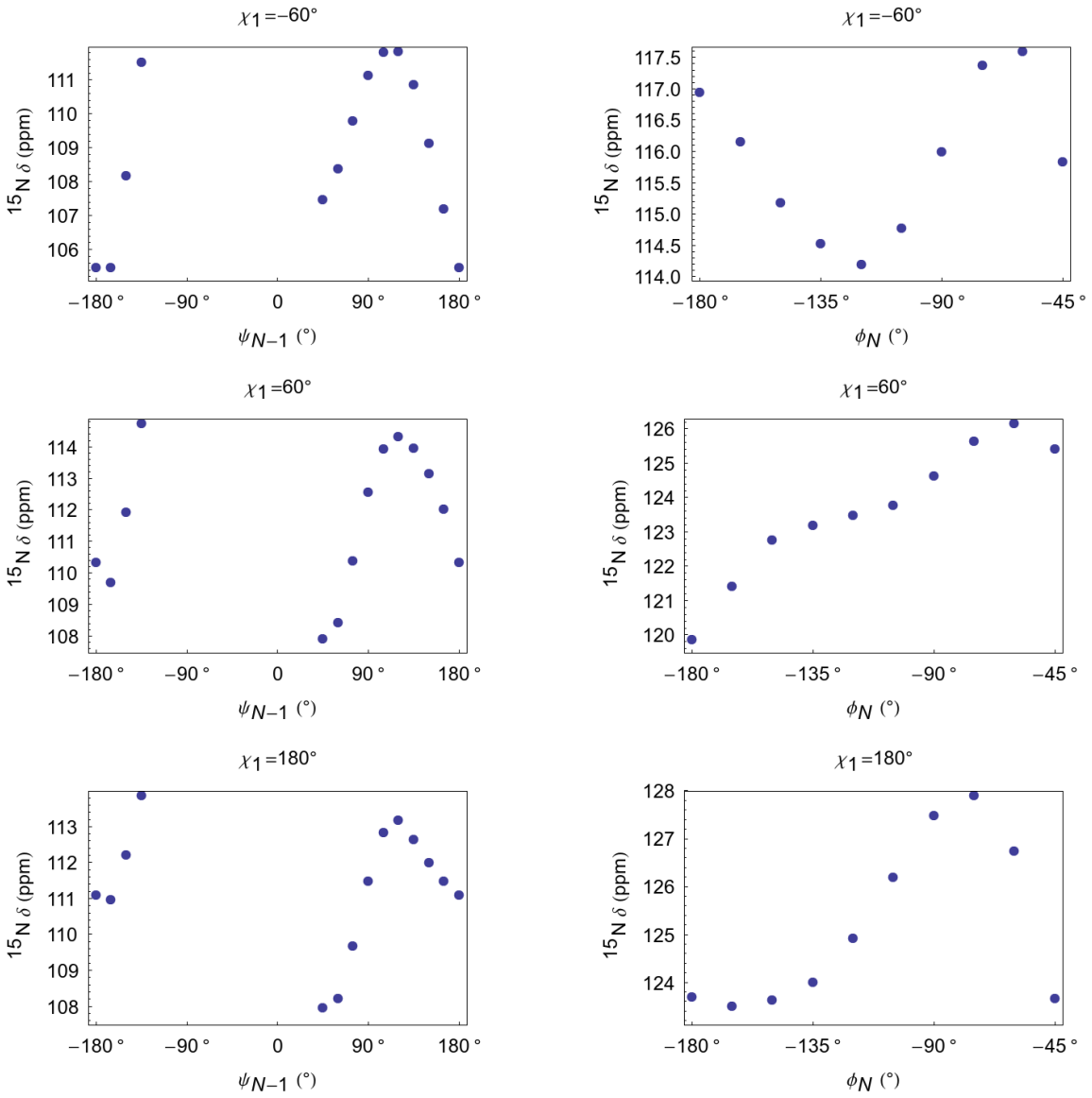


Figure J.17: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=valine) and ϕ_N (AA_N=valine). DFT calculations as described in Section 11.3.

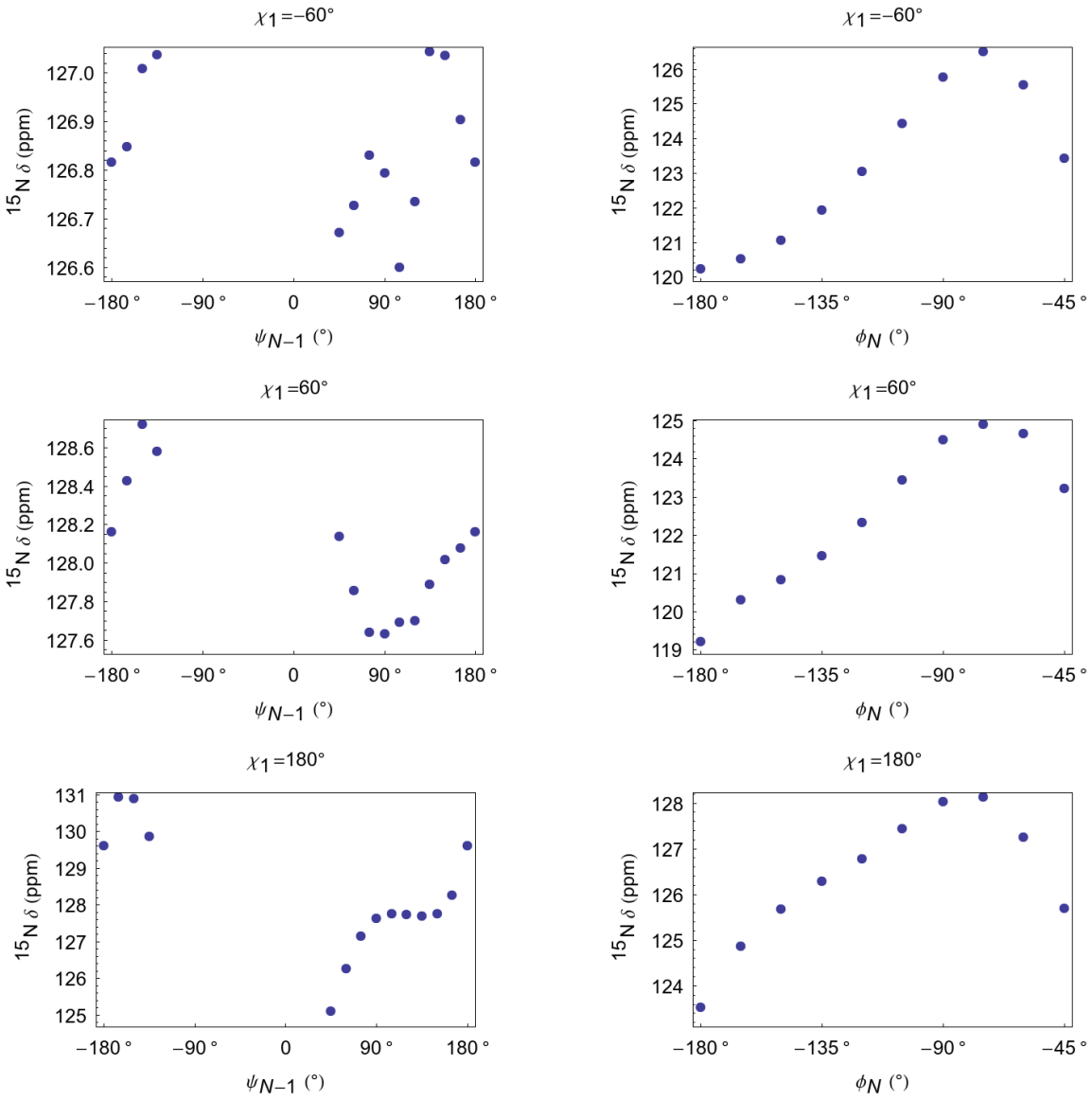


Figure J.18: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=tryptophan) and ϕ_N (AA_N=tryptophan). DFT calculations as described in Section 11.3.

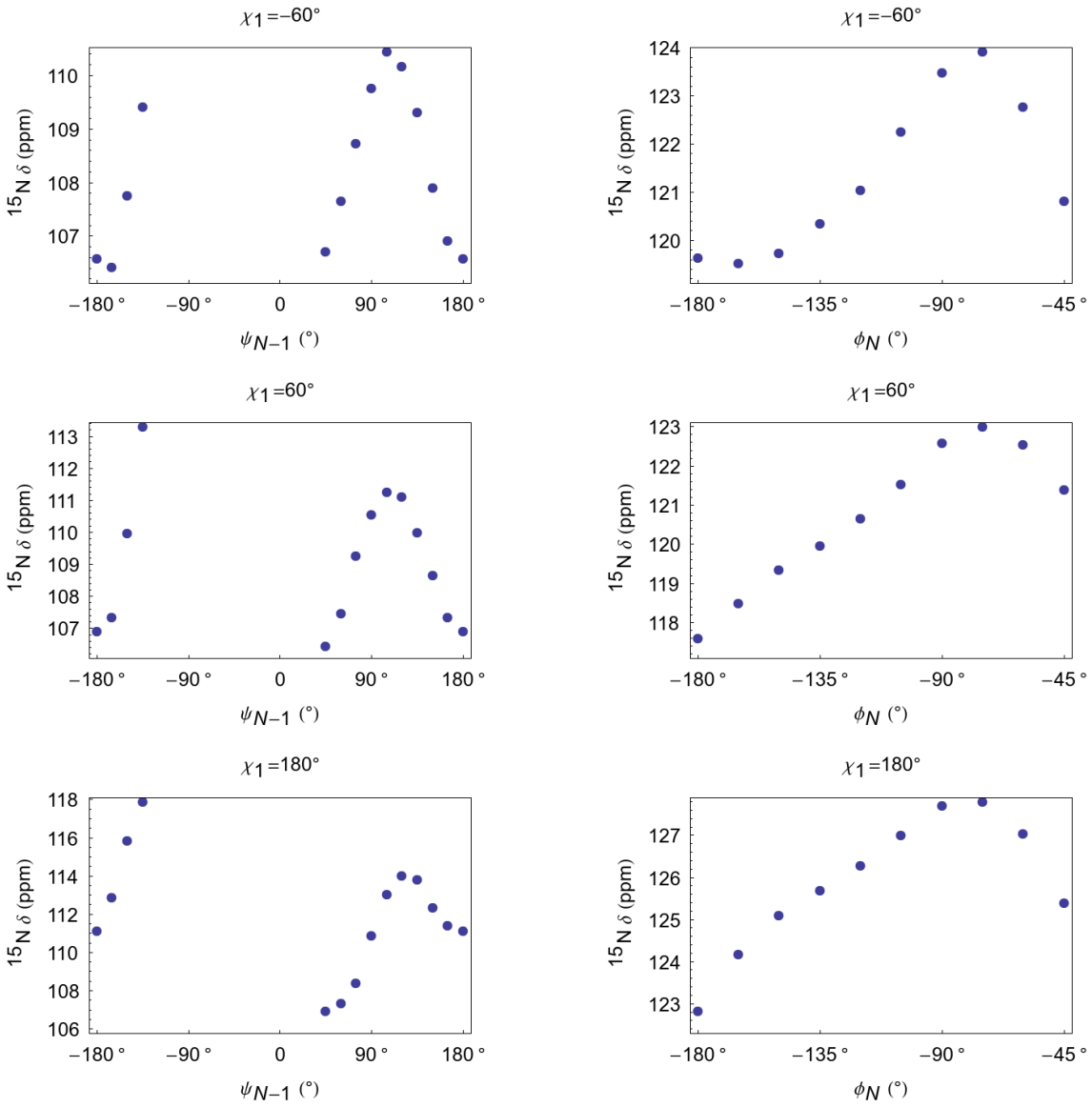


Figure J.19: Amide nitrogen chemical shift as a function of ψ_{N-1} (AA_{N-1}=tyrosine) and ϕ_N (AA_N=tyrosine). DFT calculations as described in Section 11.3.

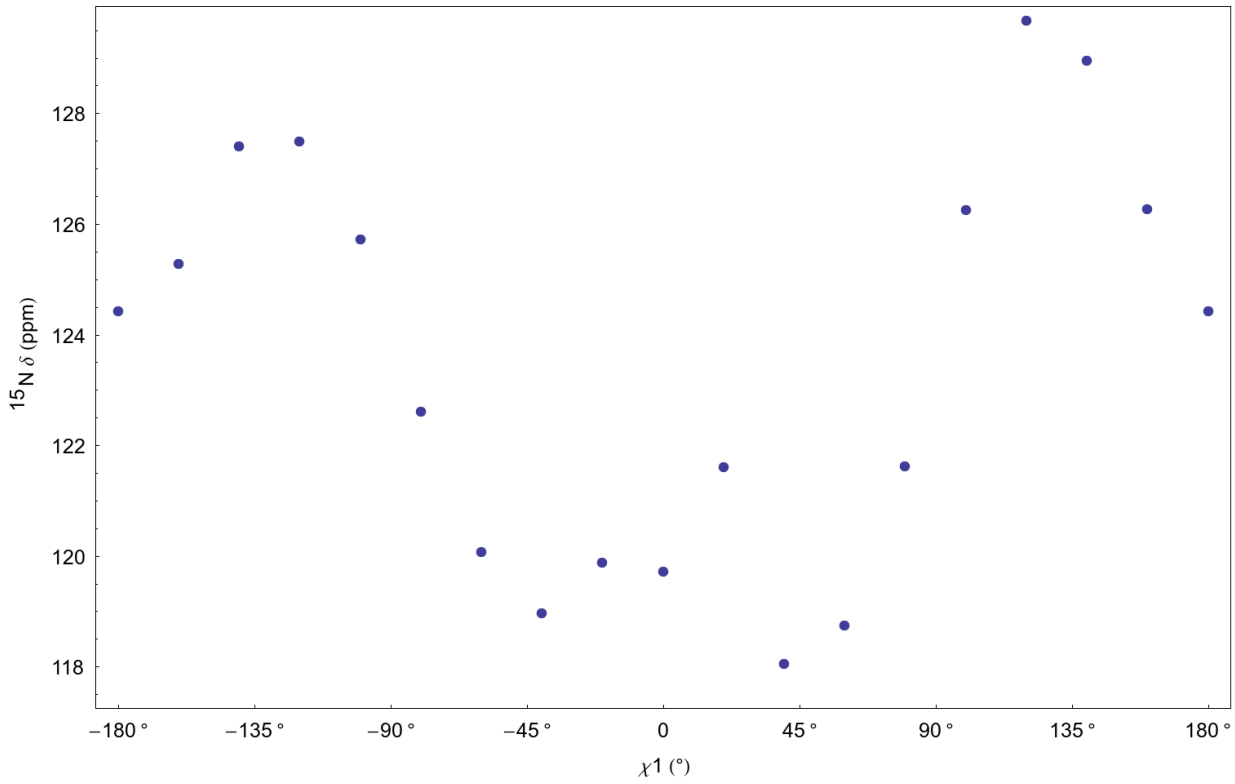


Figure J.20: Amide nitrogen chemical shift as a function of cysteine χ1. DFT calculations as described in Section 11.3.

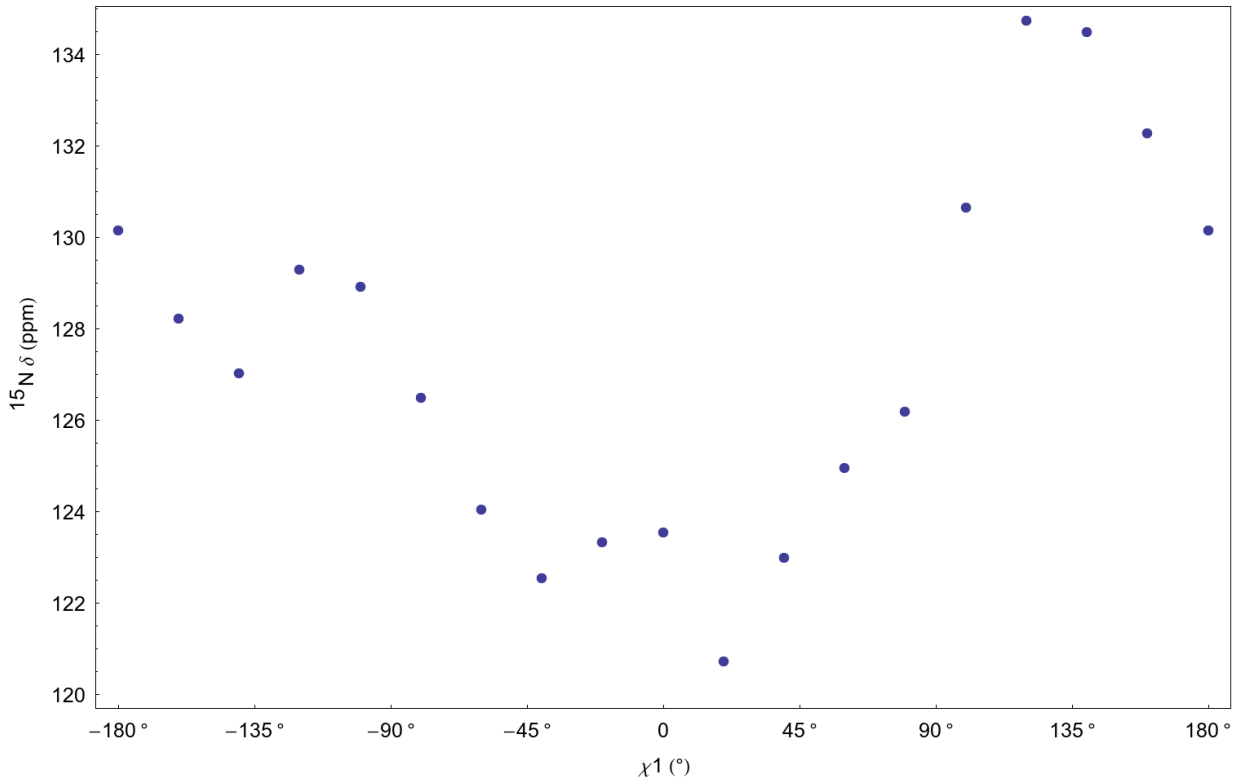


Figure J.21: Amide nitrogen chemical shift as a function of aspartate χ1. DFT calculations as described in Section 11.3.

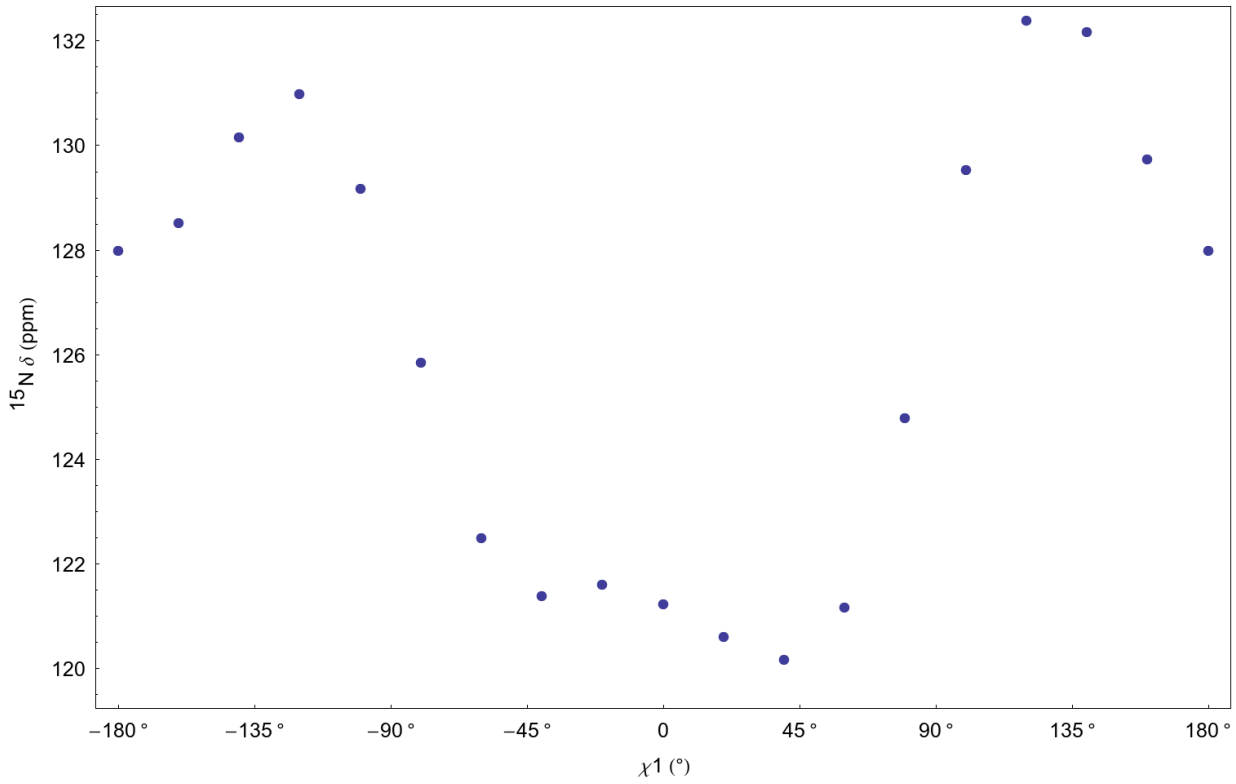


Figure J.22: Amide nitrogen chemical shift as a function of glutamate χ1. DFT calculations as described in Section 11.3.

