

Thermodynamics, kinetics and inclusion body formation
of a de novo designed protein Threefoil

by

Su Martha Ma

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Science
in
Chemistry

Waterloo, Ontario, Canada, 2014

© Su Martha Ma 2014

Author's Declaration

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

Threefoil is a small engineered protein of 141 amino acids which is a member of the beta-trefoil superfamily, with three-fold symmetry and high thermal and kinetic stability. Its primary sequence was designed based on a predicted beta-trefoil glycosidase from the halophilic Archaeon *Haloarcula marismortui*. Threefoil predominantly forms inclusion bodies when over-expressed in *Escherichia coli* at 37°C, with little to no protein soluble in the cytoplasm. Nevertheless, Threefoil is capable of adopting a native beta-trefoil structure when refolded from solubilized inclusion bodies. The focus of this thesis is on characterization of the folding and stability of Threefoil through thermodynamic and kinetic experiments for wild-type Threefoil, in addition to sugar- and metal-binding studies and characterization of Threefoil inclusion bodies. Various Threefoil mutants, designed to increase protein stability, are also characterized to probe the origins of, as well as to give insight into, the mechanism of inclusion body formation. The thermodynamic and kinetic stability of wild-type Threefoil was studied using spectral probes, mainly fluorescence, circular dichroism (CD) and dynamic light scattering (DLS). The major observed spectral changes in kinetic and thermodynamic experiments can be fit to a 2-state transition between the folded state and a denatured state containing extensive residual secondary structure. At high protein concentrations, the folding of wild-type Threefoil is complicated by protein misfolding and aggregation. As Threefoil is remarkably resistant to denaturation even at high concentrations of urea and guanidine hydrochloride (GuHCl), studies were also conducted in guanidine isothiocyanate (GuSCN), which is a much stronger denaturant than urea and GuHCl. Remarkably, the time that is required for Threefoil samples to reach equilibrium in renaturation curves is approximately 100 days, while equilibrium by denaturation in the stronger denaturant, GuSCN, requires more than two years. The expression levels of Threefoil mutants A62V, Q78I, D85P and D93P were also studied. None of the four mutants studied exhibited any pronounced increase in solubility compared to wild-type when expressed in *E. coli*.

Acknowledgements

Firstly I would like to thank Dr. Elizabeth Meiering for not only encouraging me to mature as a Master's student, but also as an individual. I greatly appreciate her guidance and her patience in allowing me to progress at my own pace throughout the course of my Master's program.

Secondly I thank the members of the Meiering group, in particular Aron Broom and Dr. Jessica Rumfeldt for their help and insight in helping to solve the (many) problems that I have had throughout my studies. I also thank Briallen Lobb and Laura Bahlmann, co-op students who have been of great help in furthering the progress of my project.

Lastly I thank my parents, for whose generous support and love I am extremely grateful.

Dedication

I dedicate this thesis to my parents, Shu Hua Li and Qin Ke Ma.

Table of Contents

Author's Declaration -----	ii
Abstract -----	iii
Acknowledgements -----	iv
Dedication -----	v
Table of Contents -----	vi
List of Figures -----	ix
List of Tables -----	xi
List of Abbreviations -----	xii
Chapter 1: Introduction -----	1
1.1: Protein folding -----	1
1.1.1: How proteins fold -----	2
1.2: Protein stability -----	4
1.2.1: Equilibrium studies of folding and unfolding -----	6
1.2.2: Kinetic studies of folding and unfolding -----	8
1.2.3: Ligand-binding on protein stability -----	11
1.3: Threefoil -----	12
1.4: Mutant selection criteria -----	14
1.5: Inclusion body formation -----	17
Chapter 2: Methods -----	22
2.1: Protein expression -----	22
2.1.1: Protein expression -----	22
2.1.2: Inclusion body isolation -----	23
2.1.3: Affinity column purification -----	23
2.1.4: Refolding and concentration of Threefoil -----	24
2.2: Folded Threefoil characterization -----	25
2.2.1: Fluorescence -----	25
2.2.2: Circular Dichroism -----	25
2.3: Threefoil kinetics -----	25
2.3.1: Refolding kinetics -----	26
2.3.2: Unfolding kinetics -----	27
2.3.3: Fitting of kinetic traces -----	27
2.4: Threefoil equilibrium studies -----	28
2.4.1: Renaturation equilibrium -----	28
2.4.2: Denaturation equilibrium -----	29
2.4.3: Fitting of equilibrium data -----	29
2.5: Metal removal from Threefoil -----	29
2.6: Inclusion body characterization -----	30
2.6.1: Fluorescence -----	30

2.6.2: Circular dichroism -----	30
2.6.3: Differential scanning calorimetry -----	31
2.6.4: ANS binding -----	31
2.7: SDS-PAGE gel analysis of Threefoil mutants -----	30
Chapter 3: Results -----	33
3.1: Residual structure in urea and GuHCl -----	34
3.2: Equilibrium stability of Threefoil -----	39
3.2.1: Refolding curves of metallated Threefoil -----	39
3.2.2: Unfolding curves of metallated Threefoil -----	42
3.2.3: Renaturation equilibrium studies of Threefoil inclusion bodies -----	46
3.3: Threefoil kinetics -----	52
3.3.1: Unfolding kinetics of metallated Threefoil -----	52
3.3.2: Refolding kinetics of metallated Threefoil -----	56
3.3.3: Kinetics of metallated Threefoil reconstructed from equilibrium curves -----	58
3.4: Kinetics of Threefoil inclusion bodies -----	58
3.4.1: Unfolding kinetics of urea-denatured inclusion bodies -----	58
3.4.2: Refolding kinetics of Threefoil inclusion bodies -----	59
3.4.3: Protein concentration dependence of refolding from urea-solubilized inclusion bodies -----	60
3.5: Chevron analysis for Threefoil -----	61
3.6: Threefoil metal-binding studies -----	64
3.6.1: Characterization of metal-free Threefoil -----	64
3.6.2: Folding and unfolding kinetics of metal-free Threefoil -----	66
3.7: Carbohydrate-binding effects on Threefoil kinetics -----	68
3.8: Characterization of Threefoil inclusion bodies -----	69
3.8.1: Differential scanning calorimetry -----	69
3.8.2: ANS binding -----	70
3.9: Preliminary Threefoil mutant analysis -----	72
3.9.1: Mutant expression -----	72
Chapter 4: Discussion -----	81
4.1: Threefoil folds according to a two-state model with residual structure in the denatured state -----	81
4.2: Threefoil folding and unfolding is remarkably slow -----	83
4.3: Metallation on Threefoil stability -----	85
4.4: Carbohydrate-binding on Threefoil stability -----	85
Chapter 5: Conclusions and future work -----	87
5.1: Equilibrium and kinetic studies -----	87
5.2: Ligand binding in Threefoil stability -----	88
5.3: Future work -----	89
References -----	90

Appendix 1	97
Appendix 2	99
Appendix 3	101
Appendix 4	103
Appendix 5	105
Appendix 6	107
Appendix 7	109

List of Figures

Figure 1.1: Folding energy landscapes and energy diagrams -----	3
Figure 1.2: A sample equilibrium curve for a two-state protein folding transition -----	7
Figure 1.3: A representative chevron plot for a protein folding according to a two-state transition between the folded and unfolded states. -----	10
Figure 1.4: Primary amino acid sequence of designed protein Threefoil -----	13
Figure 1.5: Tertiary structure of native Threefoil -----	13
Figure 1.6: Location of the ten mutants selected using prediction algorithms for Threefoil -----	16
Figure 3.1: Fluorescence spectra of the transition from folded to unfolded Threefoil -----	34
Figure 3.2: Fluorescence spectra of Threefoil IBs in 6M urea, 6M GuHCl and 6M GuSCN. ---	36
Figure 3.3: CD spectra of urea- and GuHCl-denatured inclusion body proteins, native Threefoil and Onefoil -----	37
Figure 3.4: The effect of centrifugation and filtration on Threefoil denatured in 6M urea -----	38
Figure 3.5: Equilibrium renaturation curves in GuSCN for Threefoil, monitored at 313 nm ----	41
Figure 3.6: Equilibrium denaturation curves in GuSCN for Threefoil -----	43
Figure 3.7: Equilibrium renaturation and denaturation curves for Threefoil in GuSCN -----	44
Figure 3.8: Renaturation equilibrium curves over time in urea -----	48
Figure 3.9: C_{mid} and m -values from global fitting of urea renaturation equilibrium curves, monitored at different wavelengths and by different optical probes -----	50
Figure 3.10: Representative unfolding kinetics of Threefoil in GuSCN, monitored at an emission wavelength of 313 nm -----	53
Figure 3.11: Representative refolding kinetic traces in GuSCN, monitored at an emission wavelength of 313 nm -----	57
Figure 3.12: Unfolding kinetics of 6M urea-denatured inclusion bodies in GuSCN -----	59
Figure 3.13: Refolding kinetics of urea-denatured inclusion body proteins -----	60
Figure 3.14: Dependence of refolding rate of urea-denatured inclusion bodies on protein concentration -----	61
Figure 3.15: Rates of refolding and unfolding of Threefoil in GuSCN -----	62
Figure 3.16: Folding branches of Chevron generated in urea, GuHCl and GuSCN -----	63
Figure 3.17: Comparison of fluorescence spectra of metal-free and metal-bound Threefoil ----	65

Figure 3.18: Folding and unfolding kinetic rates in the presence and absence of sodium -----	67
Figure 3.19: Folding and unfolding kinetics in the presence of lactose and sucrose -----	68
Figure 3.20: DSC scans of Threefoil in GuHCl -----	70
Figure 3.21: ANS binding to urea-denatured inclusion body proteins -----	71
Figure 3.22: SDS PAGE gel of expression of wild-type Threefoil, at 37°C, 1 mM IPTG in LB broth -----	73
Figure 3.23: SDS-PAGE gel of expression of wild-type Threefoil, at 37°C, 0.1 mM IPTG in LB broth -----	74
Figure 3.24: SDS-PAGE gel of expression of mutant A62V Threefoil, at 37°C, in LB broth ---	75
Figure 3.25: SDS-PAGE gel of expression of mutant A62V Threefoil, at 18°C, in LB broth ---	76
Figure 3.26: SDS-PAGE gel of expression of mutant Q78I Threefoil, at 37°C and 25°C, in LB broth -----	77
Figure 3.27: SDS-PAGE gel of expression of mutant Q78I Threefoil, at 18°C, in LB broth ----	77
Figure 3.28: SDS-PAGE gel of expression of mutant D85P Threefoil, at 37°C and 25°C, in LB broth -----	78
Figure 3.29: SDS-PAGE gel of expression of mutant D85P Threefoil, at 18°C, in LB broth ---	79
Figure 3.30: SDS-PAGE gel of expression of mutant D93P Threefoil, at 37°C and 25°C, in LB broth -----	80
Figure 3.31: SDS-PAGE gel of expression of mutant D93P Threefoil, at 18°C, in LB broth ---	80

List of Tables

Table 1: Summary of parameters from fitting of GuSCN denaturation and renaturation curves monitored by fluorescence (FL) -----	45
Table 2: Summary of parameters from individual fitting of urea renaturation equilibrium curves monitored by CD and fluorescence (FL) -----	51
Table 3: Comparison of Threefoil kinetic and equilibrium parameters -----	54-55

List of Abbreviations

1-Anilinonaphthalene-8-Sulfonic Acid - ANS

Ammonium acetate - CH₃COONH₄

Ammonium persulfate - APS

Ammonium phosphate - (NH₄)₃PO₄

Amyotrophic lateral sclerosis - ALS

Arbitrary units - au

Circular dichroism - CD

Differential scanning calorimetry - DSC

Dynamic light scattering - DLS

Fibronectin type III domain - fnIII

Fluorescence - FL

Guanidine hydrochloride - GuHCl

Guanidine isothiocyanate - GuSCN

Human muscle acylphosphatase - AcP

Inclusion bodies - Ibs

Isopropyl-β-D-thiogalactopyranoside - IPTG

Linear extrapolation method - LEM

Luria Broth - LB

Magnesium chloride - MgCl₂

Magnesium sulfate - MgSO₄

Mass spectrometry - MS

Nickel-nitriloacetic acid - Ni-NTA

Potassium chloride - KCl

Sodium chloride - NaCl

Sodium dodecyl sulfate - SDS

Sodium phosphate - NaPi

Solvent-accessible surface area - Δ ASA

Superoxide dismutase - SOD

Tetramethylethylenediamine - TEMED

Thioflavin-T - ThT

Tryptophan - Trp

Tyrosine - Tyr

Chapter 1: Introduction

1.1: Protein folding

Proteins are the workhorses of the biological world, carrying out a wide range of functions. From enzymes to structural cell elements to cell signalling, they are ubiquitous within the cellular infrastructure and are necessary to life. A protein's tertiary structure, its three-dimensional conformation in space, is encoded by its primary sequence, which is its linear sequence of amino acids. The protein folding problem is a unique challenge in academia and focuses on the prediction of a protein's tertiary structure based on its primary amino acid sequence. The ability to predict how a protein folds is of interest not only because at a basic biological level proteins need to fold properly in order to carry out their function, but also because there is a great range of practical applications in research and in industry. In particular, recombinant protein technology is widely used in the pharmaceutical industry^{1,2} where proteins are used in drug-delivery systems or as therapeutic drugs. Being able to design novel proteins with desired properties, and being able to predict whether they will fold properly while retaining their intended function, is an important avenue toward reducing both the cost and time needed to bring protein drugs to the market.

Hand-in-hand with the problem of protein folding is the problem of protein misfolding and aggregation. Protein misfolding can be considered as predominantly an intramolecular phenomenon, where a protein folds but fails to adopt its native, intended conformation.^{3,4} Protein aggregation, on the other hand, describes an intermolecular phenomenon where one or more misfolded proteins associate, which often involves hydrophobic interactions.

Aggregation of proteins is implicated in a wide variety of neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis (ALS).^{3,4} Aggregation is also common in recombinant protein production, for example in systems such as *Escherichia coli*⁵⁻⁷ where, due to the high volume of protein being over-expressed, the cellular quality checking and repair mechanisms become overwhelmed.⁷

Currently, although advances have been made, there is still no reliable or accurate general solution to the protein folding problem; however, advancements in experimental techniques are allowing many more protein structures to be determined, which brings us closer to elucidating the link between the sequence and structure. This fundamental issue of protein folding, therefore, is a relevant and important issue.

1.1.1: How proteins fold

Because there is, in principle, an enormous number of conformations that a protein may adopt^{3,4}, one can imagine that if a protein were to sample all of these possible conformations, the native state would never be populated. Generally, protein folding is thought of as a search across what is termed the folding energy landscape.^{8,9} The simplest folding landscape is often represented as a smooth funnel (Figure 1.1A), at the top of which is the unfolded protein. In its unfolded state the peptide chain is flexible and can sample a large number of conformations, as it lacks sufficient stabilizing interactions to form a defined structure. As the protein randomly samples the energy landscape at the top of the funnel, the chain can begin to form long-lasting stable, mainly native interactions, which lock the protein into defined conformations and lead to a narrowing of the funnel. At the bottom of the funnel is the protein's lowest-energy native state.

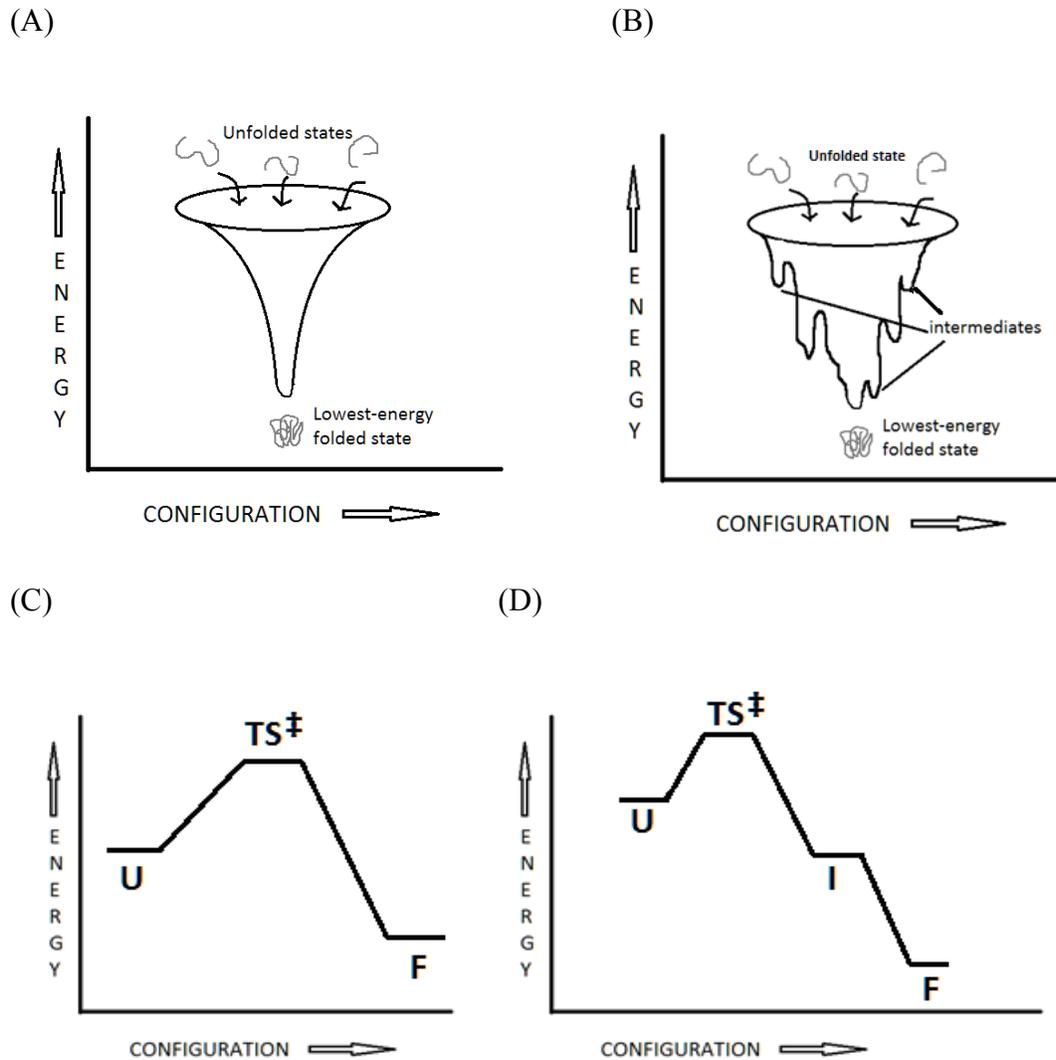


Figure 1.1: Folding energy landscapes and energy diagrams

(A) Depicted is a simple folding energy landscape. The protein can transition from the unfolded state to the folded state along the smooth walls of the funnel. In this scenario there are only two states that the protein populates: The unfolded state, which is an ensemble of interconverting conformations, and the native state, its lowest energy, folded form.

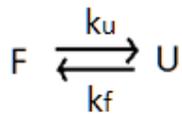
(B) Depicted is a more complex folding energy landscape. The protein can transition from the unfolded state to the folded state by following along the surface of the energy funnel. However, along the way it can fall into energy minima and transiently populate kinetic or equilibrium intermediates. Depending on the depth of the well of the local minimum, the protein can either become trapped (if the well is deep) or escape again (shallow well) and continue along to find its native state. In this scenario there are more than two states that the protein can populate: The unfolded state, which is a flexible peptide chain, and the native state, its lowest energy folded form, as well as numerous intermediate states.

(C) The free energy diagram of folding along a smooth (left) and rugged (right) energy landscape. There are many intermediates which can be populated during folding along a rugged energy landscape. The transition state (TS^\ddagger) corresponds to the highest energy state of the protein, which is short-lived and which cannot be directly observed in experiments. Two-state folding proceeds along a smooth energy landscape while multi-state folding proceeds along a more rugged energy landscape.

For many proteins the folding landscape can be rugged and dotted with local minima (referred to as ‘wells’ or ‘pits’) in which the protein can become trapped (Figure 1.1B, 1.1D). These local minima represent low-energy intermediate states into which the protein can become trapped as it folds toward its native conformation. Misfolding and subsequent aggregation often occur due to the protein falling into a local minimum along the energy landscape. If the well is shallow enough, then the protein may escape and continue along the funnel toward the native state. If the well is deep, however, the protein may become trapped in a stable intermediate state. There is evidence that intermediates can either enhance or hinder folding; however many aspects are still not understood in detail for specific proteins.

1.2: Protein stability

Similar to other chemical reactions, cooperative protein folding involves overcoming an energy barrier as the protein transitions from an unfolded state to a folded state (Figure 1.1C). Many small proteins (< 100 residues) have been observed to fold according to a two-state mechanism, where only the unfolded (U) and the folded (F) states are significantly populated over the course of the reaction, and no intermediates accumulate (Figure 1.1A, 1.1C).¹⁰



The unfolding rate constant for converting from F to U is k_u , and the folding rate constant for converting from U to F is k_f . Kinetic rate constants are related to the change in free energy from a

ground state (folded or unfolded) to the transition state:

$$K_U = [U] / [F] = k_u / k_f \quad (1.1)$$

The energetics of reversible protein unfolding can be described by the simplified Gibbs-Helmholtz equation:

$$\Delta G = \Delta H - T\Delta S \quad (1.2)$$

where T is the temperature, ΔG is the change in the Gibbs free energy, ΔH is the change in enthalpy and ΔS is the change in entropy, of the unfolded state (U) relative to the folded state (F).¹¹ The change in Gibbs free energy is the thermodynamic stability of the protein as a function of energy difference between the folded and unfolded states, which varies with solution conditions (pH, denaturant, etc.). At equilibrium, the change in Gibbs free energy of unfolding is related to the equilibrium constant for unfolding, K_U :

$$\Delta G_U = G_U - G_F = -RT \ln (K_U) \quad (1.3)$$

where R is the gas constant and T is the absolute temperature. G_U represents the Gibbs free energy of the unfolded state and G_F the Gibbs free energy of the folded state. For many small proteins of the order of 100 amino acids, the change in Gibbs free energy is on the order of 5-10 kcal/mol.¹⁰ Note that the equilibrium constant K_U is the ratio of the kinetic unfolding rate constant to the kinetic folding rate constant.

The relationship between ΔG_U and K_U is often also useful for determining whether a protein folds according to a two-state mechanism, or whether a folding intermediate is involved in the process, by measuring and comparing equilibrium and kinetics, often monitored by multiple probes of structure such as fluorescence and circular dichroism (CD).¹² If equilibrium

and kinetic data coincide then two-state behavior is implied; if they differ, intermediate(s) may accumulate during the course of three-state or multi-state folding. Often, while the observed equilibrium behavior is two-state, transient intermediate(s) may be observed in kinetics. In addition, many proteins possess residual structure in the denatured state, which can lead to the transition from the folded state to the unfolded state appearing multi-state. This will be discussed in-depth in the results.

1.2.1: Equilibrium studies of folding and unfolding

Experimentally, ΔG_U can be determined by measuring equilibrium stability curves, both of refolding and unfolding, as well as by measuring the kinetics of folding and unfolding. In this section, the equilibrium stability of a protein that undergoes a two-state transition is described; the following section deals with two-state kinetics.

In an unfolding equilibrium curve, also referred to as a denaturation equilibrium curve, protein samples are diluted into buffer containing varying amounts of denaturant and left to unfold. In a refolding equilibrium curve, or a renaturation equilibrium curve, the protein is first denatured then diluted into buffer containing again varying amounts of denaturant and allowed to refold. Once equilibrium is reached, the curves are measured by a spectral probe such as circular dichroism (CD) or fluorescence. An equilibrium curve is generated by plotting the signal at a certain wavelength (usually where signal change is greatest) against the denaturant concentration. For a two-state transition, equilibrium curves measured using different optical probes should coincide. It is useful to plot the signal as a function of a fraction of unfolded protein, which normalizes the equilibrium curves to 1 and allows for comparison between

numerous probes or of curves from both renaturation and denaturation equilibrium experiments. The coincidence of equilibrium denaturation and renaturation curves indicates that the folding process is reversible and that equilibrium has been reached. Figure 1.2 shows a sample denaturation equilibrium curve where the transition is two-state from folded to unfolded. As is characteristic of a two-state transition, the curve is sigmoidal. The fraction of the unfolded protein, designated f_U , can be calculated using the following equation:

$$f_U = \frac{Y_{obs} - Y_U}{Y_N - Y_U} \quad (1.4)$$

where Y_{obs} is the observed signal, Y_U is the signal of the unfolded state and Y_N is the signal of the native state.

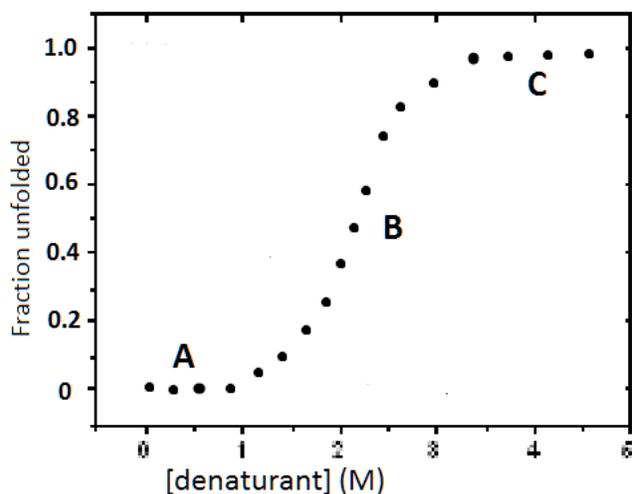


Figure 1.2: A sample equilibrium curve for a two-state protein folding transition

Three regions of the curve are indicated: A: The pre-transition region where the protein is assumed to be folded and there is no appreciable protein denaturation. B indicates the transition region where the transition from folded to unfolded occurs in a cooperative manner. C: The post-transition region where the protein is assumed to be unfolded and where there is no appreciable protein renaturation.

Data for an equilibrium curve can be fit using the linear extrapolation method (LEM) with the following equation:

$$Y = \frac{Y_N - (Y_N - (Y_U + S_U[\text{denaturant}])) * \exp(-\Delta G_u^{H2O} + m[\text{denaturant}]/RT)}{1 + \exp(-\Delta G_u^{H2O} + m[\text{denaturant}]/RT)} \quad (1.5)$$

where S_U is the urea dependence of the signal for the unfolded state, R is the gas constant, T is the temperature in Kelvin, ΔG_u^{H2O} is the Gibbs free energy of unfolding of the protein at temperature T in 0 M denaturant, and m is the denaturant dependence of ΔG_u .¹²⁻¹⁴ Of these parameters, of interest especially is the m -value, which is proportional to the change in solvent-accessible surface area (ΔASA) upon unfolding.¹⁵ The m -value can be calculated for a protein based on the number of amino acids and ΔASA as described by Meyers, Pace and Schultz, who studied a set of 45 proteins and empirically determined the relationship between the m -value and ΔASA in urea and GuHCl.¹⁵ Comparison of the calculated m -value with the experimental m -values from fitting equilibrium curve data can give insight into whether a protein folds according to a two-state mechanism or not. Deviation from the calculated m -value indicates non two-state behavior and may implicate the presence of an intermediate species¹⁵ or the presence of residual structure.

1.2.2: Kinetic studies of folding and unfolding

In this section, the kinetics of a two-state transition is discussed. ΔG_u for a two-state transition can be measured by determining K_U from the ratio of the unfolding and refolding kinetic rates, which are determined experimentally by monitoring the spectral signal change over

time of a protein as it either folds or unfolds. Often kinetics are monitored using fluorescence, tracking changes in the tertiary structure via the environment of fluorescing aromatic amino acids (tryptophan and tyrosine), or by CD, tracking changes in secondary structural elements such as alpha-helices or beta-sheets. Depending on the wavelength of measurement it is possible to monitor mainly secondary structure, mainly tertiary structure or a mixture of the two.

Kinetic rate constants for a two-state transition for a 2-state transition are typically obtained by fitting the time-dependence of the optical signal to a single-exponential equation:

$$A(t) = a*t + b + c_i*exp(-k_i*t) \quad (1.6)$$

where $A(t)$ is the measured signal as a function of time t , a is the drift, b is the offset (the signal achieved when the protein has reached equilibrium), c_i is the amplitude of the signal change and k_i is the rate constant for either folding or unfolding. The half-life ($t_{1/2}$) of a kinetic reaction is the time required for half the sample to either fold or unfold and can be calculated as follows:¹¹

$$t_{1/2} = \ln(2)/k_{obs} \quad (1.7)$$

where k_{obs} is the observed kinetic rate constant, which is the sum of the rate constants of folding and unfolding¹¹:

$$k_{obs} = k_f + k_u \quad (1.8)$$

where:

$$\ln k_u = \ln k_u^{H2O} + m_u*[D] \quad (1.9)$$

$$\ln k_f = \ln k_f^{H2O} + m_f*[D] \quad (2.0)$$

k_u^{H2O} and k_f^{H2O} are the respective unfolding and folding rate constants of the protein in the absence of denaturant (D). In high denaturant the contribution of folding kinetics to the observed rate constant is negligible. i.e., $k_u \gg k_f$, therefore $k_{obs} \sim k_u$. Conversely, in low denaturant $k_f \ll k_u$ and so $k_{obs} \sim k_f$.

For proteins folding via a 2-state transition, a plot of the natural logarithm of the kinetic rate k_{obs} against denaturant concentration yields a V-shaped curve known as a chevron plot which can be fit to an equation combining the folding and unfolding rate constants¹¹:

$$\ln k_{obs} = \ln [k_f^{H_2O} \exp(-m_f*[D]) + k_u^{H_2O} \exp(m_u*[D])] \quad (2.1)$$

The chevron plot (Figure 1.3) yields several important parameters, which are described below.

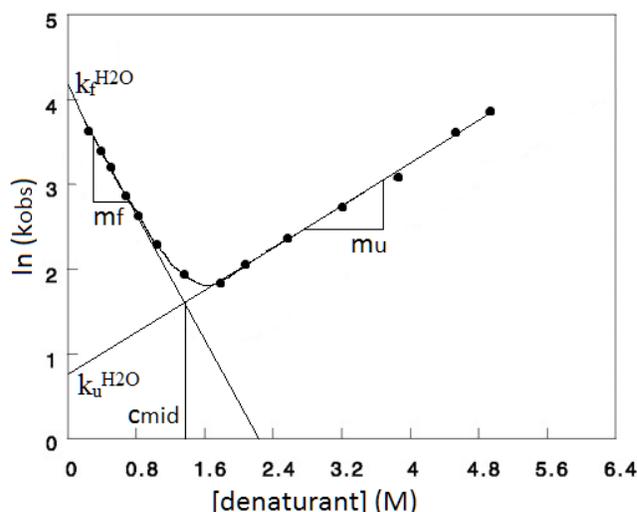


Figure 1.3: A representative chevron plot for a protein folding according to a two-state transition between the folded and unfolded states.

If a protein folds according to a two-state transition, the m -values of folding and unfolding, m_f and m_u , respectively, should sum to the m -value obtained from equilibrium studies. That is,

$$m_{eq} = m_f + m_u \quad (2.2)$$

$k_f^{H_2O}$ and $k_u^{H_2O}$ are obtained by extrapolating the unfolding branch and refolding branch of the chevron plot back to the y -axis and represent the rates of folding and unfolding, respectively, in the absence of denaturant. C_{mid} is the denaturation midpoint, indicating the denaturant concentration at which half of the proteins are denatured.

If an intermediate is populated during a kinetic experiment, often transiently in the dead time of the experiment, then the major transition observed is that of the protein folding from the intermediate state, rather than from the unfolded state. On the other hand, in an equilibrium experiment, the intermediate is not significantly populated; therefore the major transition observed is from the unfolded to folded state. A protein which exhibits two-state folding behavior will not populate an intermediate; thus, m_{eq} determined by equilibrium curves and by kinetics are expected to agree.

1.2.3: Ligand-binding on protein stability

Anywhere between one-quarter and one-third of all proteins require a metal to be bound in order to function.^{16,17} Metals can serve in a structural role or in a catalytic role and are inserted into proteins by metallochaperones.^{16,18} Metal-binding can induce a conformational change within a protein which may be stabilizing⁴⁵ or destabilizing¹⁹ depending on the protein and the metal that is bound. The mechanism of correct metal-insertion is largely an unanswered question, as is whether a protein can tolerate a certain degree of freedom in which metal is inserted.²⁰

Mismetallation is often also an issue in the production of recombinant proteins within host *E. coli* or yeast cells, as the cellular machinery is not always well equipped to cope with the over-expression of proteins.^{16,20,21} Even in cells expressing proteins at a physiological level, mismetallation can occur. In the case of proteins being mismetallated, degradation of the protein or misfolding is possible, the latter of which may become more likely in proteins requiring more than one metal bound to function, and can lead to aggregate formation.