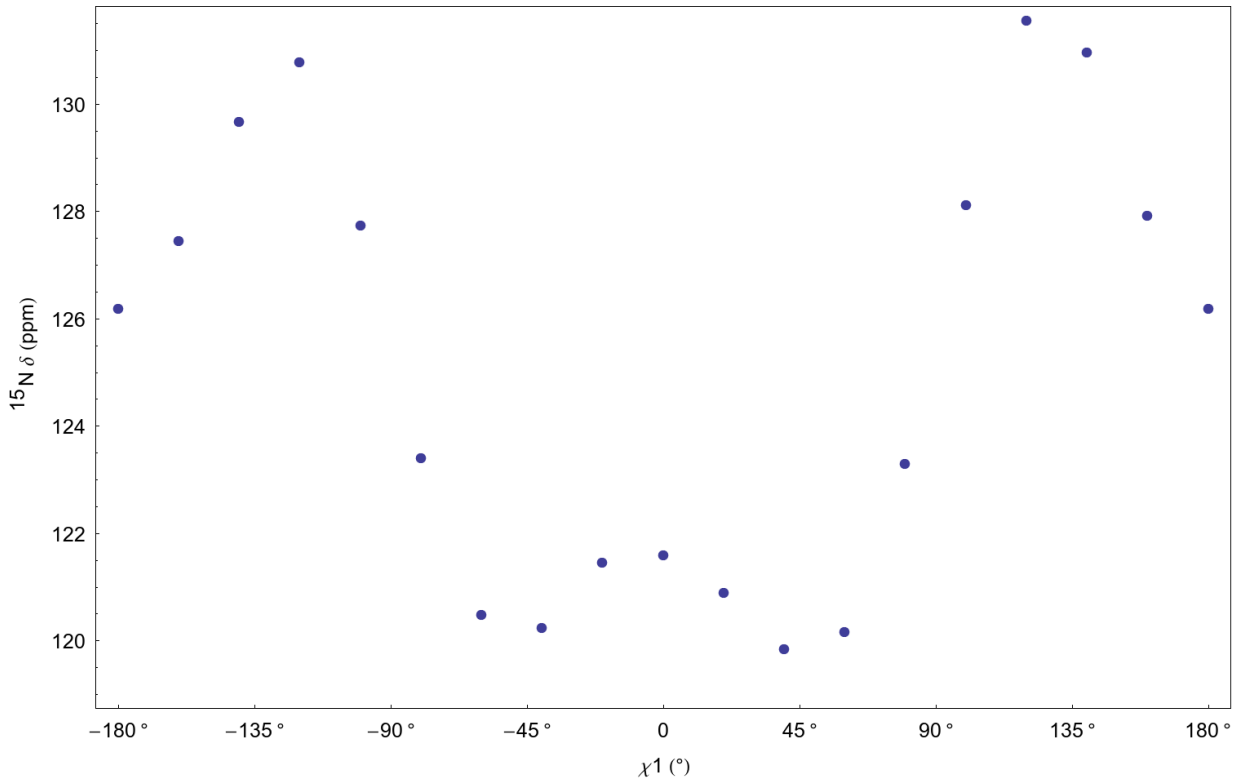


Figure J.23: Amide nitrogen chemical shift as a function of phenylalanine χ1. DFT calculations as described in Section 11.3.

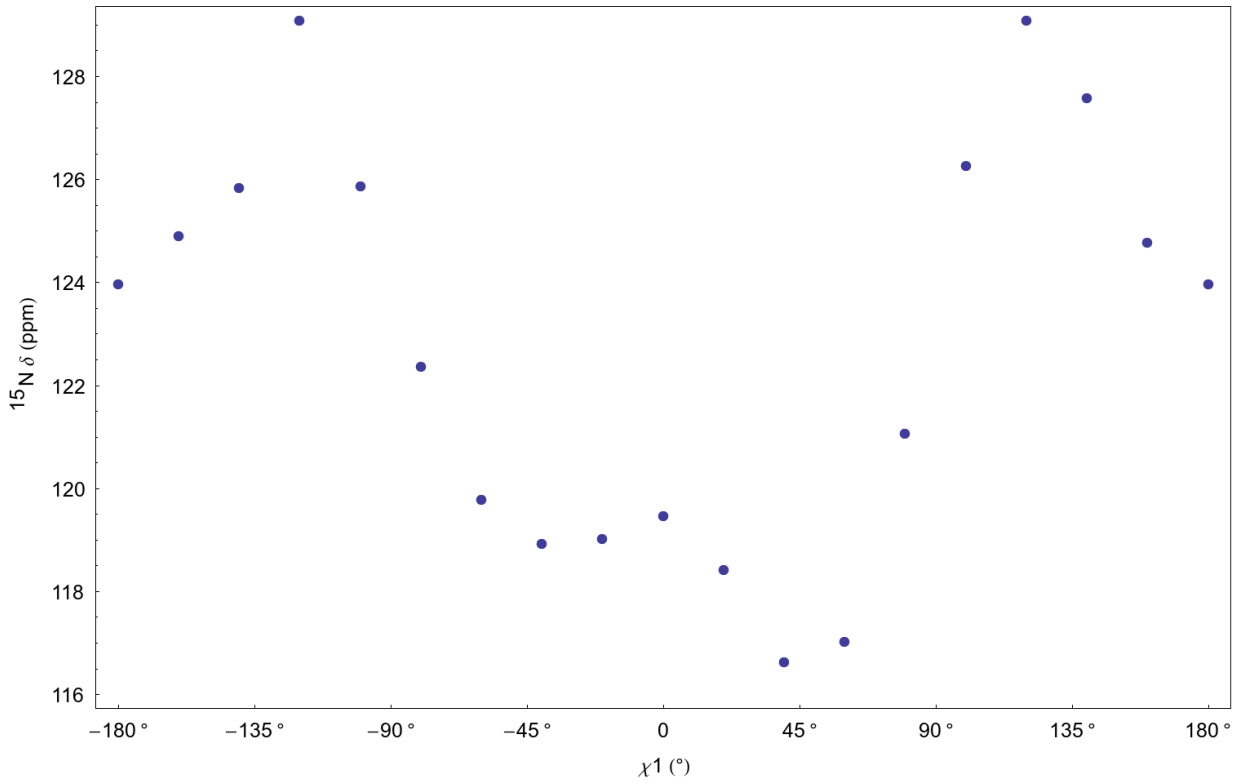


Figure J.24: Amide nitrogen chemical shift as a function of histidine χ1. DFT calculations as described in Section 11.3.

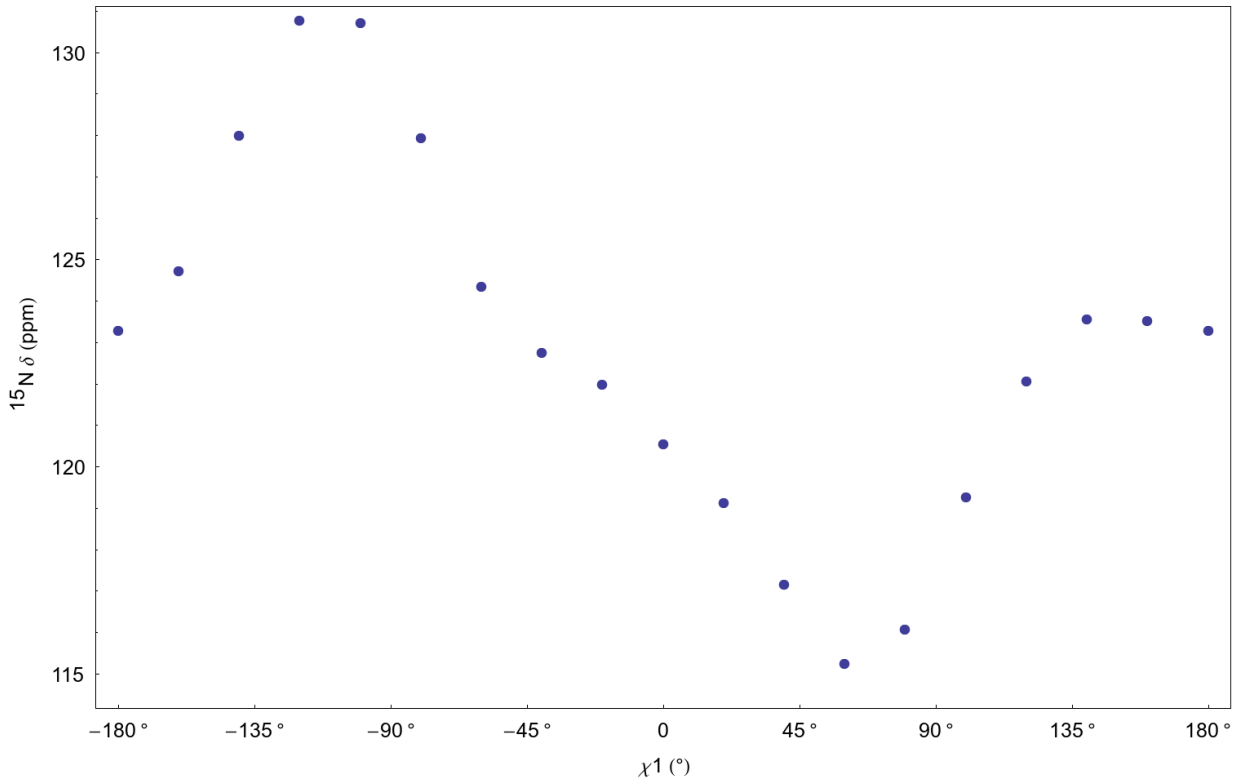


Figure J.25: Amide nitrogen chemical shift as a function of isoleucine $\chi 1$. DFT calculations as described in Section 11.3.

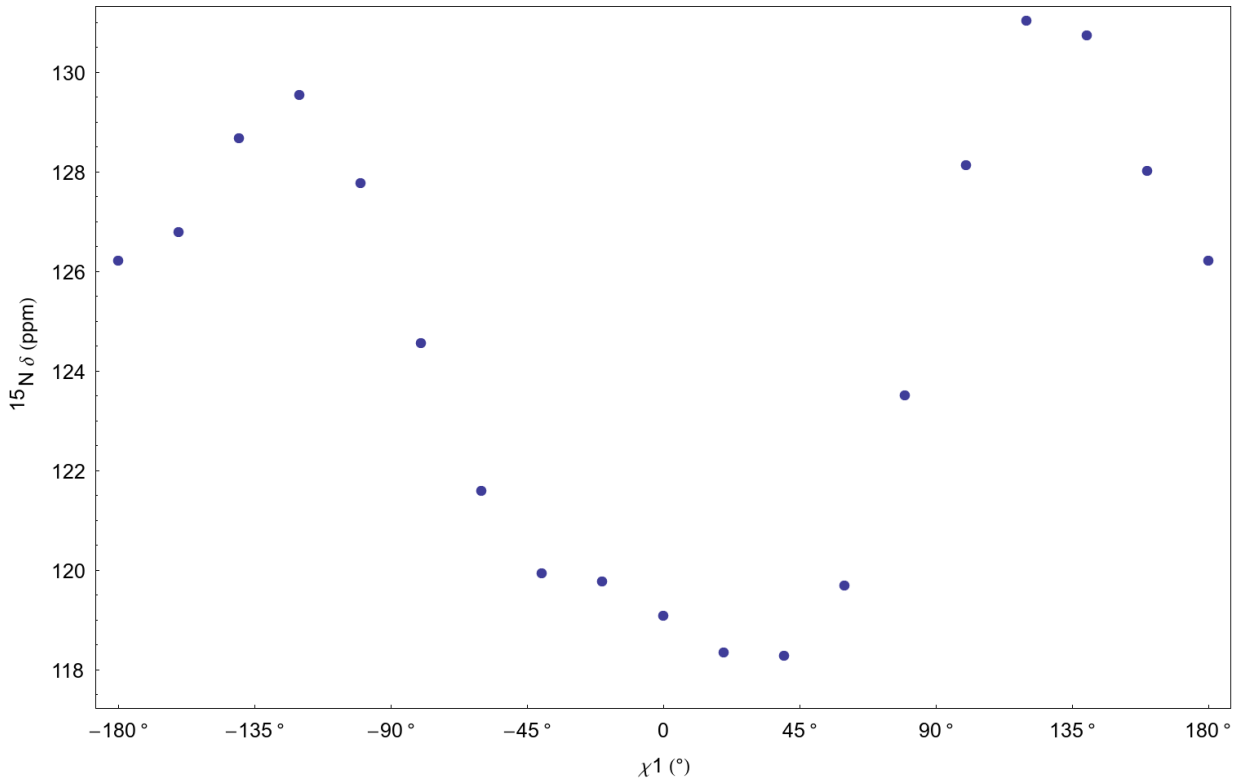


Figure J.26: Amide nitrogen chemical shift as a function of lysine χ1. DFT calculations as described in Section 11.3.

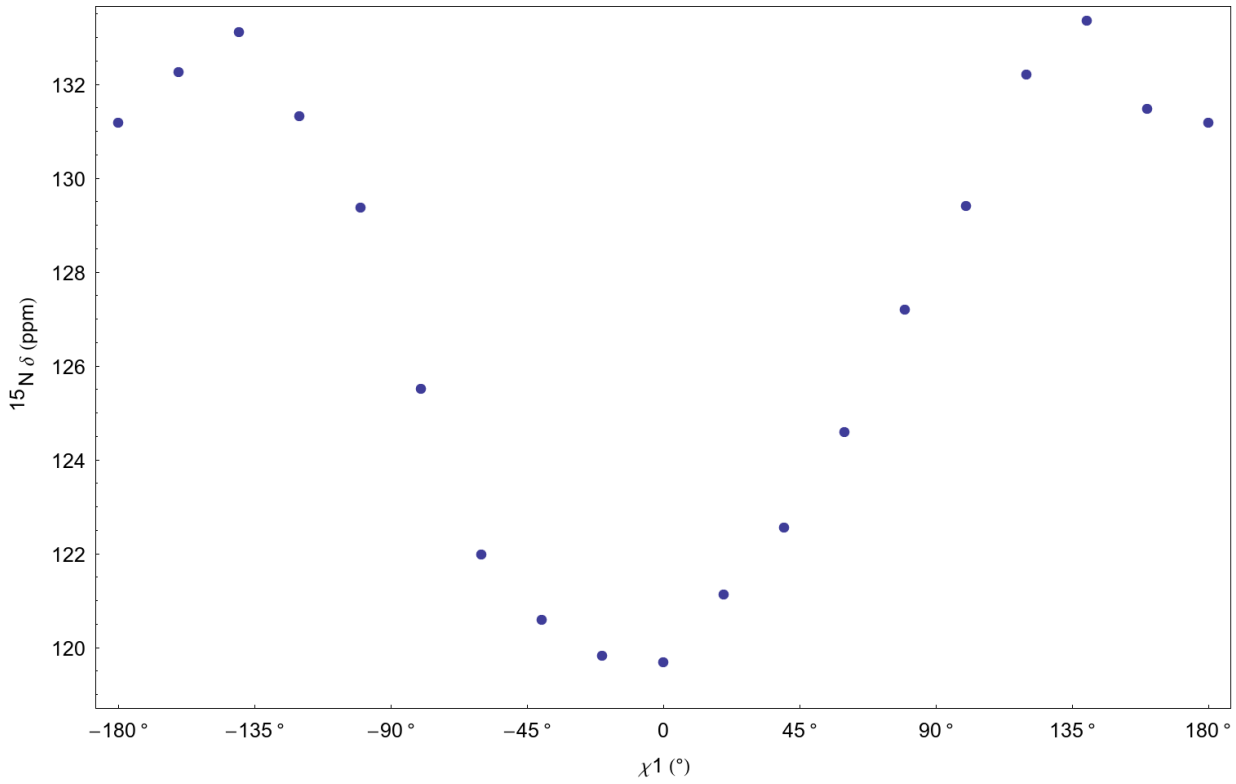


Figure J.27: Amide nitrogen chemical shift as a function of leucine $\chi 1$. DFT calculations as described in Section 11.3.

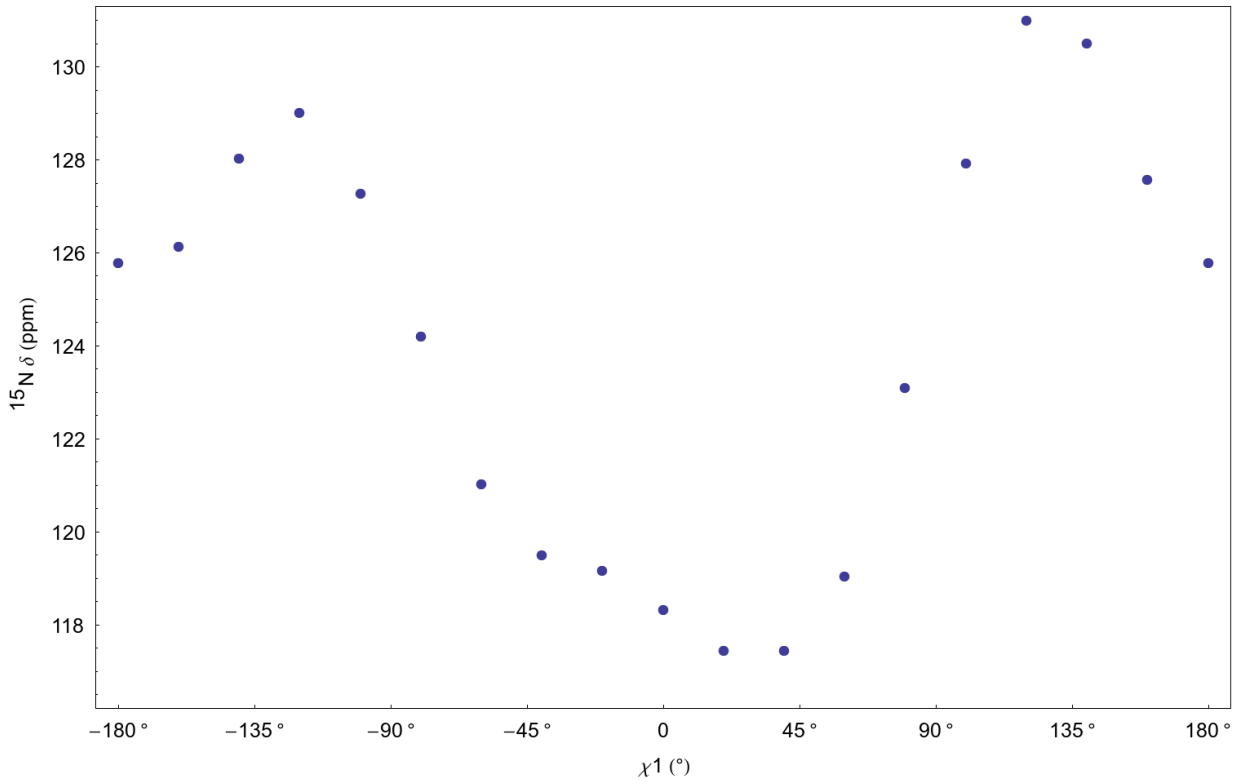


Figure J.28: Amide nitrogen chemical shift as a function of methionine $\chi 1$. DFT calculations as described in Section 11.3.

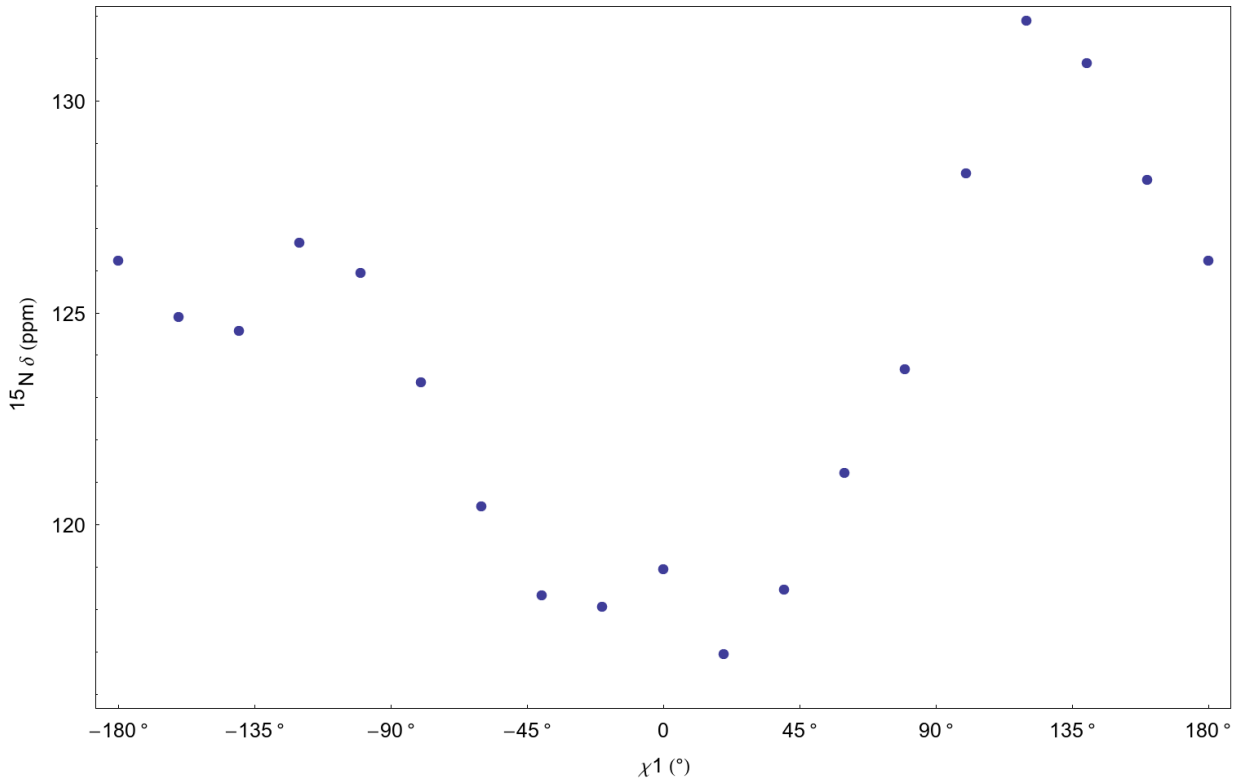


Figure J.29: Amide nitrogen chemical shift as a function of asparagine $\chi 1$. DFT calculations as described in Section 11.3.

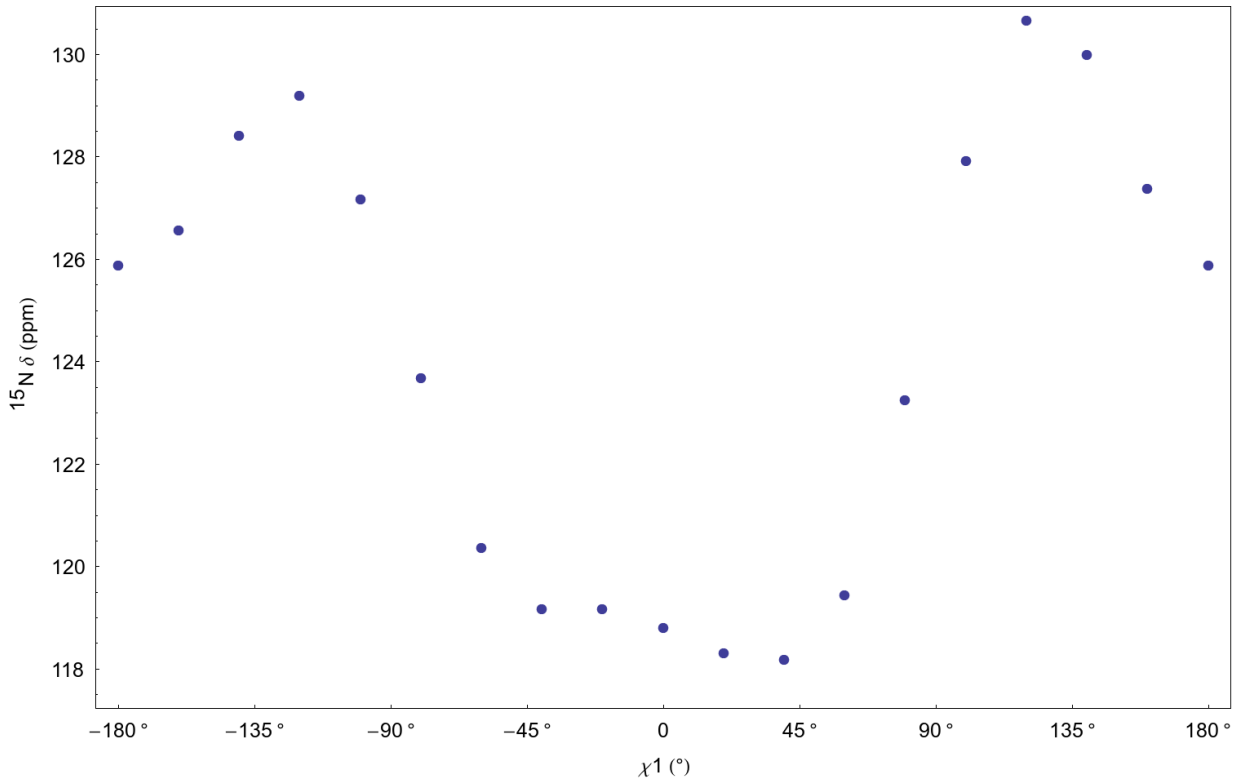


Figure J.30: Amide nitrogen chemical shift as a function of glutamine χ1. DFT calculations as described in Section 11.3.

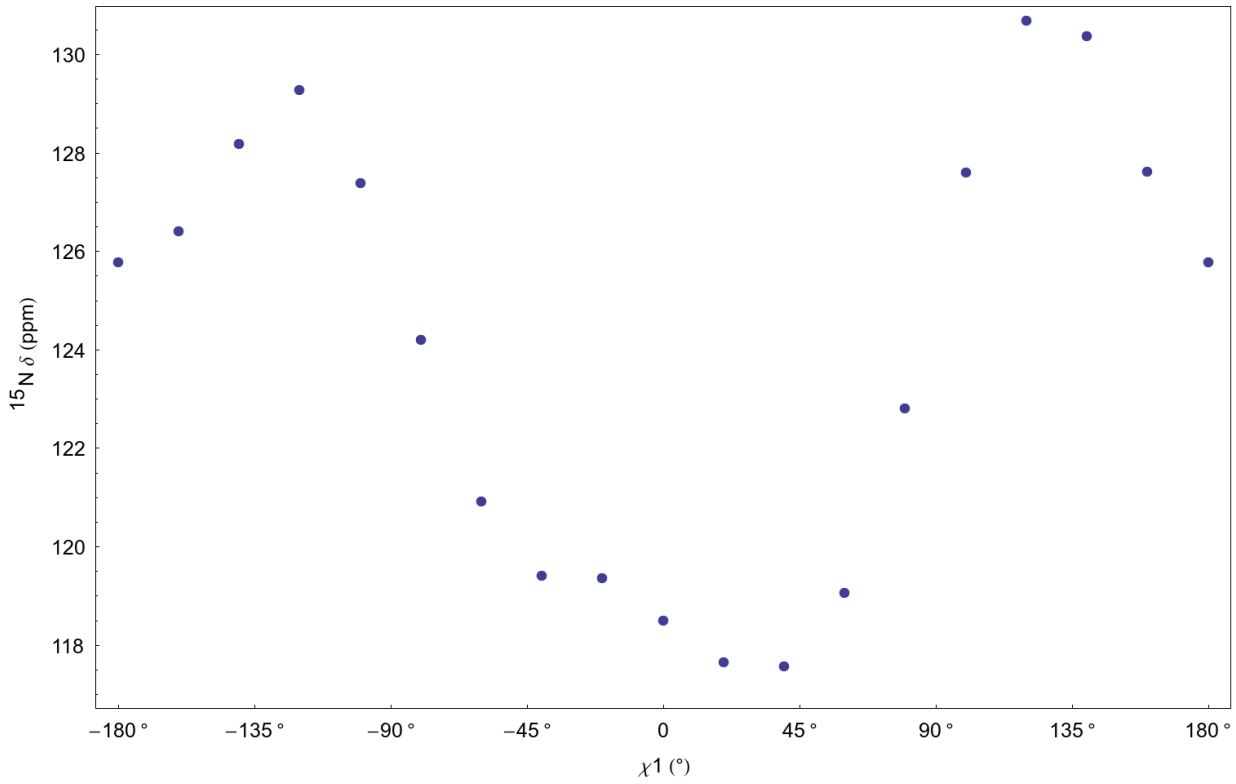


Figure J.31: Amide nitrogen chemical shift as a function of arginine χ1. DFT calculations as described in Section 11.3.

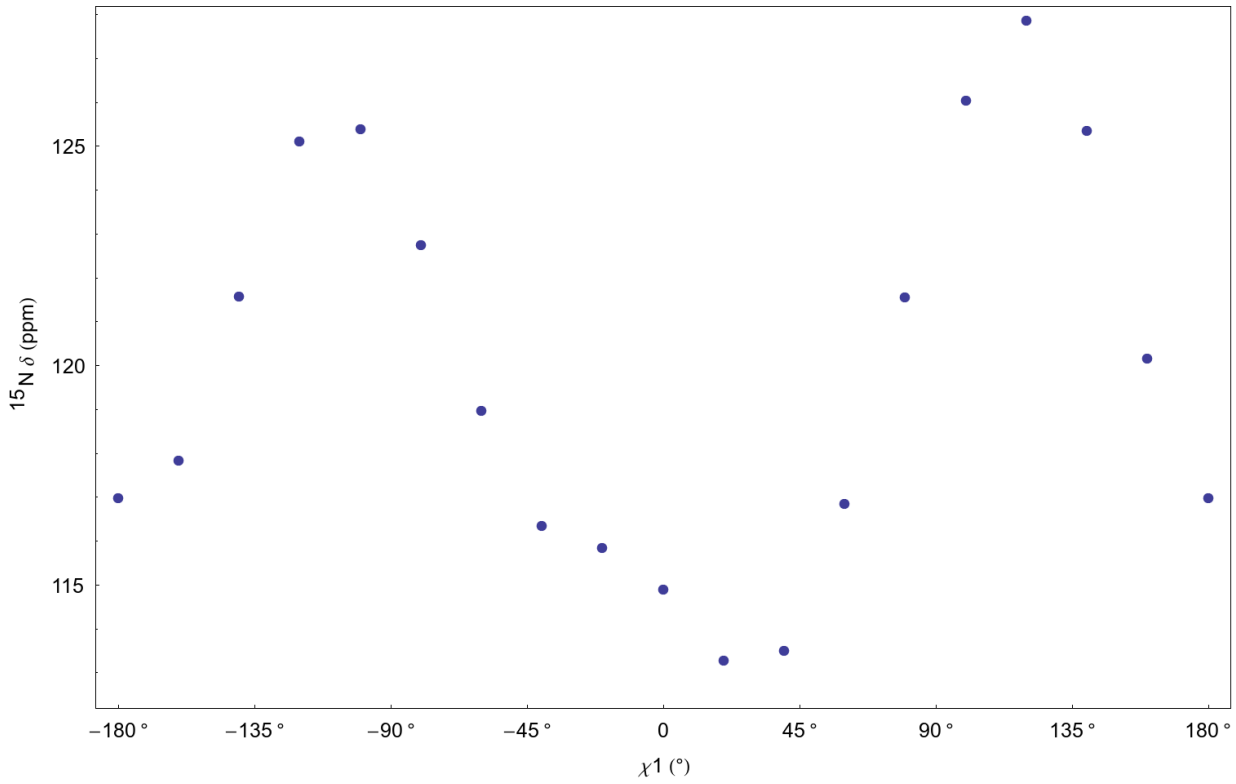


Figure J.32: Amide nitrogen chemical shift as a function of serine χ1. DFT calculations as described in Section 11.3.

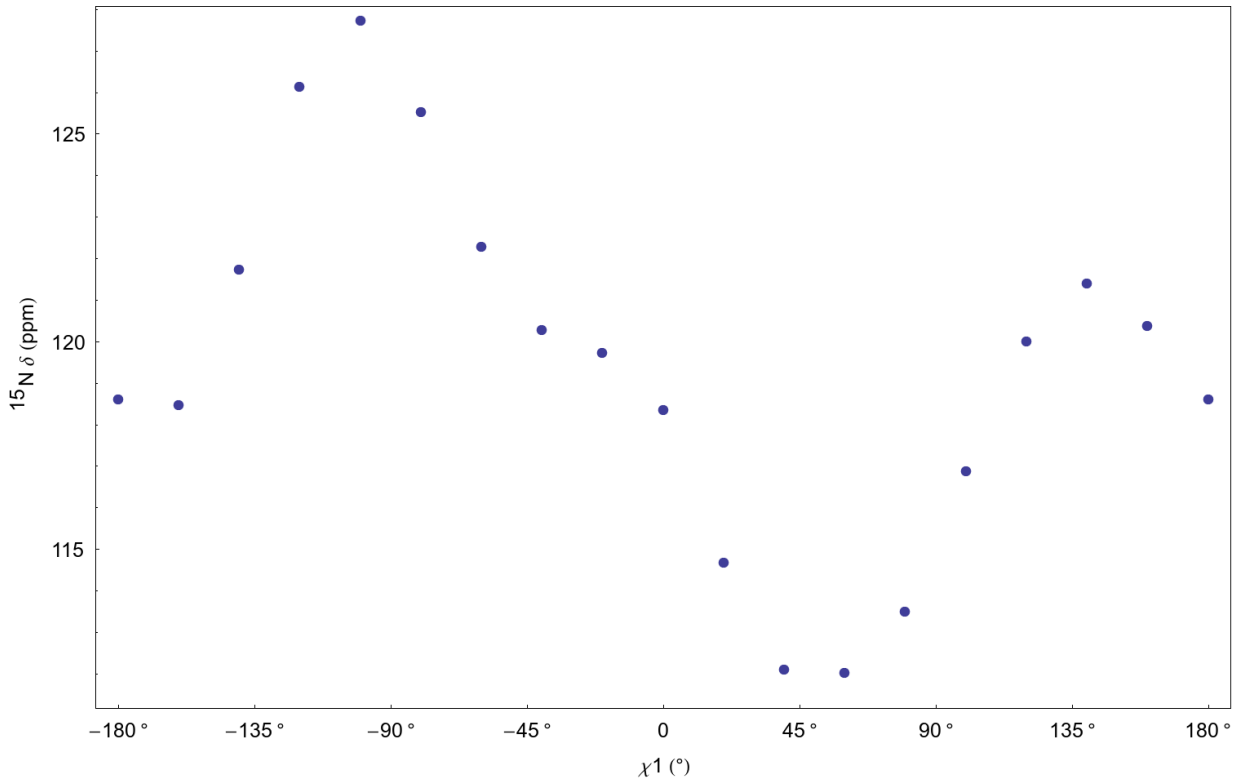


Figure J.33: Amide nitrogen chemical shift as a function of threonine $\chi 1$. DFT calculations as described in Section 11.3.

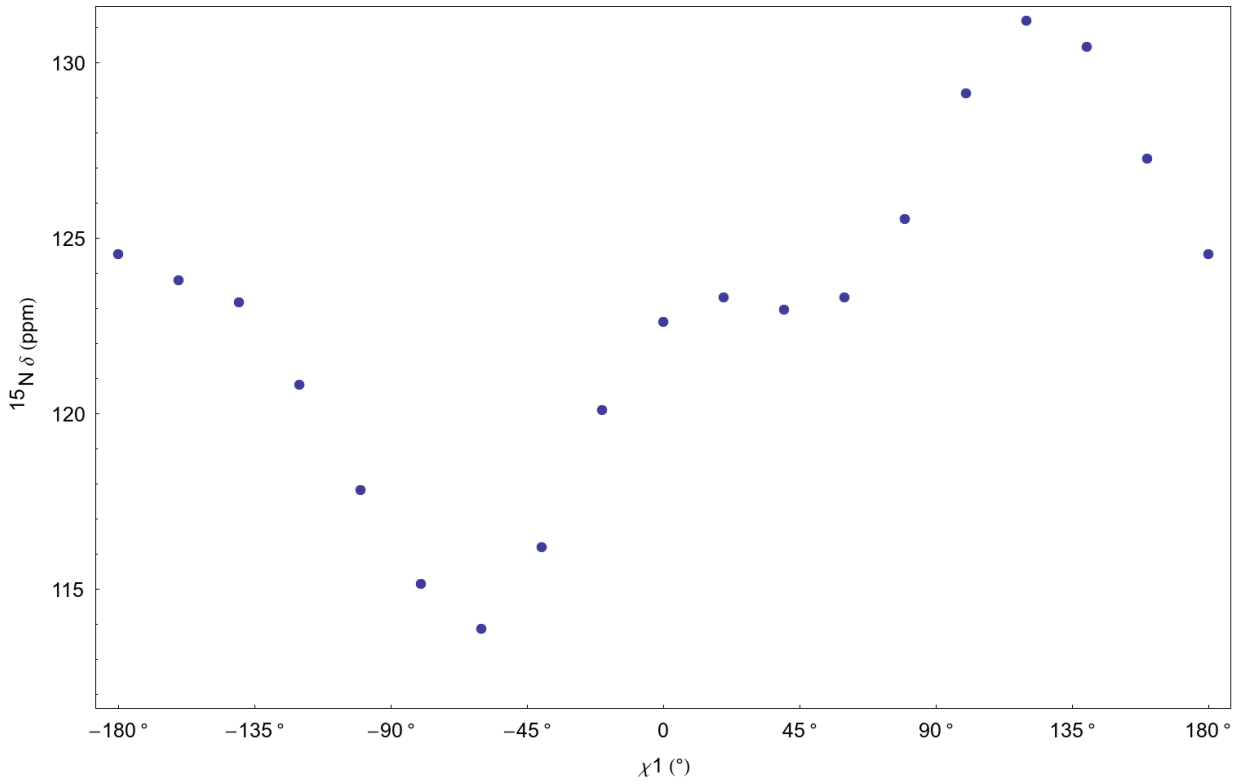


Figure J.34: Amide nitrogen chemical shift as a function of valine $\chi 1$. DFT calculations as described in Section 11.3.

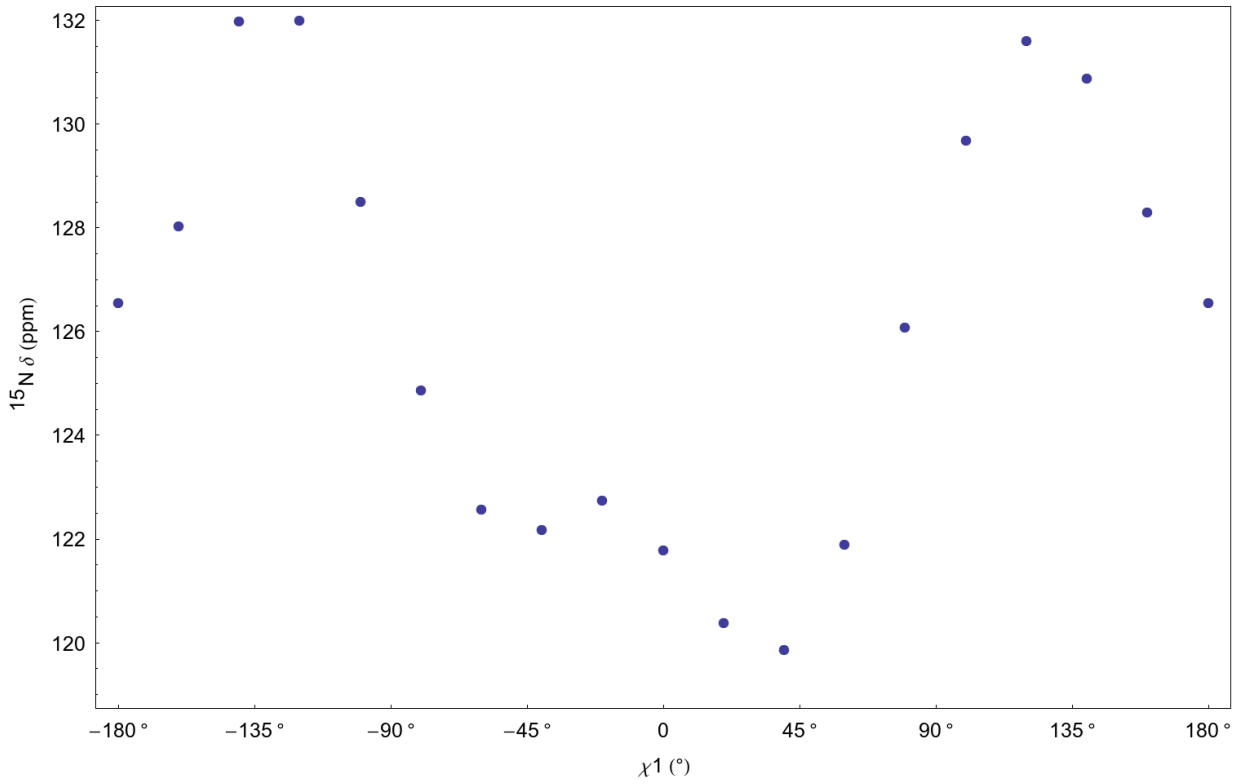


Figure J.35: Amide nitrogen chemical shift as a function of tryptophan $\chi 1$. DFT calculations as described in Section 11.3.

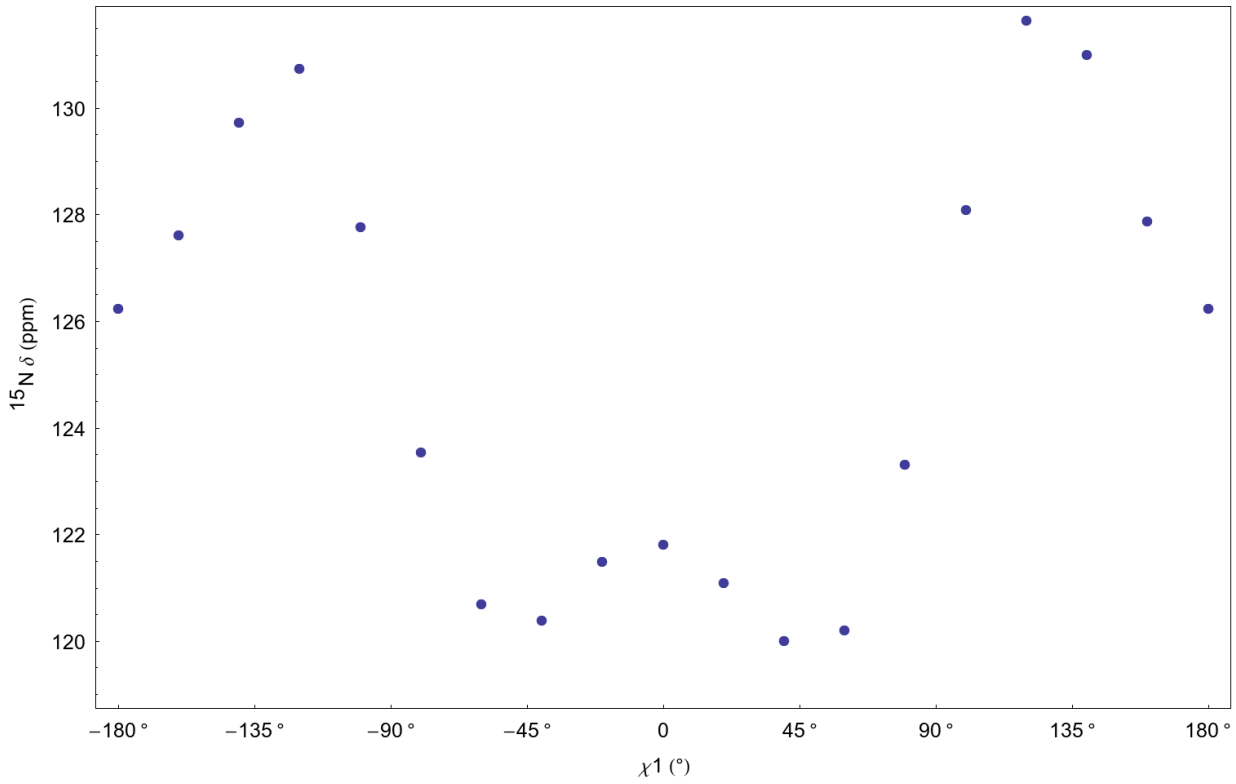


Figure J.36: Amide nitrogen chemical shift as a function of tyrosine χ1. DFT calculations as described in Section 11.3.