The stabilizing effect of sugars on the native state of globular proteins is well-documented, although a precise mechanism of stabilization is yet unknown.²²⁻²⁴ It is suggested

that the presence of sugar osmolytes such as sucrose, lactose, galactose, etc. within a solvent is stabilizing because the sugars are excluded from the surface of the cell, inducing a protein conformation which has the smallest solvent-exposed surface area.^{22,23 25} In order to also minimize the energetic cost of creating a cavity within a crowded aqueous solution, the protein conformation which is most compact, most often the native state, is the most energetically favorable.²² Carbohydrate-binding also affects the thermodynamic stability of a protein by binding preferentially to either the native state or the transition state—or both, shifting the equilibrium to favor the native state. Studies on β -lactoglobulin, for example, indicate that trehalose, sucrose, mannose and mannitol all increase protein stability.²⁶ The stabilizing effect of carbohydrate-binding is thought to increase with increasing concentrations of osmolyte in solution.²⁶

1.3: Threefoil

The rational protein design of stable, cooperatively folded proteins is a field that has attracted much interest in recent years. The ability to design a novel protein with certain selected functionalities is desirable and useful in the pharmaceutical field.^{1,2} Successfully imbuing the desired folds and functionalities onto these proteins via rational design would be a major step toward the development of these materials. Threefoil is a designed protein of 141 amino acids with pseudo three-fold primary sequence and structural symmetry, comprised of three repeating subdomains.^{27,28} The 141 amino acid sequence, excluding a poly-histidine tag at the N-terminus for affinity purification, is given in Figure 1.4.

GDGYYKLVARHSGKALDVENASTSDGANVIQYSYSGGDNQQWRLVDL

GDGYYKLVARHSGKALDVENASTSDGANVIQYSYSGGDNQQWRLVDL

GDGYYKLVARHSGKALDVENASTSDGANVIQYSYSGGDNQQWRLVDL

Figure 1.4: Primary amino acid sequence of designed protein Threefoil

Threefoil is composed of three identical repeated peptide sequences. Each line of the sequence above corresponds to one trefoil subdomain.

Threefoil's sequence was based on one of the most symmetric beta-trefoils identified in the conserved domain database of protein sequences, a predicted glycosidase from the organism *Haloarcula marismortui*, an archaeon originating from the dead sea.²⁷ The 3-dimensional structure of native Threefoil as solved by X-ray crystallography is illustrated in Figure 1.5.

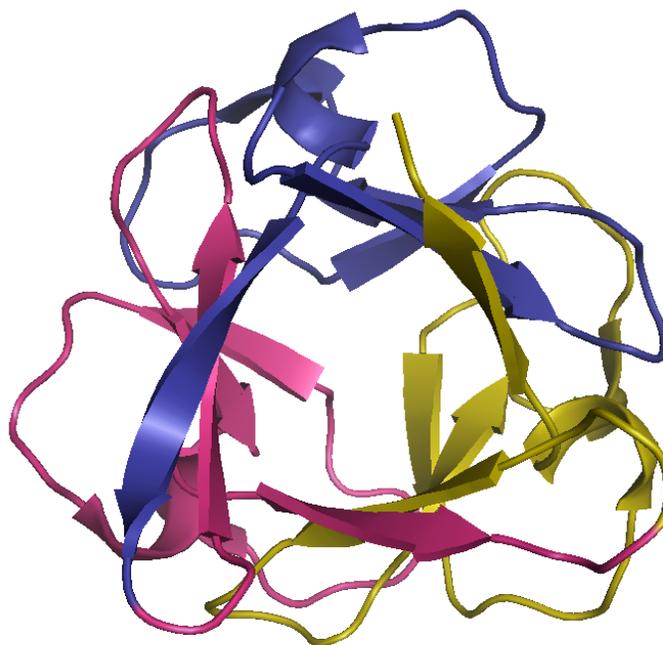


Figure 1.5: Tertiary structure of native Threefoil

The tertiary structure of native Threefoil as solved by X-ray crystallography, showing three-fold symmetry. In purple, pink and yellow are highlighted the three identical repeating units. The purple unit is the N-terminal unit, leading into the second unit in pink, and the third C-terminal unit in yellow. Image generated in PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.). Protein databank code: 3PG0.

Twenty-six of the 47 amino acids of each repeat element in the sequence coding for this glycosidase were identical, with a further eleven being conserved in two of the three units.^{27,28} In order to generate a completely symmetric protein, a BLAST search was conducted to match the initial protein sequence to thirteen of its closest relatives. The final sequence was generated by determining which amino acid is most often conserved in the sequence of its closest relatives.^{27,28} In the case where the closest relatives did not offer a unanimous choice as to the amino acid to be used, ROSETTA was used to determine the most energetically favorable amino acid for its position within the protein.²⁷

Despite changing ~33% of the amino acid residues from the template sequence, Threefoil folds cooperatively to the designed beta-trefoil structure; it also binds sodium in its core, retains carbohydrate binding functionality and shows remarkable stability, all of which are discussed in this thesis. While Threefoil can be obtained in its native state it is found almost exclusively in inclusion bodies when over-expressed in *E. coli*; however, the native protein can be obtained by refolding from denaturant-solubilized inclusion bodies. Interestingly, the single 47 amino acid repeat called Onefoil, which makes up a single subdomain of Threefoil, does not fold to the same native state as a single unit within Threefoil. Onefoil does, however, contain extensive secondary structure, as probed by circular dichroism experiments.²⁷

1.4: Mutant selection criteria

Several point mutations were made to Threefoil in order to probe the areas of the protein responsible for inclusion body formation. The process of mutant selection was carried out by Aron Broom (Ph.D. candidate) and Zachary Jacobi (co-op student) of the Meiering lab and will

be discussed briefly in this section. The primary goal was to seek out stabilizing mutations for Threefoil using currently available stability prediction algorithms which predict the change in ΔG (the $\Delta\Delta G$) upon mutation. By using numerous algorithms rather than relying on a single one, it was hoped that a consensus could be reached that best predicts a successful mutation. In addition, the shortcomings of one prediction algorithm hopefully could be made up by another. The prediction algorithms used were: FoldX, Rosetta, EGAD, MuPRO, PoPMuSiC and MultiMutate, all of which are currently available either online or by request from their respective authors.

As Threefoil is symmetrical, it was assumed that the $\Delta\Delta G$ at each position on each unit would be approximately equally stabilized. Only 18 mutations were considered at each position, excluding the wild-type amino acid that already occurs at that position and cysteine, as it tends to form disulfide bonds which can heavily influence the aggregation propensity. $\Delta\Delta G$ s from each algorithm were normalized in order to avoid skewing of data from sources which heavily favored one amino acid or the other. After normalization, mutations were chosen which had a consensus as being stabilizing by more than half of the algorithms used and the top ten mutations were ultimately decided upon. The choice to place them all on the second repeating unit of Threefoil was intentional, as it is the furthest from both the N- and the C-termini, and is thought to be important toward initiating and mediating contacts between units one and three. The ten mutations made are: D49N, K53V, A62V, E66Y, E66L, A68G, Q78I, D85P, R90L, D93P. Their positions are noted in Figure 1.6.

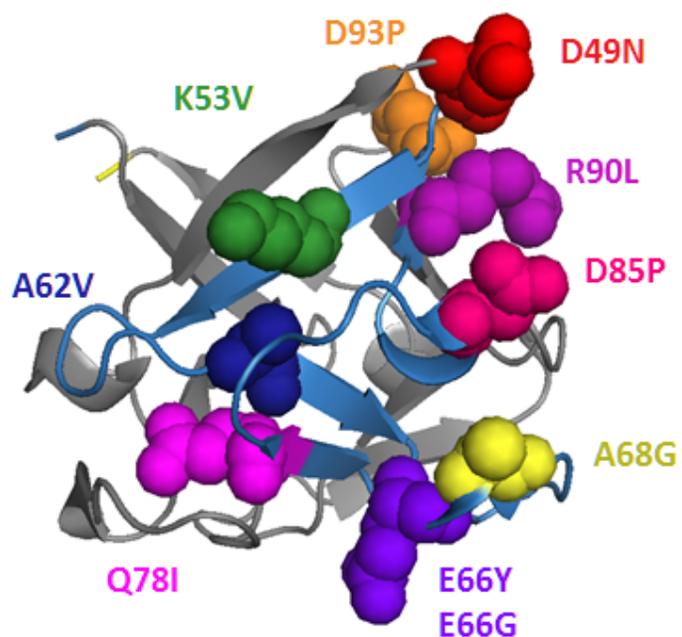


Figure 1.6: Location of the ten mutants selected using prediction algorithms for Threefoil

All of the mutants were selected to be placed on the second repeating unit (strands highlighted in light blue), which is thought to mediate contacts between the first and third unit. The N-terminal is indicated in yellow and the C-terminal is indicated in blue. Image courtesy of Aron Broom.

1.5: Inclusion body formation

Inclusion bodies (IBs), such as those observed in the over-production of recombinant proteins, have historically been considered to be amorphous rather than structured aggregates such as amyloid fibrils.^{29,30} Amyloid-like fibrils are often implicated in degenerative human disease, such as Parkinson's, Alzheimers and amyotrophic lateral sclerosis (ALS) diseases. They are also commonly found in bacterial inclusion bodies, Lewy bodies and amyloides, and are involved in the cellular functions of some organisms.³¹⁻³⁴ Amyloid fibrils display characteristic properties including binding thioflavin-T (ThT) and Congo red and being enriched in beta-sheet content and seeding the aggregation of further amyloid fibrils. Structural studies have also revealed a cross-beta core motif within amyloid and amyloid-like fibrils, consisting of beta sheets stacked perpendicular to the axis of the fibril.^{29,31,32}

In contrast to amyloid aggregates observed in disease IBs are macroscopically amorphous, although they are not completely unstructured. In addition to recombinant proteins, mutations of normally soluble proteins can also result in IB formation during over-expression in bacterial systems.³⁵⁻³⁸ First documented in 1982 by Burnett *et al.* in *E. coli* producing human insulin, IBs pose a major problem in the production of proteins for commercially available antibiotics, certain over-the-counter drugs and in industrial processes.³⁹ One of the primary reasons that IBs have evoked so much interest is the discovery that not only are IBs pure (containing as much as 95% of the protein of interest, with minor protein, RNA/DNA and chaperone contaminants), they are also biologically active to some extent.^{34,40} This specificity in aggregation, leading to highly pure IBs, has been studied extensively.

IBs are known to form through specific interactions⁴¹⁻⁴³ and display many amyloid-like characteristics. They have been shown to bind amyloid-specific dyes such as ThT and Congo red; additionally, when bound to Congo red, they cause a red shift in the IB UV spectrum.^{44,45} IBs also display a characteristic apple-green birefringence and have been shown to seed the aggregation of like, soluble proteins.⁴⁴⁻⁴⁸ It has also been found that cross- β structures within IBs do not form randomly; rather, they form from sequence-specific regions that are unique to each protein.⁴² These regions may constitute portions of α -helices, β -sheets or random loops; no correlation has been found, to date, between the position of the sequence within the protein and its tendency to form the cross- β infrastructure implicated in amyloid formation.³¹ For example, the mutations made which result in IB formation in Interleukin 1- β are found within β -sheets, β -hairpins and random loops.⁴³ Surprisingly, the β -sheet content of a protein does not appear to play a role in its propensity to aggregate or to form IBs. Interleukin 1- β , for example, is a protein composed of mainly β -sheets and loops; Interleukin 2, on the other hand, is an all- α protein, but both form IBs.

Introducing as few as one or two point mutations to an otherwise soluble protein can also induce IB formation. Protein mutations are common in neurodegenerative diseases such as ALS and Huntington's disease. In the case of ALS, for example, mutations to the protein implicated in causing the disease, superoxide dismutase (SOD), often cause significant inclusion body formation and can vastly shorten a patient's life expectancy.⁴⁹ Some mutations to SOD can cause up to 94% IB formation, while others may have no marked effect.⁴⁹ If the propensity of a protein to form β -sheets increases by mutation, then the propensity to aggregate may also increase.^{50,51} It has also been found that adding particular mutations to β -sheets can induce the formation of cross- β structures.⁵²⁻⁵⁴ Turns and loops with enough hydrophobic residues introduced may also

associate and form cross- β structures, such as in the case of interleukin-1 β , where mutations to the 97th amino acid (located on a loop connecting two β -sheets) can increase the rate of IB formation by up to 50% compared to the wild type.^{49, 51,53,55}

The exact correlation between any specific mutation and the propensity to aggregate into IBs is yet unknown; however, generally introducing hydrophobic residues decreases a protein's solubility and increases its aggregation rate.^{49-51,56,57} Most aggregation events occur early after expression, and it has been hypothesized that mutations which destabilize the fast-folding native state can increase IB formation by decreasing the folding rate, thus allowing more time for aggregation events to occur.⁵⁶ Mutations may also stabilize the folding intermediate, exposing hydrophobic residues for a longer period of time and causing them to associate into aggregates or IBs.^{50,53,56} Introducing hydrophobic mutations may also change the ordering of water or solvent molecules around exposed surfaces, possibly lowering the entropy of activation for the formation of an aggregation-prone intermediate.⁵³

The effect of hydrophobicity seems to play a large factor on protein aggregation processes, as solubility is closely linked to hydrophobicity and mutations introducing hydrophobic residues.^{30,50,53,56,57-62} For example, two proteins with the same charge but different hydrophobic amino acid content--both expressed under the same conditions--can have IB formation rates which vary up to 1000-fold.^{50,63,64} However, while hydrophobicity is linked with aggregation, not all mutations which introduce hydrophobic residues will cause a protein to form IBs⁶². Generally, the fewer number of hydrophobic stretches there are in a protein, the higher its solubility upon expression, whereas the more continuous stretches of hydrophobic residues there are in a protein, the lower its solubility.^{30,50,56,58,65}

Effects of amino acid substitutions on net or average charge also play an important role on aggregation propensity. Proteins whose charges are decreased by mutation display an increased tendency to aggregate into inclusion bodies.^{30,50,56,64,66-68} However, it must be noted that unless charge is dramatically altered, the effect on overall aggregation rate may not be pronounced, as in the case of human muscle acylphosphatase (AcP) where mutations to the protein resulting in charges ranging from +1 to +5 yielded little difference in aggregation rates between mutants.⁵⁰ Between different proteins and under the same experimental conditions, charge may play a larger role in aggregation as in the case of AcP and HypF-N (E. coli HypF N-terminal domain) where the latter aggregates at a rate of nearly 1000 times the former, with the same overall charge.^{50,56,64,66}

It is now generally agreed that IBs possess certain characteristics similar to amyloid, such as ThT and Congo red binding, birefringence under cross-polarized light and having extensive cross- β structures within their cores. It is also thought that polypeptides generally possess an intrinsic propensity to form fibrils and to aggregate⁶⁹⁻⁷¹; however, the question of why some aggregate into IBs upon mutation, while others do not, is yet to be fully answered. A variety of methods have been developed in order to characterize IBs; however, we are still limited in our understanding of protein folding and aggregation.

Although the structure and physical properties of IBs have been documented and studied, there is still a lack of high resolution data. The causes of IB formation--whether due to crowding in the bacterial cytosol, stress placed upon the cell by over-expression of proteins, the inability of chaperones to mediate proper folding or otherwise--are also poorly understood. Many questions regarding the mechanism of aggregate formation in general remain. First and foremost is the issue of why proteins form aggregates, amyloid or inclusion bodies, whether it is a protective

mechanism that prevents misfolded proteins from harming the cell, or whether it is a harmful side-effect itself of misfolding. The same protein can be capable of giving rise to different types of aggregates, either completely amorphous or ordered, such as in the case of the tobacco mosaic coat virus protein and the amyloid β protein.⁷² The rate of folding also varies across the range of proteins. Each folding or aggregation pathway can have two or more component phases, each with its own kinetics and each affected by a different combination of factors. How does a protein 'choose' among all possible conformations when folding, such that one protein may be completely soluble, while another may aggregate into inclusion bodies? These problems must be solved in order to fully understand protein aggregation into IBs.

Threefoil when expressed in host *E. coli* at 37°C or 25°C forms almost exclusively inclusion bodies. This is desirable for protein purification in that the IBs may be solubilized in denaturant and isolated in high purity, then refolded to obtain the native protein. Because it is a small, globular, highly symmetric protein that can fold to its native state, it acts as a good model system toward understanding IB formation pertinent to over-expression of proteins, for therapeutic or biotechnological purposes, and to aggregation in disease and in general.

Chapter 2: Methods

2.1: Protein expression

Threefoil mutant expression and characterization was carried out with Briallen Lobb and Laura Bahlmann, both co-op students in the Meiering group.

2.1.1: Protein expression

Threefoil and Threefoil mutants were transformed into the electrocompetent *E. coli* strain BL21 (DE3) (Invitrogen) using the pET-28a expression vector incorporating genes for Threefoil (or mutants) with a poly-histidine tag used in later protein purification steps and for kanamycin antibiotic resistance in its sequence. Transformation was carried out by adding approximately 1 µg of the vector to 40 µL of electrocompetent cells. After addition of the vector, the mixture was incubated on ice for 2 minutes, then transferred to a chilled 0.1 cm gap, 1 mL electroporation cuvette (BioRad laboratories, Inc.) and subjected to 1.8 kV for ~4 ms (BioRad E. Coli Pulser, BioRad Laboratories, Inc.). Immediately after electroporation, 1 mL of SOC medium (20 g/L Bactotryptone, 5 g/L Bacto yeast extract, 10 mM sodium chloride (NaCl), 2.5 mM potassium chloride (KCL), 10 mM magnesium chloride (MgCl₂), 10 mM magnesium sulfate (MgSO₄), 20 mM glucose) was added to the cuvette and the cells were resuspended thoroughly before incubation for 1 hr at 37°C, 225 rpm. The transformed cells were then plated on Luria Broth (LB; 20 g/L Bacto-tryptone, 5 g/L yeast extract, 10 g/L NaCl) and agar (15 g/L) plates containing kanamycin (30 mg/mL) and incubated upside-down, overnight at 37°C. Single colonies were selected for re-streaking on LB/agar/kanamycin (30 mg/mL) plates and incubated again at 37°C, upside-down, overnight. One colony was then picked out from the re-streaked

plate and incubated in LB containing kanamycin (30 mg/mL) and grown overnight at 37°C, 225 rpm. The overnight LB culture was diluted 1 in 100 into LB containing kanamycin (30 mg/mL) and grown to an optical density at 600 nm of between 0.6-0.8 absorbance units prior to induction.

Induction was carried out with 1mM isopropyl- β -D-thiogalactopyranoside (IPTG) and the resulting cells allowed to express protein for 24 hours at 37°C, 225 rpm before being harvested by centrifugation (4000 x g, 20 min., 4°C).

2.1.2: Inclusion body isolation

Centrifuged cells were lysed by resuspension in a small amount (<50 mL) of ice-cold lysis buffer (300 mM sodium chloride (NaCl), 100 mM sodium phosphate (NaPi), 10 mM Tris, pH 8). The resuspended cells were then passed three times through a homogenizer (EmulsiFlex-C5, Avestin, Inc.) at 17000 psi to ensure complete lysis of the cells. DNase (50 μ g/mL) and MgCl₂ (5 mM) were added after lysis and the mixture was incubated, with shaking, on ice for 30 min. The inclusion bodies were then centrifuged (4000 x g, 20 min, 4°C), washed once with the lysis buffer and centrifuged again and stored at -80°C until ready to be used.

2.1.3: Affinity column purification

Threefoil was expressed with a poly-histidine tag (composed of ten histidines in total) for the purpose of purification. It has been shown that the poly-histidine tag does not interfere in protein folding and does not affect the native state structure of the protein.⁷³ At high pH the histidines become deprotonated, making them ideal for immobilization on an affinity column such as the nickel column used in this study. This method of protein purification has the

advantage of producing a high yield of pure protein, as impurities in the sample that do not bind to the column can be washed away. At low pH, then, the histidines are re-protonated and the protein elutes off of the column.

Folded Threefoil was obtained by solubilising the inclusion bodies in urea and purifying the protein via immobilized metal affinity chromatography (IMAC), using a nickel-nitrilotriacetic acid (Ni-NTA, QIAGEN) IMAC column. Briefly, the inclusion bodies were solubilised in buffered urea (6M urea, 300 mM NaCl, 100 mM NaPi, 10 mM Tris, pH 8) and allowed to sit for 1-3 days at room temperature in order for the proteins to be released from the inclusion bodies. A Ni-NTA IMAC column was prepared for binding the poly-His tag by running binding buffer (6M urea, 300 mM NaCl, 100 mM NaPi, 10 mM Tris, pH 8) through at a rate of 2 mL/min. The urea-solubilized inclusion bodies were then run through the column at a rate of 2 mL/min and washed with wash buffer (6M urea, 300 mM NaCl, 100 mM NaPi, 10 mM Tris, pH 6.3). Elution was carried out at low pH in elution buffer (6M urea, 300 mM NaCl, 100 mM NaPi, 10 mM Tris, pH 4.5) at a rate of 1 mL/min. The resulting protein solution, containing purified unfolded protein from inclusion bodies, was then stored at -80°C until further use.

2.1.4: Refolding and concentration of Threefoil

The unfolded inclusion body proteins in urea, obtained after Ni-NTA column purification, were dialysed to refold the protein to its native state. Dialysis was carried out using a 3.5 kDa cutoff regenerated cellulose membrane (Spectrapore) at a protein concentration of between 10-20 μ M (diluted to this concentration in wash buffer), as higher than 20 μ M the protein tended to aggregate while refolding. The protein was placed into dialysis buffer (300 mM NaCl, 100 mM NaPi, pH 6.6) and allowed to dialyse for 16 hours; longer than that, the protein

has a tendency to heavily aggregate. After dialysis the protein was filtered through a 0.2 μm filter to remove any remaining aggregates.

The dialysed, filtered protein was subsequently concentrated using a 50 mL stirred cell Amicon unit (EMD Millipore Corp.) with either a 3.5 kDa or 10 kDa cutoff filter (EMD Millipore Corp.). In order to reduce the concentration of urea remaining in the sample after dialysis, the protein was exchanged into buffer (300 mM NaCl, 100 mM NaPi, pH 6.6) until the concentration of urea was less than 0.1 mM. The protein was then stored at -80°C until further use.

2.2: Folded Threefoil characterization

2.2.1: Fluorescence

Fluorescence spectra of the dialysed, folded protein were taken on a Jobin Yvon Fluorolog-322 spectrofluorometer (Horiba, Ltd.) using a 1 cm path-length quartz cuvette. The excitation wavelength was 280 nm and the emission wavelengths measured were from 300-380 nm. The excitation slit width was 1 nm for equilibrium measurements. The emission slit widths ranged from 3-10 nm, depending on protein concentration. All spectra were measured at 25°C .

2.2.2: Circular Dichroism

CD spectra of the dialysed, folded protein were taken on a JASCO J-715 Circular Dichroism Spectropolarimeter (Jasco Inc.) with a 0.1 cm path length quartz cuvette, in buffer (300 mM NaCl, 100 mM NaPi, pH 6.6). The experimental parameters were as follows: 0.1 nm step size, 50 nm/min scan rate, 4 sec integration time and 4 averaged scans.

2.3: Threefoil kinetics

Plate reader folding and unfolding kinetic experiments were conducted with Aron Broom of the Meiering group.

2.3.1: Refolding kinetics

Refolding kinetics in very low concentrations of denaturant were slow enough to be monitored by manual mixing, and were measured by fluorescence on a SpectraMax M5 fluorescence plate reader (Molecular Devices LLC) using a 96-well clear-bottom plate. The excitation wavelength was 280 nm and the emission wavelength measured was 313 nm, where the largest signal change was observed to occur over the time span of the experiment. Protein concentrations used were typically between 3-4 μM . Due to instability in temperature control near 25°C, kinetic experiments involving sugar- and metal-binding were conducted at 27°C, where there was more reliable temperature control.

Natively folded protein was unfolded in 3M GuSCN (3M GuSCN, 100 mM NaPi, 300 mM NaCl, pH 6.6) and allowed to sit for 3 days to allow the protein to unfold completely. In the case of refolding from inclusion body proteins, the unfolded stock protein used was that taken directly off of the Ni-NTA column, in elution buffer as described in 2.1.3.

Refolding kinetics were conducted by rapidly diluting the protein into refolding buffer (300 mM NaCl, 100 mM NaPi, pH 6.6), containing varying concentrations of denaturant, and mixed by pipetting. In the case of sugar-binding studies, the refolding buffer also contained between 0.5-50 mM of the sugars studied (sucrose and lactose). In the case of metal-binding kinetic studies, the concentration of metal studied (NaCl) ranged between 0-300 mM and a

metal-free refolding buffer was used composed of 100 mM ammonium phosphate $((\text{NH}_4)_3\text{PO}_4)$, pH 6.6.

2.3.2: Unfolding kinetics

Unfolding kinetics in urea and guanidine hydrochloride (GuHCl) were too slow to be measured over the course of even several hours at $>6\text{M}$ denaturant, so it was necessary to use guanidine isothiocyanate (GuSCN), a much more powerful denaturant. Unfolding kinetics of native Threefoil as well as of urea-denatured Threefoil were measured by fluorescence using the same experimental parameters as in refolding kinetics (2.3.1), by manual mixing.

Briefly, natively folded protein (or urea-denatured inclusion body protein) was diluted into unfolding buffer, containing varying concentrations of GuSCN and an unfolding buffer (300 mM NaCl, 100 mM NaPi, pH 6.6) and manually mixed. In the case of sugar-binding studies, the unfolding buffer also contained between 0.5-50 mM of the sugars studied (sucrose and lactose). In the case of metal-binding kinetic studies, the concentration of metal studied (NaCl) ranged between 0-300 mM and a metal-free unfolding buffer was used composed of 100 mM ammonium phosphate $((\text{NH}_4)_3\text{PO}_4)$, pH 6.6.

2.3.3: Fitting of kinetic traces

Kinetics traces of refolding and unfolding were fit to a single exponential (Equation 1.6) without a linear drift. Threefoil was found to photobleach when kinetic experiments were conducted on the Fluorolog 3-22 spectrophotometer, as the signal for kinetics would systematically decrease over time where the signal was not expected to change (i.e at equilibrium). Photobleaching was suspected as the signal was found to recover after the

fluorimeter shutter was closed for a period of time (several minutes). Further evidence for photobleaching lies in the signal for a natively folded Threefoil sample decreasing over time and recovering after the shutter was closed for several minutes. As no kinetics are expected for the natively folded protein, the signal was not expected to decrease. Photobleaching was significantly reduced by reducing the excitation slit widths from 1 nm to 0.5 nm, and was not found to contribute significantly to the signal change on the plate reader. The fitting software used for all kinetic experiments was Biokine 32 (Biologic, Version 4.65).

2.4: Threefoil equilibrium studies

2.4.1: Renaturation equilibrium

Renaturation equilibrium samples were prepared by diluting unfolded protein into refolding buffer (100 mM NaPi, 300 mM NaCl, pH 6.6) containing varying concentrations of denaturant. Unfolded protein was prepared by diluting natively folded protein 2-fold into 6M GuSCN and allowing the protein-denaturant solution to sit at room temperature for 3 days to allow the protein to completely unfold. Renaturation equilibrium curves for inclusion body proteins were made from stock unfolded protein obtained directly from the Ni-NTA column, in 6M urea.

Renaturation equilibrium samples were monitored by fluorescence and by CD over the course of 8 months, which is the approximate time needed for the samples to reach equilibrium. Fluorescence spectra were taken of each sample and an equilibrium curve was constructed by plotting the signal at either 313 nm or 370 nm against the concentration of denaturant of each sample. The fluorescence parameters were as described in section 2.2.1. CD spectra for urea and GuHCl equilibrium curves were also taken for each sample and an equilibrium curve was

constructed in a similar fashion to that for fluorescence, save that the signals at 215 nm and 230 nm were plotted versus denaturant concentration. Parameters for the CD spectra measurements were as described in 2.2.2.

2.4.2: Denaturation equilibrium

Denaturation equilibrium samples were prepared by diluting natively folded protein into unfolding buffer (as described in 2.3.2) containing varying concentrations of denaturant. Because Threefoil does not reach equilibrium when denatured in urea or GuHCl in a practical time frame, only denaturation equilibrium in GuSCN was studied.

Samples were measured over the course of 3 years by fluorescence, using the same fluorimeter parameters as for the renaturation equilibrium studies.

2.4.3: Fitting of equilibrium data

Equilibrium data were fit in OriginPro 5.0 (OriginLab Corp.) according to the linear extrapolation method (LEM) to a two-state transition between the native and the denatured states (Equation 1.5). The pre-transition baseline of the equilibrium curves was assumed to have a slope of 0 (unless otherwise stated) and the post-transition baseline was allowed to vary.

2.5: Metal removal from Threefoil

Removal of the metal from natively folded Threefoil for kinetics and equilibrium studies was achieved through extensive exchange (12-15 times, from a volume of 200 mL to 5 mL for each exchange) into ammonium phosphate buffer (100 mM $(\text{NH}_4)_3\text{PO}_4$) containing ammonium acetate (2mM $\text{CH}_3\text{COONH}_4$, pH 8.5), using an Amicon device and a 3.5 kDa cutoff filter.

The protein thus produced was analysed for metal content by both denatured and native mass spectrometry (MS) at the on-site mass spectrometry facility at the University of Waterloo. Spectra were taken on a Micromass Q-TOF Ultima Global mass spectrometer (MS Vision). The buffer used in native MS consisted of the exchange buffer described above, with the addition of $\text{CH}_3\text{COONH}_4$ to a final concentration of 10 mM or 20 mM. The buffer used in denatured MS consisted of 1:1 acetonitrile:water and 0.2% (v/v) formic acid. The expected mass was calculated using the amino acid sequence of Threefoil and compared to the obtained experimental mass.

2.6: Inclusion body characterization

Inclusion bodies were characterized by fluorescence, CD, ANS binding and differential scanning calorimetry (DSC) as proteins eluted off of the Ni-NTA column in 6M urea and elution buffer as described in 2.1.3. Inclusion bodies obtained after harvesting cells were used directly as is for Fourier transform infra-red spectroscopy.

2.6.1: Fluorescence

Fluorescence spectra of unfolded inclusion body proteins were taken using the same parameters as described in 2.2.1.

2.6.2: Circular dichroism

CD spectra of unfolded inclusion body proteins were taken using the same parameters as described in 2.2.2.

2.6.3: Differential scanning calorimetry

Differential scanning calorimetry (DSC) measurements were carried out using a MicroCal LLC VP-DSC calorimeter (MicroCal Inc.). Samples were scanned at a rate of 1°C/min from 25°C to 95°C, unless otherwise specified, for both up and down scans. Buffer/buffer baseline scans were taken and subtracted from protein/buffer scans and the resulting data was normalized for protein concentration. All samples were degassed, for 10 min with stirring, prior to loading into the DSC calorimeter. Scan rate dependence measurements in 0 M denaturant were performed by Aron Broom of the Meiering group.

2.6.4: ANS binding

1-Anilinonaphthalene-8-Sulfonic Acid (ANS, Sigma) binding to unfolded proteins from inclusion bodies was studied by fluorescence. Sample preparation was carried out by addition of ANS (25 μ M) to 10 μ M unfolded inclusion body protein in buffered 6M urea, 10 μ M natively folded protein unfolded in 3M GuSCN (unfolded control) and 10 μ M natively folded protein (folded control) respectively. ANS (25 μ M) was also added to each of the matching buffers without protein, for background signal subtraction. Samples were allowed to incubate in ANS for 10 minutes prior to measurement. The ANS-bound samples were excited at 360 nm and signal emission was monitored between 400 and 600 nm.