J.3 pWT Adnectin Amide Nitrogen Chemical Shifts & Dihedral Histograms

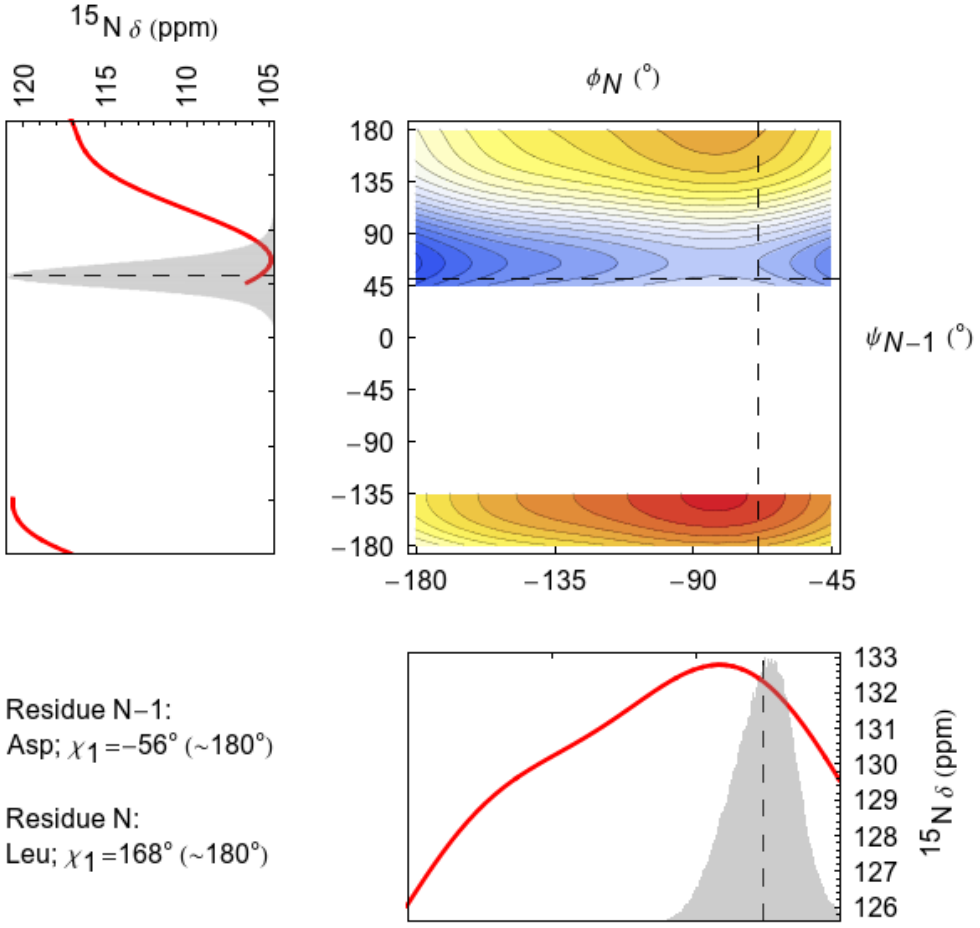


Figure J.37: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues D7 and L8 (i.e., N=8) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

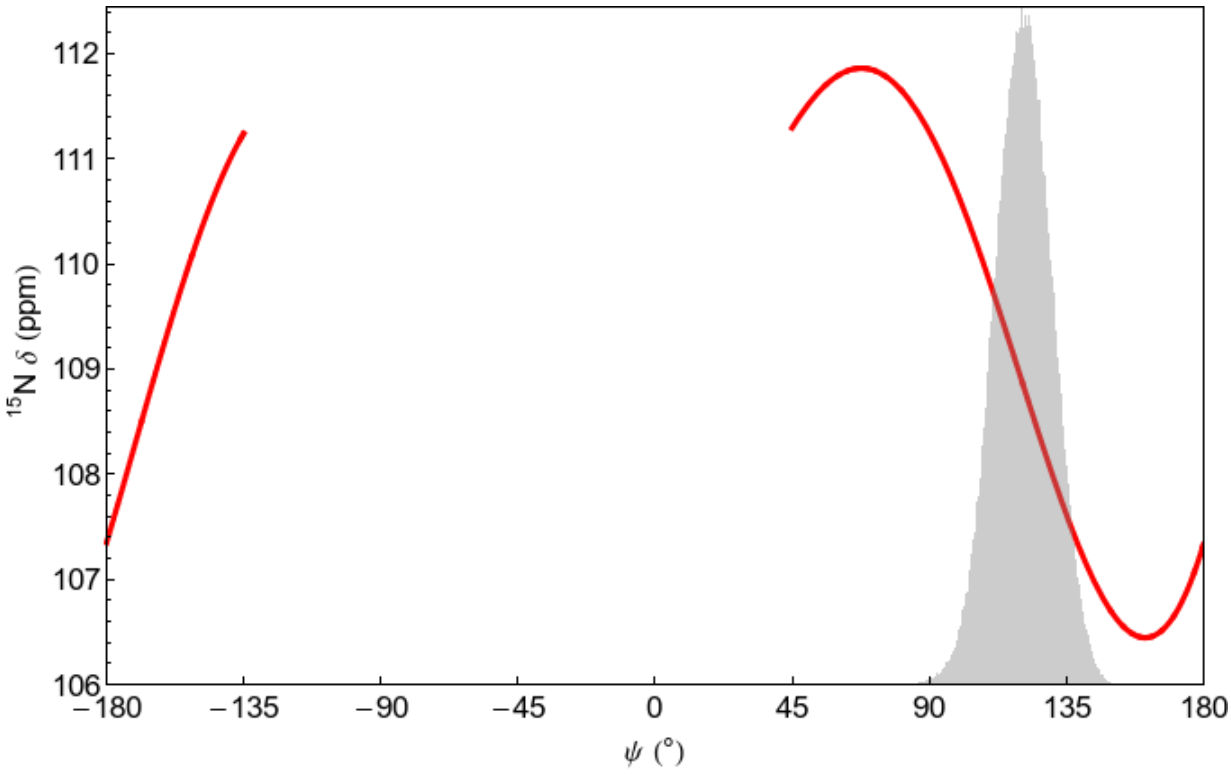


Figure J.38: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue L8 (i.e., N=8) in a 100 ns MD simulation.

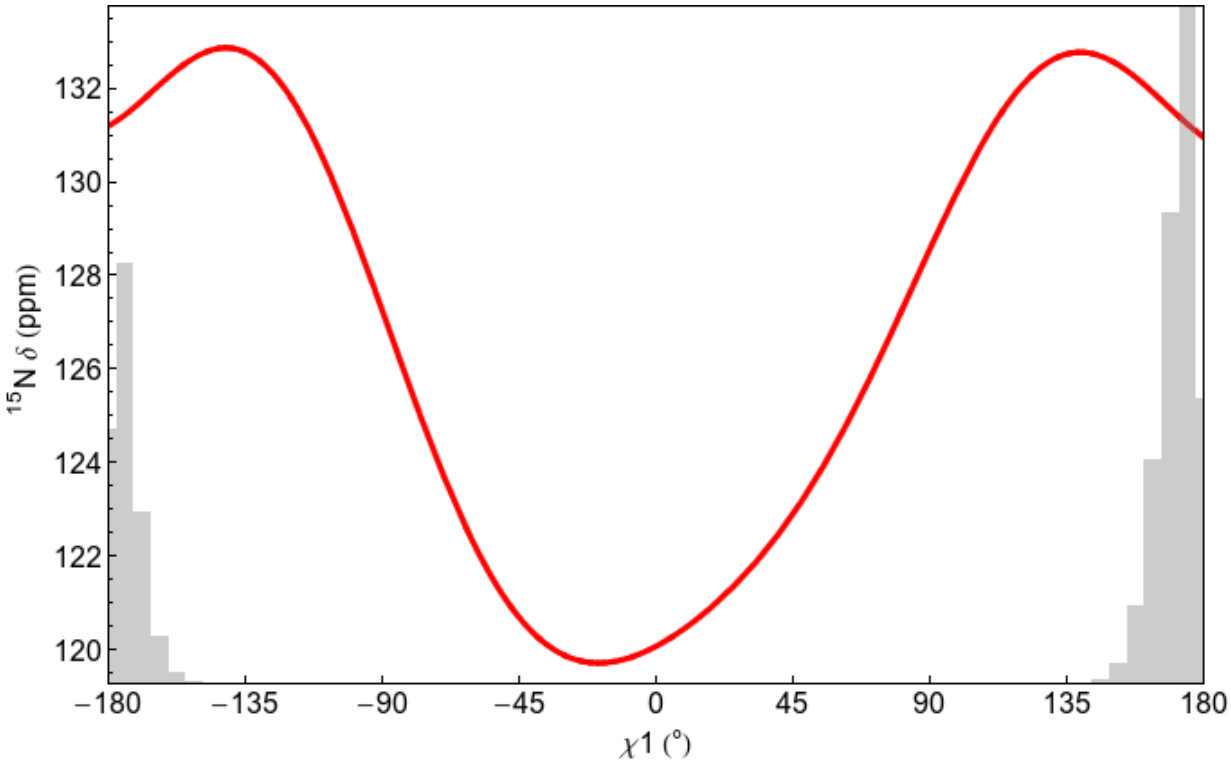


Figure J.39: A smooth curve (red) representing DFT calculations (Section J.2) of leucine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue L8 (i.e., N=8) in a 100 ns MD simulation.

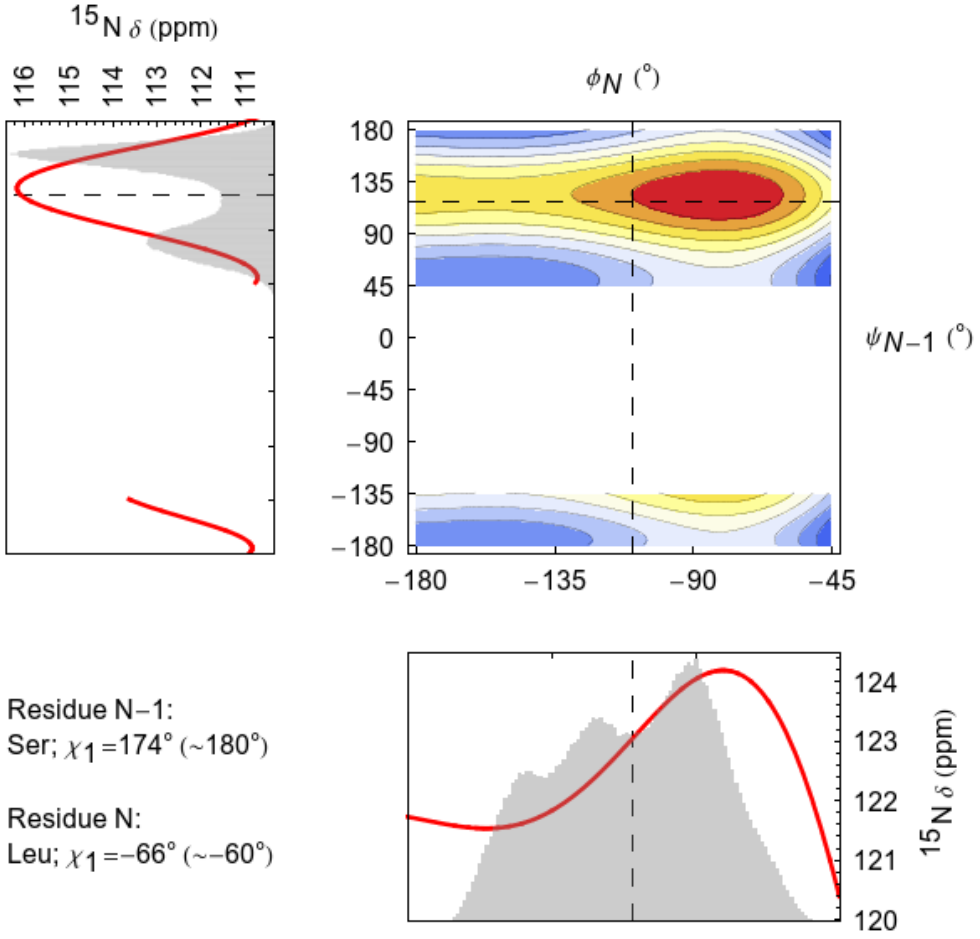


Figure J.40: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues S17 and L18 (i.e., N=18) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

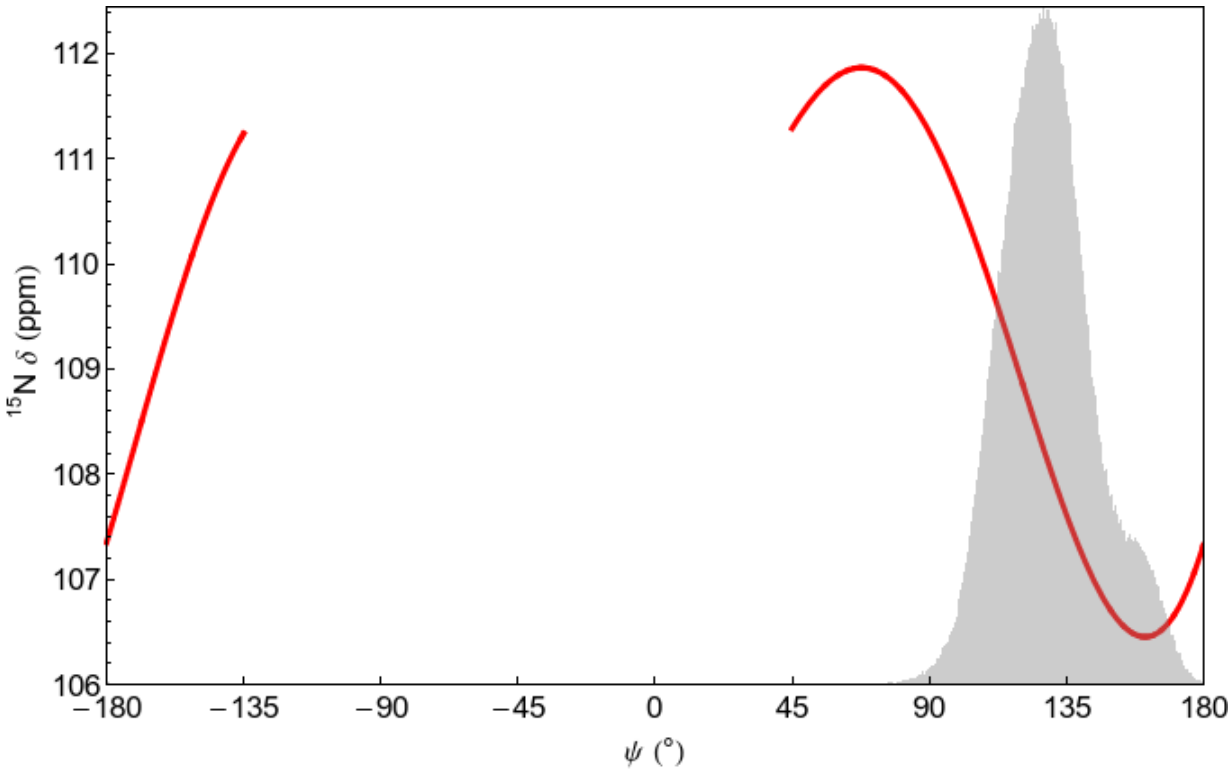


Figure J.41: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue L18 (i.e., N=18) in a 100 ns MD simulation.

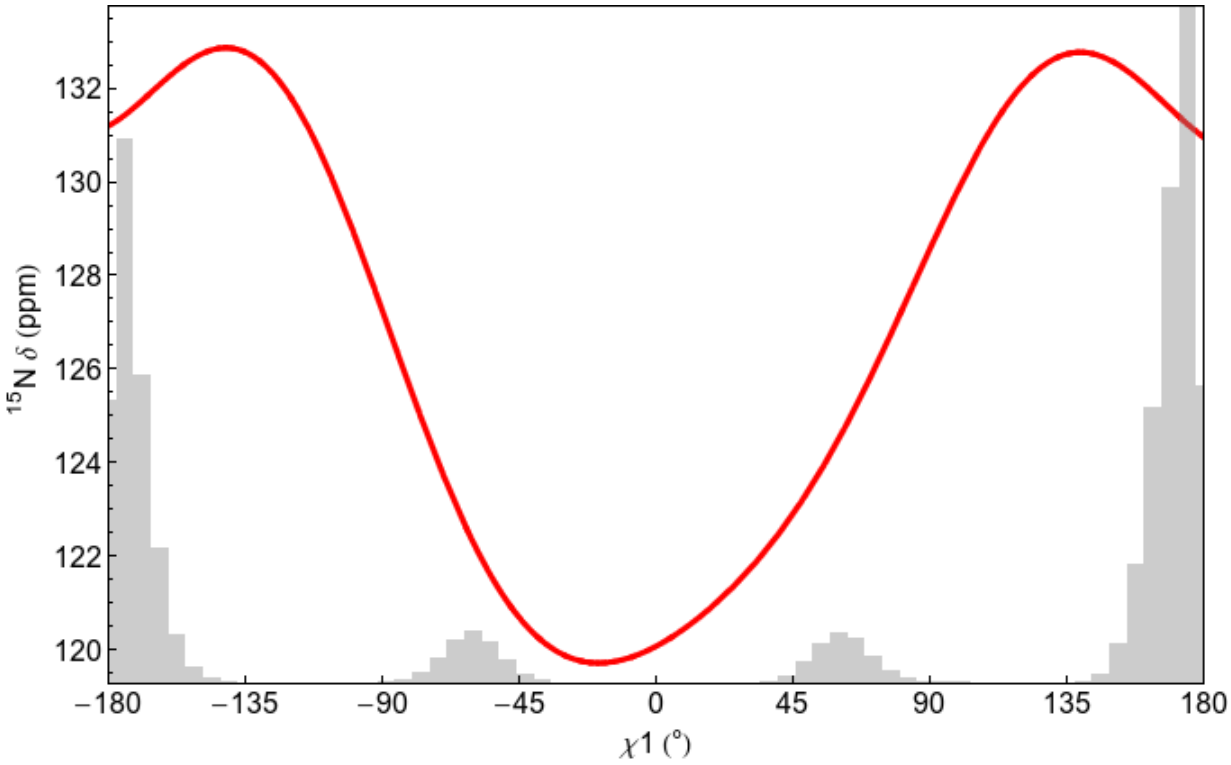


Figure J.42: A smooth curve (red) representing DFT calculations (Section J.2) of leucine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue L18 (i.e., N=18) in a 100 ns MD simulation.

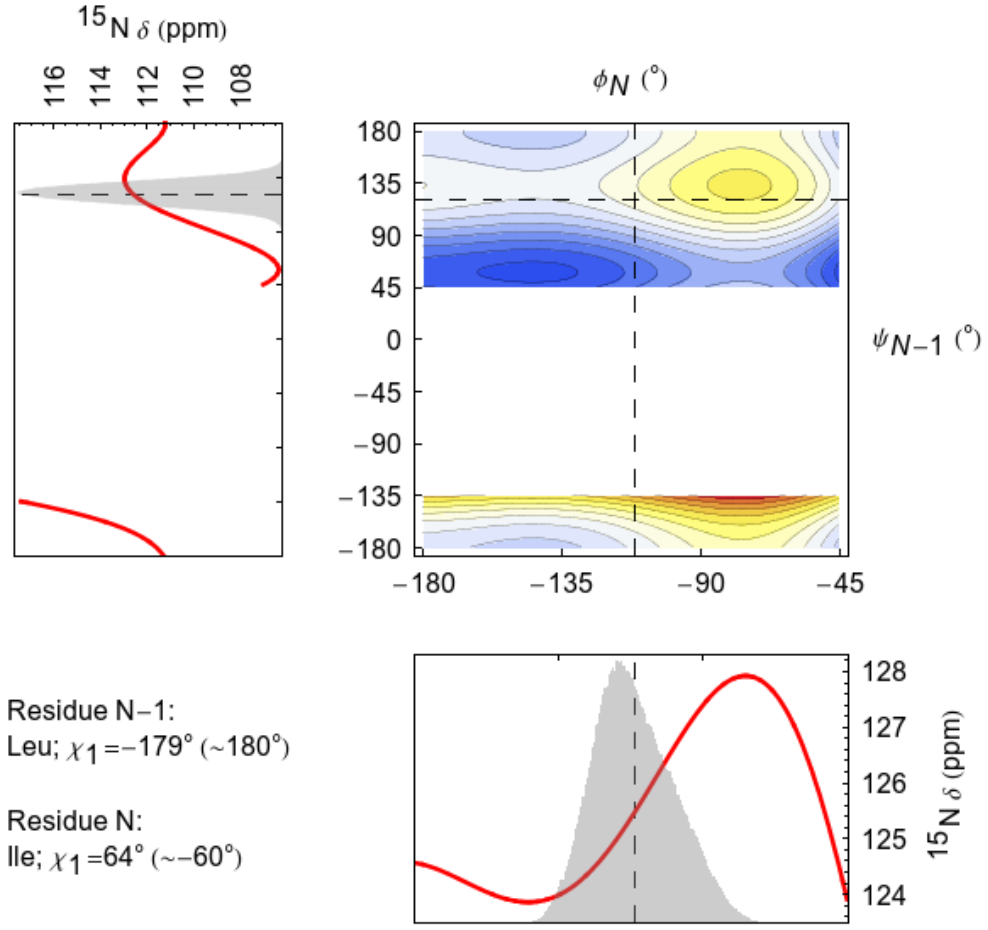


Figure J.43: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues L19 and I20 (i.e., N=20) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

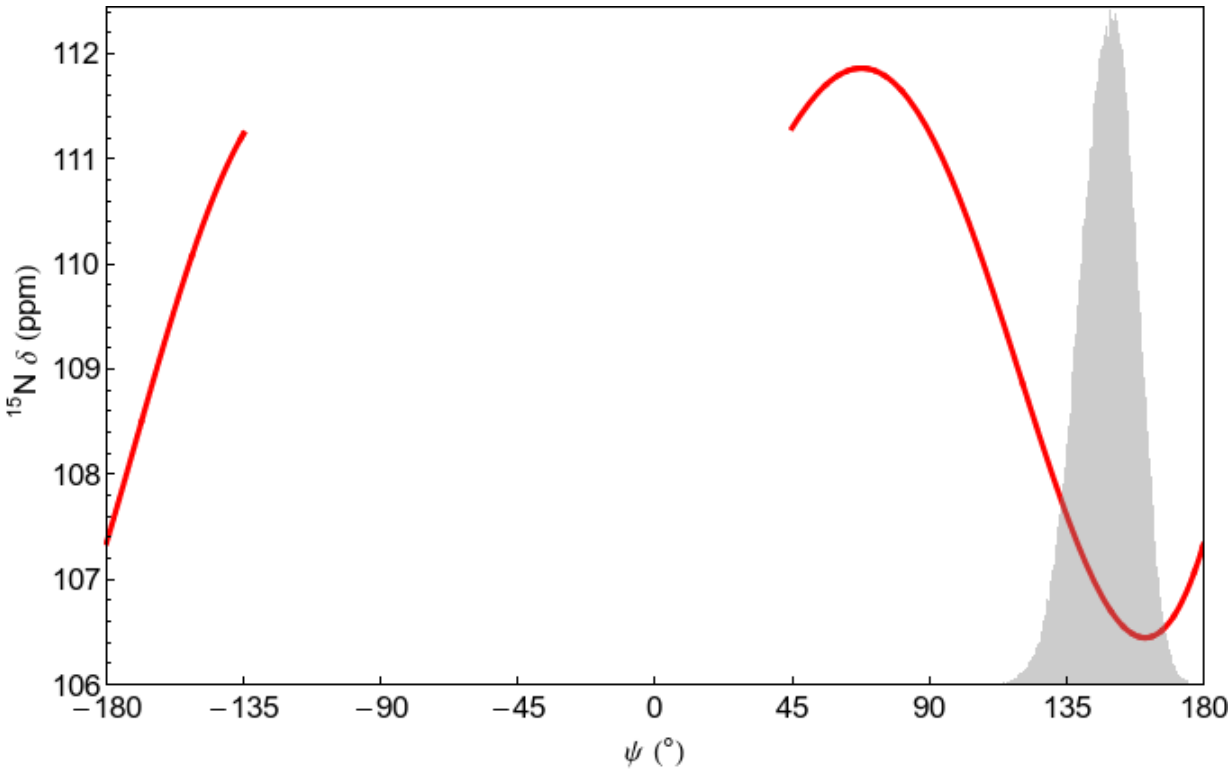


Figure J.44: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue I20 (i.e., N=20) in a 100 ns MD simulation.