2.7: SDS-PAGE gel analysis of Threefoil mutants

SDS-PAGE gels used in the analysis of Threefoil wild-type and mutant aggregation propensity consist of a 12% running (lower) gel consisting of 40% acrylamide (BioShop Canada), 1.5 M Tris (BioShop Canada) pH 8.8, 10% sodium dodecyl sulfate (SDS, BioShop

Canada), 10% ammonium persulfate (APS, BioShop Canada) and 1% Tetramethylethylenediamine (TEMED, BioShop Canada). A 5% stacking (upper) gel was also used consisting of 40% acrylamide (BioShop Canada), 0.5 M Tris (BioShop Canada) pH 6.8, 10% sodium dodecyl sulfate (SDS, BioShop Canada), 10% ammonium persulfate (APS, BioShop Canada) and 1% Tetramethylethylenediamine (TEMED, BioShop Canada). SDS-PAGE gels were run at 175V for 2 hours with cooling, under constant voltage.

Chapter 3: Results

Structural transitions were monitored using mainly tryptophan fluorescence, with a small contribution from tyrosine fluorescence at the wavelength of excitation used. Threefoil contains three tryptophans which are buried in the core. Unfolding of Threefoil results in the solvent-exposure of the tryptophans (Figure 1.4), causing a red-shift in the signal of maximum wavelength. Comparison of the folded and unfolded spectra indicates that this exposure decreases the fluorescence signal due to quenching by the solvent and also shifts the maximum wavelength of emission toward longer wavelengths. Figure 3.1 shows the change in fluorescence spectra of Threefoil upon addition of GuSCN.

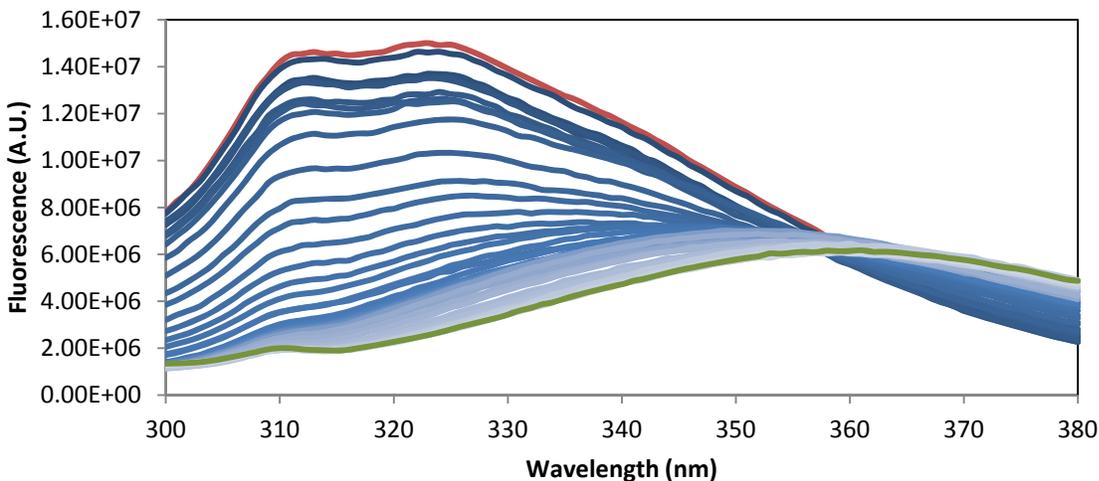


Figure 3.1: Fluorescence spectra of the transition from folded to unfolded Threefoil

The folded spectrum for tryptophan fluorescence for Threefoil (red) is characterized by peaks at 313 nm and 323 nm, when excited at 280 nm. At 310 nm there may be a contribution to fluorescence from tyrosine residues, as well as contribution from the water Raman signal; however, the major contribution to fluorescence is from tryptophan residues, at 323 nm.

The unfolded spectrum is shown in green and shows a wavelength maximum at ~361 nm. The presence of a small peak at ~310 nm in the unfolded Threefoil spectrum corresponds to the water Raman peak as well as a small contribution from tyrosine emission.

There is a clear red-shift from the folded spectrum to unfolded spectrum, indicating solvent-exposure of tyrosine residues. Spectra in between (dark blue to light blue) represent Threefoil in various stages of unfolding and show a gradual red-shift of the spectrum, as well as a decrease in fluorescence intensity, as the tryptophan fluorescence becomes quenched by the buffer. Spectra were acquired in 100 mM NaPi, 300 mM NaCl pH 6.6 buffer on a Fluorolog 3-22 spectrophotometer, at 25°C. Excitation wavelength was 280 nm, with respective excitation and emission slit widths of 1 and 5 nm, using gratings with a blazing of 500 nm.

3.1: Residual structure in urea and GuHCl

In 6M urea, solubilized Threefoil IBs were found to have extensive secondary residual structure and elements of residual tertiary structure. (Figures 3.2 and 3.3) The fluorescence emission maximum for the spectrum of protein unfolded in urea is ~10 nm blue-shifted from that of protein unfolded in 6M GuSCN. Threefoil is believed to be almost fully denatured in 6M GuSCN because its fluorescence spectrum compares closely with the spectrum of pure L-

tryptophan in 6M GuSCN (Figure 3.2). Presumably in its fully denatured state, Threefoil's tryptophans would likewise be completely solvent-exposed.

The maximum fluorescence in 6M urea is ~350 nm, compared to the same protein in 6M GuSCN, which has a maximum fluorescence at ~360 nm (Figure 3.2). By CD, in 6M urea there is quite a noticeable amount of secondary structural elements that remain, as evidenced by the peak at ~230 nm which is sensitive to the environment of aromatic side-chains as well as backbone structure (Figure 3.3).²⁸ Onefoil is used for comparison, because it has no tertiary structure but contains extensive β -structure,²⁸ which appears between 216-218 nm in the CD spectrum. It is clear that Threefoil inclusion bodies even when denatured in 6M urea also contains a significant secondary structure. Interestingly Threefoil IBs in 6M GuHCl retains as much secondary structure as in 6M urea, despite GuHCl being a stronger denaturant than urea (Figure 3.3); however, there is less residual structure as seen by fluorescence (Figure 3.2).

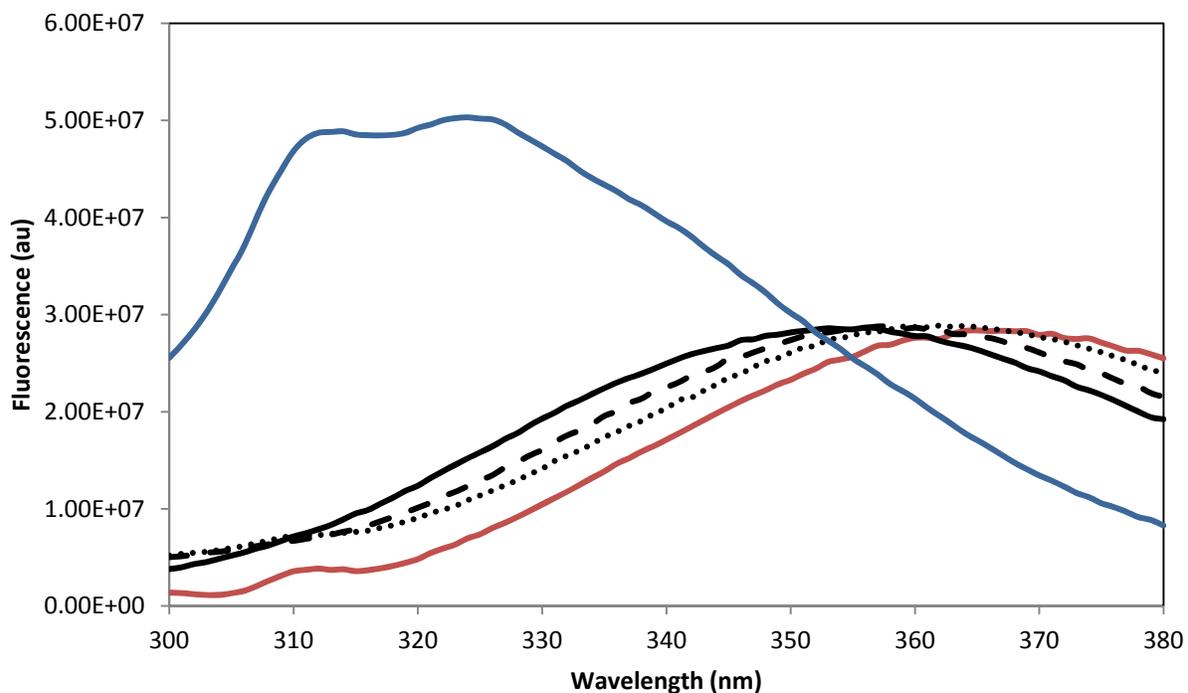


Figure 3.2: Fluorescence spectra of Threefoil IBs in 6M urea, 6M GuHCl and 6M GuSCN

The maximum fluorescence signal for urea-denatured inclusion body proteins (solid line) occurs at ~350 nm, while the maximum signal for 6M GuHCl-denatured Threefoil IBs (dashed line) occurs at ~357 nm. In 6M GuSCN (dotted line) in the most strongly denaturing conditions, the maximum fluorescence occurs at 363 nm. Under the same denaturing conditions in GuSCN, L-tryptophan (red) has a maximum fluorescence at 367 nm. Folded Threefoil is also given as reference (blue). This demonstrates that even at high urea and GuHCl concentrations the proteins still possess residual structure, since in both denaturants the spectrum is blue-shifted compared to GuSCN, indicating that the tryptophan residues being monitored are not completely solvent-exposed, and that the protein is not fully unfolded.

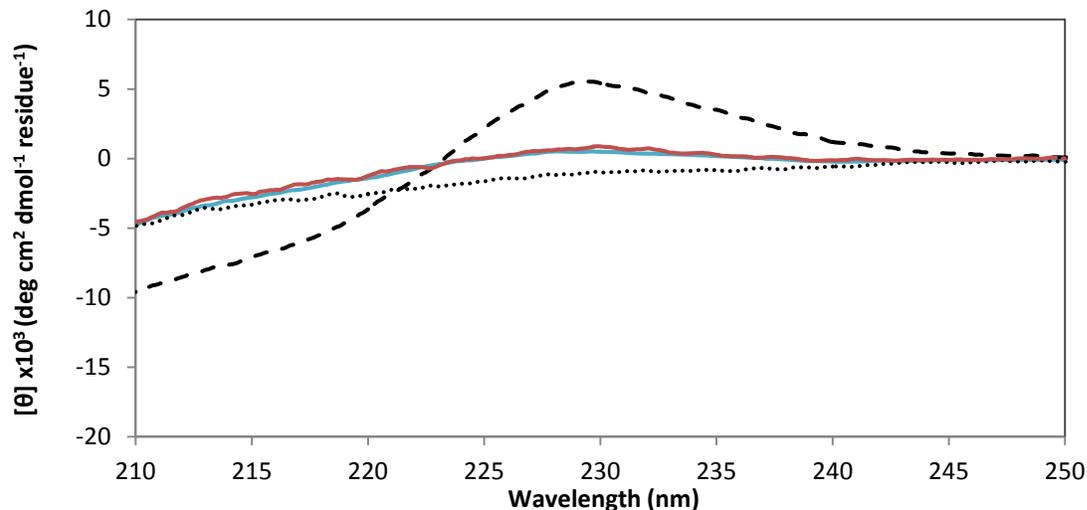


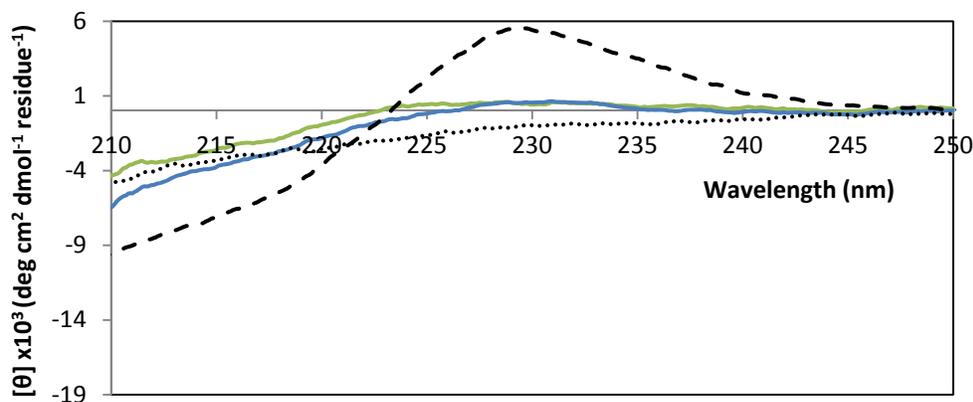
Figure 3.3: CD spectra of urea- and GuHCl-denatured inclusion body proteins, native Threefoil and Onefoil

Onefoil (dotted line) is used as an unfolded reference in the absence of usable unfolded protein spectra in GuSCN, which absorb light at the wavelengths of interest at high concentrations. Native Threefoil (black dashed line) is also included as reference. In 6M urea (blue solid line), Threefoil shows evidence of extensive secondary structure. The signal at 218 nm indicates the presence of extensive β -sheet structure, similar to Onefoil, and the positive signal at 230 nm indicates that the aromatic side-chains are not completely solvent-exposed. Likewise for inclusion bodies denatured in 6M GuHCl (solid red spectrum), there is extensive secondary beta structure. Spectra of folded Threefoil and Onefoil courtesy of Aron Broom.^{27,28}

It should also be noted that Threefoil resolubilized in 6M urea tend to contain small soluble aggregates, which can be removed by centrifugation and filtration (Figure 3.4A). The presence of such aggregates affects the CD spectrum of Threefoil (Figure 3.4A), and can further be detected by dynamic light scattering (DLS) to have a larger hydrodynamic radius than protein that has been centrifuged and filtered (Figure 3.4B). Estimated according to the empirical equation determined by Wilkins *et. al.*⁷⁴ for hydrodynamic radii estimated from NMR methods of a range of proteins in the native and in highly denatured states, a hydrodynamic diameter of ~ 7.4 nm is expected for Threefoil from DLS measurements. As seen in Figure 3.4B, Threefoil samples which have not been centrifuged or filtered appear have a diameter of ~ 10 nm, *versus* ~ 7.5 nm for the diameter of centrifuged and filtered protein, indicating the presence of aggregates in untreated samples. Upon centrifugation and filtration there is $\sim 5\%$ decrease in the

protein concentration, suggesting that the major observed optical changes may nevertheless correspond to non-aggregated protein.

(A)



(B)

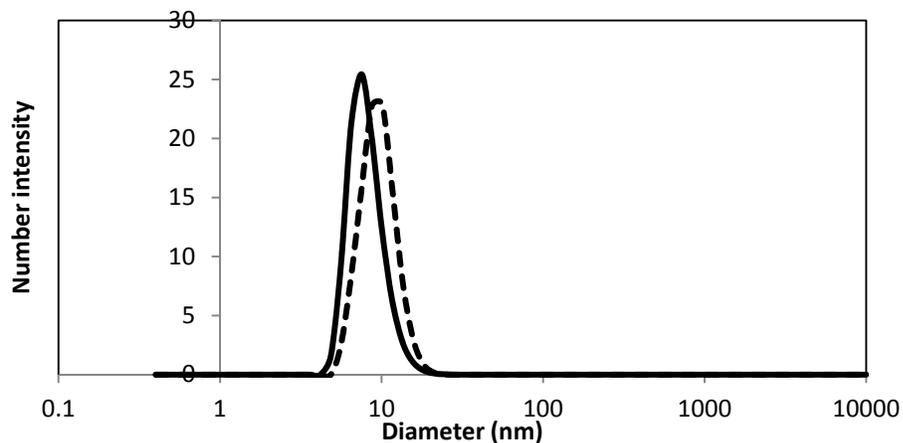


Figure 3.4: The effect of centrifugation and filtration on Threefoil denatured in 6M urea

(A) CD spectrum of protein that has (green spectrum) and hasn't (blue spectrum) been centrifuged and filtered. It is evident that while filtration and centrifugation doesn't affect the signal of the aromatic residues near 230 nm, it does affect the beta-sheet signal near 218 nm. The aggregates that form in 6M urea-denatured inclusion body samples appear to have more beta-sheet content.

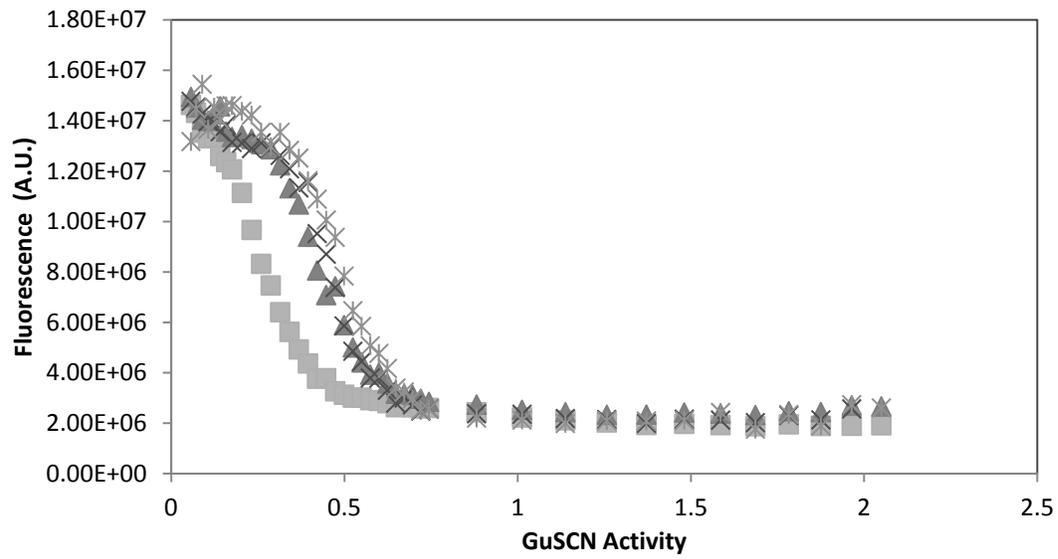
(B) DLS scans of protein prior to centrifugation and filtration (dashed spectrum) and after centrifugation and filtration (solid spectrum). When aggregates are in solution, the maximum diameter is larger, at approximately 10 nm, versus approximately 7.5 nm for centrifuged and filtered protein.

3.2: Equilibrium stability of Threefoil

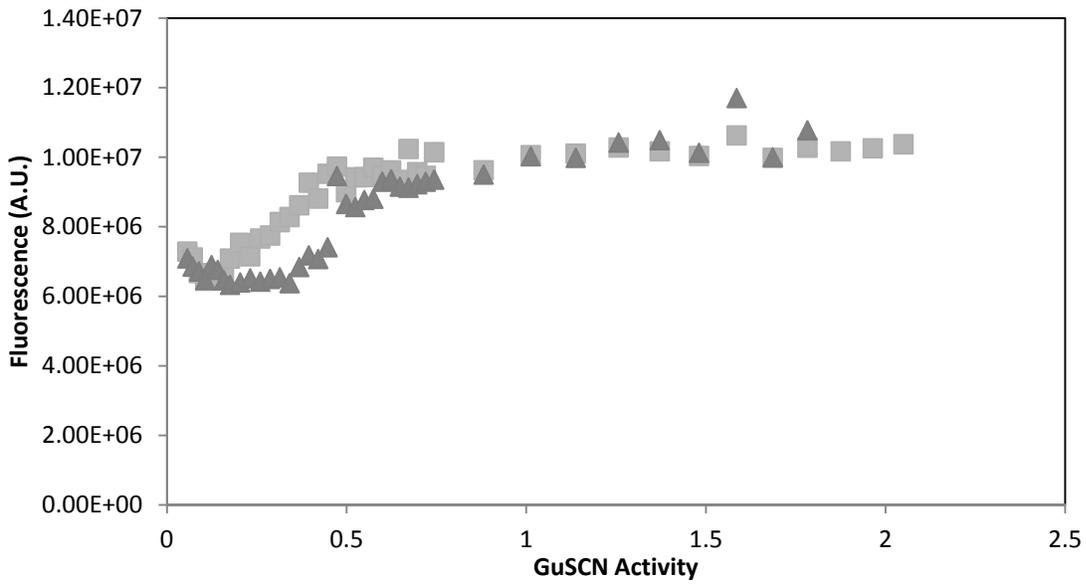
3.2.1: Refolding curves of metallated Threefoil

Refolding equilibrium of Threefoil was studied by denaturing native Threefoil in GuSCN, then diluting the unfolded protein into refolding buffer containing varying concentrations of GuSCN. Equilibrium was established very slowly, on the order of ~ 100 days for renaturation. For denaturation, at the end of two years equilibrium was still not reached. The equilibrium curve was fit according to a model for a reversible two-state transition between the native folded state to the unfolded state, using the linear extrapolation method, according to Equation 1.5 (Figure 3.5). At equilibrium the midpoint of the transition (C_{mid}) is 0.53 ± 0.02 GuSCN activity units (D), the denaturant-dependence of dG (m) is 8.77 ± 1.33 kcal/(mol $\cdot D$) and the ΔG° value is 4.65 ± 0.73 kcal/mol. Values from fitting are also summarized in Table 1. The curves monitored at 370 nm were not included in the fitting, as they were generally too scattered to obtain reliable numbers.

(A)



(B)



(C)

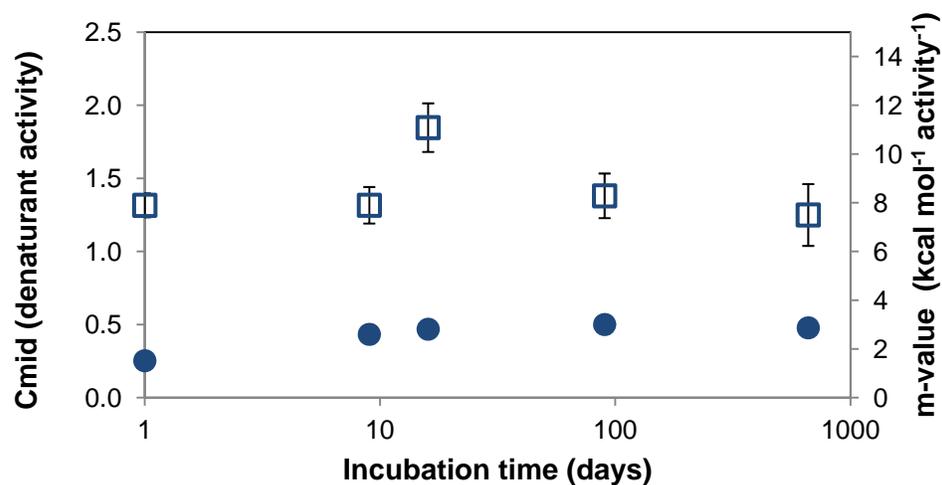


Figure 3.5: Equilibrium renaturation curves in GuSCN for Threefoil, monitored at 313 nm

Equilibrium of renaturation in GuSCN was established after approximately 100 days, but curves were followed for two years. The data fit fairly well to a two-state transition using the linear extrapolation method. A sloping pre-transition baseline and a non-sloping post-transition baseline was assumed.

Curves are represented as follows: (A) Fluorescence monitored at 313 nm at (■) 24 hours; (▲) 10 days; (×) 17 days; (*) 660 days.

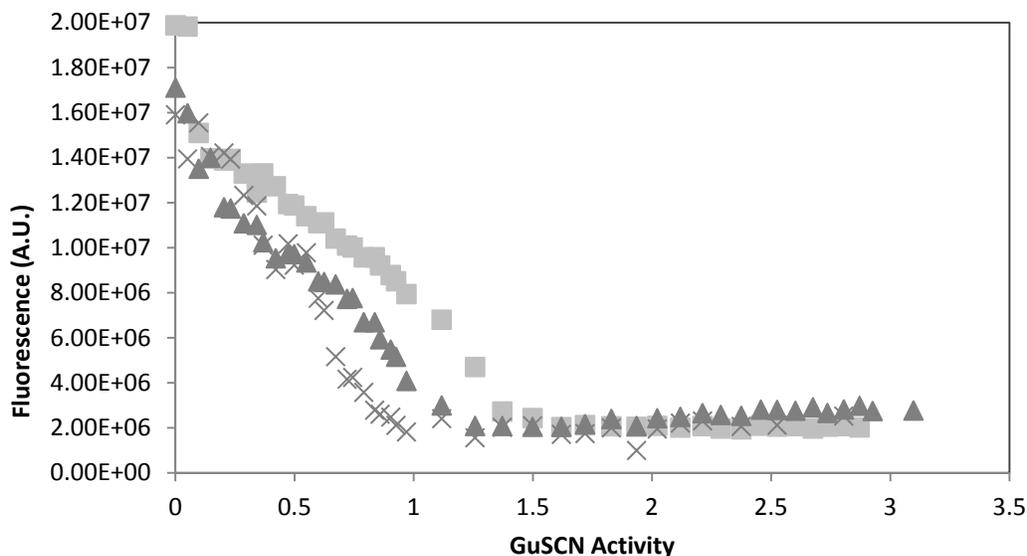
(B) Fluorescence monitored at 370 nm at (■) 24 hours; (▲) 15 days.

C: Cmid and m-values obtained from fitting equilibrium curves monitored at 313 nm to equation 1.5, assuming a sloping pre-transition baseline but a non-sloping post-transition baseline. Closed circles represent Cmid values; open squares represent m-values. The equilibrium Cmid is 0.53 ± 0.02 GuSCN activity units (D), and the equilibrium m-value is 8.77 ± 1.33 kcal/(mol*D).

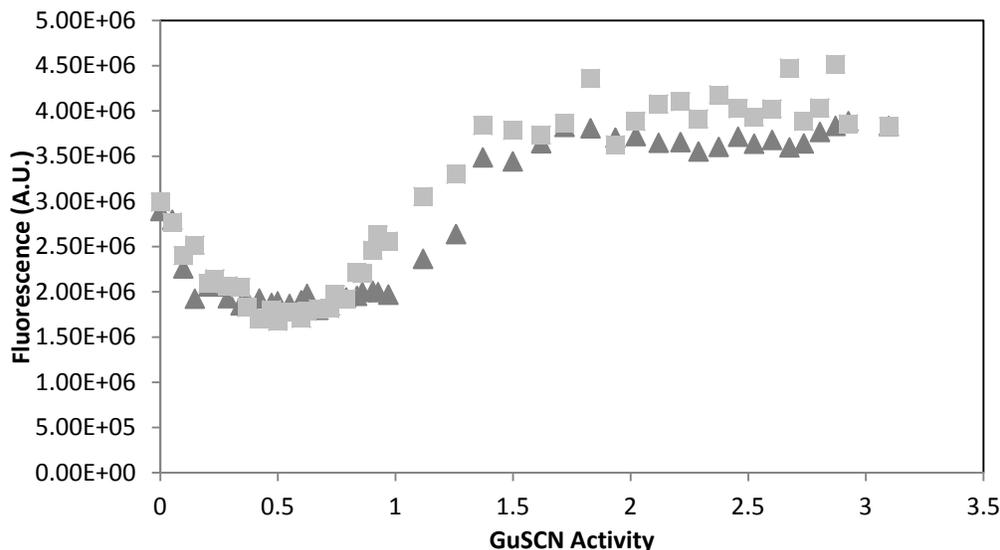
3.2.2: Unfolding curves of metallated Threefoil

Unfolding equilibrium of Threefoil was studied by denaturing native Threefoil in unfolding buffer containing various concentrations of GuSCN. Equilibrium was not established even after 3 years, as the signal continued to change over time. The equilibrium denaturation curve was also fit according to a two-state model with a sloping pre-transition baseline and a non-sloping post-transition baseline, using the linear extrapolation method, according to Equation 1.5 (Figure 3.6). The midpoint of the transition (C_{mid}) was 0.72 ± 0.04 GuSCN activity units (D), the denaturant-dependence of $dG(m)$ was 8.23 ± 1.41 kcal/(mol*D) and the ΔG° value was determined to be 5.96 ± 1.06 kcal/mol. Values from fitting are also summarized in Table 1. The curves monitored at 370 nm were not included in the fitting, as they were generally too scattered to obtain reliable numbers.

(A)



(B)



(C)

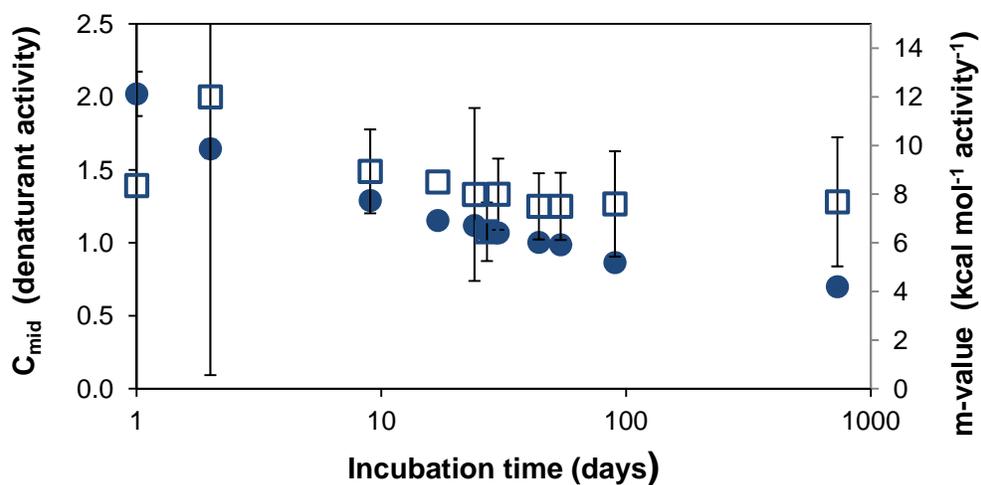


Figure 3.6: Equilibrium denaturation curves in GuSCN for Threefoil

Equilibrium of denaturation in GuSCN was not established even after 2 years. The data fit fairly well to a two-state transition using the linear extrapolation method. A sloping pre-transition baseline and a non-sloping post-transition baseline was assumed.

Curves are represented as follows: A: Fluorescence measured at 313 nm at (■) 9 days; (▲) 54 days; (×) 730 days.

B: Fluorescence monitored at 370 nm at (■) 9 days; (▲) 44 days

C: C_{mid} and m -values obtained from fitting equilibrium curves monitored at 313 nm to equation 1.5, assuming a sloping pre-transition baseline but a non-sloping post-transition baseline. Closed circles represent C_{mid} values; open squares represent m -values. The C_{mid} is 0.72 ± 0.04 GuSCN activity units (D), and the m -value is 8.23 ± 1.41 kcal/(mol*D).

After approximately two years the midpoint of the renaturation and denaturation curves do not coincide (Figure 3.7), due to the fact that the denaturation equilibrium samples have not yet reached equilibrium.

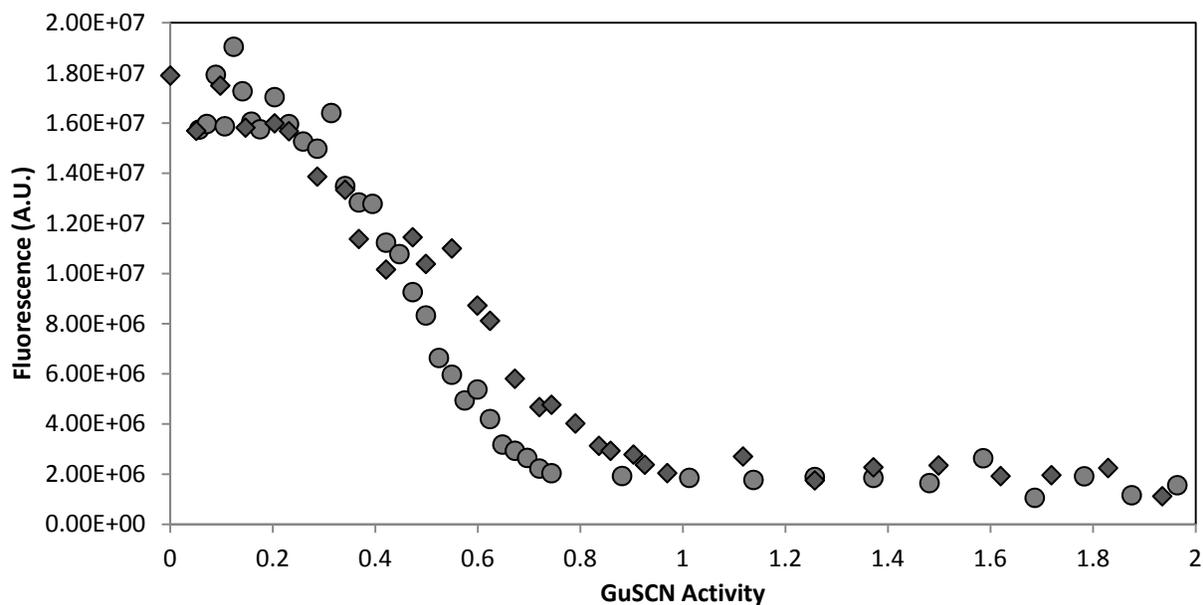


Figure 3.7: Equilibrium renaturation and denaturation curves for Threefoil in GuSCN

At the conclusion of two years, the curves do not coincide, due to the fact that the denaturation samples have not yet reached equilibrium. Curves were scaled for the post-transition baseline to show the relative positions of the midpoints. (●) Renaturation curve at 2 years; (◆) Denaturation curve at 2 years.

Denaturant		Probe	Time	Cmid	error	m-value	error	dG	error
			(Days)	(D) ^a		(kcal/(mol*D))		(kcal/mol)	
GuSCN	Denaturation	FL 313	1	2.02	0.15	8.35	11.66	16.88	23.60
			2	1.64	0.06	12.00	11.44	19.73	18.83
			9	1.29	0.02	8.94	1.73	11.53	2.23
			17	1.15	0.03	8.50	0.00	9.80	0.23
			24	1.12	0.04	7.99	3.55	8.94	3.99
			27	1.08	0.03	6.45	1.20	6.95	1.31
			30	1.07	0.02	8.00	1.46	8.54	1.57
			44	1.00	0.02	7.50	1.36	7.52	1.37
			54	0.99	0.02	7.50	1.38	7.40	1.37
			90	0.86	0.04	7.60	2.17	6.57	1.89
			730	0.70	0.05	7.68	2.65	5.37	1.90
				Average					
GuSCN	Renaturation	FL 313	1	0.25	0.02	7.90	0.49	2.00	0.20
			9	0.43	0.01	7.90	0.75	3.42	0.34
			16	0.47	0.01	11.08	1.00	5.20	0.47
			90	0.50	0.01	8.29	0.92	4.15	0.47
			660	0.52	0.02	8.69	1.84	4.52	0.97
				Average					3.86
Eqm value	Denaturation	FL 313	730	0.72	0.04	8.23	1.41	5.96	1.06
	Renaturation	FL 313	660	0.53	0.02	8.77	1.33	4.65	0.73

Table 1: Summary of parameters from fitting of GuSCN denaturation and renaturation curves monitored by fluorescence (FL)

^aD denotes GuSCN activity, which is calculated from GuSCN concentration according to equation 3.1.

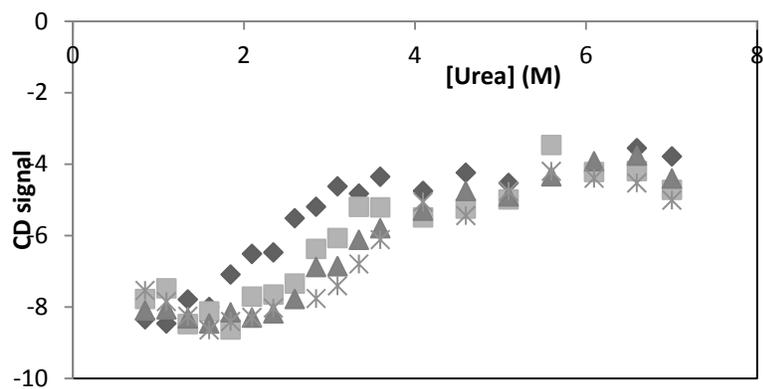
3.2.3: Renaturation equilibrium studies of Threefoil inclusion bodies

Due to the incredibly long amount of time that it takes Threefoil to reach equilibrium, it was necessary to monitor the samples over a period of several months. Renaturation equilibrium was monitored by two spectral probes, fluorescence and CD, at wavelengths which report on the tertiary and secondary structures respectively of Threefoil during folding (Figure 3.8). The parameters (C_{mid} and m -value) obtained from globally fitting urea renaturation curves monitored by CD and fluorescence (Figure 3.9) agree within error, suggesting that the transition is cooperative and two-state, when refolding from proteins isolated from inclusion bodies with residual structure in the denatured state. Due to scatter in the data causing large errors in the fitted values, curves were globally fit to Equation 1.5 in order to find the trend in m -values, with shared sloping pre- and post-transition baselines and varying m -values and C_{mid} (Appendix 1).

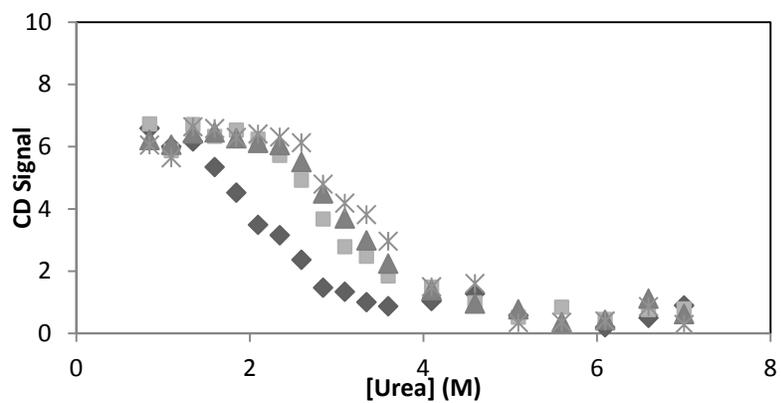
Within the same optical probe, monitoring the folding equilibrium at two different wavelengths also yielded similar m -values and C_{mid} values, which again points to the transition being two-state with residual structure. The curves were fit to a two-state transition assuming a sloping pre-transition baseline as well as a sloping post-transition baseline. Values from fitting curves individually (Appendix 2) are summarized in Table 2, in order to obtain a reliable C_{mid} value near equilibrium.

While renaturation equilibrium in GuHCl was monitored by CD and fluorescence as well, the curves were too scattered to be accurately fit. However, folding kinetics (discussed in later sections) could be reconstructed from the equilibrium curves, since the kinetics at any given protein concentration depends only on the signal change with time.

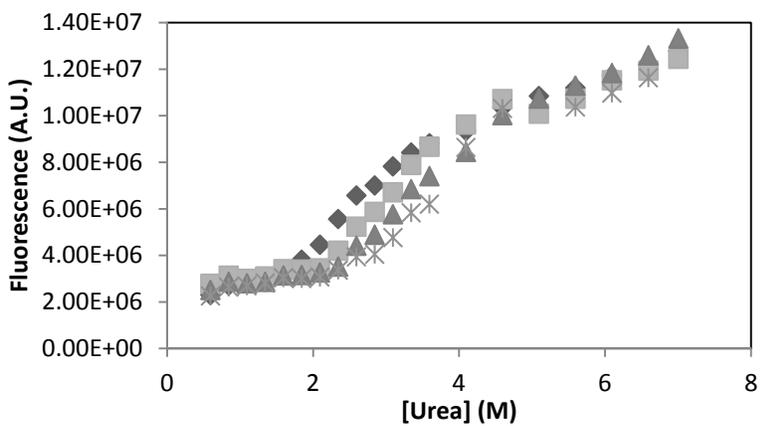
(A)



(B)



(C)



(D)

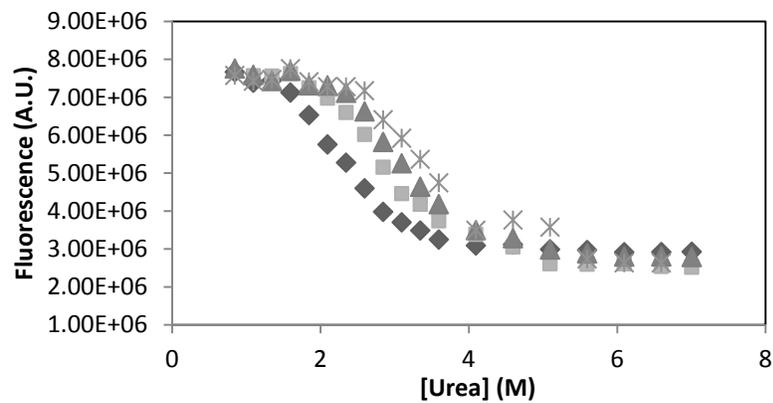
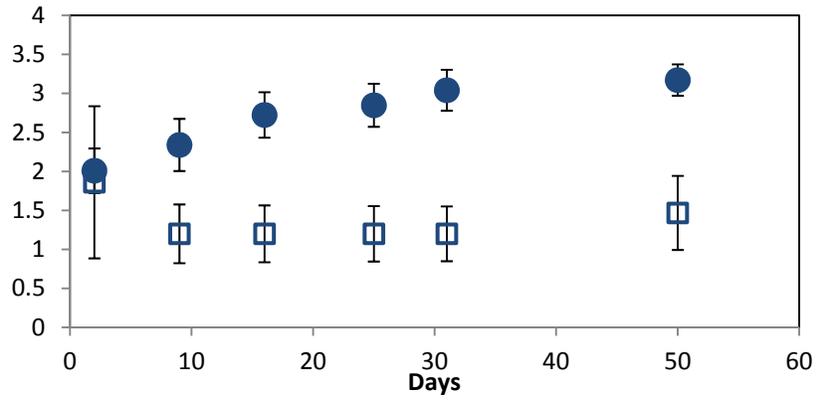


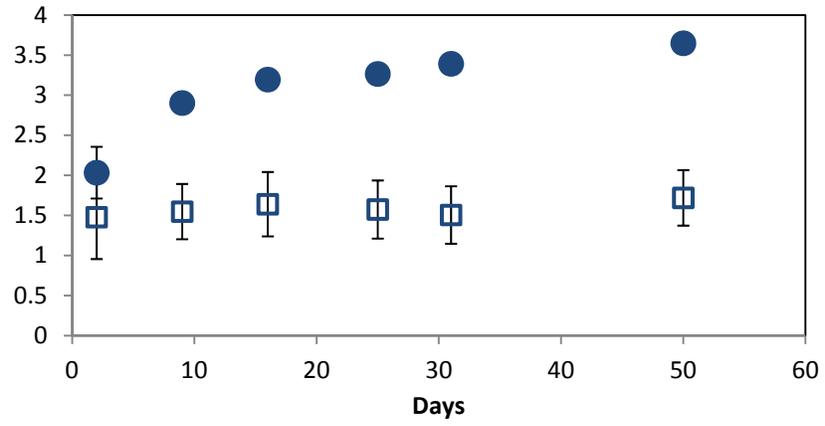
Figure 3.8: Renaturation equilibrium curves over time in urea

(A) Curves monitoring CD signal at 215 nm which reports on secondary beta-sheet structure. (B) Curves monitoring CD signal at 230 nm reporting on tertiary structure, of the aromatic residues and their exposure to solvent. (C) Curves monitoring fluorescence signal at 370 nm. (D) Curves monitoring fluorescence signal at 313 nm. Curves shown were measured on: (♦) Day 2; (■) Day 9; (▲) Day 16; (*) Day 31

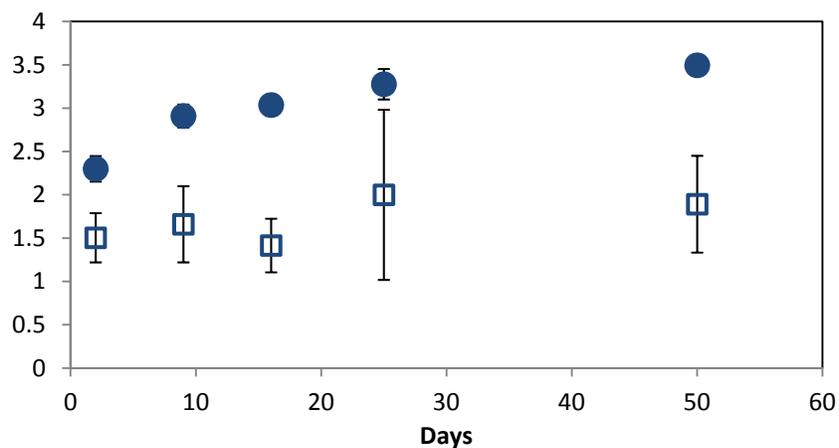
(A)



(B)



(C)



(D)

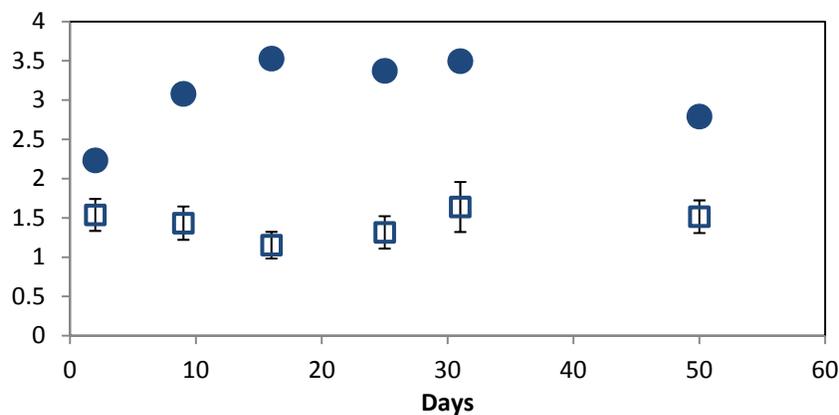


Figure 3.9: Cmid and m-values from global fitting of urea renaturation equilibrium curves, monitored at different wavelengths and by different optical probes

Closed circles represent m-values; open squares represent Cmid values. All curves were fit to a two-state transition with sloping pre- and post-transition baselines according to equation 1.5.

(A) Fitting for curves monitoring CD signal at 215 nm which reports on secondary beta-sheet structure. (B) Fitting for curves monitoring CD signal at 230 nm reporting on tertiary structure, of the aromatic residues and their exposure to solvent. (C) Fitting for curves monitoring fluorescence signal at 370 nm. (D) Fitting for curves monitoring fluorescence signal at 313 nm.

Denaturant		Probe	Time (Days)	Cmid (M)	error	m-value (kcal mol ⁻¹ M ⁻¹)	error	dG (kcal/mol)	error
Urea	Renaturation	CD 215	2	2.46	0.33	2.12	1.47	5.22	3.69
			9	2.42	0.36	1.63	1.01	3.96	2.51
			16	2.71	0.14	1.71	0.51	4.61	1.40
			25	2.73	0.13	2.00	0.65	5.47	1.79
			31	3.10	0.19	1.61	0.65	4.99	2.04
			50	3.07	0.28	1.38	0.71	4.24	2.21
		Average						4.75	0.58
Urea	Renaturation	CD 230	2	2.04	0.08	1.42	0.27	2.91	0.57
			9	2.85	0.21	1.66	0.52	4.72	1.51
			16	2.97	0.09	1.51	0.24	4.47	0.72
			25	3.22	0.09	1.35	0.21	4.35	0.69
			31	3.05	0.37	1.15	0.59	3.52	1.84
			50	3.64	0.14	1.73	0.53	6.29	1.94
		Average						4.38	1.15
Urea	Renaturation	FL 313	2	2.30	0.03	1.63	0.11	3.76	0.27
			9	2.89	0.04	1.30	0.12	3.75	0.36
			16	3.16	0.03	1.59	0.14	5.04	0.45
			25	3.58	0.15	0.99	0.24	3.53	0.87
			31	3.46	0.08	1.66	0.35	5.73	1.21
			50	3.63	0.08	1.82	0.38	6.61	1.37
		Average						4.74	1.26
Urea	Renaturation	FL 370	2	2.45	0.10	1.55	0.32	3.80	0.80
			9	2.83	0.09	1.70	0.36	4.81	1.03
			16	3.34	0.12	1.26	0.24	4.21	0.83
			25	3.28	0.07	2.45	0.67	8.02	2.22
			31	3.42	0.09	1.84	0.46	6.28	1.58
			50	3.69	0.13	1.19	0.23	4.41	0.86
		Average						5.26	1.60
Eqm value^a									
Urea	Renaturation	CD 215		3.29	0.10	1.74	0.27	5.72	0.91
Urea	Renaturation	CD 230		3.53	0.10	1.47	0.21	5.19	0.76
Urea	Renaturation	FL 313		3.49	0.10	1.50	0.30	5.23	1.07
Urea	Renaturation	FL 370		3.50	0.10	1.66	0.46	5.82	1.61

Table 2: Summary of parameters from individual fitting of urea renaturation equilibrium curves monitored by CD and fluorescence (FL)

^aEquilibrium m-values are given as the average across all m-values from fitting, while equilibrium Cmid values were obtained from fitting the measured Cmid values to a single exponential equation.

3.3: Threefoil kinetics

3.3.1: Unfolding kinetics of metallated Threefoil

Threefoil is remarkably resistant to unfolding, even in the presence of strong denaturants such as guanidine hydrochloride (GuHCl). Guanidine isothiocyanate (GuSCN), is a denaturant that is ~3.5 times as powerful as GuHCl and ~7 times as powerful as urea.^{12,75} GuSCN was required in order to obtain unfolding kinetics in a practical amount of time at 25°C. Even in 4M GuSCN, i.e. in extremely strongly denaturing conditions, the kinetics are still so slow that they could be measured by fluorescence using manual mixing techniques.

The major transition of kinetic unfolding experiments fit well to a single exponential process (Figure 3.10), consistent with a single kinetic process. Unfolding kinetics were not obtainable in GuSCN monitored by CD, as GuSCN absorbs light strongly in the spectral range studied. More representative unfolding kinetic traces are given in Appendix 3.

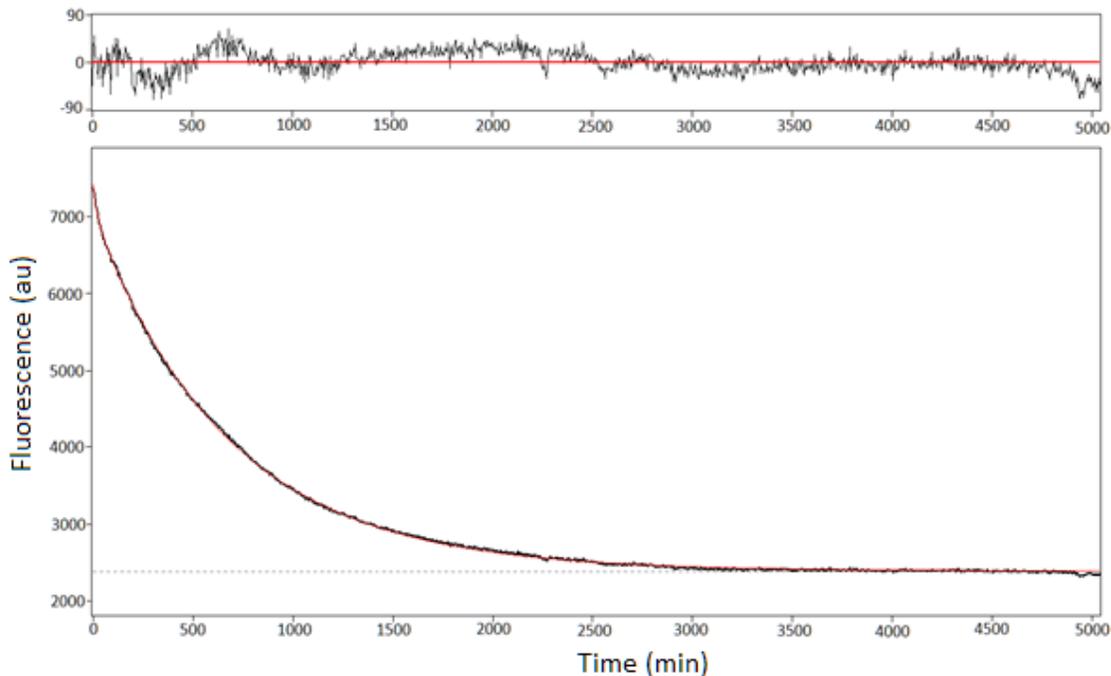


Figure 3.10: Representative unfolding kinetics of Threefoil in GuSCN, monitored at an emission wavelength of 313 nm

Threefoil unfolding in 2.55M GuSCN. The red line represents the line of fit, by fitting the data to a double exponential decay equation without a linear drift; however, the phase with the faster rate and smaller amplitude is believed to be an artefact, which may be related to temperature variation in the plate reader and/or photobleaching. Inset shows the residual of fitting, a.k.a. the deviation of the data from the line of fit.

For the major phase, the kinetic unfolding rate was 0.00147 min^{-1} , the amplitude was 4637.8 A.U.; for the minor phase, the kinetic unfolding rate was 0.0421 min^{-1} and the amplitude was 379.1 A.U. (arbitrary units). The offset was 2393.1 A.U.

At 3M GuSCN the protein requires days to unfold fully, but at concentrations of 4.5M GuSCN and higher, Threefoil unfolding occurs on the order of minutes. Even so it is clear that in such a strong denaturant, Threefoil still resists unfolding, as it takes approximately ten minutes in 5.268M GuSCN (described further below) to obtain more than five half-lives of unfolding. This illustrates Threefoil's remarkable stability against unfolding in the presence of strong denaturants and also that Threefoil has among the slowest unfolding rates reported for most super-stable proteins (Table 3).

Protein	function	# amino acids	Denaturant	Conditions	Temperature (°C)	m (kcal/(mol*M))	Cmid (M)	dG (kcal/mol)	kf (1/s)	mf (kcal/(mol*M))	T _{1/2} (s)	ku (1/s)	mu (kcal/(mol*M))	T _u (1/2) (s)	BT	mu'	Reference		
3fo1	sugar binding	141	Urea	pH 6.6, eqm	25	1.50	3.50	5.21	1.83E-04	1.70	3.79E+03								
			GuHCl	pH 6.6, eqm	25		1.75E-04	4.29											
			GuSCN	pH 6.6, kin	25	9.11	0.73	6.69	8.70E-05	5.88	7.97E+03	1.10E-09	3.23	6.13E+08	0.65	22.34			
				pH 6.6, eqm	25	8.77	0.53	4.65											
Hisactophilin	actin binding	118	Urea	pH 7.7, eqm	20	2.20	5.68	11.34										1	
				Myristoylated	20	2.18	4.38	8.16											
				Non-myristoylated	20														
			Urea	pH 7.7, kin	20														
IL-1B	receptor binding	153		Myristoylated	20	1.87	5.65	9.53	1.27E+03	1.18	5.46E-04	1.30E-04	0.68	5.33E+03			0.68		
				Non-myristoylated	20	1.87	4.68	8.25	2.25E+01	1.18	3.08E-02	2.00E-05	0.68	3.47E+04					
					25	6.74	1.35	9.10											
					25	4.80	1.06	5.09	3.31E+00	6.05	2.09E-01	6.90E-04	0.47	1.00E+03	0.95	0.97	2		
FGF-1	receptor and heparin binding	146	GuHCl		25	4.13	1.53	6.32	5.30E+00	3.37	1.31E-01	1.19E-04	0.76	5.82E+03	0.82	1.57	2		
					25	2.99	3.70	11.06	1.34E+02	2.51	5.17E-03	1.07E-06	0.48	6.48E+05	0.84	0.99	2		
TNfn3	cell development	96	Urea	pH 5.0, eqm	25	1.26	5.31	6.68										3	
			GuSCN	pH 5.0, eqm	25	9.42	0.72	6.78											
			GuSCN	pH 5.0	25	6.38	1.47	9.38	2.40E+02									4	
			Urea	pH 6.3, eqm	25	1.92	4.58	8.79											
Barnase	ribonuclease activity	110	GuHCl	pH 6.3, eqm	25	4.40	2.00	8.80										5	
			GuSCN	pH 6.3, eqm	25	14.31	0.77	11.00											
GFP	bioluminescence	238	GuHCl	pH 7.5, kin	25				1.00E-02		6.93E+01	2.10E-11		3.30E+10			6		
Top7	none	120	GuHCl			2.22	6.54	~6	2.45E+02								7		
										1.20E+01									
									1.60E+00										

Table 3: Comparison of Threfoil kinetic and equilibrium parameters

- Smith, M.T., Meissner, J., Esmonde, S., Wong, H.J., Meiring, E.M. 2010. *PNAS*. **107**(49):20952-20957.
- Lee, J., Blaber, S.L., Dubey, V.K., Blaber, M. 2011. *J. Mol. Biol.* **407**:744-763.
- Hamil, S.J., Meekehof, A.E., Clarke, J. 1998. *Biochem.* **37**(22):8071-8079.
- Cota, E., Clarke, J. 2000. *Prot. Sci.* **9**(1):112-120.
- Clarke, J., Fersht, A.R. 1993. *Biochem.* **32**(16):4322-4329.
- Fukuda, H., Arai, M., Kuwajima, K. 2000. *Biochem.* **39**(39):12025-12032.
- Watters, A.L., Deka, P., Corrent, C., Callender, D., Varani, G., Sosnick, T., Baker, D. 2007. *Cell*. **128**:613-624.

Protein	function	# amino acids	Denaturant	Conditions	Temperature (°C)	m	Cmid	dG	kf	mf	Tr(1/2)	ku	mu	Td(1/2)	BT	mu'	Reference
					(°C)	(kcal/(mol*M))	(M)	(kcal/mol)	(1/s)	(kcal/(mol*M))	(s)	(1/s)	(kcal/(mol*M))	(s)			
a-lytic protease	protease	198	GuHCl		25				6.00E-12		1.16E+11	1.20E-07		5.78E+06			8
Consensus Ankyrin repeat proteins	none																
N1C		91	GuHCl	2-state kin	25	2.60	1.40	3.70	6.38E+02		1.09E-03	3.40E+00		2.04E-01			9
N2C		2x91	GuHCl	2-state kin	25	4.00	1.35	5.40	7.90E+02		8.77E-04	6.30E-04		1.10E+03			
N3C		3x91	GuHCl	3-state kin	25	2.30	4.10	9.20	4.00E+02		1.73E-03	1.00E-08		6.93E+07			
N4C		4x91	GuHCl	Eqm		4.40	3.79	16.70									
N5C		5x91	GuHCl	Eqm		3.00	5.60	16.80									
N6C		6x91	GuHCl	Eqm		~8	>8										
HIV gp41	Glycoprotein	34	Urea		25	2.4 (per trimer)	11.54	27.70	1.30E+15	2 (per mol trimer)	1.00E-15	1.10E-05	0.5 (per mol trimer)	6.30E+04	0.80	0.50	10
SIV gp41	Glycoprotein	36	Urea		25	3.6 (per trimer)	5.25	18.90	1.10E+10	2.1 (per mol trimer)	6.30E-11	5.70E-04	1.4 (per mol trimer)	1.22E+03	0.60	1.40	
Lpp-56 trimer	Membrane lipoprotein	56	GuHCl		25	3.6 (per trimer)	5.25	18.90	6.80E+05	0.5 (per mol trimer)	1.02E-06	3.80E-10	2.7 (per mol trimer)	1.82E+09	0.16	5.57	
AaFd1	Electron transfer				25									3.47E+11			
AaFd5	Electron transfer		GuHCl		25									3.47E+11			11
ORF56			GuHCl	2-state kinetics	25									3.85E+06			
PRCP	Peptidase		GuHCl	2-state kinetics	25				9.30E-02					3.47E+14			
TkRNase Hill	Ribonuclease		GuHCl	2-state kinetics	50				7.80E-01					1.39E+07			
AmFd	Electron transfer		GuSCN	2-state kinetics	25									1.16E+09		6.79	
SLPI	Protease inhibitor	107	GuSCN	Eqm, pH 7	20	6.10	3.10	18.90									12
Cutinase	lypolytic enzyme	197	Urea	Eqm pH 4.5	25	3.35	4.56	1.36									13
			GuHCl		25	2.40	2.34	7.69	1.10E+00	-1.58	6.30E-01	1.77E-01	0.25	3.92E+00	0.86	0.25	14
			GuSCN		25	4.93	1.18	7.90	1.34E+02	-6.93	5.16E-03	2.94E+00	0.08	2.35E+01	0.99	0.17	
						3.07	0.82	3.42	1.07E+00			5.10E+00	0.49			3.39	

Table 3: Comparison of Threefold kinetic and equilibrium parameters

8. Sohl, J.L., Jaswal, S.S., Agard, D.A. 1998. *Nature*. **395**(6704):817-819.
9. Wetzel, S.K., Settanni, G., Kenig, M., Binz, H.K., Pluckthun, A. 2008. *J. Mol. Biol.* **376**(1):241-257.
10. Galvagnion, C., Smith, M.T., Broom, A., Vassal, K.A., Meglej, G., Gaspar, J.A., Stathopoulos, P.B., Cheyne, B., Meiering, E.M. 2009. *Biochem.* **48**(13):2891-2906.
11. Luke, K.A., Higgins, C.L., Wittung-Stafshede, P. 2007. *FEBS J.* **274**(16):4023-4033.
12. Moczygmba, C., Guldry, J., Jones, K.L., Gomes, C.M., Teixeira, M., Wittung-Stafshede, P. 2001. *Prot. Sci.* **10**(8):1539-1548.
13. Lin, C. C.-J., Lu, B.Y., Chang, J.Y. 2006. *BBA Prot. Proteomics*. **1764**(7):1286-1291.

3.3.2: Refolding kinetics of metallated Threefoil

Refolding of Threefoil is also extremely slow from GuHCl or GuSCN-denatured protein stock (Figure 3.11) and generally requires more than five hours to gain any appreciable number of half-lives even at denaturant concentrations below 0.05M GuSCN. Complete refolding appears to be achieved only after 20-24 hours even at low denaturant (e.g. <0.05M) concentrations.

Refolding is also complicated by aggregation—at high protein concentrations, generally above 8-10 μ M, aggregation becomes plainly visible to the eye when the unfolded protein is diluted into refolding buffer. Such an occurrence generally hinders monitoring refolding kinetics by CD, as a high protein concentration is needed to obtain sufficient signal to noise ratio for CD. Because Threefoil does not denature completely in urea, refolding kinetics could not be obtained in urea. More representative folding kinetic traces are given in Appendix 4.

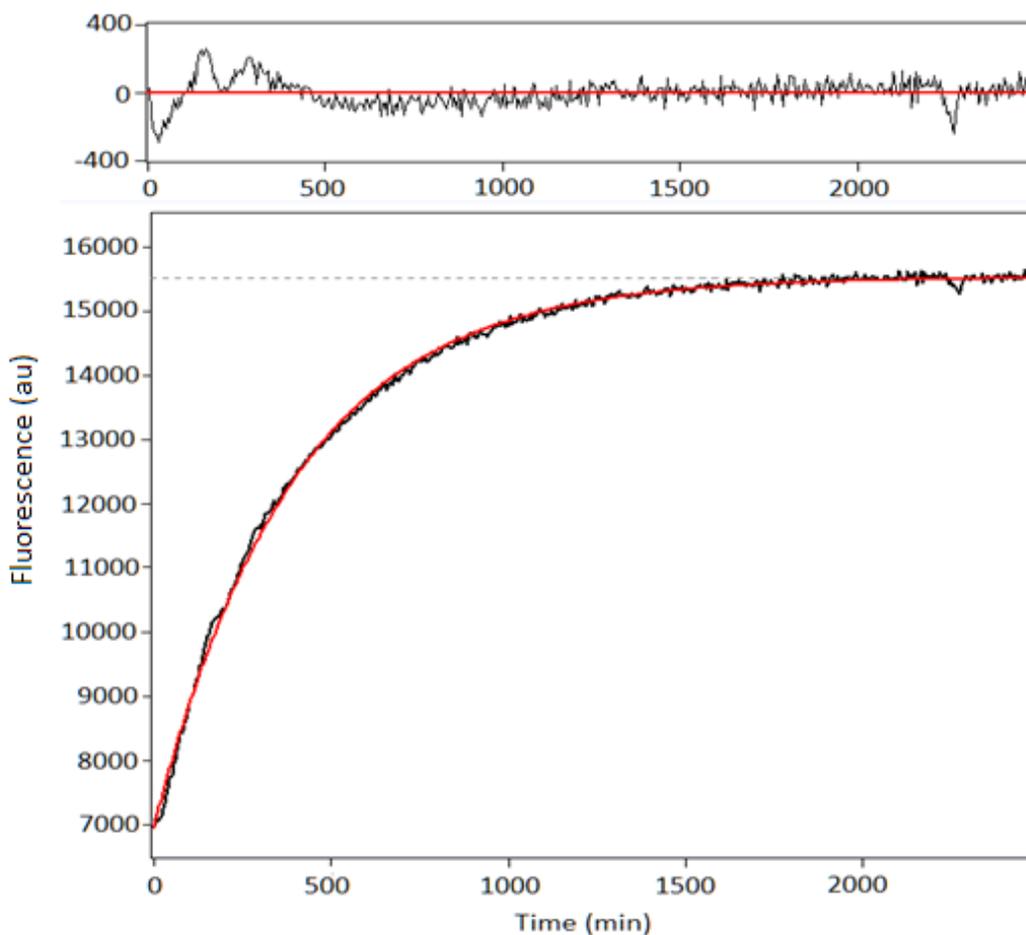


Figure 3.11: Representative refolding kinetic traces in GuSCN, monitored at an emission wavelength of 313 nm

Threefoil refolding in 0.08M GuSCN. The red line represents the line of fit, for fitting the data to a single exponential equation without a linear drift. Inset shows the residual of fitting, or the deviation of the data from the line of fit.