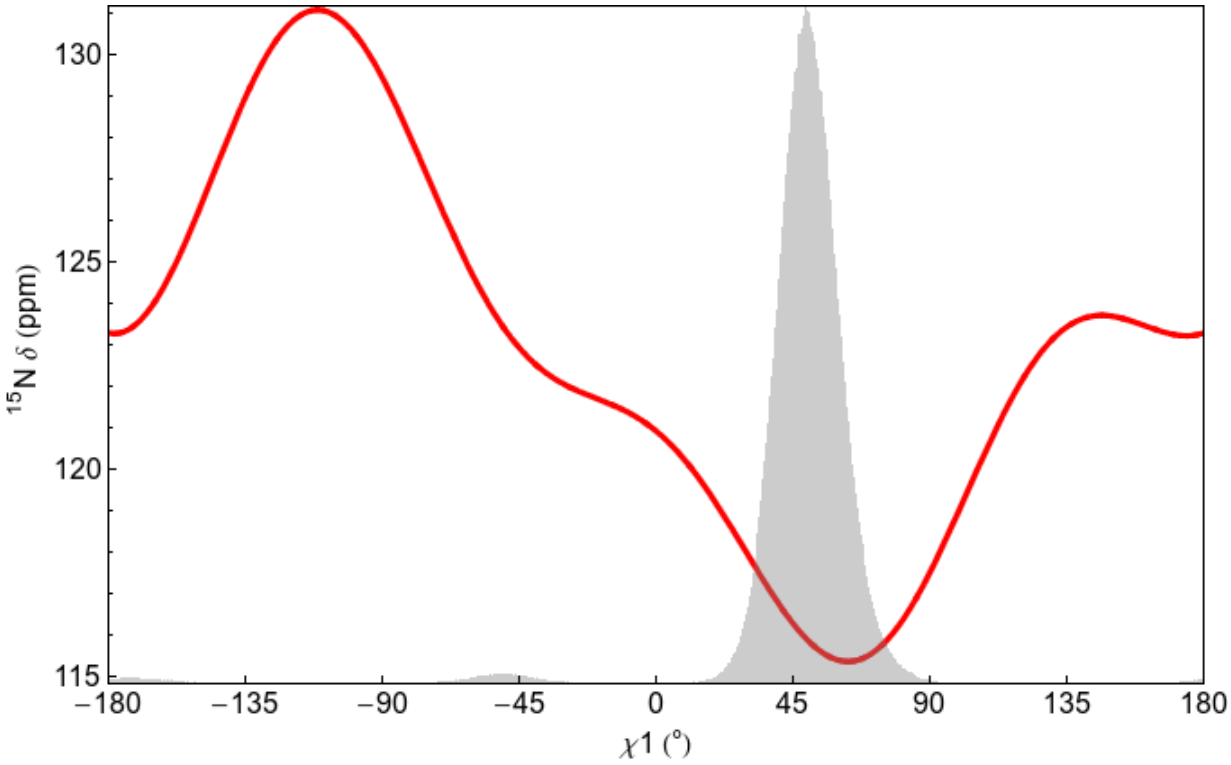


Figure J.45: A smooth curve (red) representing DFT calculations (Section J.2) of isoleucine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue I20 (i.e., N=20) in a 100 ns MD simulation.

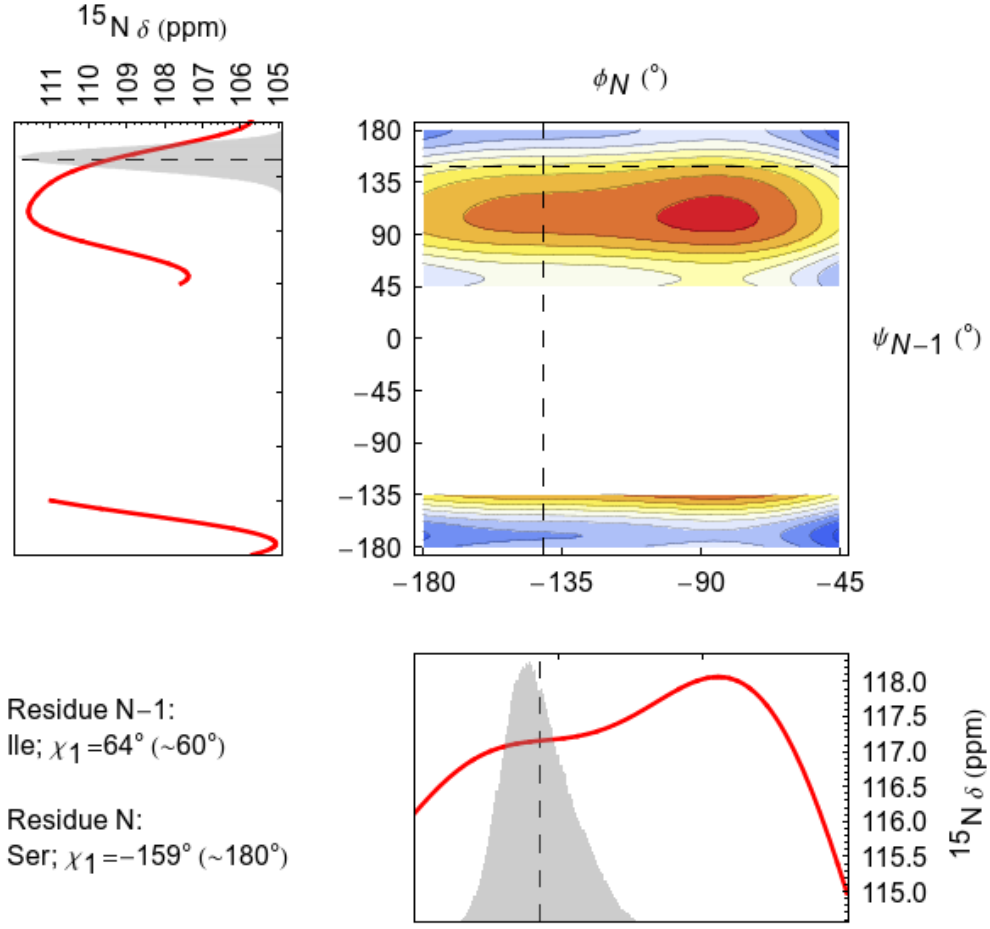


Figure J.46: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues I20 and S21 (i.e., N=21) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

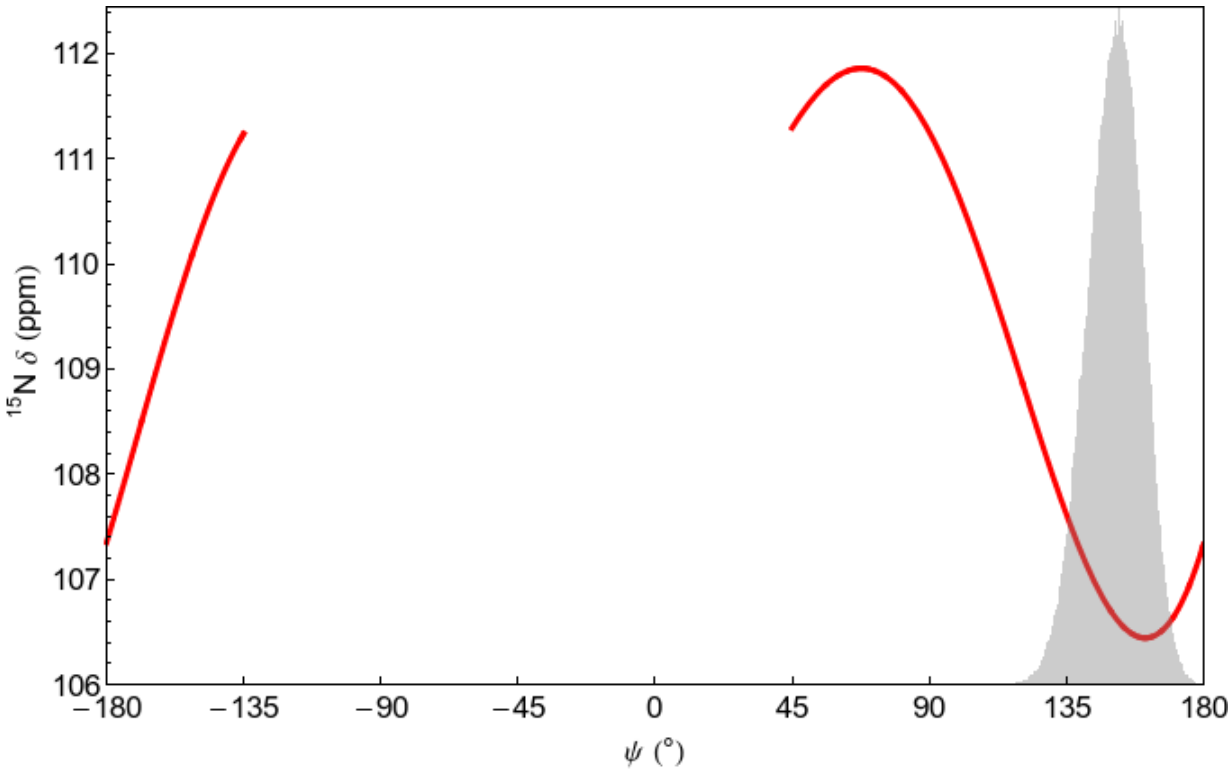


Figure J.47: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue S21 (i.e., N=21) in a 100 ns MD simulation.

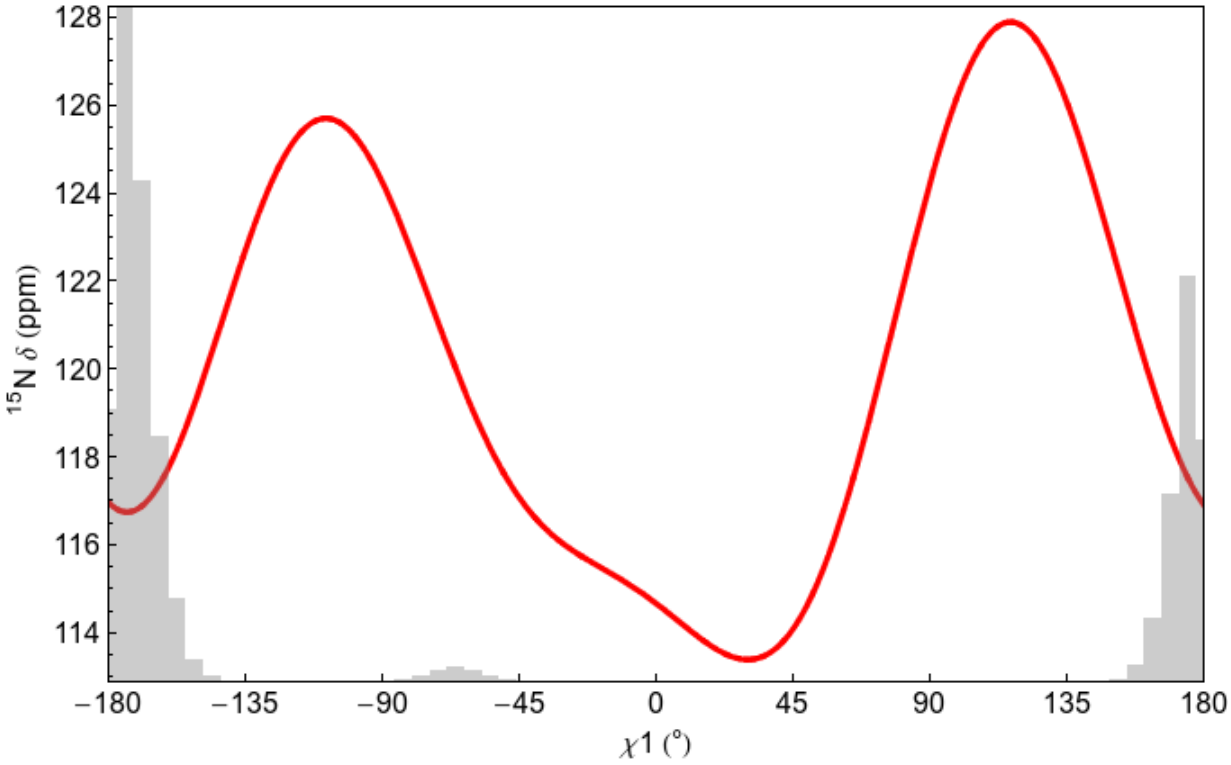


Figure J.48: A smooth curve (red) representing DFT calculations (Section J.2) of serine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue S21 (i.e., N=21) in a 100 ns MD simulation.

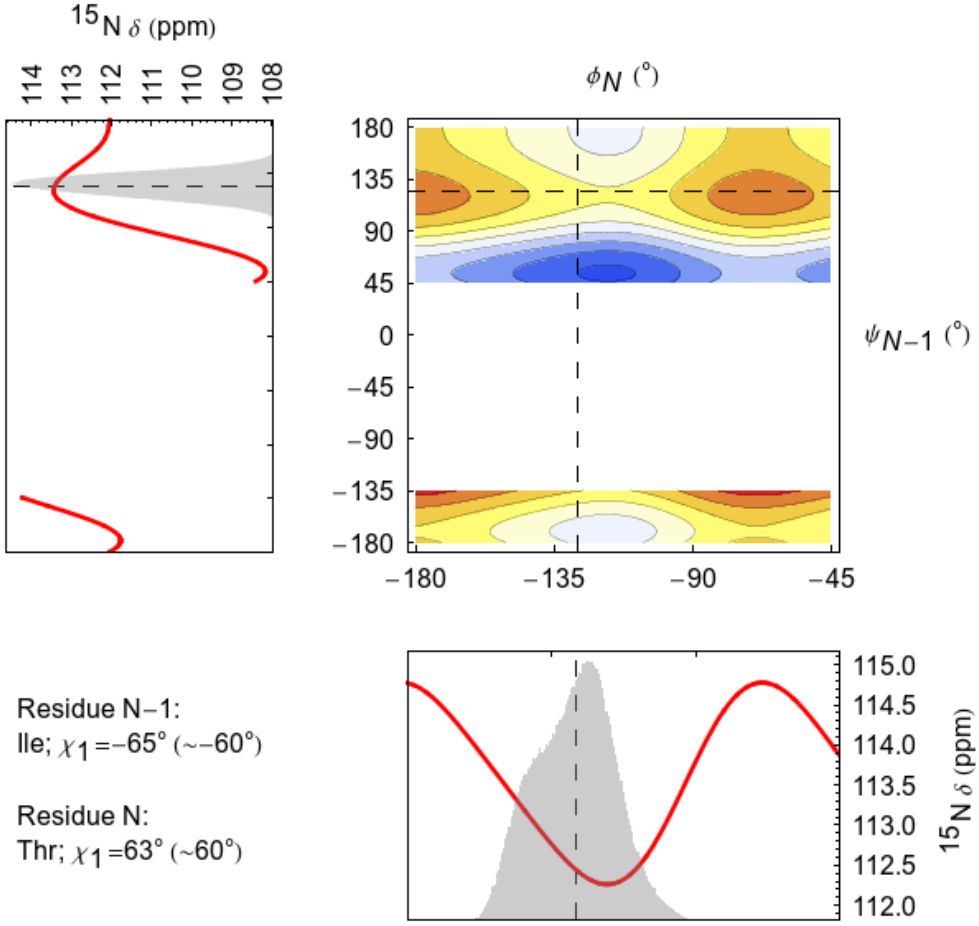


Figure J.49: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues I34 and T35 (i.e., N=35) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

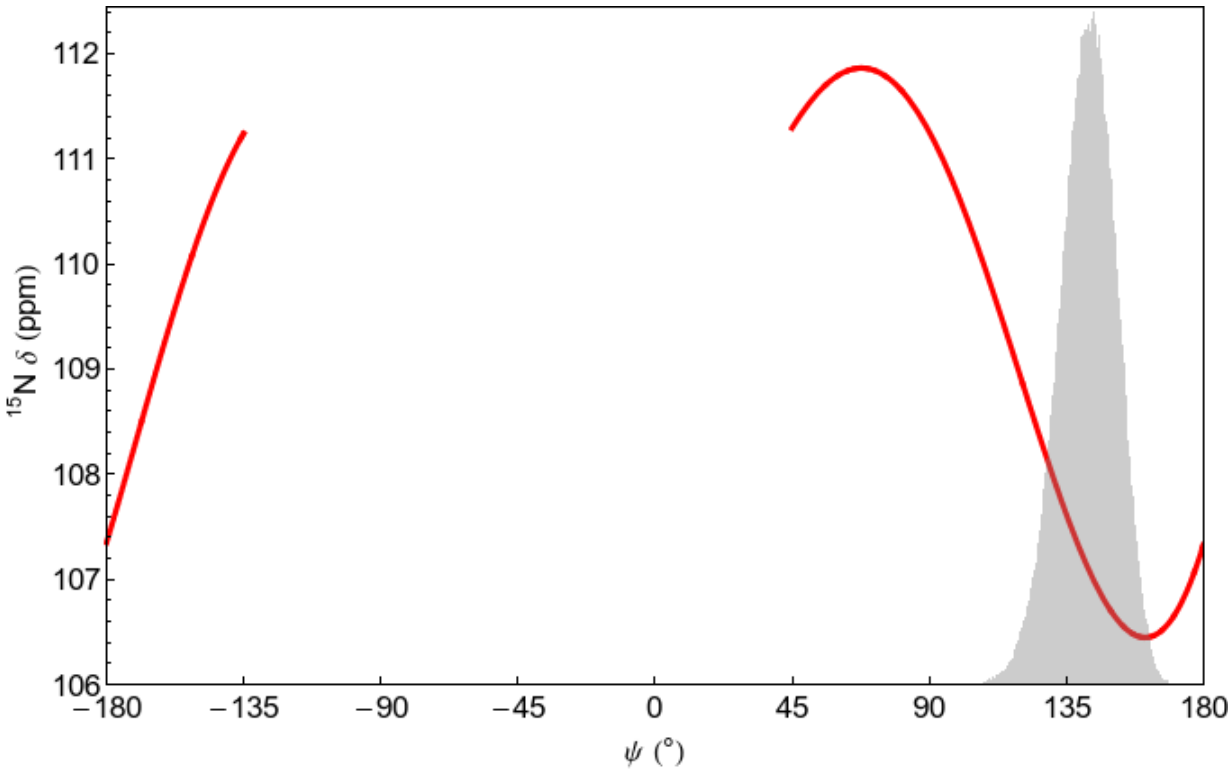


Figure J.50: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue T35 (i.e., N=35) in a 100 ns MD simulation.

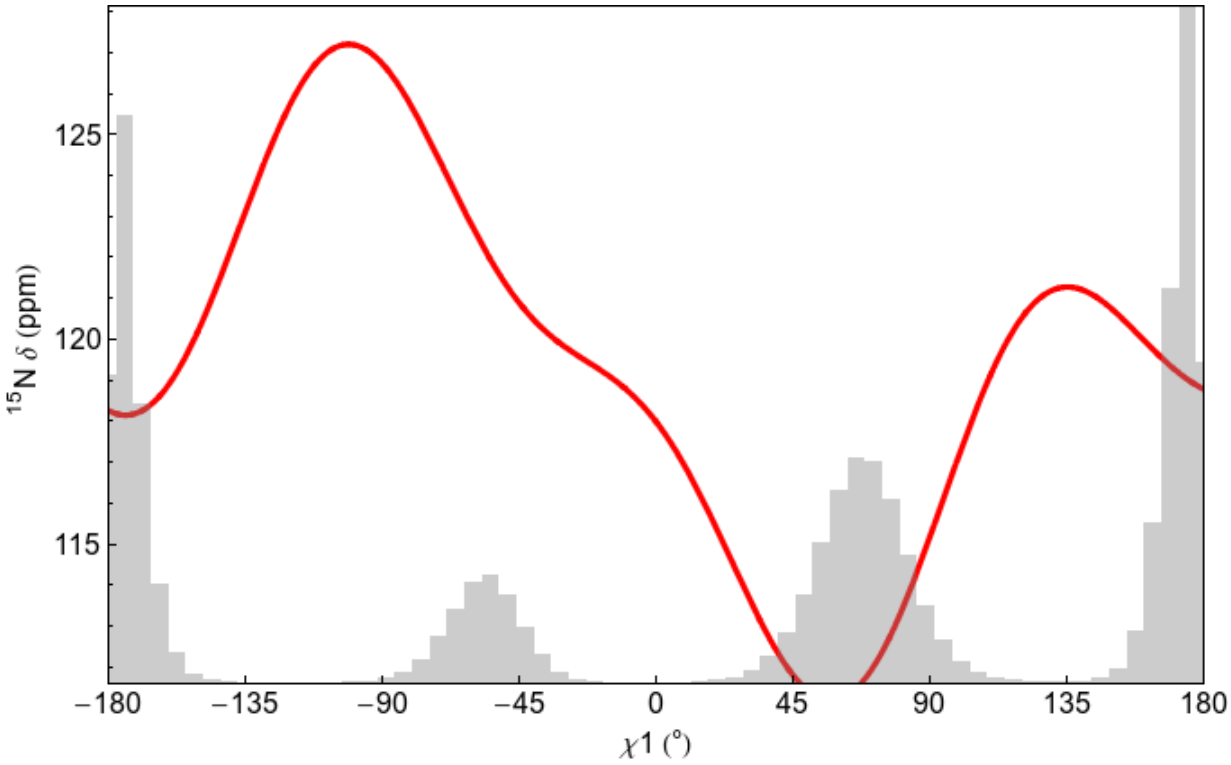


Figure J.51: A smooth curve (red) representing DFT calculations (Section J.2) of threonine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue T35 (i.e., N=35) in a 100 ns MD simulation.