The kinetic refolding rate was 0.00253 min^{-1} , the amplitude was -8573.4 A.U. (arbitrary units) and the offset was 15534.8 A.U. Deviation from the line of fit in the residual may reflect initial artefact of decreasing fluorescence due to temperature variation/photobleaching, as also observed in unfolding (Figure 3.10).

The observed kinetic refolding rates may have contributions from protein refolding and from protein aggregation, the latter of which is easily seen to occur when a higher concentration of unfolded protein is rapidly diluted into buffer. However, the data are well-fit to a single exponential, and the rates are independent of protein concentration (see section 3.3.3) suggesting

that the rate of aggregation does not markedly affect the observed rate of folding, and that the latter is indeed the major transition observed.

3.3.3: Kinetics of metallated Threefoil reconstructed from equilibrium curves

Kinetics at intermediate concentrations, which become extremely slow, were reconstructed from equilibrium curves by measuring the fluorescence of equilibrium curve samples over a period of months. The individual equilibrium curves were normalized for the unfolded baseline prior to taking data from each time point. In this case the signal at 313 nm was taken rather than 370 nm, as the fluorescence signal change was greatest at 313 nm, giving better signal-to-noise ratio and less scattered data. Each kinetic trace generated was fit to a single exponential equation with no drift. It was not possible to generate unfolding equilibrium curves in urea, as Threefoil does not denature in urea²⁷; therefore, this analysis was carried out only for GuSCN unfolding equilibrium curve and for refolding curves for urea, GuHCl and GuSCN.

3.4: Kinetics of Threefoil inclusion bodies

3.4.1: Unfolding kinetics of urea-denatured inclusion bodies

Since denatured Threefoil inclusion body proteins contained secondary and tertiary structure, it was of interest to determine whether this partially unfolded state would unfold further with any measurable kinetics in the presence of a stronger denaturant, in order to give insight into the stability and nature of the residual structure. At the GuSCN concentrations tested, no measurable kinetics were obtained (Figure 3.12).

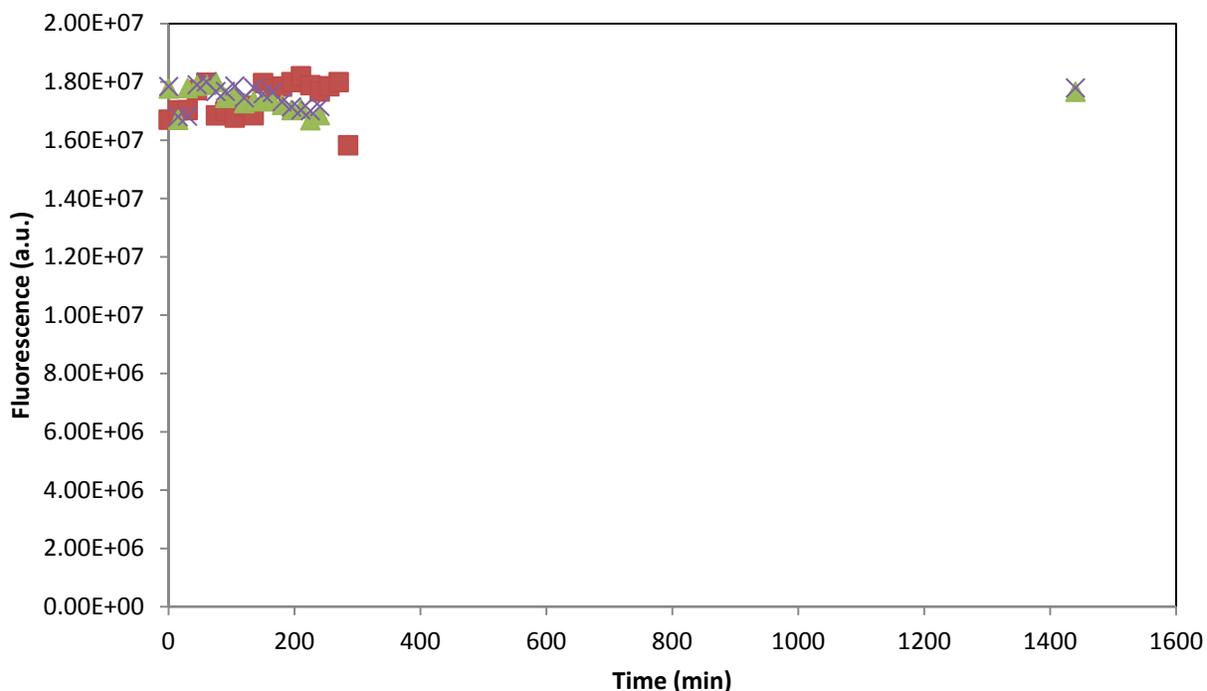


Figure 3.12: Unfolding kinetics of 6M urea-denatured inclusion bodies in GuSCN

Emission is monitored at 370 nm for unfolding kinetics. For concentrations of 2.5 M (■), 4.55 M (▲) and 5.88 M (×) GuSCN, no kinetics occur. Final protein concentration was 1.5 μ M after dilution into unfolding buffer containing GuSCN. Kinetic traces were monitored at 25°C.

3.4.2: Refolding kinetics of Threefoil inclusion bodies

Refolding from a urea-denatured state is slow, requiring at least 18-20 hours for the protein to fully refold, and yields proteins with extensive secondary structure. Generally in order to obtain measurable refolding kinetics it was necessary again to take staggered measurements over the course of ~24 hours. However, the presence of residual structure does not appear to affect the ability of the protein to fold to its native state (Figure 3.13).

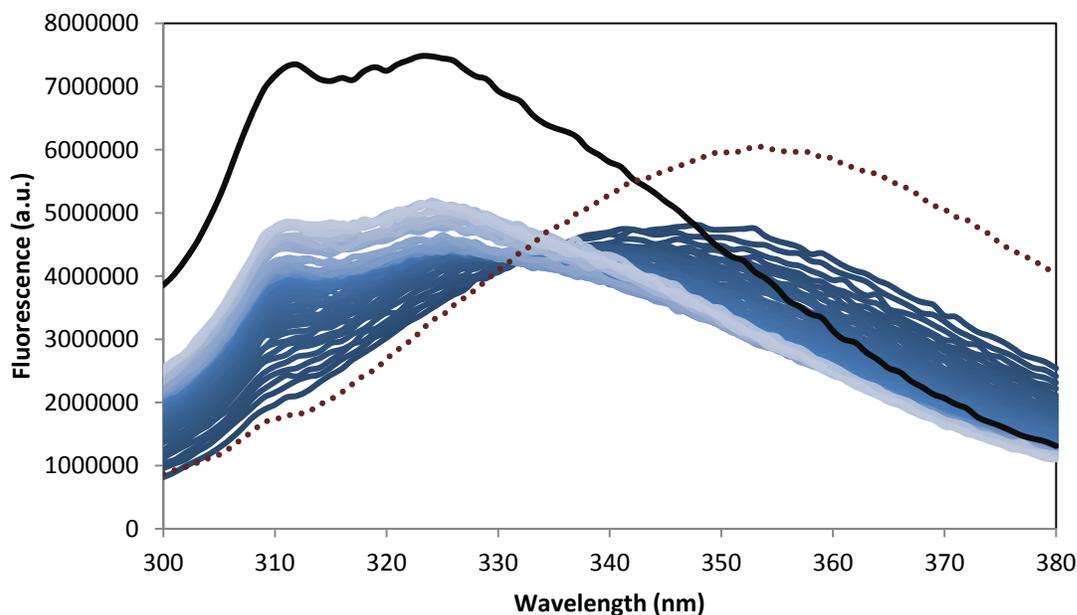


Figure 3.13: Refolding kinetics (spectra over time) of urea-denatured inclusion body proteins

Protein folding into a final concentration of 0.106 M urea was monitored by measuring spectra continuously with time. The 6M urea-denatured IB unfolded spectrum (dotted spectrum) is given as a reference for the starting state of the protein. The spectrum measured at 18 hours (black solid line) indicates that the protein folds to a native-like conformation. Blue spectra represent the course of protein refolding, from denatured (darker spectra) to native (lighter spectra). The excitation wavelength was 280 nm; emission wavelengths were monitored from 300-380 nm. Spectra were taken continuously over the course of 5 hours, and then a single scan was taken at 18 hours.

3.4.3: Protein concentration dependence of refolding from urea-solubilized inclusion bodies

The Threefoil refolding rate is not known to have a protein concentration dependence when folding from a state with residual structure, as seen in Figure 3.14 below. Refolding kinetics were measured by fluorescence monitored at two different wavelengths, 313 nm and 370 nm, by manual mixing techniques. The protein concentrations covered a span of roughly 100-fold, from 0.5 μM to 5.3 μM , and no dependence of folding rate on protein concentration was observed.

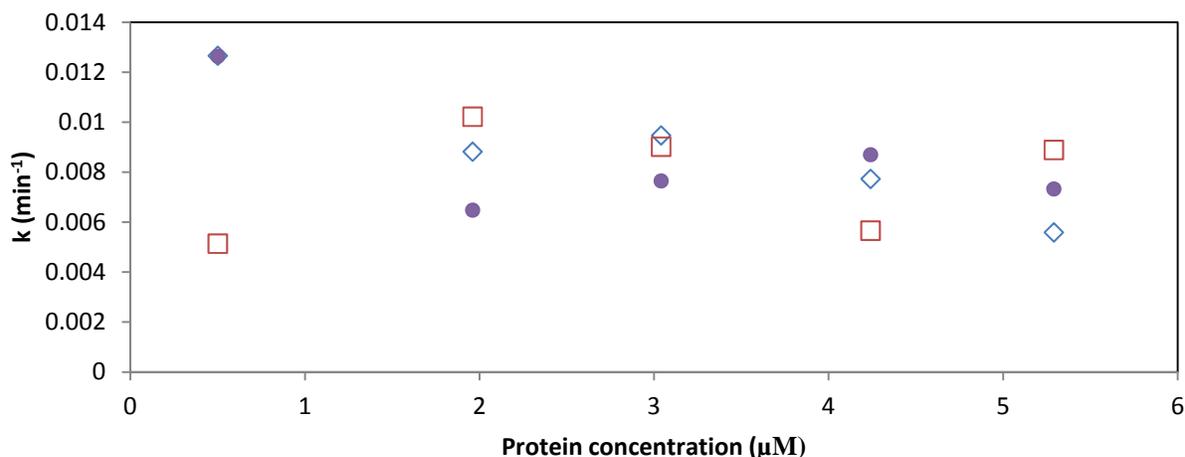


Figure 3.14: Dependence of refolding rate of urea-denatured inclusion bodies on protein concentration

(◊) Refolding rate monitored at 313 nm. (● and ◻) Refolding rate monitored at 370 nm, two independent trials. The refolding rates have no dependence on the protein concentration and are largely coincident between the two fluorescence wavelengths monitored.

3.5: Chevron analysis for Threefoil

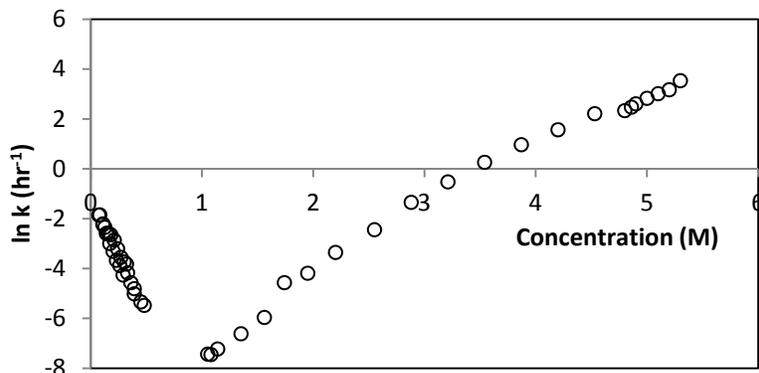
Although unfolding kinetic rates could only be obtained in GuSCN, refolding kinetic rates were obtained in all 3 denaturants studied (urea, GuHCl and GuSCN). It was necessary to convert GuSCN concentration (M) to units of GuSCN activity (D) as the rates plotted against concentration units were observed to have nonlinear behaviour (Figure 3.15A). The relationship between the free energy of unfolding and the denaturant concentration is nonlinear at high concentrations,^{12,75} which gives an unreliable extrapolation in the chevron curve to rates in 0M denaturant. When rates are plotted against denaturant activity (Figure 3.15B), linearity is restored, giving a more accurate extrapolation.^{12,75} This technique was similarly used by Cota and Clarke⁸⁸ for their interpretation of the folding results of the fibronectin type III (fnIII) domain in GuSCN. Conversion between concentration and activity uses the following

empirically derived equation:

$$D = [M] * (C_{0.5}/(C_{0.5} + [M])) \quad (3.1)$$

where D is the GuSCN activity unit, M is the GuSCN concentration and $C_{0.5}$ is the denaturant constant, 6.47 M.⁷⁵

(A)



(B)

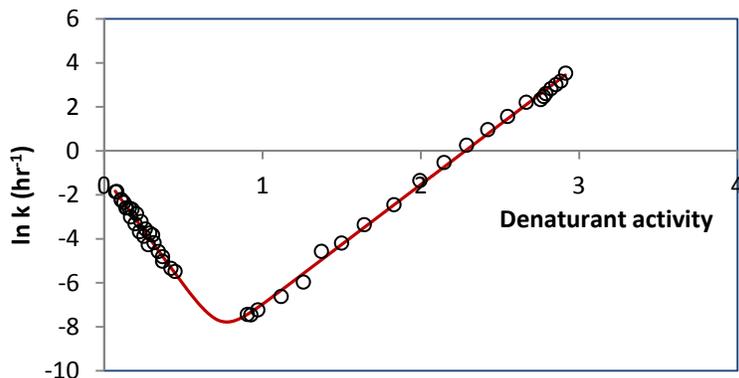


Figure 3.15: Rates of refolding and unfolding of Threefoil in GuSCN

Refolding and unfolding rates were obtained for the fluorescence signal at 313 nm. Both refolding and unfolding kinetic rates were fit to a single exponential decay (equation 1.6) with a slope of 0.

(A) Rates plotted against denaturant (GuSCN) concentration are non-linear at high denaturant concentrations.

(B) Rates plotted against denaturant activity are linear at high denaturant concentrations. The red line is a two-state line of fit for the chevron plot.

Figure 3.16 shows the folding branches in all 3 denaturants, generated by combining kinetic data which includes both faster kinetics in which fluorescence was measured continuously in real time and kinetics reconstructed from renaturation equilibrium curve samples monitored over many months. The linearity of extrapolation and the same extrapolated refolding rate to 0 M denaturant for urea, GuHCl and GuSCN indicates that in all three denaturants, the same transition is monitored and that in low denaturant, folding occurs from comparable denatured states. These results support that in all cases folding proceeds via a two-state cooperative transition from a denatured state with residual structure to the native state, with no intermediate.

Results from fitting chevron plots are given in Table 3.

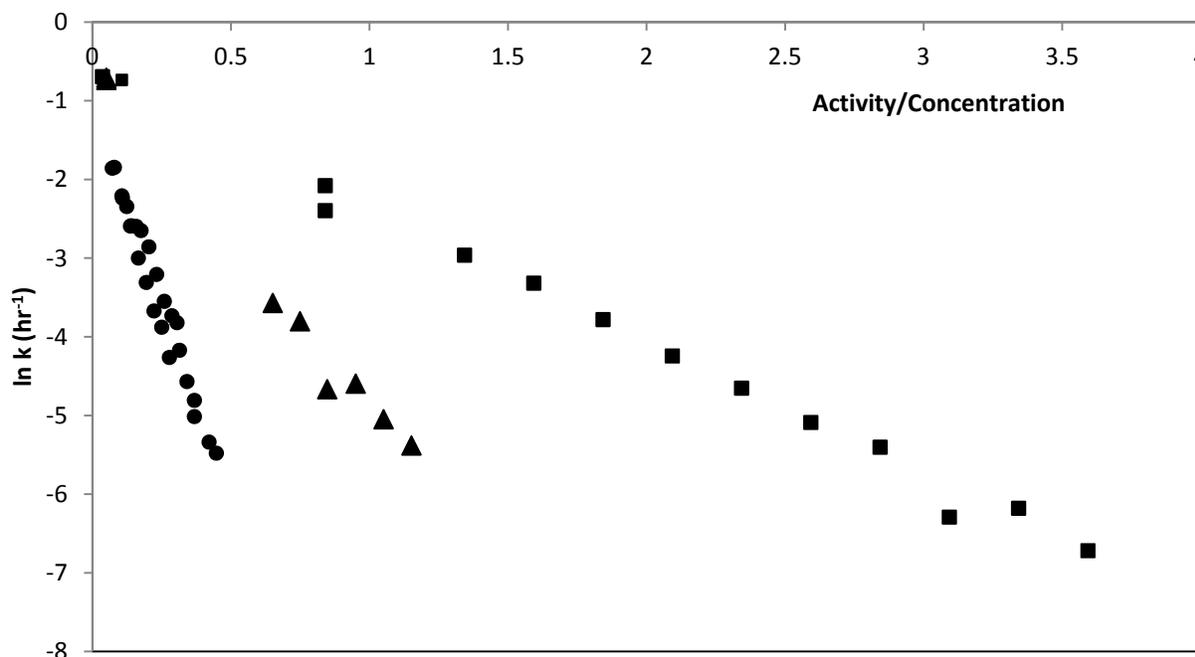


Figure 3.16: Folding branches of Chevron generated in urea, GuHCl and GuSCN

(■) Urea refolding kinetics; (▲) GuHCl refolding kinetics; (●) GuSCN refolding kinetics
 All three folding branches extrapolate to approximately the same rate (i.e, to the same $k_f^{\text{H}_2\text{O}}$), consistent with two-state folding.

3.6: Threefoil metal-binding studies

In general, metal-binding stabilizes proteins.^{17,76-78} The preferential binding of a metal to the folded state is stabilizing if it decreases the energy of the native state, thus shifting the equilibrium of native to unfolded protein in favor of the native protein. The stabilizing effect of metal-binding is also seen in the case of Threefoil, which binds sodium preferentially over other monovalent and divalent metals of similar size: Ca^{2+} , Mg^+ , K^+ , as determined both experimentally and also proposed previously from the crystal structure. The cavity for metal-binding is $\sim 2.5 \text{ \AA}$ across, and each of Mg^{2+} , K^+ and Ca^{2+} could possibly be coordinated as they are all smaller than 2.5 \AA in diameter.¹⁷ The glycosidase sequence template from which Threefoil was designed came from *Haloarcula marismortui*, a halophilic archaeon found in the dead sea where external sodium concentrations can reach up to 4M.^{27,28} Therefore it is most likely that sodium is the metal coordinated to Threefoil, given that it is found in abundance in the environment in which the organism thrives.

3.6.1: Characterization of metal-free Threefoil

In order to extract the metal from Threefoil, it was necessary to exchange Threefoil extensively into sodium-free buffer (as described in chapter 2.5). From mass spectrometry (MS) it was determined that the metal had been exchanged out of Threefoil, as the calculated mass of Threefoil (based on its amino acid sequence without a metal bound) matched the experimentally obtained peak in the native MS spectrum. The calculated mass was 18034 a.m.u. versus the experimentally determined mass of 18033 a.m.u. Interestingly the same metal-free, native peak was seen in the denatured MS spectrum as well, which speaks to Threefoil's high stability in even the strongly acidic denaturing conditions used. MS spectra are included in Appendix 5.

Removal of the metal does not affect the protein's tertiary structure to a large degree, as probed by fluorescence (Figure 3.17). Although the metal-free protein suffers a decrease in fluorescence signal, peaks appear at 313 nm and 323 nm that also appear in the native spectrum. This suggests that the tryptophans remain in the same environment in the metal-free protein as in the metallated protein; however, the protein may not be as tightly packed as in the native state, resulting in slight solvent-exposure of the tryptophan residues leading to quenching of the fluorescence signal.

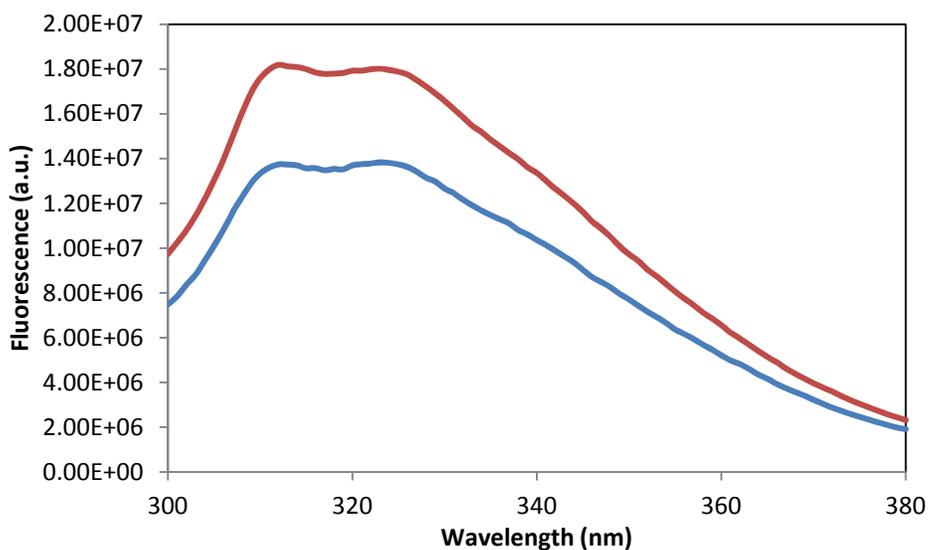


Figure 3.17: Comparison of fluorescence spectra of metal-free and metal-bound Threefoil

Threefoil with metal-bound (red spectrum) appears more folded in comparison to Threefoil with the metal removed (blue spectrum). With the metal removed the protein spectrum still appears native-like, however the absolute fluorescence is lower, indicating that removal of the metal may affect the tertiary structure of the protein, resulting in a more loosely packed protein in which the tryptophan residues are somewhat more solvent exposed than in the metal-bound native protein. Spectra were excited at 280 nm; protein concentration for both metal-bound and metal-free spectra was 1.25 μM . The buffer conditions were as follows: Metal-free protein, 100 mM ammonium phosphate; metal-bound protein, 100 mM sodium phosphate, 300 mM sodium chloride.

3.6.2: Folding and unfolding kinetics of metal-free Threefoil

The absence of sodium from the protein and from the unfolding buffer increases the rate of unfolding, suggesting that sodium may stabilize either the native state or the transition state or both. Metal-binding can shift the folding equilibrium by binding with higher affinity to the native state, and therefore affect the equilibrium stability. Bound metals that coordinate with nearby amino acid residues within the metal-binding site also may form covalent links which can also stabilize the protein.

In GuSCN, the unfolding kinetics were sufficiently slow to be obtainable by manual mixing techniques. The major transition was best fit to a single exponential equation and accounted for more than 90% of the amplitude of change. In the case for Threefoil, coordination of a metal (sodium) greatly enhances the stability, as in the absence of metal (Figure 3.18) the protein unfolds faster in denaturant. As opposed to the dramatic enhancement in kinetic unfolding rate, however, metal-binding does not affect the kinetic rate of refolding out of GuSCN. In the absence or presence of metal the refolding kinetic rate remains largely unchanged, which suggests that metal-binding may preferentially stabilize the native state over the transition or unfolded states (Figure 3.18). This also suggests that the metal-binding site is not formed in the transition state. Representative folding and unfolding kinetic traces are given in Appendix 6.

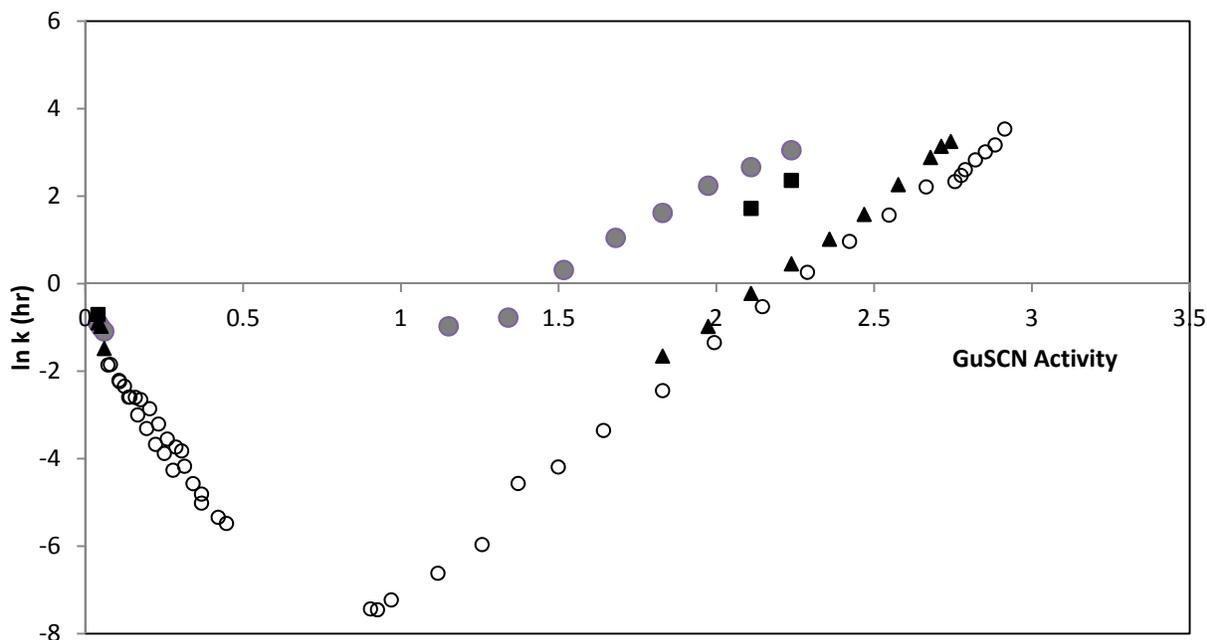


Figure 3.18: Folding and unfolding kinetic rates in the presence and absence of sodium

Kinetic studies were performed on a SpectraMax M5 fluorescence plate reader (Molecular Devices LLC), monitored at 319 nm, in GuSCN at 27°C. Open circles indicate rates of folding and unfolding of Threefoil in the presence of 300 mM NaCl, at 25°C. Closed circles (●) represent kinetics in the absence of sodium. Closed squares (■) represent kinetics in the presence of 3 mM sodium. Closed triangles (▲) represent kinetics in the presence of 300 mM lactose.

Unfolding in the absence of sodium results in a faster unfolding rate, which suggests that the protein without metal bound is destabilized. When the protein is unfolded in the presence of 300 mM sodium, however, the unfolding kinetic rate decreases, suggesting that the protein is stabilized. The presence of absence of sodium has no marked effect on the kinetic rate of folding.

Ch. 3.7: Carbohydrate-binding effects on Threefoil kinetics

Carbohydrate-binding to Threefoil appears to have a stabilizing effect; however, Threefoil binds carbohydrates selectively. Sucrose-binding has little effect on kinetic rate of refolding or unfolding, while lactose-binding has a marked effect. This can be rationalized by the fact that Threefoil preferentially binds carbohydrates with galactose subunits.²⁸ As sucrose does not contain a galactose subunit, presumably it does not bind to Threefoil, while lactose, composed of a galactose and a glucose subunit, does bind. Lactose-binding was found to increase the rate of refolding and decrease the rate of unfolding (Figure 3.19), indicating partial formation of the binding site in the transition state. Representative folding and unfolding kinetic traces are given in Appendix 7.

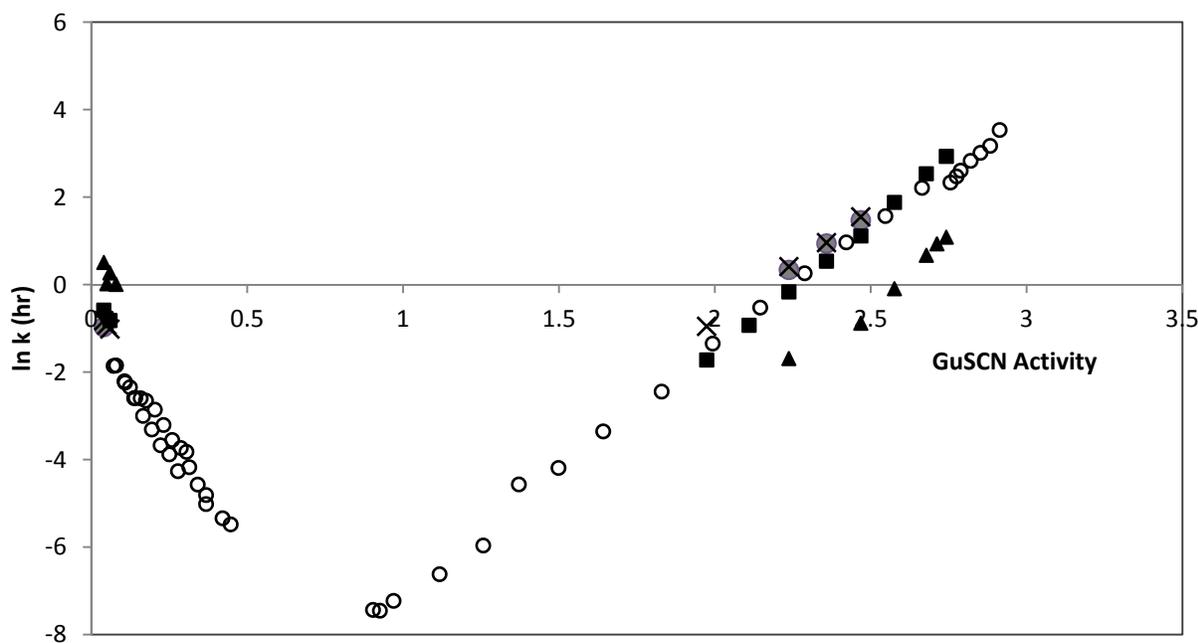


Figure 3.19: Folding and unfolding kinetics in the presence of lactose and sucrose

Kinetic studies were performed on a SpectraMax M5 fluorescence plate reader (Molecular Devices LLC), monitored at 313 nm, in GuSCN at 27°C: Closed circles (●) represent kinetics in the presence of 0.5 mM lactose. Closed squares (■) represent kinetics in the presence of 5 mM lactose. Closed triangles (▲) represent kinetics in the presence of 50 mM lactose. Crosses (×) represent kinetics in 50 mM sucrose. Open circles indicate kinetics of Threefoil in the absence of carbohydrates, monitored at 25°C.

3.8: Characterization of Threefoil inclusion bodies

3.8.1: Differential scanning calorimetry

It is noteworthy that Threefoil displays remarkable thermal stability, as probed by preliminary differential scanning calorimetry (DSC) experiments. Figure 3.20 shows native Threefoil DSC scans. It was necessary to conduct DSC experiments in denaturant because native 3foil does not unfold thermally in a reversible manner and, upon cooling, tends to aggregate. In order to attempt to prevent these problems, scans were measured in the presence of GuHCl. In the presence of concentrations of GuHCl below 0.25 M, there is little effect of denaturant on the apparent T_m , $T_{m,app}$, and protein aggregation is decreased; nevertheless, unfolding remained irreversible as no endotherm was observed upon cooling and rescanning the sample. Experiments were also conducted in 3 M and 6 M GuHCl. In all concentrations of GuHCl shown, the protein did not aggregate upon cooling; however, there was no reversibility of folding, presumably because the rate of folding is so slow that in the time frame of the experiment, the protein does not properly refold.