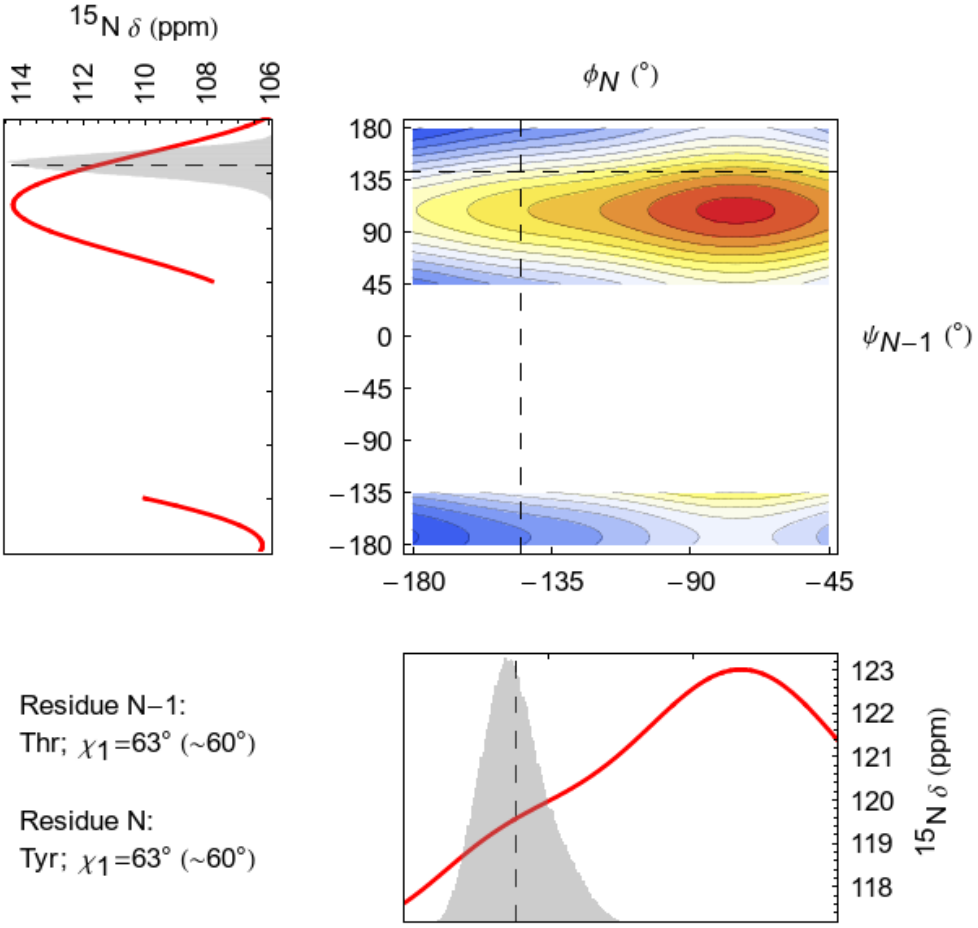


Figure J.52: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues T35 and Y36 (i.e., N=36) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

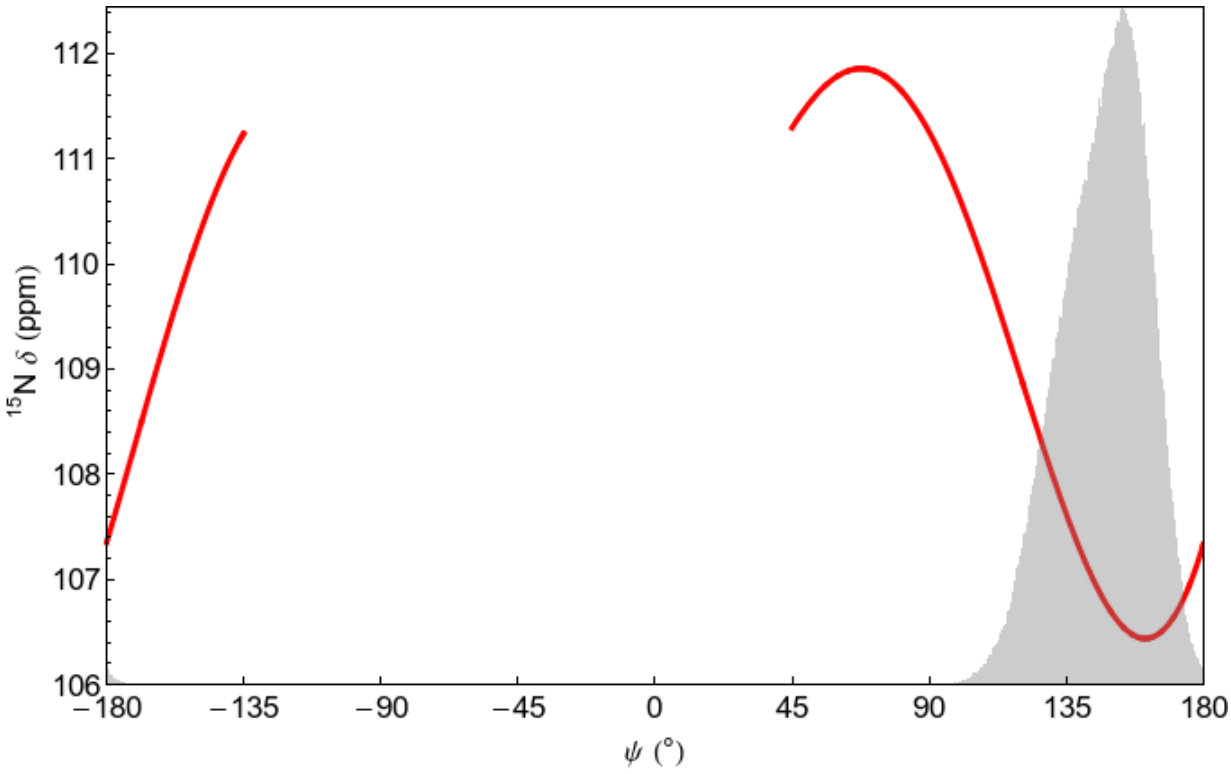


Figure J.53: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue Y36 (i.e., N=36) in a 100 ns MD simulation.

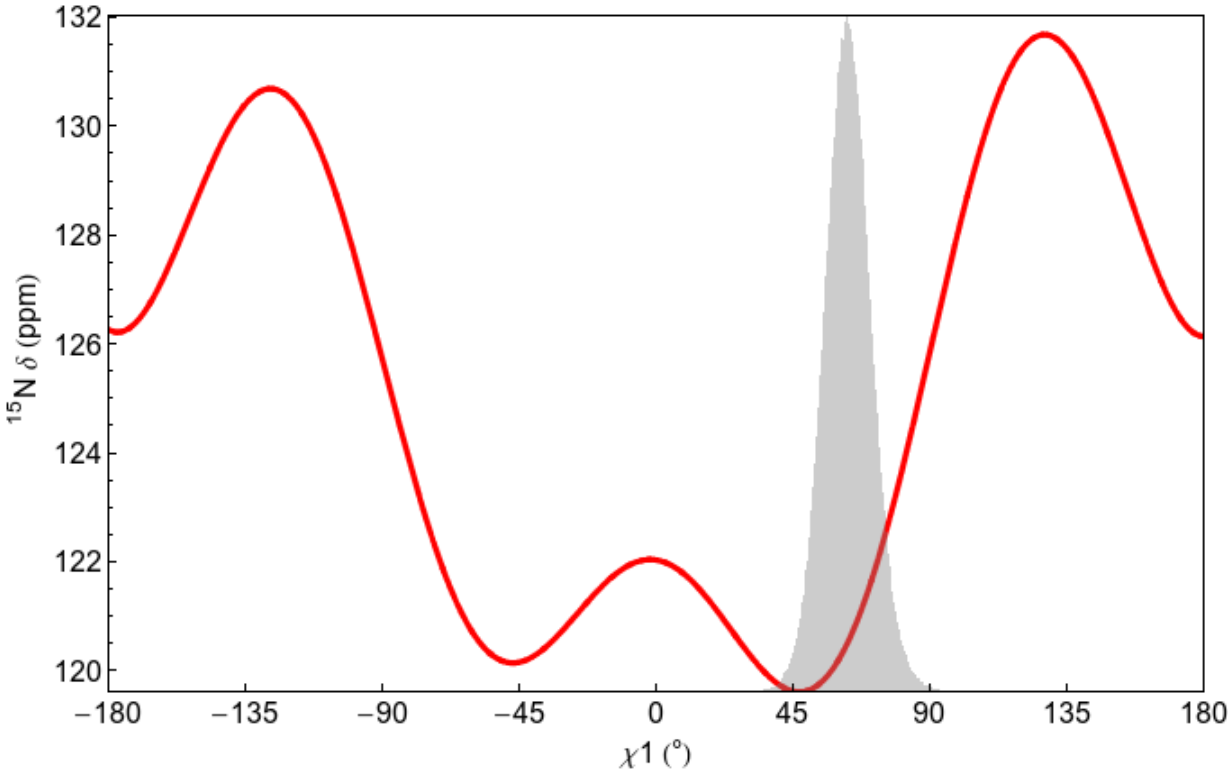


Figure J.54: A smooth curve (red) representing DFT calculations (Section J.2) of tyrosine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue Y36 (i.e., N=36) in a 100 ns MD simulation.

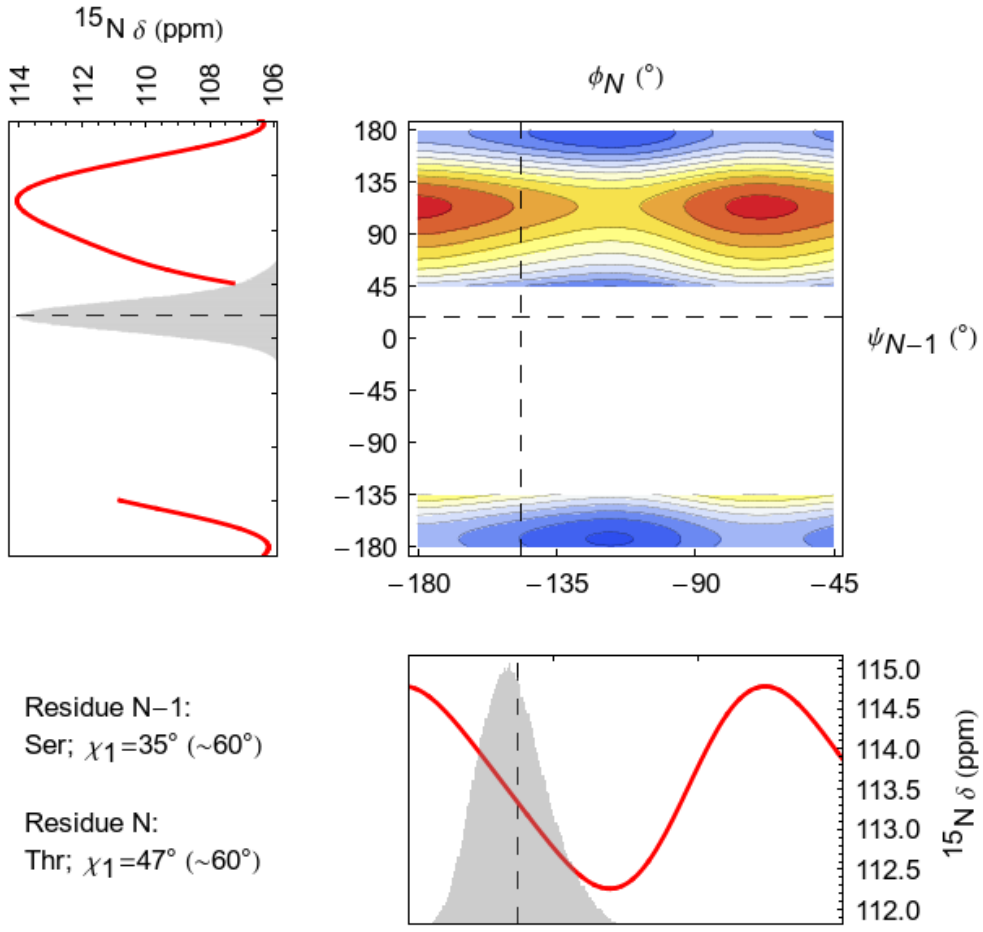


Figure J.55: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues S55 and T56 (i.e., N=56) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

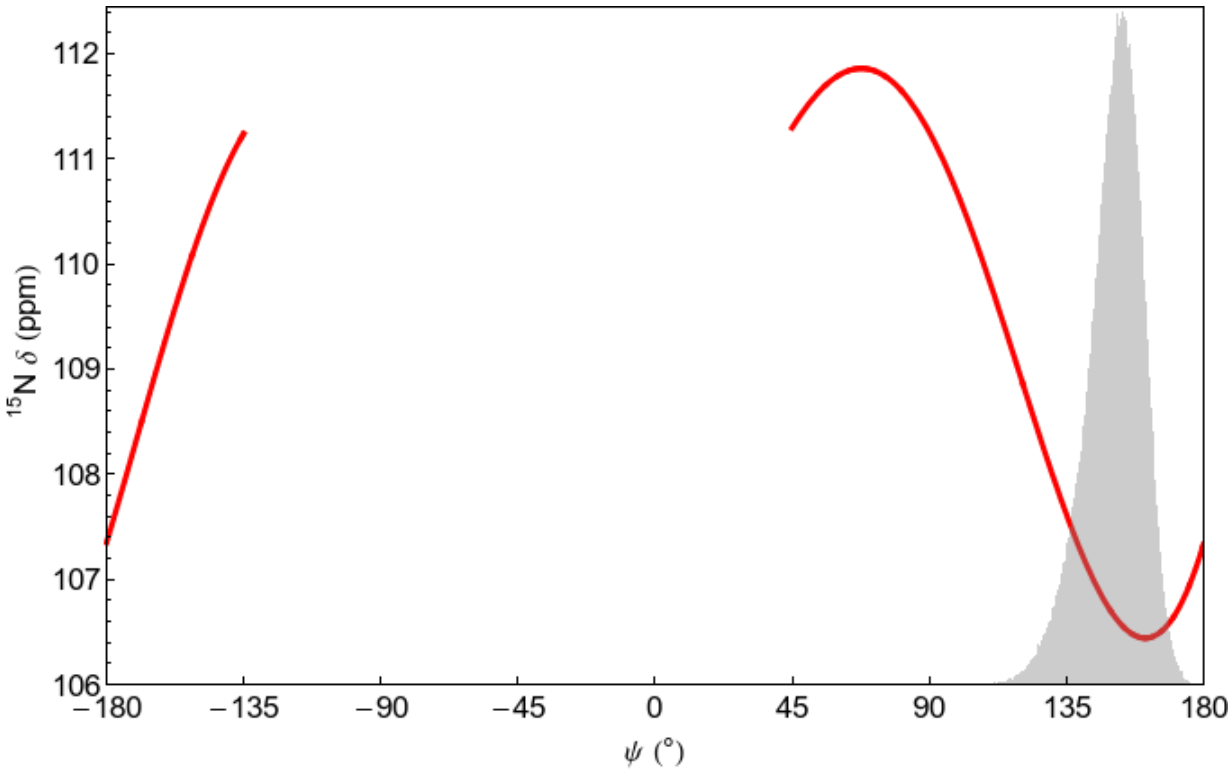


Figure J.56: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue T56 (i.e., N=56) in a 100 ns MD simulation.

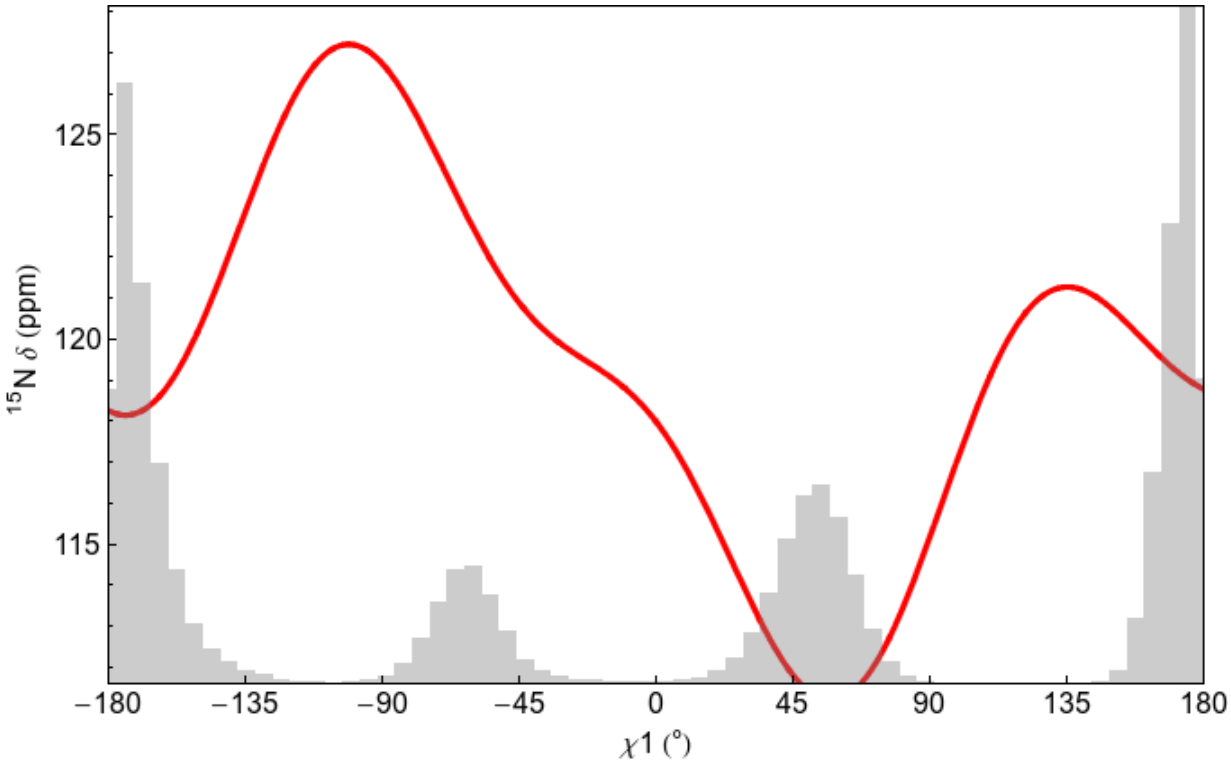


Figure J.57: A smooth curve (red) representing DFT calculations (Section J.2) of threonine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue T56 (i.e., N=56) in a 100 ns MD simulation.