Native Threefoil displays a scan-rate dependence of apparent melting temperature, $T_{m,app}$, indicating that unfolding is under kinetic control rather than under thermodynamic control.^{79,80} (i.e. Unfolding is dependent on the difference in energy barrier between the native and unfolded states.⁷⁹) At a scan rate of 1°C/minute, the $T_{m,app}$ is 96°C whereas at a scan rate of 0.5°C/min, $T_{m,app}$ is 94°C for native protein in the absence of denaturant. Scan-rate dependence of the apparent T_m has been well documented in a host of proteins as a reflection of kinetic control of unfolding^{79,80}, such as in Glucosamine-6-phosphate Deaminase which also denatures irreversibly.⁸⁰ In the presence of denaturant, GuHCl concentrations of 0.25 M, 3 M and 6 M

yield respective $T_{m \text{ app}}$ of 96°C, 82°C and 74°C. The $T_{m \text{ app}}$ decreasing with increasing denaturant concentration, indicates that the protein is destabilized in high denaturant.

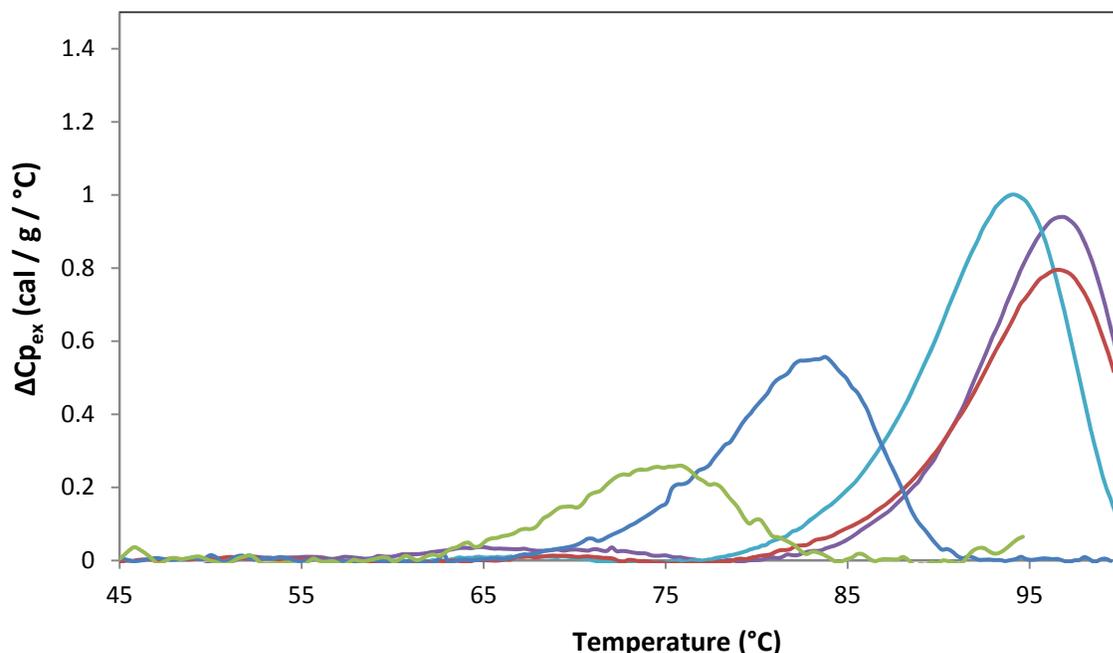


Figure 3.20: DSC scans of Threefoil in GuHCl

There is a dependence of T_m on the scan rate (purple, 1°C/min and light blue, 0.5°C/min; 0.2 mg/mL protein, 0 M denaturant), indicating that denaturation is under kinetic control. With increasing concentrations of denaturant, the excess heat capacity diminishes and the T_m decreases, both pointing to the protein being destabilized.

Curves scanned in the presence of denaturant, at 1°C/min, with 0.3 mg/mL protein are as follows: Red: 0.25 M GuHCl; dark blue: 3 M GuHCl; green: 6 M GuHCl.

3.8.2: ANS binding

In order to probe the nature of the residual structure of urea-denatured inclusion body proteins, ANS binding was carried out and monitored by fluorescence. ANS is used as a probe for the molten globule state (or molten-globule-like states) and is known to preferentially bind to exposed hydrophobic groups, with an enhancement in ANS fluorescence signal as well as a

marked blue-shift of the wavelength of maximum fluorescence.⁸¹ As the ANS is thought to screen exposed aromatic residues from the solvent, the result of ANS-binding is a change in the wavelength of emission.⁸¹ ANS binding to Threefoil inclusion body proteins in urea resulted in a relatively small increase in fluorescence and little blue-shift (Figure 3.21), indicating little binding to the denatured state. This indicates that the residual structure is not molten globule-like.

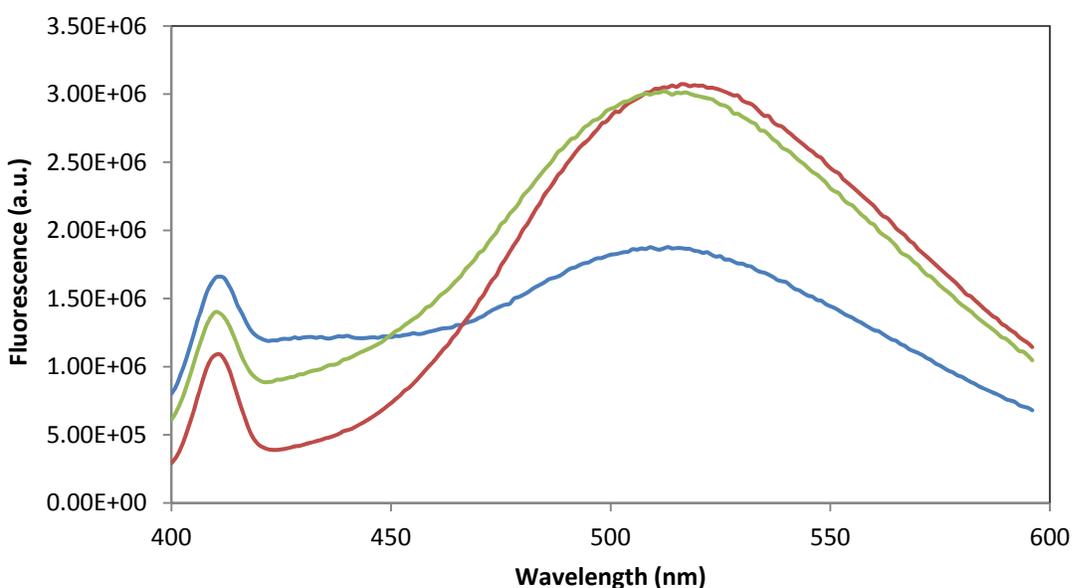


Figure 3.21: ANS binding to urea-denatured inclusion body proteins

ANS binding to the 6 M urea-denatured inclusion body proteins with residual structure (green) yielded a slight fluorescence enhancement and a slight (~2 nm) blue-shift. This suggests that the residual structure is not molten globule-like. The excitation wavelength was 360 nm. The folded reference spectrum (blue) contains native protein. The unfolded reference (red) contained folded Threefoil denatured in 3 M GuSCN. The concentration of ANS used was 25uM and the concentration of Threefoil used was 10 uM. ANS binding experiments were carried out at 25°C.

3.9: Preliminary Threefoil mutant analysis

Mutants of Threefoil discussed in this and the preceding chapter (Chapter 1.4) were expressed, isolated and purified according to the procedure outlined in Chapter 2.1. Data for protein expression tests were obtained with Laura Bahlmann and Briallen Lobb, both co-op students of the Meiering group.

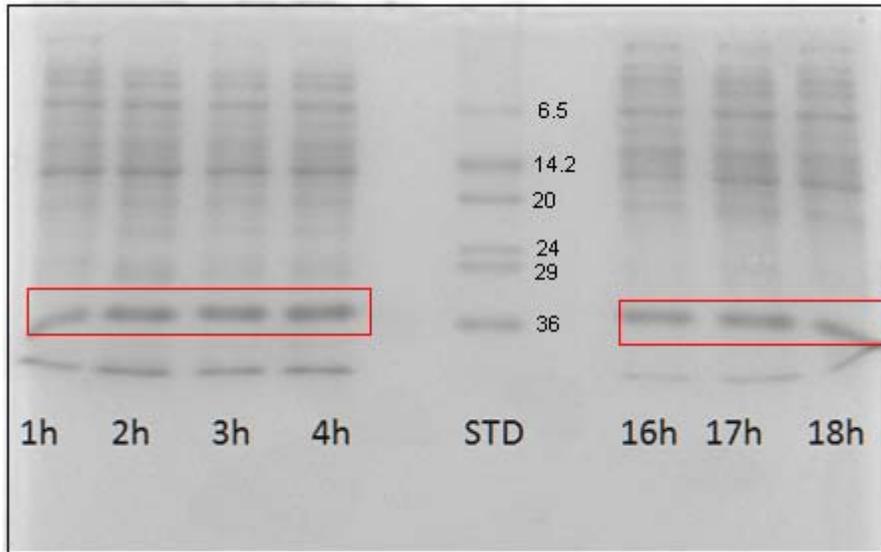
3.9.1: Mutant expression

As mentioned previously all of the mutants were able to be expressed and purified in significant amounts, suggesting that effects of mutation on the stability of the protein did not similarly affect expression rate. Of the ten mutants, only four are discussed in-depth in this thesis: A62V, Q78I, D85P and D93P. At lower growth temperatures and lower IPTG induction levels, more soluble protein was observed; however, the total amount of soluble protein produced remained in such small amount as to be impossible to isolate in any appreciable quantity. Particularly interesting to note is that despite all mutations being selected for to their stabilizing effect on the protein, none appeared to significantly affect the solubility of the protein. Discussed in this and later sections are the four mutants whose denaturation kinetics were studied.

Figure 3.22 shows SDS-PAGE gels of the wild-type Threefoil, expressed under normal conditions of 37°C and 1 mM IPTG in LB broth. Under such conditions the protein is nearly 100% insoluble and requires isolation from inclusion bodies in order to obtain native protein. What appears to be soluble protein (Figure 3.22 A) is actually contamination from the pellet, as the intensity of the Threefoil band does not increase over time, as expected. In the pellet fraction,

the band intensity does increase over time, reaching a maximum at approximately 16 hours (denoted 'O/N' for overnight in Figure 3.22 B).

(A)



(B)

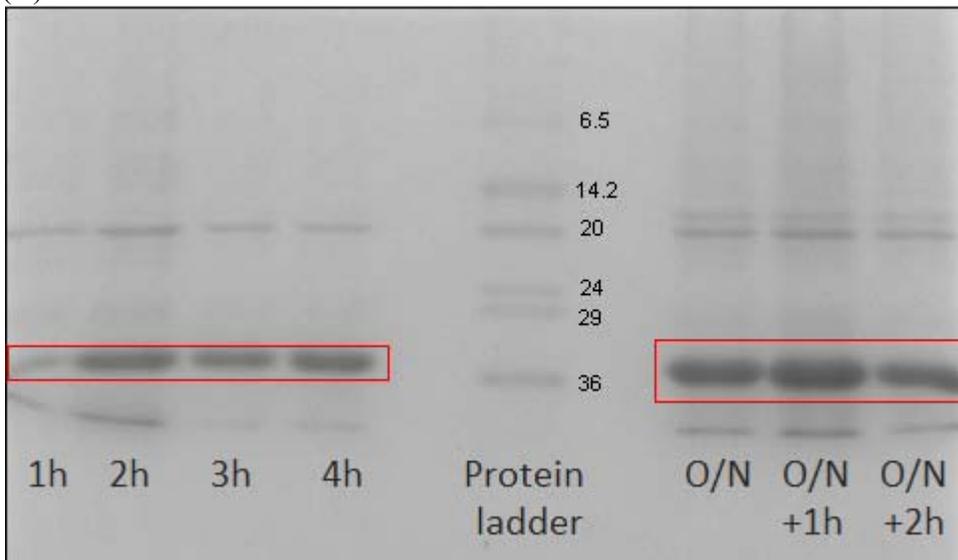


Figure 3.22: SDS PAGE gel of expression of wild-type Threefoil, at 37°C, 1 mM IPTG in LB broth

(A) The supernatant fraction from a typical Threefoil sample after homogenization. (B) The pellet fraction from a typical Threefoil sample after homogenization. Outlined in red is the Threefoil protein band. 'STD' denotes the protein standard ladder (SigmaMarker low range protein ladder, SIGMA). What appears to be soluble protein in the supernatant is most likely contamination from the pellet, as the protein band intensity in (A) does not change over the course of 18 hours, as is expected.

At the same temperature, 37°C, but at 0.1 mM IPTG, Threefoil solubility appears to improve (Figure 3.23A) as the band intensity increases between hours one and three. It is likely that the supernatant becomes saturate with protein after hour three, as the band intensity after that does not change noticeably. The majority of the protein, however, still goes into inclusion bodies (Figure 3.23B).

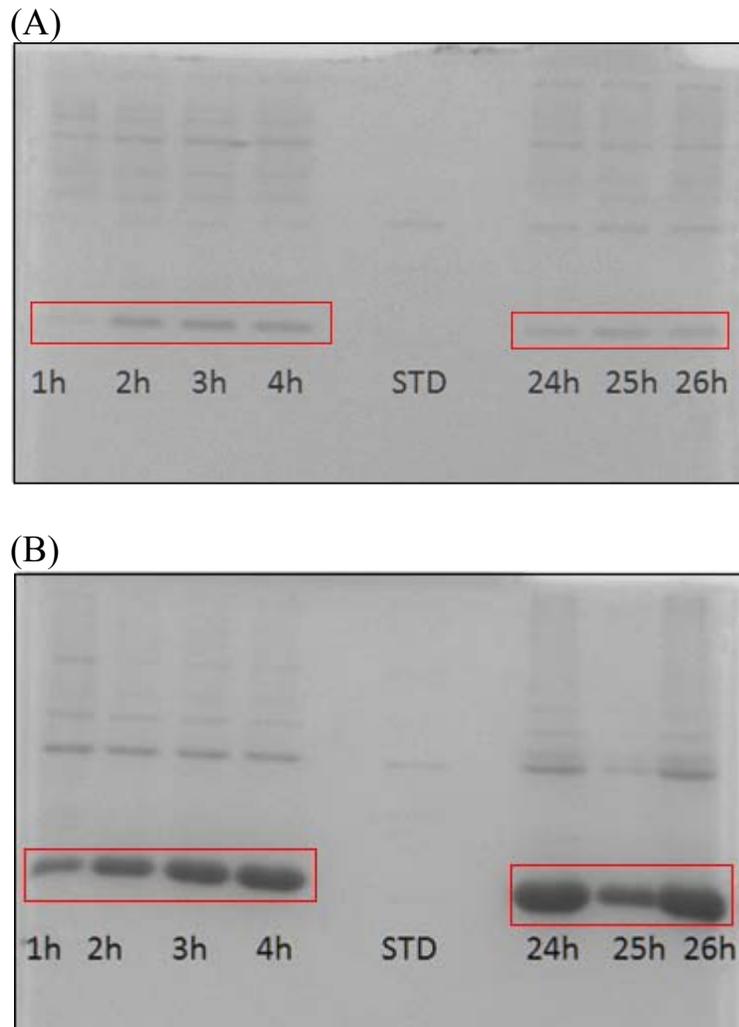


Figure 3.23: SDS-PAGE gel of expression of wild-type Threefoil, at 37°C, 0.1 mM IPTG in LB broth

(A) The supernatant fraction from a typical Threefoil sample after homogenization. (B) The pellet fraction from a typical Threefoil sample after homogenization. Outlined in red is the Threefoil protein band. ‘STD’ denotes the protein standard ladder. There appears to be more soluble protein being produced over time, as seen in the band intensity increasing in (A); however, it is likely that two to three hours after induction the supernatant is saturated with protein. Most of the protein can still be found in the pellet fraction.

Mutant A62V was expressed at 37°C in LB broth and induced using 1 mM, 0.35 mM and 0.1 mM IPTG respectively (Figure 3.24). With decreasing IPTG concentrations the solubility of the protein increased; however again most of the protein aggregated into inclusion bodies within the cells, suggesting that this mutation was not noticeably stabilizing with respect to protein expression. At a lower growth temperature of 18°C much more of the protein became soluble (Figure 3.25), indicating that slower growth rates and lower IPTG levels promote solubility, as with slower expression, the cellular machinery isn't as overwhelmed by protein overexpression and the proteins have more time to fold properly and avoid aggregation.

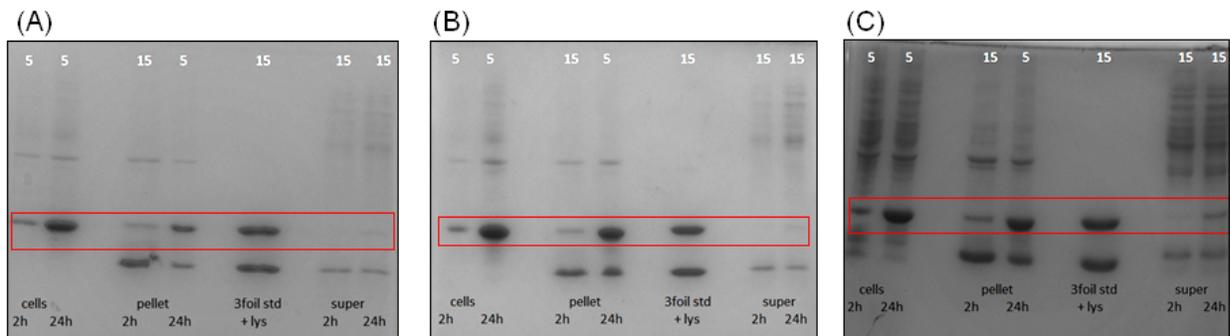


Figure 3.24: SDS-PAGE gel of expression of mutant A62V Threefoil, at 37°C, in LB broth

(A) Induction at 1 mM IPTG. There is virtually no soluble protein in the supernatant fraction (denoted 'super') and almost all of the protein goes into inclusion bodies and can be found in the pellet. (B) Induction at 0.35 mM IPTG. Slightly more soluble protein is observed at 24 hours but again most of the protein is found in the pellet. (C) Induction at 0.1 mM IPTG. A noticeable amount of soluble protein is observed at 24 hours in the supernatant fraction. 'lys' denotes a Lysozyme standard that was run along with the samples. 'Threefoil std' denotes a Threefoil standard run alongside the samples in the place of a protein standard. Numbers at the top of the gel represent the amount of sample loaded into each lane (in μL).

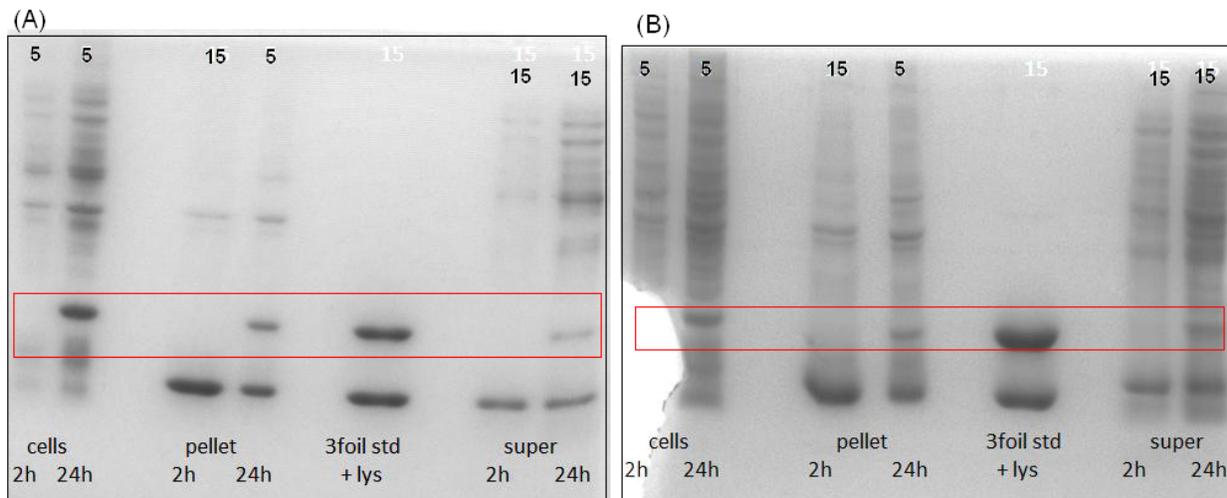


Figure 3.25: SDS-PAGE gel of expression of mutant A62V Threefoil, at 18°C, in LB broth

(A) Induction at 0.35 mM IPTG. Most of the protein is still in the pellet fraction; however, here approximately one-fifth of the protein is soluble here as opposed to that grown at 37°C at the same IPTG concentration. (B) Induction at 0.1 mM IPTG. Up to one-quarter of the protein is soluble and found in the supernatant, compared to protein expressed under the same conditions at 37°C. ‘lys’ denotes a Lysozyme standard that was run along with the samples. ‘Threefoil std’ denotes a Threefoil standard run alongside the samples in the place of a protein ladder. Numbers at the top of the gel represent the amount of sample loaded into each lane (in μ L).

For mutant Q78I, at 37°C and all three IPTG concentrations in LB broth, there was no appreciable amount of soluble protein produced (Figure 3.26 A, B, C). At 25°C slightly more protein was found in the soluble fraction but most were found in inclusion bodies (Figure 3.26 D, E, F). At 18°C, at 0.35 mM and 0.1 mM IPTG (Figure 3.27) it appears as if virtually no soluble protein was obtained; however, the poor quality of the gel prevents accurate analysis. Overall these expression tests suggest that the mutation is sufficiently destabilizing that the protein goes almost exclusively into inclusion bodies regardless of IPTG concentration or temperature.

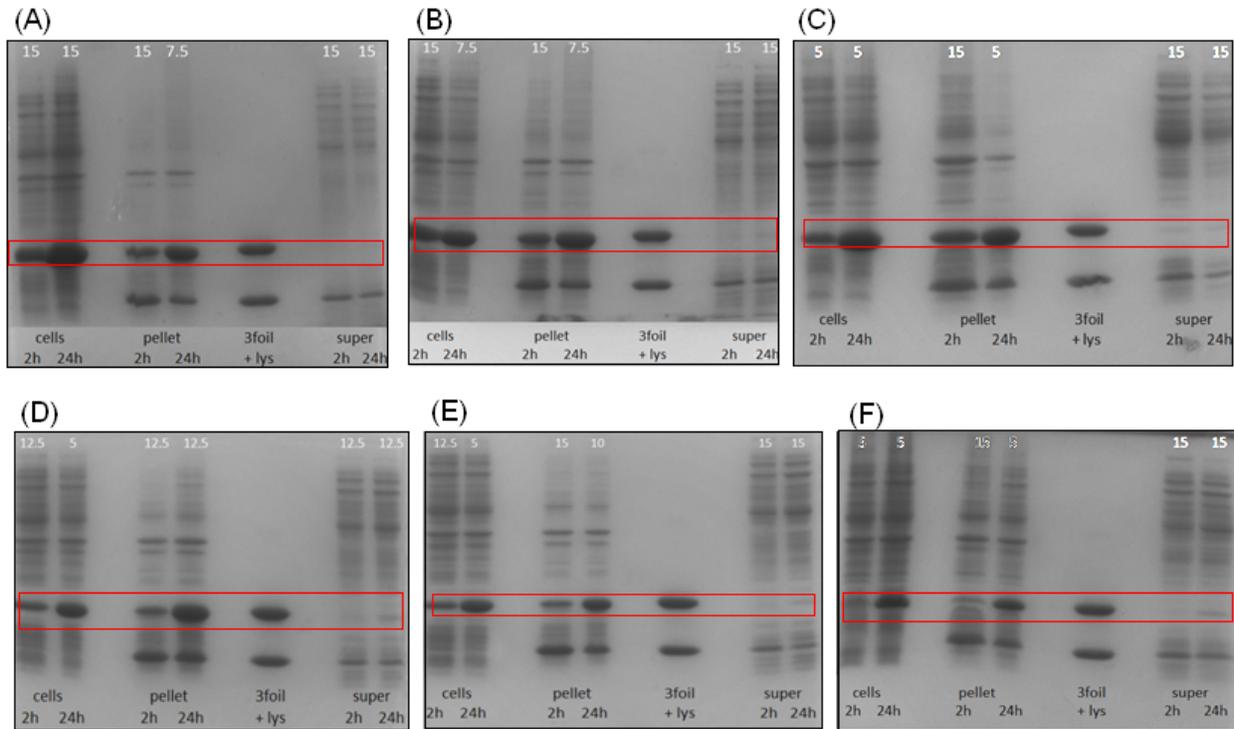


Figure 3.26: SDS-PAGE gel of expression of mutant Q78I Threefoil, at 37°C and 25°C, in LB broth

(A) 37°C, 1 mM IPTG. (B) 37°C, 0.35 mM IPTG. (C) 37°C, 0.1 mM IPTG. (D) 25°C, 1 mM IPTG. (E) 25°C, 0.35 mM IPTG. (F) 25°C, 0.1 mM IPTG. Slightly more Threefoil becomes soluble when grown at 25°C than at 37°C; however, almost all of the protein is found within the pellet. Numbers at the top of the gel represent the amount of sample loaded into each lane (in μL).

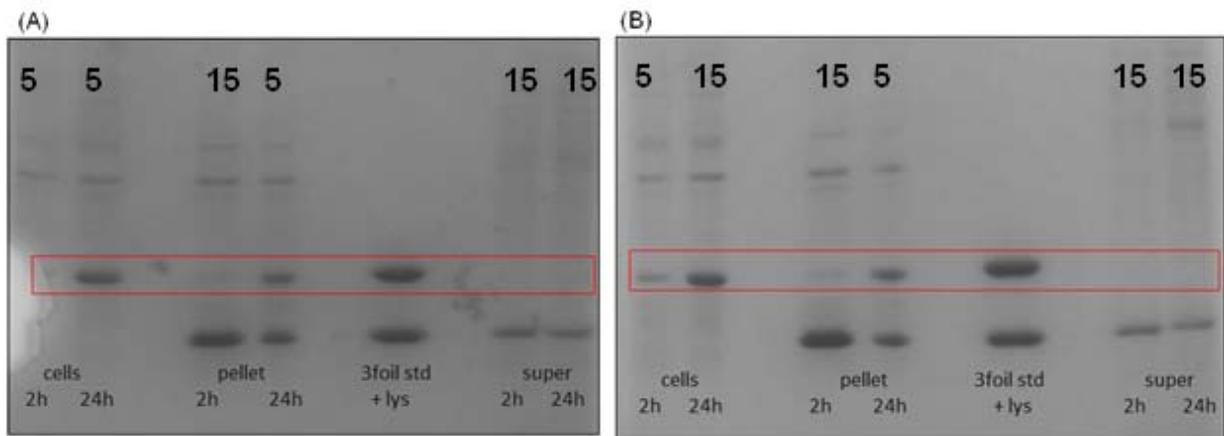


Figure 3.27: SDS-PAGE gel of expression of mutant Q78I Threefoil, at 18°C, in LB broth

(A) 0.35 mM IPTG. (B) 0.1 mM IPTG. No soluble protein can be found within the supernatant. However, it is possible that due to the poor quality of the gels, what soluble protein there is simply cannot be seen. Numbers at the top of the gel represent the amount of sample loaded into each lane (in μL).

Threefoil mutant D85P occurs in a 3_{10} helix, which is thought to contribute some instability to the protein due to unfavorable geometry of the hydrogen bonds between residues. From the expression testing of D85P, the mutation to a proline, which would conceivably introduce conformational rigidity to the 3_{10} helix, has no marked effect on solubility of the protein. At 37°C or 25°C there is little to no soluble protein obtained even at the lowest concentration of IPTG (Figure 3.28). Similarly for proteins expressed at 18°C, no increase in solubility was noticed (Figure 3.29).

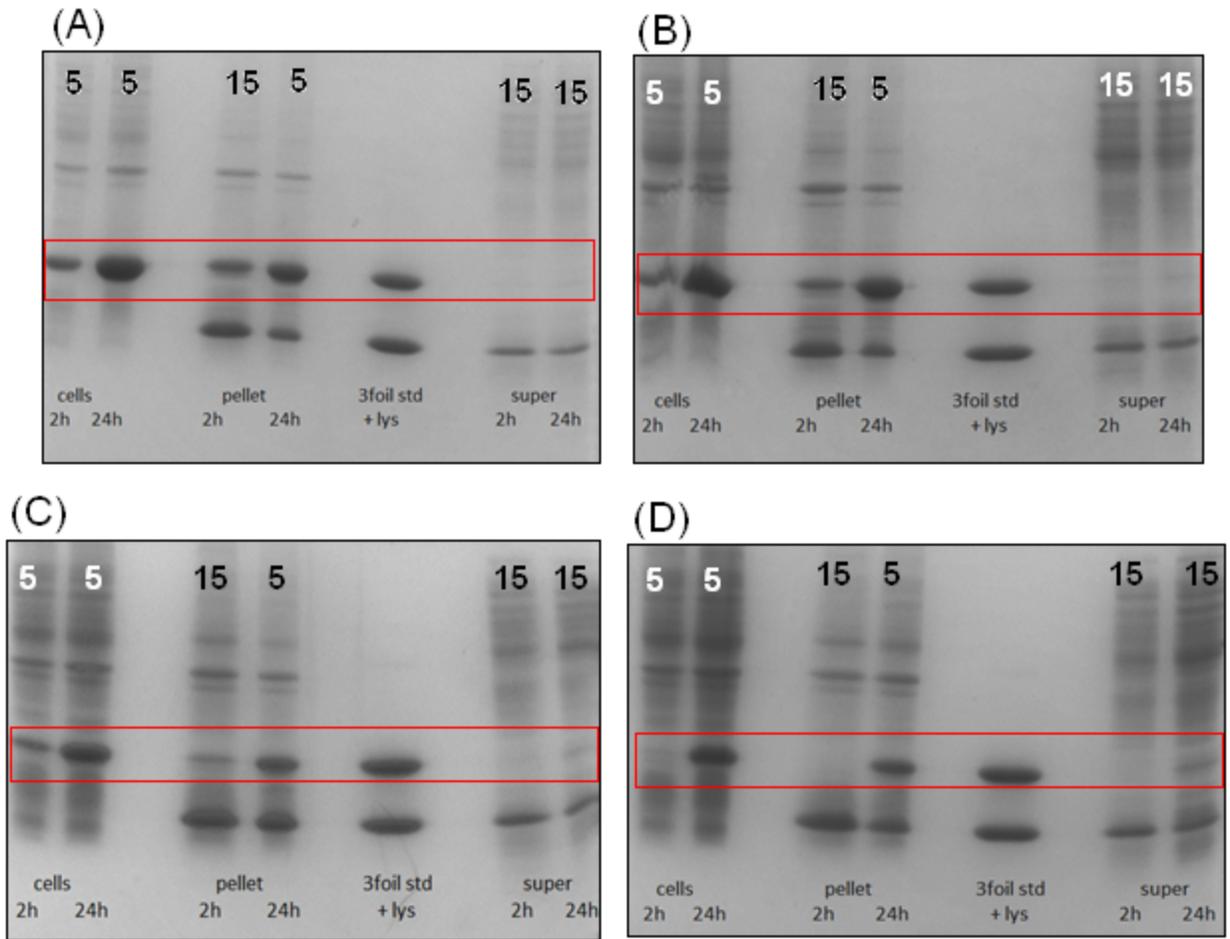


Figure 3.28: SDS-PAGE gel of expression of mutant D85P Threefoil, at 37°C and 25°C, in LB broth
 (A) 37°C, 0.35 mM IPTG. (B) 37°C, 0.1 mM IPTG. (C) 25°C, 0.35 mM IPTG. (D) 25°C, 0.1 mM IPTG.