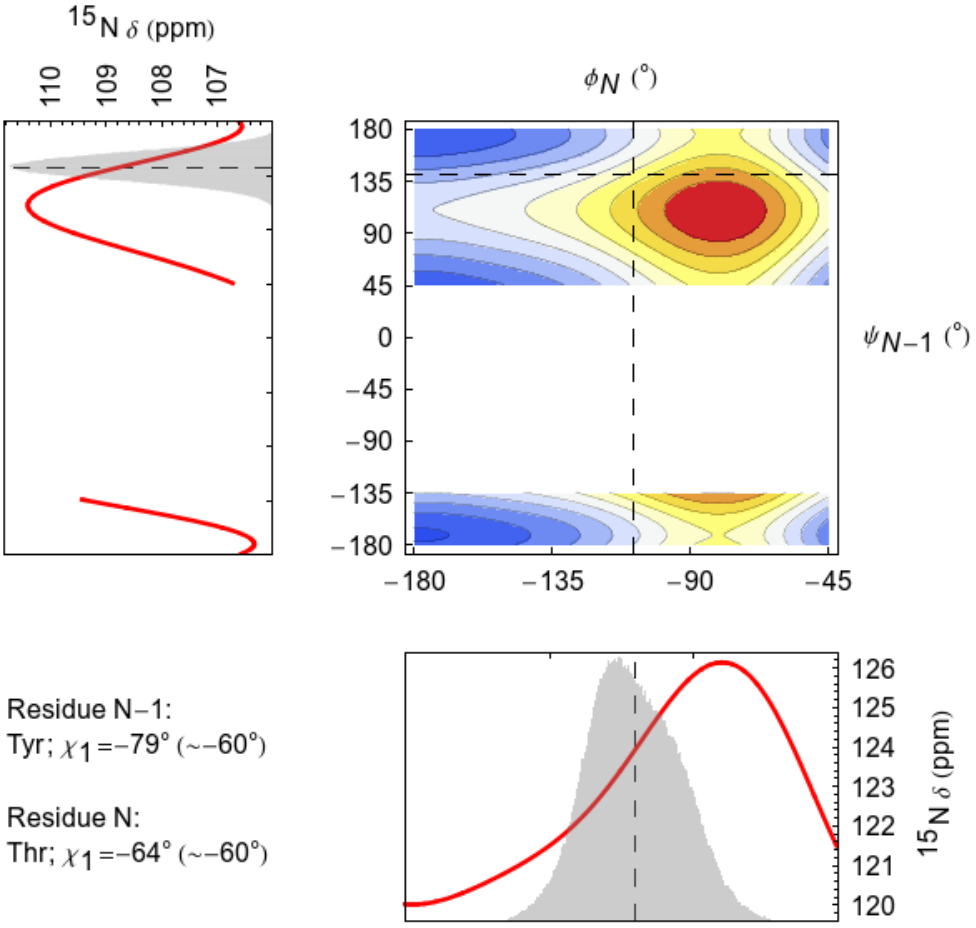


Figure J.58: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues Y68 and T69 (i.e., N=69) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

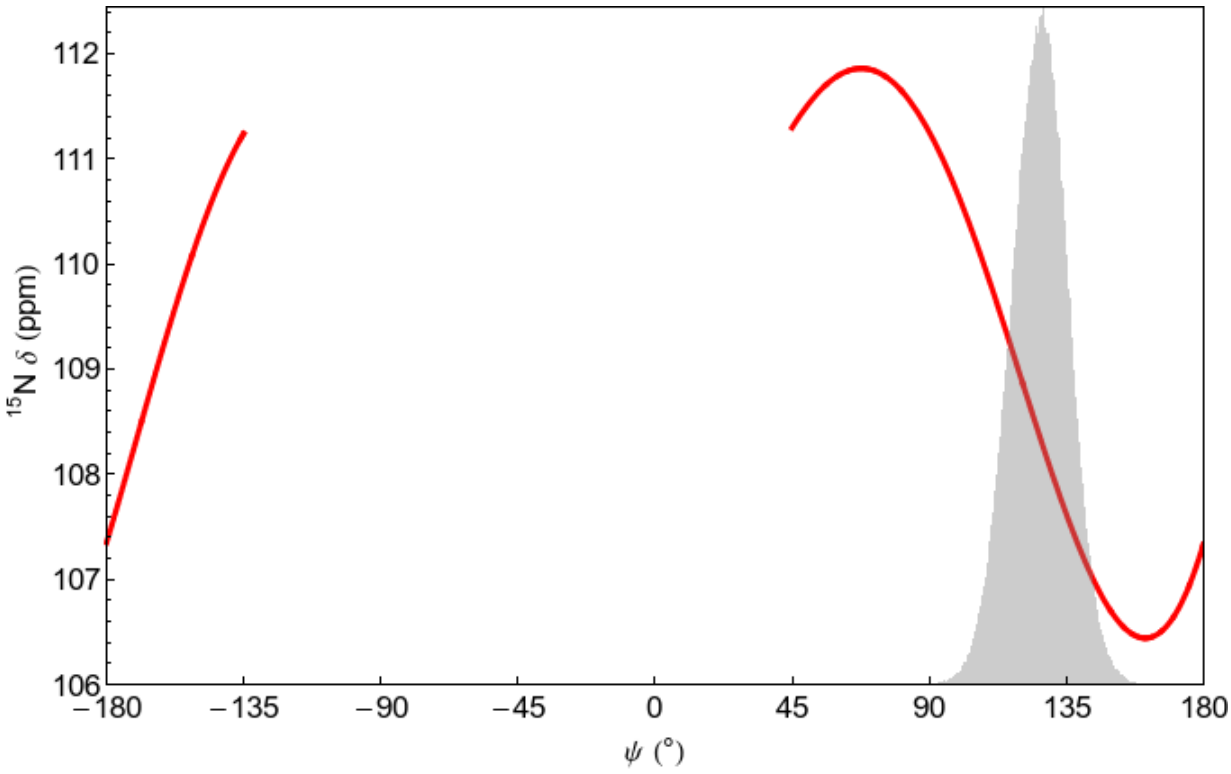


Figure J.59: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue T69 (i.e., N=69) in a 100 ns MD simulation.

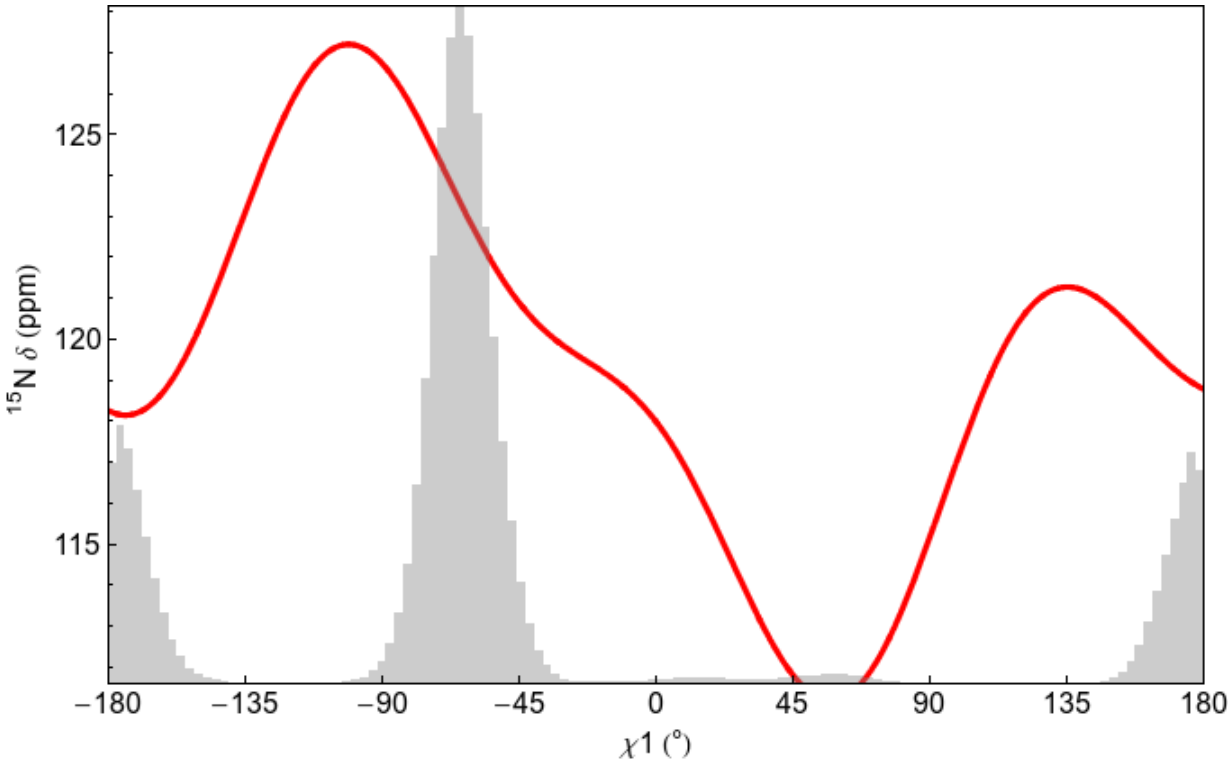


Figure J.60: A smooth curve (red) representing DFT calculations (Section J.2) of threonine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue T69 (i.e., N=69) in a 100 ns MD simulation.

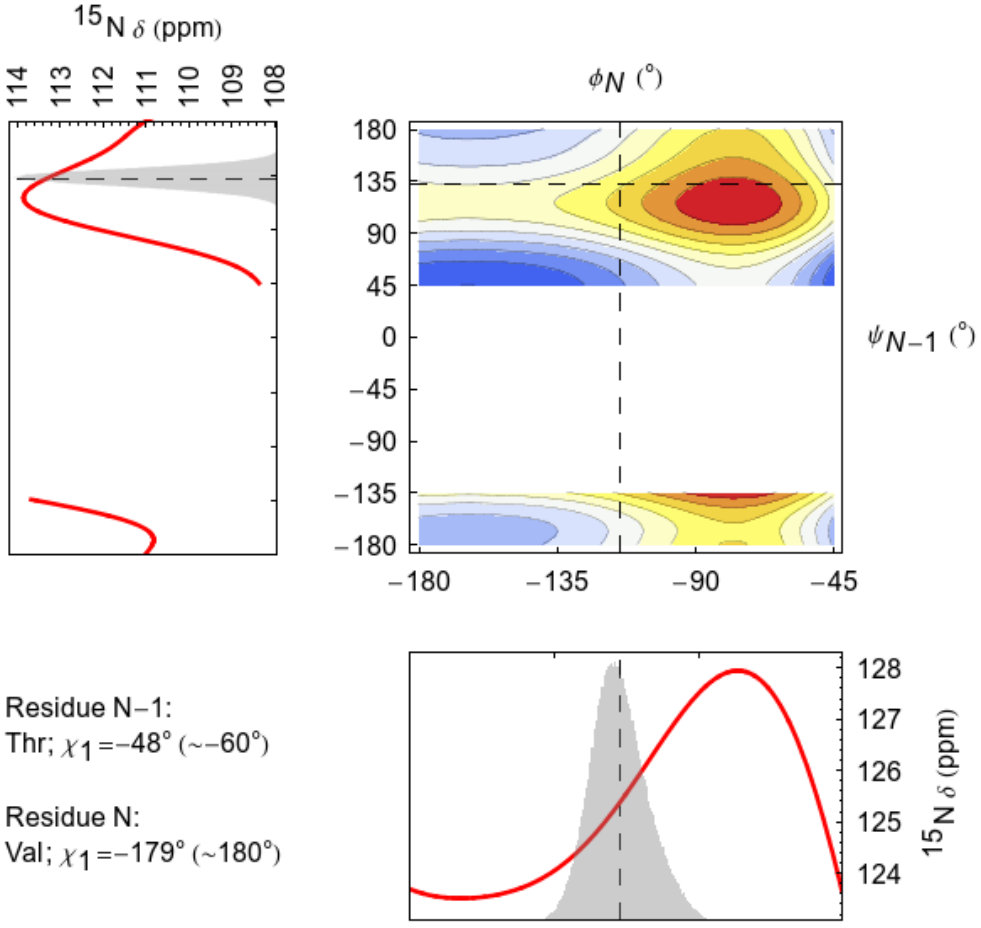


Figure J.61: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues T71 and V72 (i.e., N=72) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

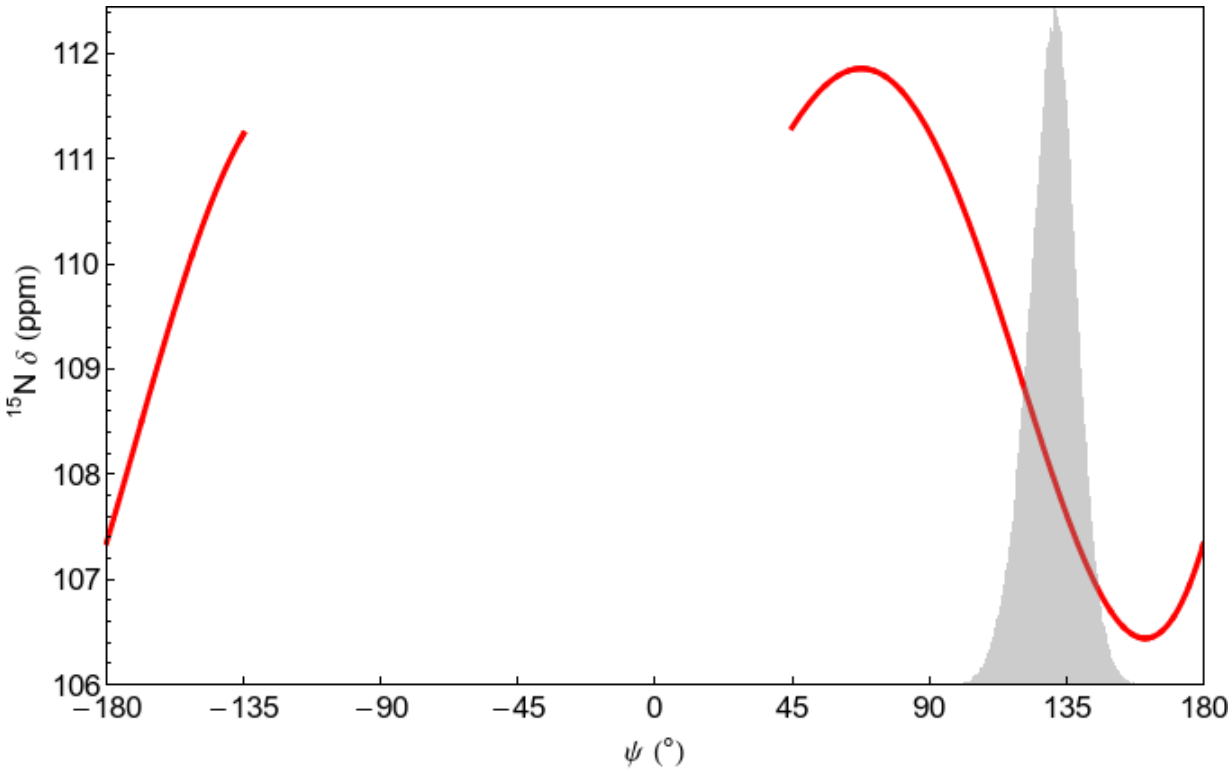


Figure J.62: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue V72 (i.e., N=72) in a 100 ns MD simulation.

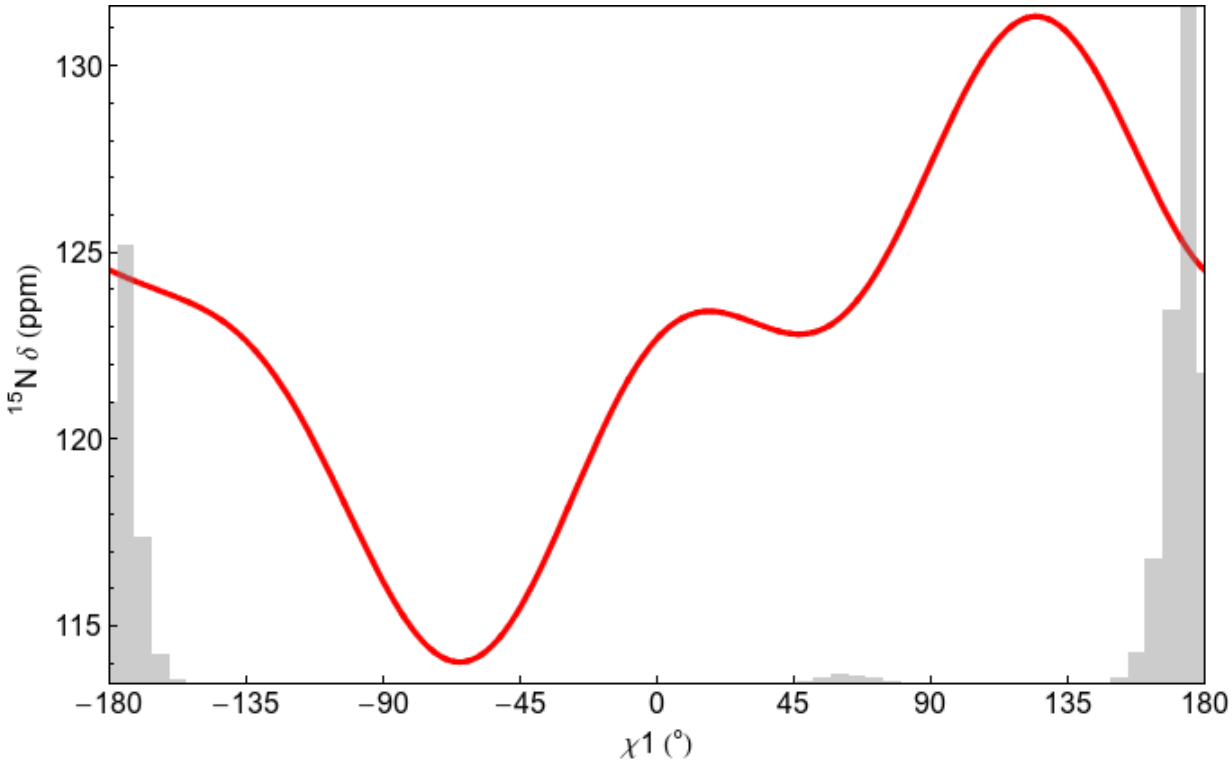


Figure J.63: A smooth curve (red) representing DFT calculations (Section J.2) of valine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue V72 (i.e., N=72) in a 100 ns MD simulation.

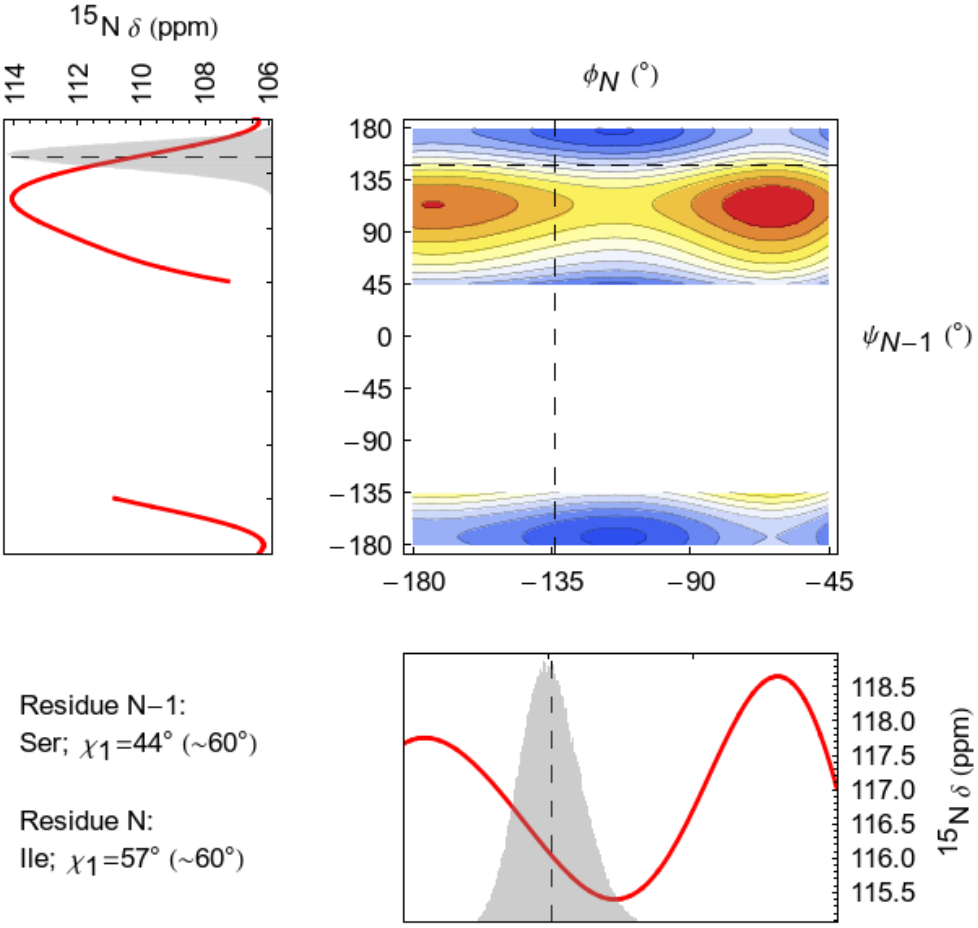


Figure J.64: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues S89 and I90 (i.e., N=90) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

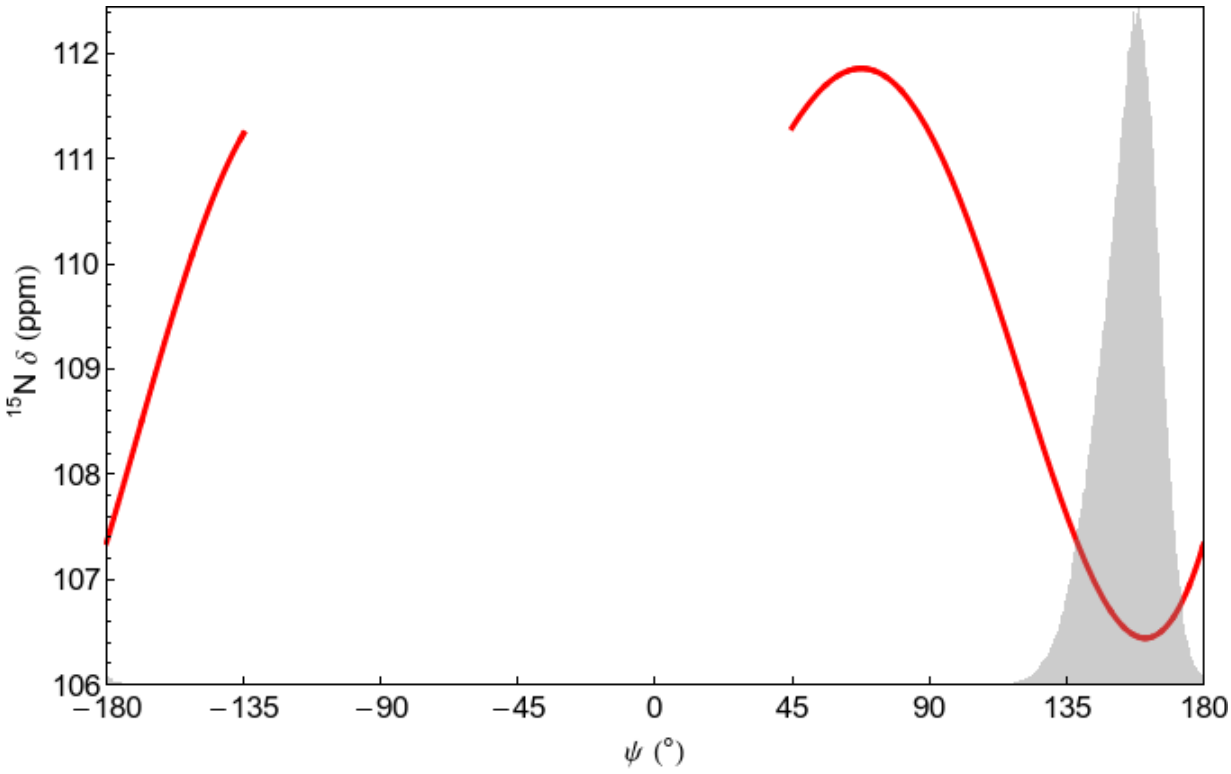


Figure J.65: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue I90 (i.e., N=90) in a 100 ns MD simulation.