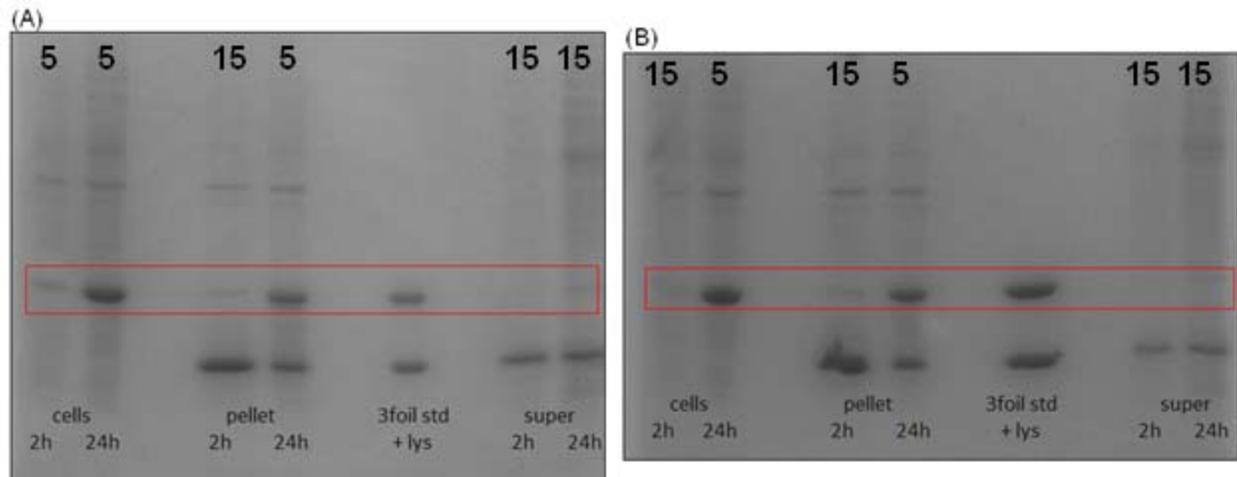


Figure 3.29: SDS-PAGE gel of expression of mutant D85P Threefoil, at 18°C, in LB broth

(A) 0.35 mM IPTG. (B) 0.1 mM IPTG.

The last mutant studied, D93P, appears slightly more soluble than the previous three mutations even at 37°C (Figure 3.30). At 25°C there is a noticeable band in the SDS-PAGE gel corresponding to the Threefoil mutant (Figure 3.30 D, E, F) at 24 hours after induction. At 18°C the band intensity corresponding to soluble protein is decreased (Figure 3.31). This suggests that either the previous samples had contamination from the pellet fraction in the supernatant fraction, giving a misleading result, or that the SDS-PAGE gel run for 18°C is unreliable. It should be noted that the intensities for all of the bands corresponding to the *E. coli* host proteins are reduced in the 18°C gel, which points to the latter explanation being more plausible.

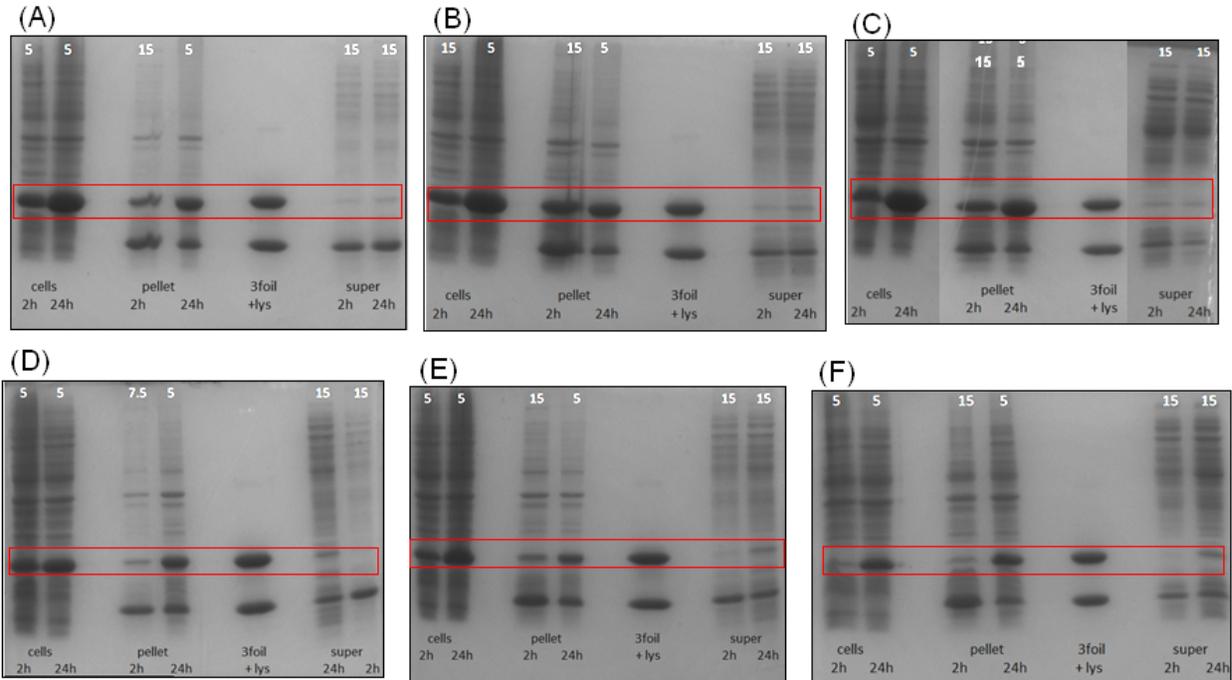


Figure 3.30: SDS-PAGE gel of expression of mutant D93P Threefoil, at 37°C and 25°C, in LB broth

(A) 37°C, 1 mM IPTG. (B) 37°C, 0.35 mM IPTG. (C) 37°C, 0.1 mM IPTG. (D) 25°C, 1 mM IPTG. (E) 25°C, 0.35 mM IPTG. (F) 25°C, 0.1 mM IPTG. While most of the protein remains in the pellet as inclusion bodies, more soluble protein is expressed for D93P. At 24 hours at 25°C, for all of the IPTG concentrations, the band intensity remains approximately the same, which points to the supernatant being saturated with protein. This indicates that more of soluble D93P is produced.

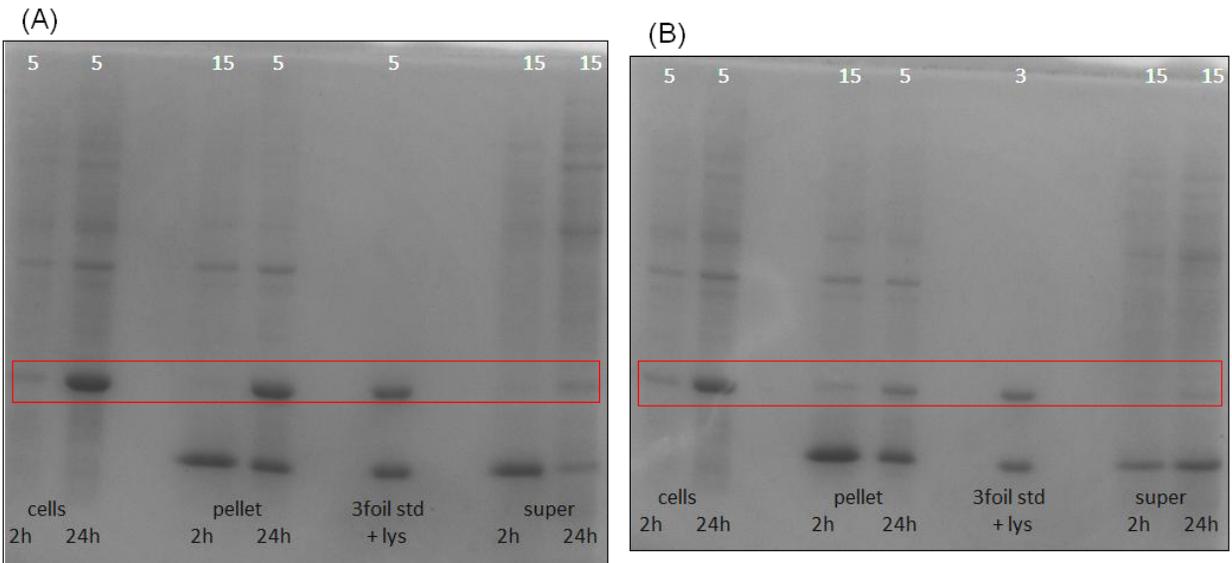


Figure 3.31: SDS-PAGE gel of expression of mutant D93P Threefoil, at 18°C, in LB broth

(A) 0.35 mM IPTG. (B) 0.1 mM IPTG.

Chapter 4: Discussion

4.1: Threefoil folds according to a two-state model with residual structure in the denatured state

In the early '80s it was generally assumed that a protein which is unfolded in strong denaturant adopts a random coil conformation. However, numerous studies have shown that even in its denatured form, a protein can retain substantial residual structure.^{82,83,84}

The results presented in this thesis show that in the case for Threefoil, there is extensive secondary and tertiary residual structure in the denatured state in urea and GuHCl. Threefoil will only denature in the strongest denaturant, GuSCN (Figure 3.2). In 3M GuSCN approximately 3 days are required for the protein to unfold completely, which speaks to Threefoil's remarkable stability. In urea, even at 6M, native Threefoil does not unfold to any great extent.²⁷ Threefoil inclusion bodies, however, are less resistant to denaturant and will denature in 6M urea to a state with residual structure (Figures 3.2, 3.3). The denatured proteins contain extensive secondary structural elements as well as retaining some protection of tryptophan. Fluorescence monitoring of Threefoil's tryptophan residues reveals that the residues are not completely solvent-exposed when the protein is denatured in urea, as the maximum fluorescence emission is approximately 10 nm blue-shifted from that expected of fully unfolded protein (Figure 3.2). This indicates that the tryptophans are still semi-shielded from solvent.

Similarly, in the presence of high concentrations of GuHCl, Threefoil also retains extensive secondary structure (Figure 3.3), as much as inclusion bodies denatured in 6M urea. However there is less residual tertiary structure in GuHCl (Figure 3.2), presumably because

GuHCl is a stronger denaturant than urea and can induce a loss of tertiary structure without affecting the secondary structure.

The presence of residual structure does not appear to hinder folding to the native state. When refolding from inclusion bodies with residual structure, Threefoil appears to mostly fold correctly to its native form despite a slight propensity toward aggregation. It is possible that the presence of residual structure acts as a scaffold for refolding, guiding the protein toward its correct native conformation. A similar result is observed for urea denaturation of the chaperonin GroEL. When GroEL is denatured in urea concentrations higher than 3.1 M, leading to complete unfolding of the protein without any residual structure, it is unable to adopt its native conformation upon refolding out of urea.⁸²

Equilibrium renaturation curves of inclusion bodies solubilized in urea, monitored by two different optical probes and at two different wavelengths per optical probe, suggest that the folding transition is indeed two-state with residual structure present in the denatured state (Table 2). Within error, renaturation m -values and C_{mid} values between CD and fluorescence agree. Within the same optical probe, m -values and C_{mid} values also agree within error. The experimental m -value, approximately 1.6 kcal/(mol*M) is lower than the calculated m -value of 1.7 kcal/(mol*M), because of the presence of residual structure in the denatured state. Folding from a denatured state with residual structure to the native state does not represent the full transition; therefore the m -value will be lower than expected.

Equilibrium denaturation and renaturation curves in GuSCN do not coincide at the midpoint, since Threefoil denaturation has not reached equilibrium at the time of measurement. However, equilibrium is reached for protein renaturation, after approximately 100 days. The m -values between renaturation and denaturation curves agree also within error, with a value of

approximately 8.5 kcal/(mol*D). The m-value obtained from equilibrium curves also agrees within error with the m-value obtained from fitting the GuSCN chevron plot (table 3), suggesting again that Threefoil folds according to a 2-state mechanism. In both branches (folding and unfolding) of the GuSCN Chevron there is a lack of rollover (or deviation from the two-state line of fit) which is characteristic of Hammond behavior, and a narrow transition state region. Experimental m-values (8.5 kcal/(mol*D)) in GuSCN also disagree with calculated m-values (11.8 kcal/(mol*D)). There does not appear to be an intermediate formed during kinetic experiments and the equilibrium curves are well-fit to a two-state transition. One possible explanation is that the denatured state of proteins in 3 M GuSCN contains secondary structural elements, as in urea-denatured proteins. Since CD spectra in GuSCN are not possible to obtain, there is no direct evidence for there being residual secondary structure. However, Lapanje *et. al.* have shown by optical rotation and viscosity studies that proteins denatured in GuSCN concentrations up to 4 M GuSCN do retain some elements of residual structure.^{85,86} Therefore, it is possible that proteins denatured in 3 M GuSCN may retain some residual secondary structure, in which case the experimental m-values are not expected to agree with calculated m-values.

4.2: Threefoil folding and unfolding is remarkably slow

Compared to other proteins of similar size, and compared within the same denaturant (urea, GuHCl or GuSCN) Threefoil folding is remarkably slow (Table 3). The Tanford β value, β_T , is a measure of the average degree of solvent-exposure in the transition state relative to that of the denatured state from the native state.¹⁰ Threefoil has a β_T value of 0.65, suggesting that the transition state has approximately 65% burial of the solvent exposed surface area. While the transition state can still be considered native-like, it is still less so than other β trefoils (Table 3).

In order to compare the solvent exposure of Threefoil upon unfolding in different denaturants, an m_u' value (Table 3) was calculated based on the empirically derived, relative denaturing strengths of GuHCl and GuSCN to urea. From comparing m_u' values it is clear that Threefoil unfolds with a much larger exposure to solvent than other the other proteins, as the m -value is proportional to the change in solvent accessible surface area upon unfolding.

Threefoil also resistant to denaturation by the most commonly used denaturants, urea and GuHCl, at high concentrations. This speaks to Threefoil's kinetic stability, that it requires one of the strongest chemical denaturants available, GuSCN, in order to denature in a practical time frame to do experiments. Compared to other β -trefoil proteins, Threefoil folding and unfolding is extremely slow and one must look to the thermophilic proteins, which must possess high kinetic stability in temperatures exceeding 80°C in order to resist thermal denaturation at high heat, to find comparable rates.

Refolding tends to be accompanied by aggregation, which is visible by eye at protein concentrations of 10 μ M and above. At such high protein concentrations a white precipitate visibly forms immediately upon addition of protein to refolding buffer, which slowly dissipates over time, indicating that the protein initially aggregates and then disaggregates. The formation of aggregates at early time points and the eventual disaggregation suggests that aggregation of Threefoil involves protein in its unfolded state, rather than the native protein. However, the major folding transition as monitored by CD and tryptophan fluorescence can be well-fit to a two-state model, despite the presence of aggregation.

Chevron analysis of Threefoil in urea, GuHCl and GuSCN (Figure 3.16) indicate that the folding branch for all three extrapolate to essentially the same point in 0M denaturant, which is

consistent with a two-state model of folding. This furthermore indicates that folding occurs from comparable denatured states whether in urea, GuHCl or GuSCN.

4.3: Metallation on Threefoil stability

Metal-binding can modulate the stability of proteins by either coordinating residues to facilitate folding to the native state or creating non-native contacts that can hinder folding. Different metals can also have markedly different effects on the stability of a protein. For example, in the tumour suppressor protein p53 DNA-binding domain, binding of Mg^{2+} and Zn^{2+} both increased protein stability.¹⁸ In mutants of the tumor suppressor protein S100A2, however, binding to Zn^{2+} destabilized the mutants while binding to Ca^{2+} stabilized the mutants.¹⁹

In Threefoil, metal-binding (sodium) increases the stability of the native state, as reflected in the increase in the rate of unfolding when the metal is removed (Figure 3.18). Refolding in the presence or absence of metal does not noticeably affect the kinetic rate, suggesting that the stabilizing effect on the native state is larger than on the transition state or the unfolded state, and also that the metal-binding site is not formed in the transition state.

4.4: Carbohydrate-binding on Threefoil stability

Carbohydrate-binding to the native protein can be expected to increase its thermodynamic stability and alter its folding and unfolding kinetics.²²⁻²⁴ Carbohydrate binding can be stabilizing if the carbohydrate binds preferentially to the native state over the unfolded state, lowering the energy barrier for folding. However, depending on how much the

carbohydrate also binds to the transition state compared to the folded and unfolded states, there can be effects on both folding and unfolding kinetics.²⁵

In Threefoil, binding to sucrose does not produce a marked effect on protein stability, whereas binding to lactose does (Figure 3.19). This is explained by the fact that the protein upon which the Threefoil sequence is based is a glycosidase which binds galactose. Therefore it is not surprising that Threefoil also binds galactose, or any other carbohydrate containing a galactose subunit, such as lactose which is composed of galactose and glucose subunits. Carbohydrate binding increases the rate of refolding and decreases the rate of unfolding, which indicates that the native state and the transition states are stabilized preferentially over the unfolded state, with possibly the native state being stabilized more than the transition state. These results also suggest that the carbohydrate-binding site may be partially formed in the transition state.

Chapter 5: Conclusions and future work

5.1: Equilibrium and kinetic studies

Threefoil forms almost exclusively in inclusion bodies but was able to be isolated in high yield and purity. When refolded from inclusion bodies, which have residual secondary structure in the denatured state in urea and GuHCl, as probed by CD and fluorescence respectively (Figures 3.2, 3.3), Threefoil is able to adopt a properly folded native conformation. Equilibrium and kinetic studies suggest that denatured Threefoil IBs fold cooperatively in a two-state transition with residual structure. From refolding equilibrium studies of IBs denatured in urea and GuHCl, it was shown that by different optical probes (CD and fluorescence) and between different wavelengths monitored for the same optical probe (215 nm and 230 nm for CD; 313 nm and 370 nm for fluorescence) the *m*-values all agree within experimental error (Table 3), which is behavior characteristic of a two-state transition.

Native Threefoil equilibrium and kinetic stability was studied by fluorescence in the presence of GuSCN, a strong denaturant, as the protein does not denature in urea and is resistant to denaturation in GuHCl. It was shown that native Threefoil also folds according to a two-state model with residual structure in the denatured state, as equilibrium *m*-values matched kinetic *m*-values in GuSCN, within error. Folding in all denaturants is complicated by competition with aggregation, which is visible to the naked eye at higher protein concentrations (above 10 μ M); however, the major transition corresponding to protein folding follows a two-state model and there is no rate-dependence of folding on protein concentration (Figure 3.14).

The thermodynamic stability of Threefoil was studied by fluorescence-monitored GuSCN denaturation and renaturation curves (Figure 3.6, 3.7). The time frame of achieving equilibrium by refolding is staggeringly long. Renaturation and denaturation curves are not coincident at the conclusion of two years' time, mainly due to the fact that the denaturation samples have not yet reached equilibrium (Figure 3.7). However, both renaturation and denaturation curves can be fit to a two-state model, and the transition in both cases is cooperative. On average more than 100 days is required for Threefoil to reach equilibrium by folding; however, equilibrium by unfolding does not occur even after two years.

5.2: Ligand binding in Threefoil stability

It was additionally shown that metal-binding, in particular sodium, has an overall stabilizing effect (Figure 3.18), as does carbohydrate binding (Figure 3.19). Both involve preferential binding to the native state over the unfolded state, which lowers the energy of, and shifts the equilibrium toward, the native state. Interestingly Threefoil is only stabilized in the presence of lactose; sucrose has no pronounced effect on stability. The stabilization by lactose is not unexpected as Threefoil's template sequence was based off of a glycosidase which bound galactose²⁷, and lactose is composed of one such galactose unit; therefore, it is both interesting and a testament to the success of Threefoil's design that it has retained its sugar-binding functionality. However, the lack of effect on the protein folding rate of sucrose is contrary to many studies in which proteins are stabilized by sucrose that do not bind it, due to molecular crowding effects,²²⁻²⁴ and bears more investigation.

Ten mutants of Threefoil were also successfully expressed and isolated in their native form from the treatment of inclusion bodies. These mutations were screened and selected by first

using six prediction algorithms to calculate the most stabilizing mutation, in terms of the $\Delta\Delta G$, at each location in the sequence. The wild-type amino acid was not included in these calculations, as was cysteine, to avoid complications from the formation of disulfide bonds which affect aggregation propensity. The mutations did not affect solubility to a great extent, indicating that the mutations also have no pronounced effect on the propensity of the protein to form inclusion bodies.

5.3: Future work

Due to the length of time that is required to carry out studies of both kinetic and thermodynamic stability, it was not always feasible to repeat experiments. Therefore in future it is desirable to carry on the work presented here and to conduct repeat measurements of the equilibrium experiments discussed in order to ensure reliability of the data. In addition the mutants must be studied in more detail with respect to their stability in comparison to wild-type Threefoil. In order to be able to draw conclusions regarding their impact on not only stability but on aggregation propensity and inclusion body formation, experiments of unfolding/refolding kinetics and denaturation/renaturation equilibrium must be undertaken on both wild-type and mutant Threefoil. The role of metallation and carbohydrate-binding as well would be interesting to study. In particular, mutations at sites which affect the carbohydrate-binding capability of the protein will reveal the relation of function to stability—whether compromising the functional sugar-binding site will affect stability of the protein.

References

1. Graumann, K., Premstaller, A. 2006. Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnol. J.* **1**: 164-186.
2. Ferrer-Miralles et. al. 2009. Microbial factories for recombinant pharmaceuticals. *Microbial Cell Factories.* **8**:17.
3. Dobson, C. M. 2004. Principles of protein folding, misfolding and aggregation. *Semin. Cell. Dev. Biol.* **15**: 3-16.
4. Dobson, C.M. 2003. Protein folding and misfolding. *Nature.* **426**: 884-890.
5. Gregoire, S., Irwin, J., Kwon, I. 2012. Techniques for monitoring protein misfolding and aggregation in living cells. *Korean J. Chem. Eng.* **29**(6): 693-702.
6. Jurgen B., Breitenstein, A., Urlacher, V., Buttner, K., Lin, H., Hecker, M., Schweder, T., Neubauer, P. 2010. Quality control of inclusion bodies in *Escherichia coli*. *Microbial Cell Factories.* **9**:41.
7. Gasser B., Saloheimo, M., Rinas, U., Dragosits, M., Rodriguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M.L., Mattanovich, D., Villaverde, A. 2008. Protein folding and conformational stress in microbial cells producing recombinant proteins: A host comparative overview. *Microbial Cell Factories.* **7**:11.
8. Frauenfelder, H., Sligar, S.G., Wolynes, P.G. 1991. The energy landscapes and motions of proteins. *Science.* **254**(5038): 1598-1603.
9. Granata D., Camilloni, C., Vendruscolo, M., Laio, A. 2013. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *PNAS*. doi:10.1073/pnas.1218350110.
10. Jackson, S. E. 1998. How do small single-domain proteins fold? *Fold Des* **3**: R81-91.
11. Fersht, A. 1999. Structure and mechanism in protein sciences: a guide to enzyme catalysis and protein folding. *W. H Freeman and Company*.
12. Pace, C. N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266-280.
13. Liu, C., Chu, D., Wideman, R.D., Houliston, R.S., Wong, H.J., Meiering, E.M. 2001. Thermodynamics of denaturation of Hisactophilin, a β -trefoil protein. *Biochem.* **40**(13): 3817-3827.

14. Tanford, C. 1970. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* **24**: 1–95.
15. Meyers, J.K., Pace, C.N., Scholtz, J.M. 1995. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Prot. Sci.* **4**(10): 2138-2148.
16. Tottey, S., Waldron, K.J., Firbank, S.J., Reale, B., Bessant, C., Sato, K., Cheek, T.R., Gray, J., Banfield, M.J., Dennison, C., Robinson, N.J. 2008. Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature.* **455**: 1138-1142.
17. Rosenzweig, A.C. 2002. Metallochaperones: Bind and deliver. *Chem. & Biol.* **9**: 673-677.
18. Xue, Y., Wang, S., Feng, X. 2009. Effect of metal ion on the structural stability of tumor suppressor protein p53 DNA-binding domain. *J. Biochem.* **146**(2): 193-200.
19. Botelho H.M., Koch, M., Fritz, G., Gomes, C.M. 2009. Metal ions modulate the folding and stability of the tumor suppressor protein S100A2. *FEBS J.* **276**: 1776-1786.
20. Stubbe, J., Cotruvo Jr, J.A. 2012. Metallation and mismetallation of iron and manganese proteins *in vitro* and *in vivo*: The class I ribonucleotide reductases as a case study. *Metallomics.* **4**: 1020-1036.
21. Maret, W. 2010. Metalloproteomics, metalloproteomes, and the annotation of metalloproteins. *Metallomics.* **2**: 117-125.
22. Graziano, G. 2012. How does sucrose stabilize the native state of globular proteins? *Int. J. Biol. Macromol.* **50**: 230-235.
23. Kim, Y.S., Jones, L.S., Dong, A., Kendrick, B.S., Chang, B.S., Manning, M.C., Randolph, T.W., Carpenter, J.F. 2003. Effects of sucrose on conformational equilibria and fluctuations within the native-state ensemble of proteins. *Prot. Sci.* **12**: 1252-1261.
24. Lee, J.C., Timasheff, S.N. 1981. The stabilization of proteins by sucrose. *J. Biol. Chem.* **256**(14): 7193-7201.
25. Sancho J, Meiering EM, Fersht AR. Mapping transition states of protein unfolding by protein engineering of ligand-binding sites. *J Mol Biol* 1991, 221, 1007-1014.
26. Saadati, Z., Bordbar, A.-K. 2008. Stability of β -Lactoglobulin A in the presence of sugar osmolytes estimated from their guanidinium chloride-induced transition curves. *Prot. J.* **27**:455-460
27. Broom, Robert Aron. 2010. *From peptides to proteins: Exploring modular evolution through the beta-trefoil fold*. Master's thesis, The University of Waterloo.

28. Broom A., Doxey, A.C., Lobsanov, Y.D., Berthin, L.G., Rose, D.R., Howell, P.L., McConkey, B.J., Meiering, E.M. 2012. Modular evolution and the origins of symmetry: Reconstruction of a three-fold symmetric globular protein. *Structure*. **20**: 161-171.
29. Teng, P.K., Eisenberg, D. 2009. Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng. Des. Sel.* **22**(8):531-536.
30. Wang, L. 2009. Towards revealing the structure of bacterial inclusion bodies. *Prion*. **3**(3):139-145.
31. Desplats, P., Lee, H.J., Bae, E.J., Patrick, C., Rockenstein, E., Crews, L., Spencer, B., Masliah, E., Lee, S.J. 2009. Inclusion formation and neuronal cell death through neuron-to-neuron transmission of alpha-synuclein. *PNAS*. **106**(31):13010-13015.
32. Kirschner, D.A., Abraham, C., Selkoe, D.J. 1986. X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in Alzheimer disease indicates cross-beta conformation. *Proc. Natl. Acad. Sci. USA*. **83**:503-507.
33. Rodríguez-Carmona *et al.* 2010. Isolation of cell-free bacterial inclusion bodies. *Microb. Cell Fact.* **9**(71)
34. Ventura, S., Villaverde, A. 2006. Protein quality in bacterial inclusion bodies. *Trends Biotechnol.* **24**:179-185.
35. Frere, J.M. 1995. Beta-lactamases and bacterial resistance to antibiotics. *Molec. Microbiol.* **16**(3):385-395.
36. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P.S., Palzkill, T. 1996. Amino acid sequence determinants of β -lactamase structure and activity. *J. Mol. Biol.* **258**:688-703.
37. Calloni, G., Zoffoli, S., Stefani, M., Dobson, C.M., Chiti, F. 2005. Investigating the effects of mutations on protein aggregation in the cell. *J. Biol. Chem.* **280**(11):10607-10613.
38. Winkelmann, J., Calloni, G., Campioni, S., Mannini, B., Taddei, N., Chiti, F. 2010. Low-level expression of a folding-incompetent protein in *Escherichia coli*: Search for the molecular determinants of protein aggregation *in vivo*. *J. Mol. Biol.* **398**:600-613.
39. Manning, M.C., Chou, D.K., Murphy, B.M., Payne, R.W., Katayama, D.S. 2010. Stability of protein pharmaceuticals: An update. *Pharm. Res.* **27**(4):544-575.
40. García-Fruitós, E., González-Montalbán, N., Morell, M., Vera, A., Ferraz, R.M., Arís, A., Ventura, S., Villaverde, A. 2005. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb. Cell. Fact.* **4**(27).

41. Fink, A.L. 1998. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des.* **3**:9-23.
42. Speed, M.A., Wang, D.I.C., King, J. 1996. Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nat. Biotech.* **14**:1283-1287.
43. Chrnyk, B.A., Evans, J., Lillquist, J., Young, P., Wetzel, R. 1993. Inclusion body formation and protein stability in sequence variants of interleukin-1beta. *J. Biol. Chem.* **268**(24):18053-18061.
44. Klunk, W.E., Jacob, R.F., Mason, R.P. 1999. Quantifying amyloid by congo red spectral shift assay. *Method Enzymol.* **309**:285-305.
45. Buell, A.K., Dobson, C.M., Knowles, T.P., Welland, M.E. 2010. Interactions between amyloidophilic dyes and their relevance to studies of amyloid inhibitors. *Biophys. J.* **99**:3492-3497.
46. Carrio, M., González-Montalbán, N., Vera, A., Villaverde, A., Ventura, S. 2005. Amyloid-like properties of bacterial inclusion bodies. *J. Mol. Biol.* **347**:1025-1037.
47. Howie, A.J., Brewer, D.B. 2009. Optical properties of amyloid stained by Congo red: History and mechanisms. *Micron.* **40**:285-301.
48. de Groot, N.S., Sabate, R., Ventura, S. 2009. Amyloids in bacterial inclusion bodies. *TiBS.* **34**(8):408-416.
49. Leinweber, B., Barofsky, E., Barofsky, D.F., Ermilov, V., Nylin, K., Beckman, J.S. 2004. Aggregation of ALS mutant superoxide dismutase expressed in *Escherichia coli*. *Free Rad. Biol. Med.* **36**(7):911-918.
50. Calamai, M., Taddei, N., Stefani, M., Ramponi, G., Chiti, F. 2003. Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins. *Biochem.* **42**:15078-15083.
51. Calamai, M., Chiti, F., Dobson, C.M. 2005. Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophys. J.* **89**:4201-4210.
52. Liu, C., Sawaya, M.R., Eisenberg, D. 2011. β_2 -microglobulin forms three-dimensional domain-swapped amyloid fibrils with disulfide linkages. *Nat. Struct. Mol. Biol.* **18**(1):49-55.
53. Knappik, A., Pluckthun, A. 1995. Engineered turns of a recombinant antibody improve its *in vivo* folding. *Prot. Eng.* **8**(1):81-89.

54. Carrio, M.M., Villaverde, A. 2002. Construction and deconstruction of bacterial inclusion bodies. *J. Biotechnol.* **96**:3-12.
55. Hoskins, J., Lovell, S., Blundell, T.L. 2006. An algorithm for predicting protein-protein interaction sites: Abnormally exposed amino acid residues and secondary structure elements. *Prot. Sci.* **15**:1017-1029.
56. de Groot, N.S., Ventura, S. 2006. Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *J. Biotechnol.* **125**:110-113.
57. Martinez-Alonso, M., Gonzalez-Montalban, N., Garcia-Fruitos, E., Villaverde, A.. 2009. Learning about protein solubility from bacterial inclusion bodies. *Microbial Cell Fact.* **8**(4).
58. de Groot, N.S., Ventura, S. 2006. Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *J. Biotechnol.* **125**:110-113.
59. Hosia, W., Bark, N., Liepinsh, E., Tjernberg, A., Persson, B., Hallen, D., Thyberg, J., Johansson, J., Tjernberg, L. 2004. Folding into a β -hairpin can prevent amyloid fibril formation. *Biochem.* **43**:4655-4661.
60. Vagenende, V., Yap, M.G.S., Trout, B.L. 2009. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochem.* **48**:11084-11096.
61. Idicula-Thomas, S., Balaki, P.V. 2005. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in *Escherichia coli*. *Prot. Sci.* **14**:582-592.
62. Wu, W., Xing, L., Zhou, B., Lin, Z. 2011. Active protein aggregates induced by terminally attached self-assembling peptide ELK16 in *Escherichia coli*. *Microbial Cell Fact.* **10**(9).
63. Calloni, G., Taddei, N., Plaxco, K.W., Ramponi, G., Stefani, M., Chiti, F. 2003. Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding. **330**:577-591.
64. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C.H., Edwards, A.M. 2000. Structural proteomics: Prospects for high throughput sample preparation. **73**:339-345.
65. Diaz, A.A., Tomba, E., Lennarson, R., Richardson, R., Bagajewicz, M.J., Harrison, R.G. 2010. Prediction of protein solubility in *Escherichia coli* using logistic regression. *Biotechnol. Bioeng.* **105**:374-383.
66. Saunders, H.M., Bottomley, S.P. 2009. Multi-domain misfolding: Understanding the aggregation pathway of polyglutamine proteins. *Prot. Eng. Des. Sel.* **22**(8);447-451.