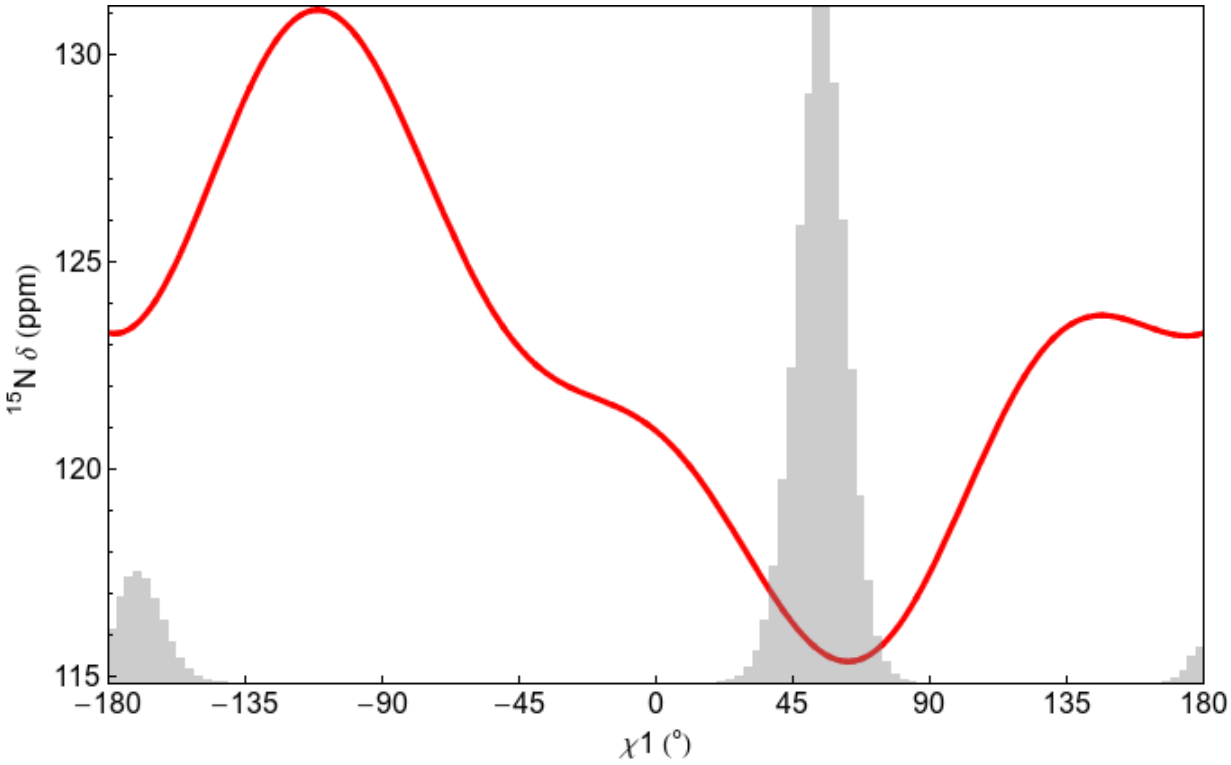


Figure J.66: A smooth curve (red) representing DFT calculations (Section J.2) of isoleucine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue I90 (i.e., N=90) in a 100 ns MD simulation.

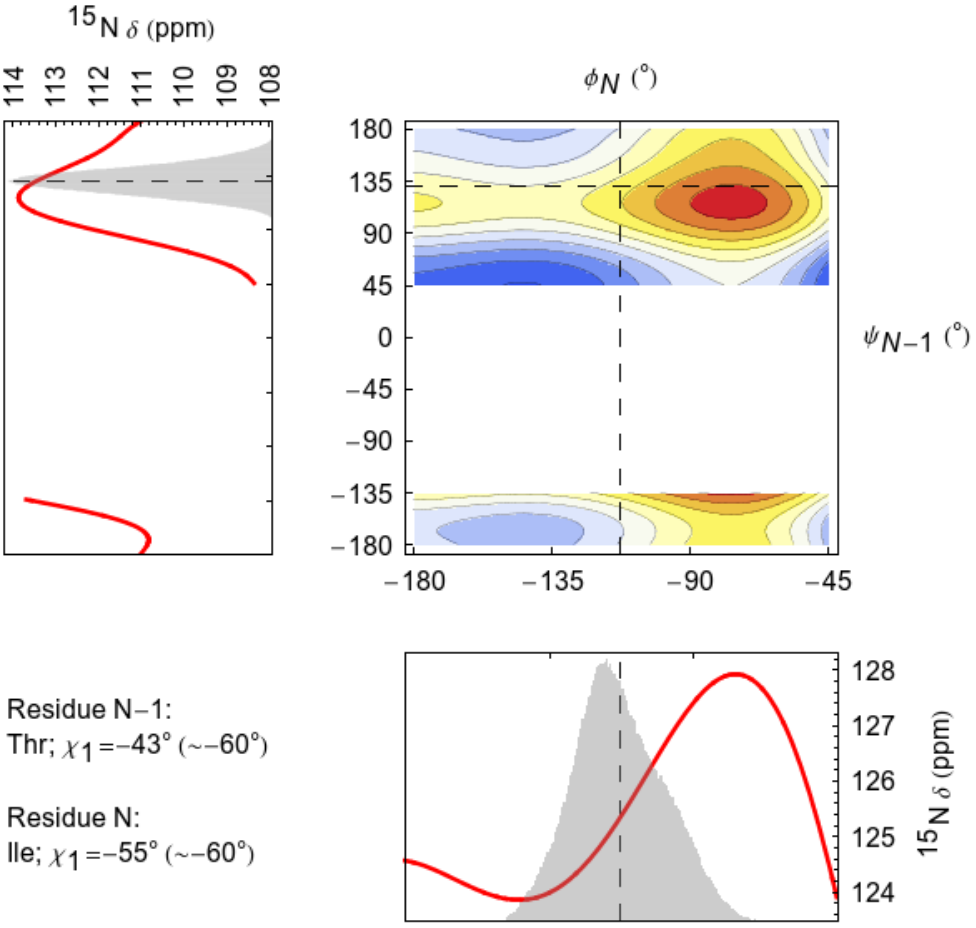


Figure J.67: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues T58 and I59 (i.e., N=59) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

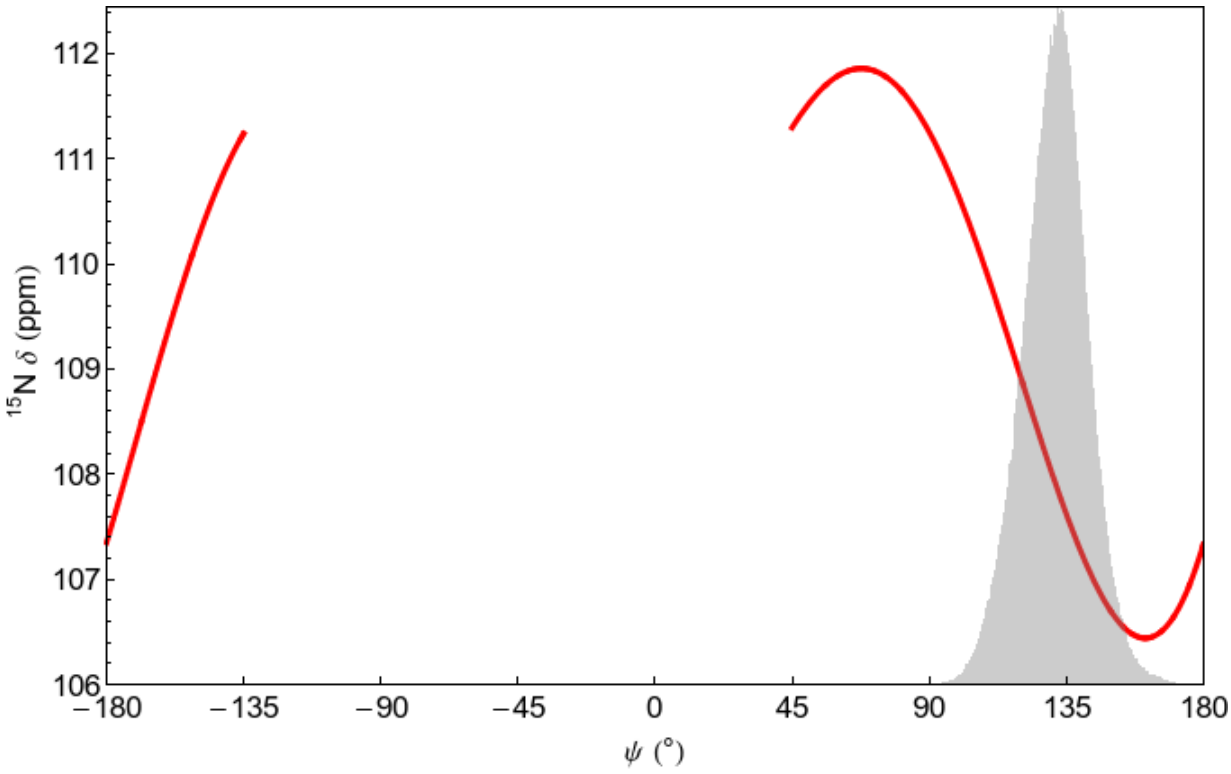


Figure J.68: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue I59 (i.e., N=59) in a 100 ns MD simulation.

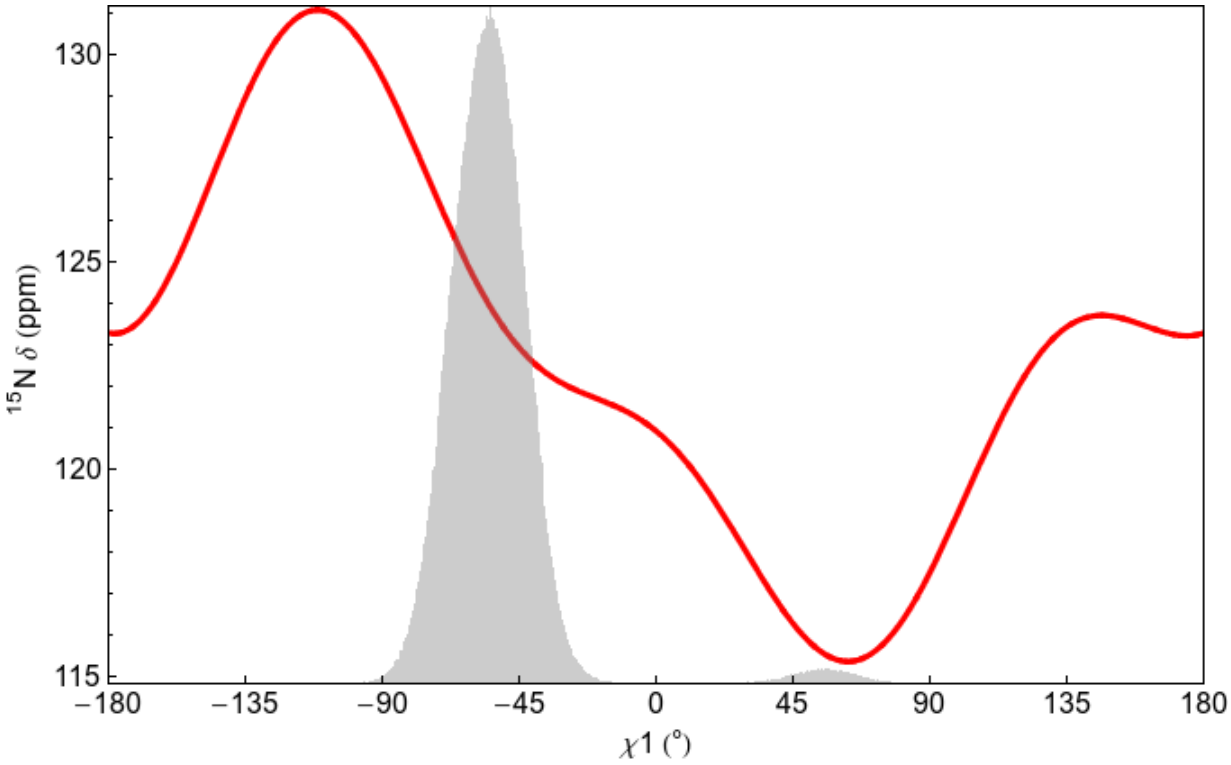


Figure J.69: A smooth curve (red) representing DFT calculations (Section J.2) of isoleucine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue I59 (i.e., N=59) in a 100 ns MD simulation.

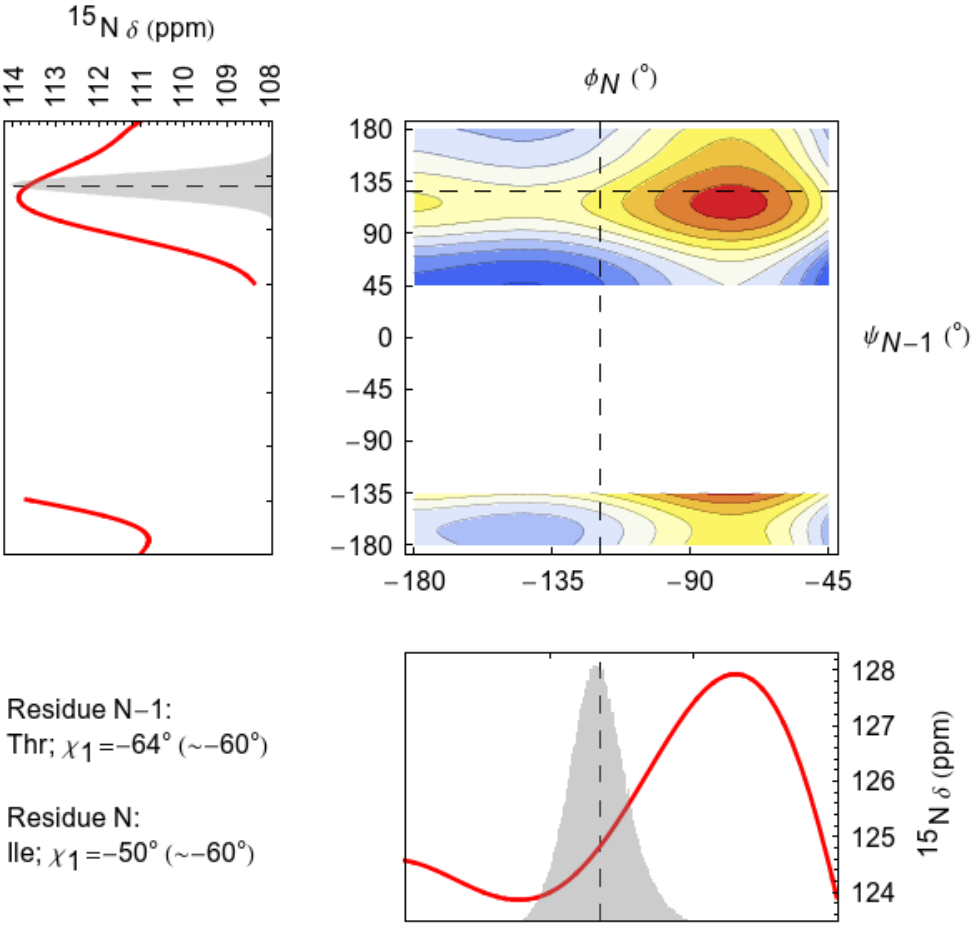


Figure J.70: Smooth curves (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as functions of ψ_{N-1} (left) and ϕ_N (bottom). The superimposed histograms (gray) illustrate the sampling of ψ_{N-1} and ϕ_N for residues T69 and I70 (i.e., N=70) in a 100 ns MD simulation. The contour plot shows the ψ_{N-1}/ϕ_N combinations predicted to result in the smallest (blue) to largest (red) amide nitrogen chemical shifts. Dashed lines indicate the mean backbone dihedral angles observed in simulation. $\chi_{1,N-1}$ and $\chi_{1,N}$ side chain dihedrals are approximated (for the DFT calculations) as the closest of 180°, 60°, or -60°to the χ_1 angles observed in the ^{10}Fn3 domain from the 1FNF crystal structure.

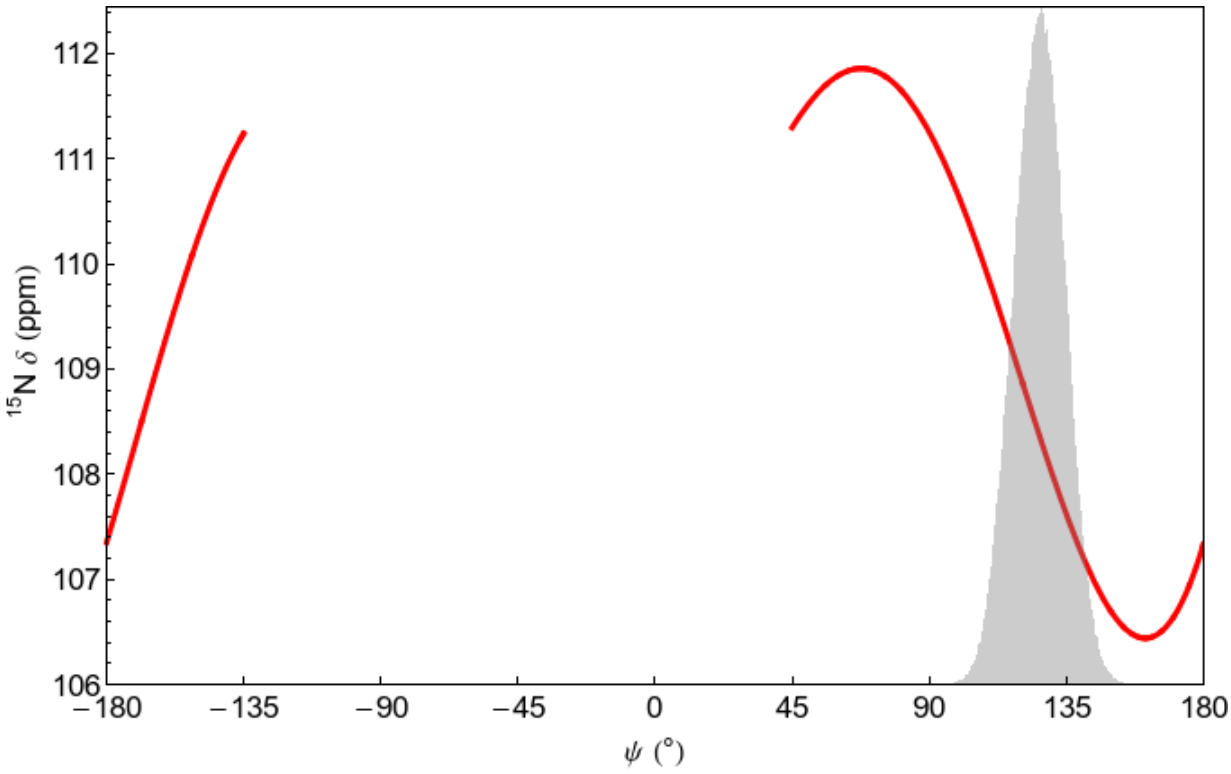


Figure J.71: A smooth curve (red) representing DFT calculations (Section J.2) of amide nitrogen chemical shift as a function of ψ_N. The superimposed histogram (gray) illustrates the sampling of ψ_N for residue I70 (i.e., N=70) in a 100 ns MD simulation.

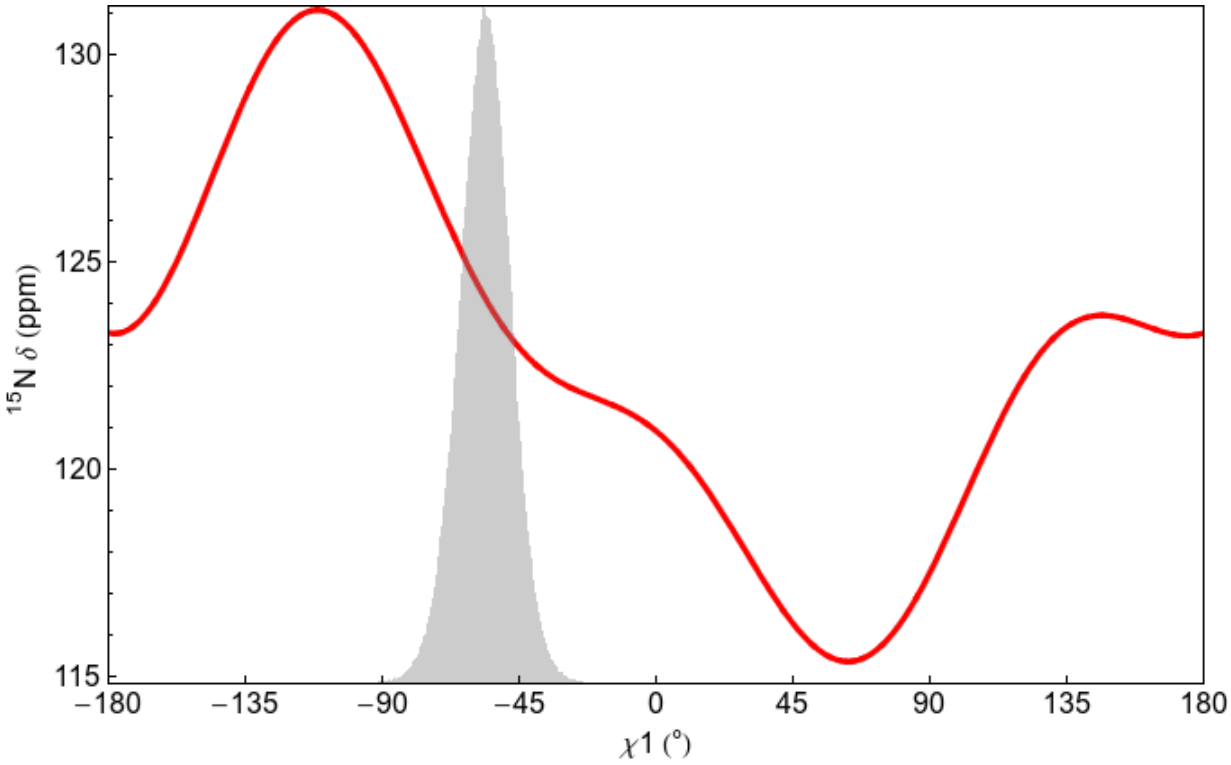


Figure J.72: A smooth curve (red) representing DFT calculations (Section J.2) of isoleucine amide nitrogen chemical shifts as a function of χ_1 (χ_2=180°). The superimposed histogram (gray) illustrates the sampling of $\chi_{1,N}$ for residue I70 (i.e., N=70) in a 100 ns MD simulation.