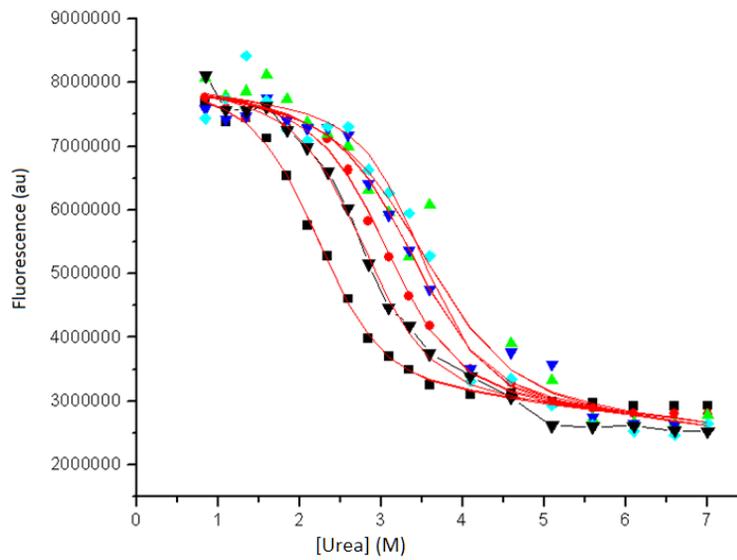
67. Khakshoor, O., Lin, A.J., Korman, T.P., Sawaya, M.R., Tsai, S.C., Eisenberg, D., Nowick, J.S. 2010. X-ray crystallographic structure of an artificial β -sheet dimer. *JACS*. **132**:11622-11628.
68. Koschorreck, M., Fischer, M., Barth, S., Pleiss, J. 2005. How to find soluble proteins: A comprehensive analysis of alpha/beta hydrolases for recombinant expression in *E. coli*. *BMC Genomics*. **6**(49).
69. Bodner, R.A., Housman, D.E., Kazantsev, A.G. 2006. New directions for neurodegenerative disease therapy. *Cell Cycle*. **5**(14):1477-1480.
70. Routledge, K.E., Tartaglia, G.G., Platt, G.W., Vendruscolo, M., Radford, S.E. 2009. Competition between intramolecular and intermolecular interactions in an amyloid-forming protein. *J. Mol. Biol.* **389**(4):776-786.
71. Jahn, T.R., Radford, S.E. 2008. Folding versus aggregation: Polypeptide conformations on competing pathways. *ABB*. **469**:100-117.
72. Garcia-Fruitos, E., González-Montalbán, N., Morell, M., Vera, A., Ferraz, R.M., Arís, A., Ventura, S., Villaverde, A. 2005. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb. Cell. Fact.* **4**(27).
73. Carson, M., Johnson, D.H., McDonald, H., Brouillette, C., Delucas, L.J. 2007. His-tag impact on structure. *Acta Cryst.* **D43**: 295-301.
74. Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A., Smith, L.J. 1999. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochem.* **38**(50):16424-16431.
75. Cota, E., Clarke, J. 2000. Folding of beta-sandwich proteins: three-state transition of a fibronectin type III module. *Prot. Sci.* **9**(1):112-20.
76. Pace, C.N. 1990. Conformational stability of globular proteins. *Trends Biochem. Sci.* **15**: 14-17.
77. Kardinahl, S., Anemuller, S., Schafer, G. 2000. The hyper-thermostable Fe-superoxide dismutase from the archaeon *Acidianus ambivalens*: Characterization, recombinant expression, crystallization and effects of metal exchange. *Biol. Chem.* **381**: 1089-1101.
78. Villalonga, M.L., Reyes, G., Villalonga, R. 2004. Metal-induced stabilization of trypsin modified with α -oxoglutaric acid. *Biotechnology Letters*. **26**: 209-212.

79. Pace, C.N., Laurents, D.V., Erickson, R.E. 1992. Urea denaturation of barnase: pH dependence and characterization of the unfolded state. *Biochem.* **31**(10): 2728-2734.
80. Hernandez-Arana, A., Rojo-Dominguez, R. 1993. Differential scanning calorimetry of the irreversible denaturation of *Escherichia coli* glucosamine-6-phosphate deaminase. *Biochem.* **32**: 3644-3648.
81. Semisotnov, G.V., Rodionova, N.A., Razgulyaev, O.I., Uversky, V.N., Gripas, A.F., Gilmanshin, R.I. 1991. Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers.* **31**(1): 119-128.
82. Gorovits, B.M., Seale, J.W., Horowitz, P.M. 1995. Residual structure in urea-denatured chaperonin GroEL. *Biochem.* **34**: 13928-13933.
83. Pace, C.N., Laurents, D.V., Erickson, R.E. 1991. Urea denaturation of barnase: pH dependence and characterization of the unfolded state. *Biochem.* **31**:2728-2734.
84. From, N.B., Bowler, B.E. 1998. Urea denaturation of Staphylococcal Nuclease monitored by fourier transform infrared spectroscopy. *Biochem.* **37**:1623-163.
85. Lapanje, S. 1971. Denaturation of globular proteins by guanidine thiocyanate. I. Optical rotation in aqueous guanidine thiocyanate solutions. *BBA.* **243**:349-356.
86. Lapanje, S. 1971. Denaturation of globular proteins by guanidine thiocyanate. II. Intrinsic viscosities in 4.5 M guanidine thiocyanate solutions. *BBA.* **243**:357-365.

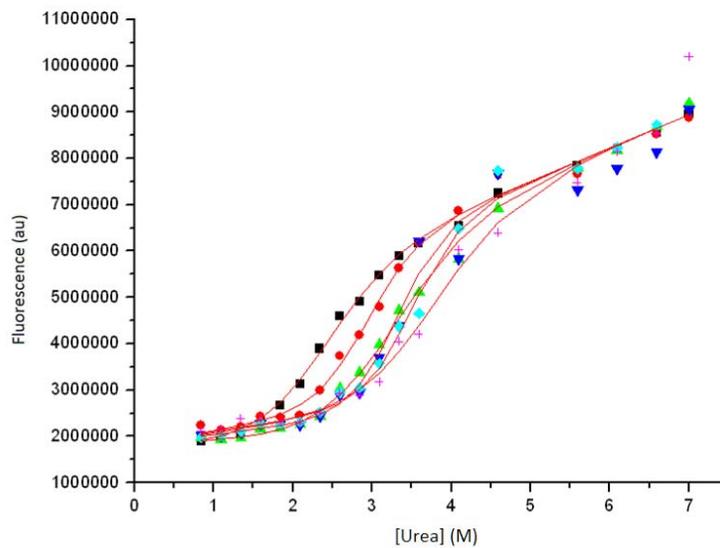
Appendix

Appendix 1: Global fitting of urea renaturation curves

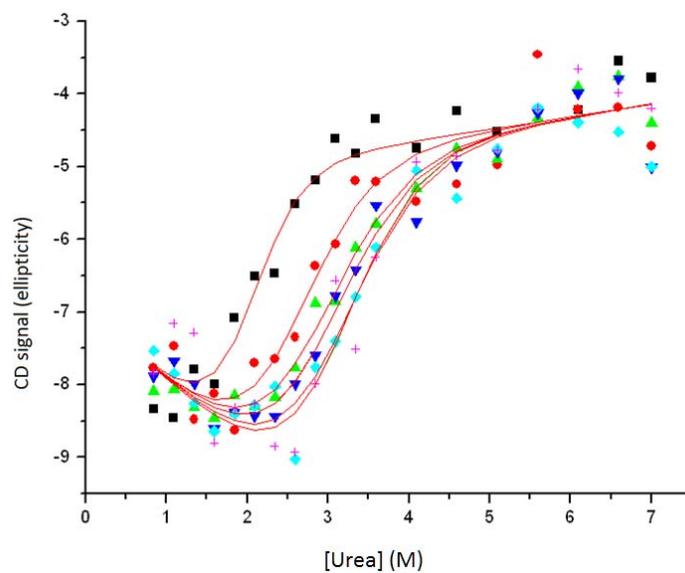
(A) Urea renaturation curves monitored by fluorescence at 313 nm. From left to right: 2, 9, 16, 25, 31, 50 days.



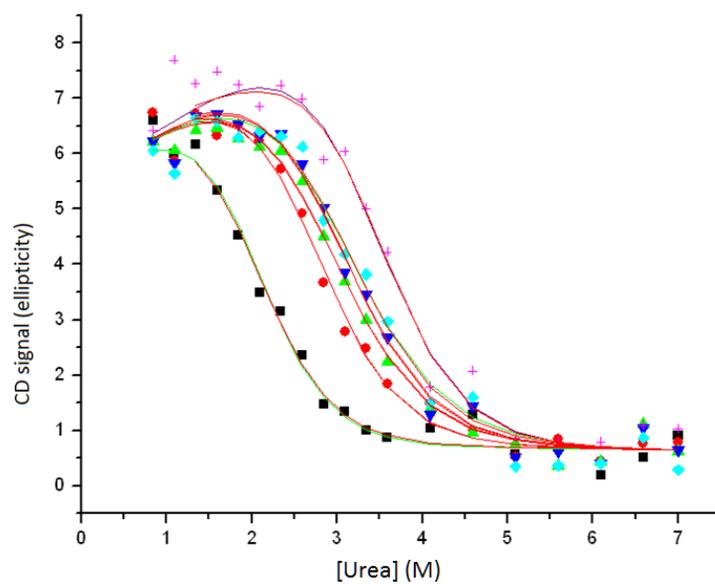
(B) Urea renaturation curves monitored by fluorescence at 370 nm. From left to right: 2, 9, 16, 25, 31, 50 days.



(C) Urea renaturation curves monitored by CD at 215 nm. From left to right: 2, 9, 16, 25, 31, 50 days.

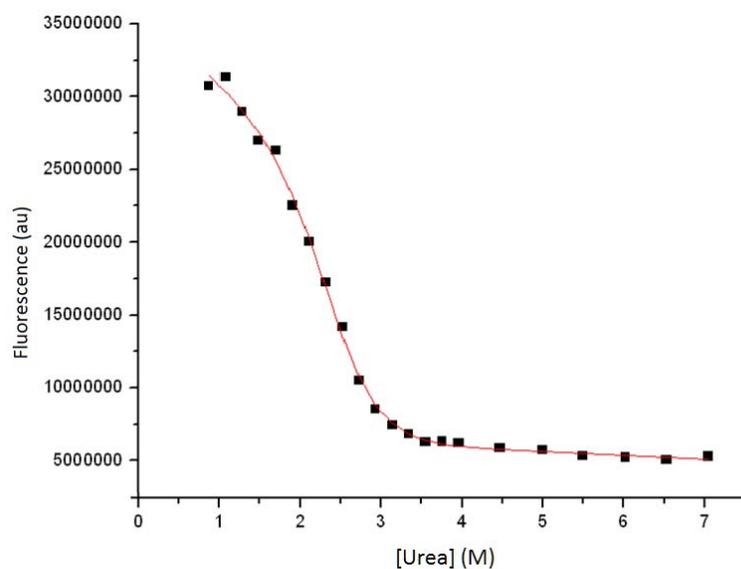


(D) Urea renaturation curves monitored by CD at 230 nm. From left to right: 2, 9, 16, 25, 31, 50 days.

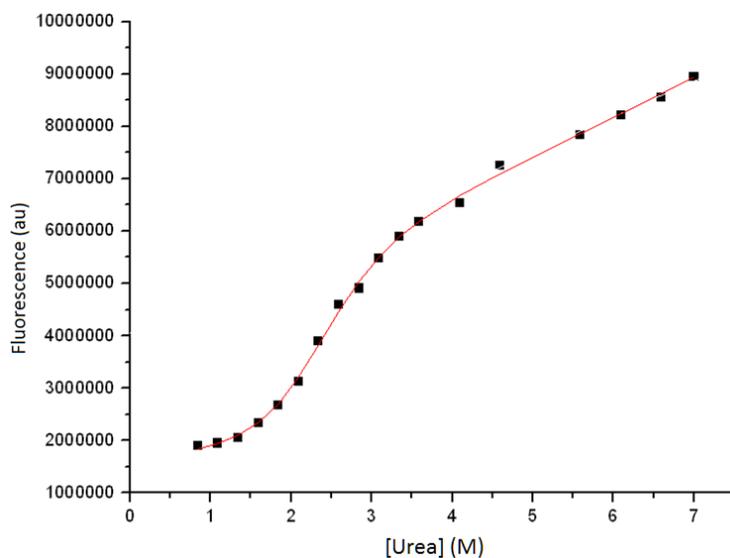


Appendix 2: Individual fitting of urea renaturation curves

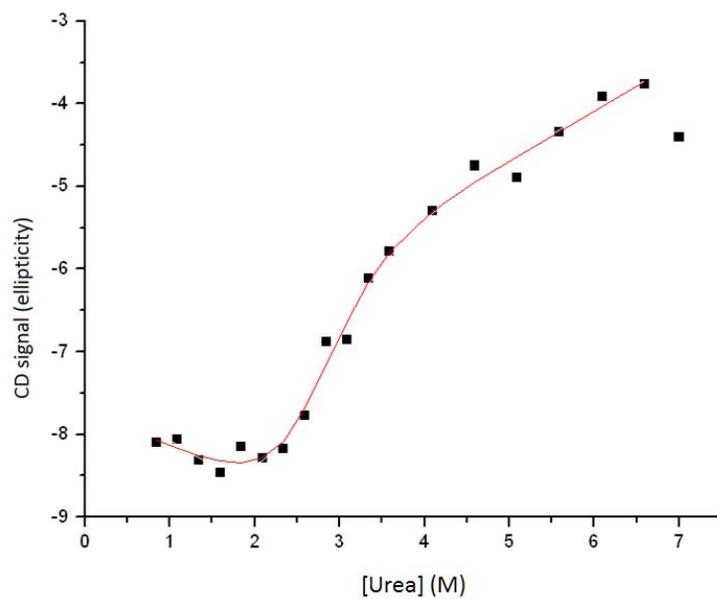
(A) Urea renaturation curve monitored by fluorescence at 313 nm at day 2.



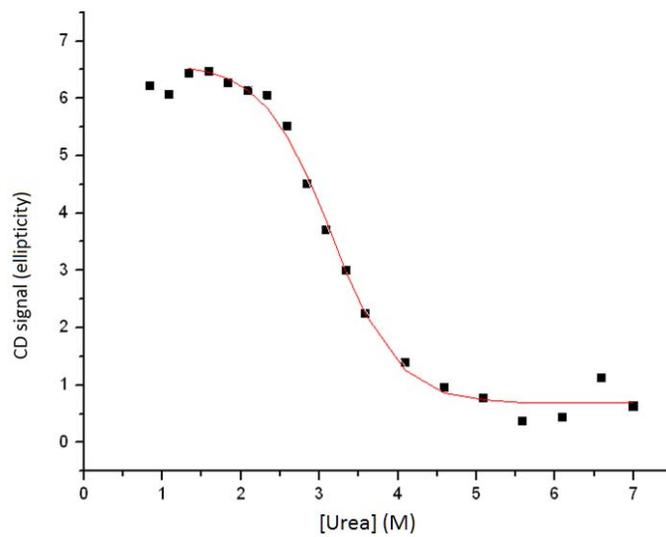
(B) Urea renaturation curve monitored by fluorescence at 370 nm at day 9.



(C) Urea renaturation curve monitored by CD at 215 nm at day 16.

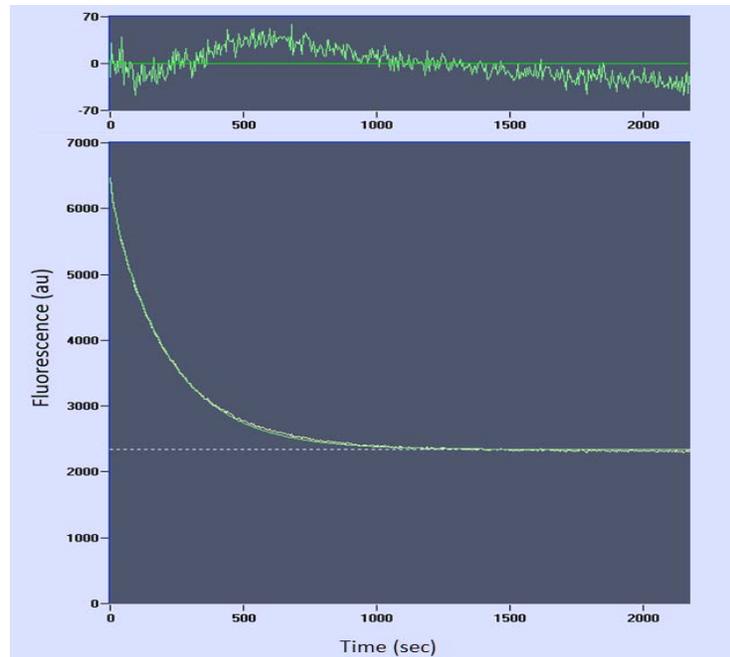


(D) Urea renaturation curve monitored by CD at 230 nm at day 16.

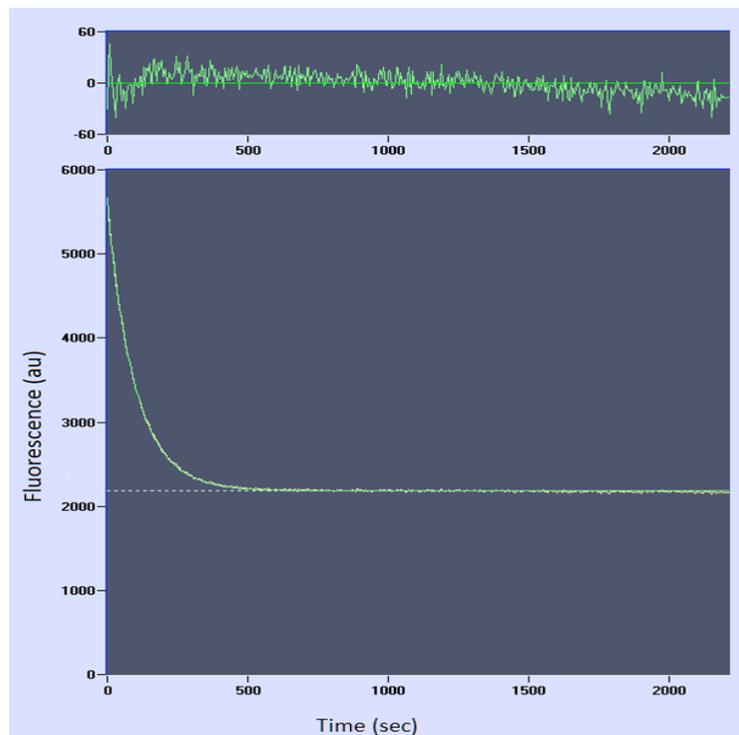


Appendix 3: Unfolding kinetic traces, measured by plate reader (SpectraMax M5 fluorescence plate reader, Molecular Devices LLC)

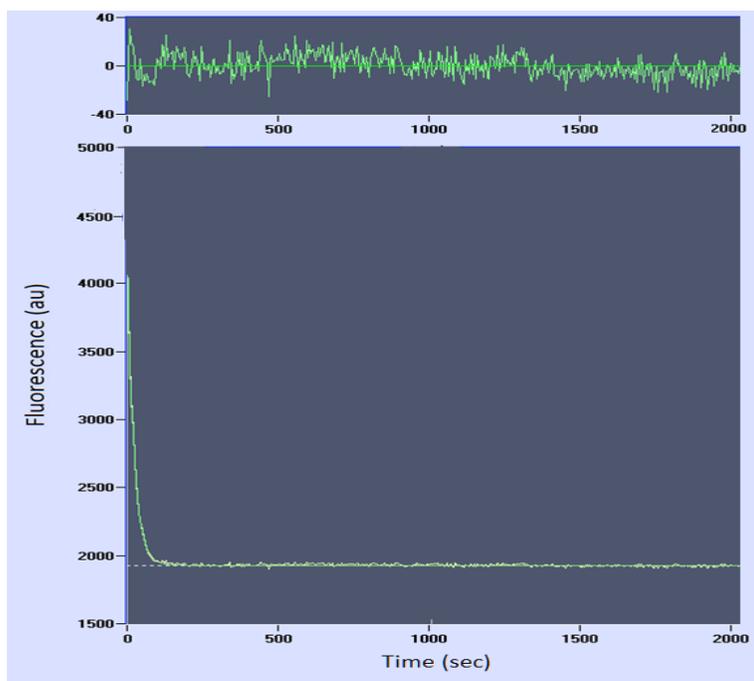
(A) Threefoil unfolding in 2.88 M GuSCN, fit to a double exponential with no drift.



(B) Threefoil unfolding in 3.21 M GuSCN, fit to a double exponential with no drift.

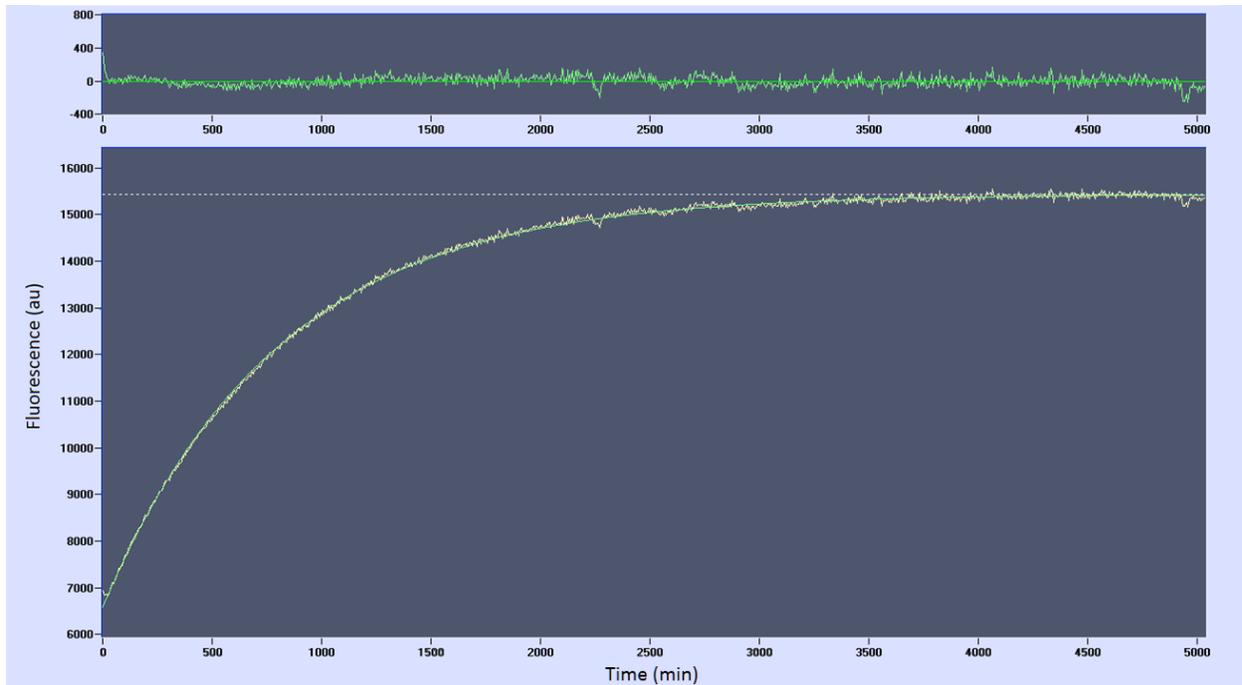


(C) Threefoil unfolding in 3.21 M GuSCN, fit to a double exponential with no drift.

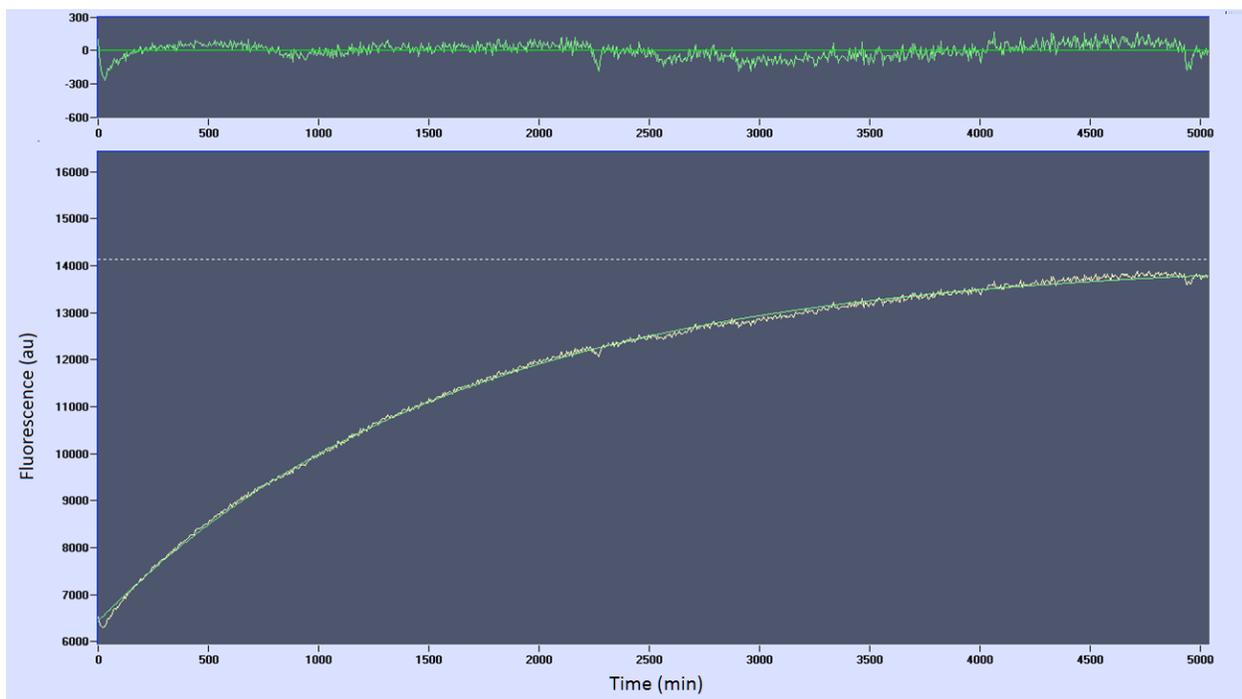


Appendix 4: Folding kinetic traces, measured by plate reader (SpectraMax M5 fluorescence plate reader, Molecular Devices LLC)

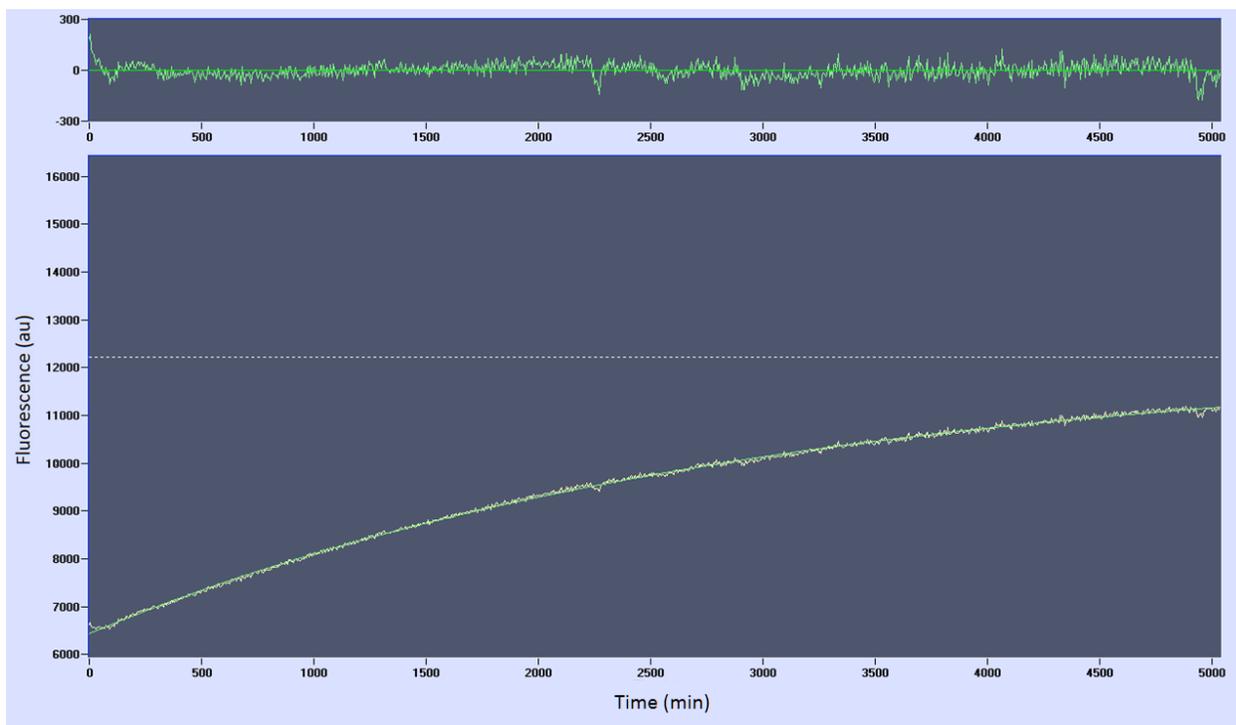
(A) Threefoil folding to 0.14 M GuSCN, fit to a single exponential with no drift.



(B) Threefoil folding to 0.2 M GuSCN, fit to a single exponential with no drift.

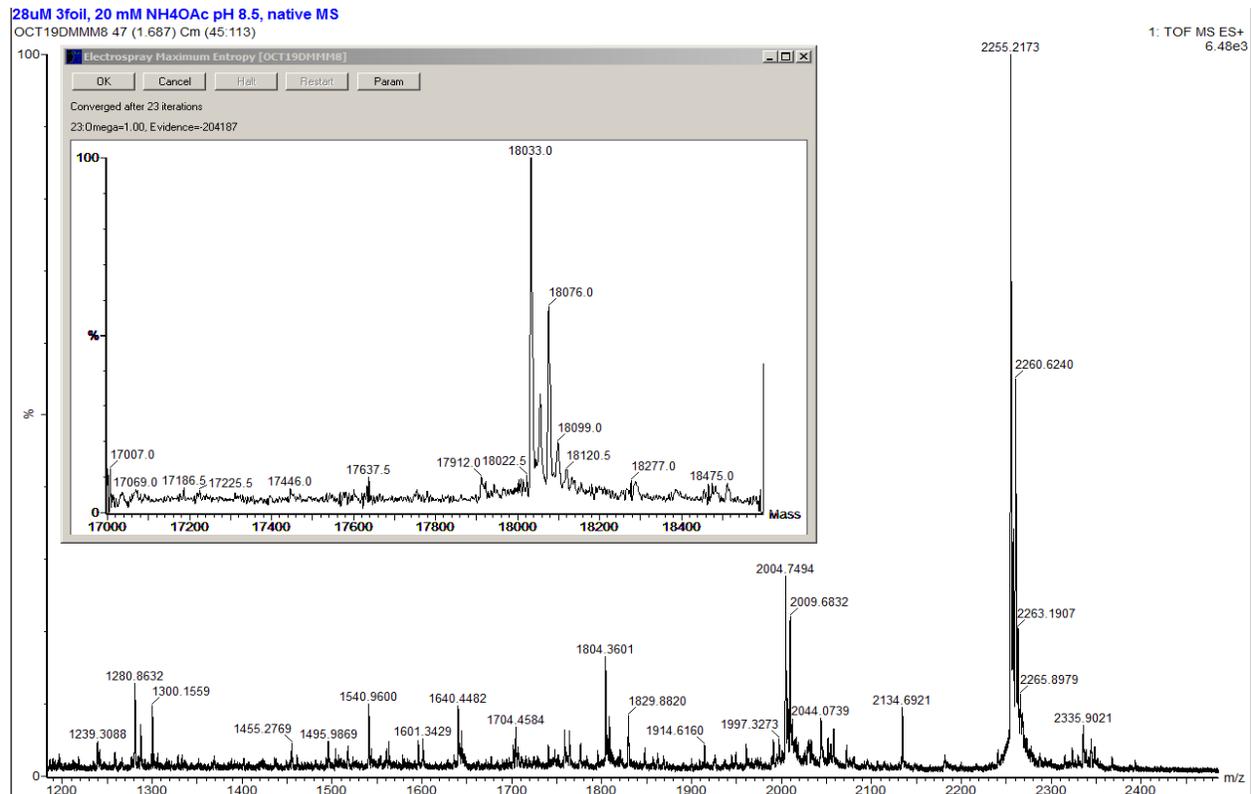


(C) Threefoil folding to 0.26 M GuSCN, fit to a single exponential with no drift.

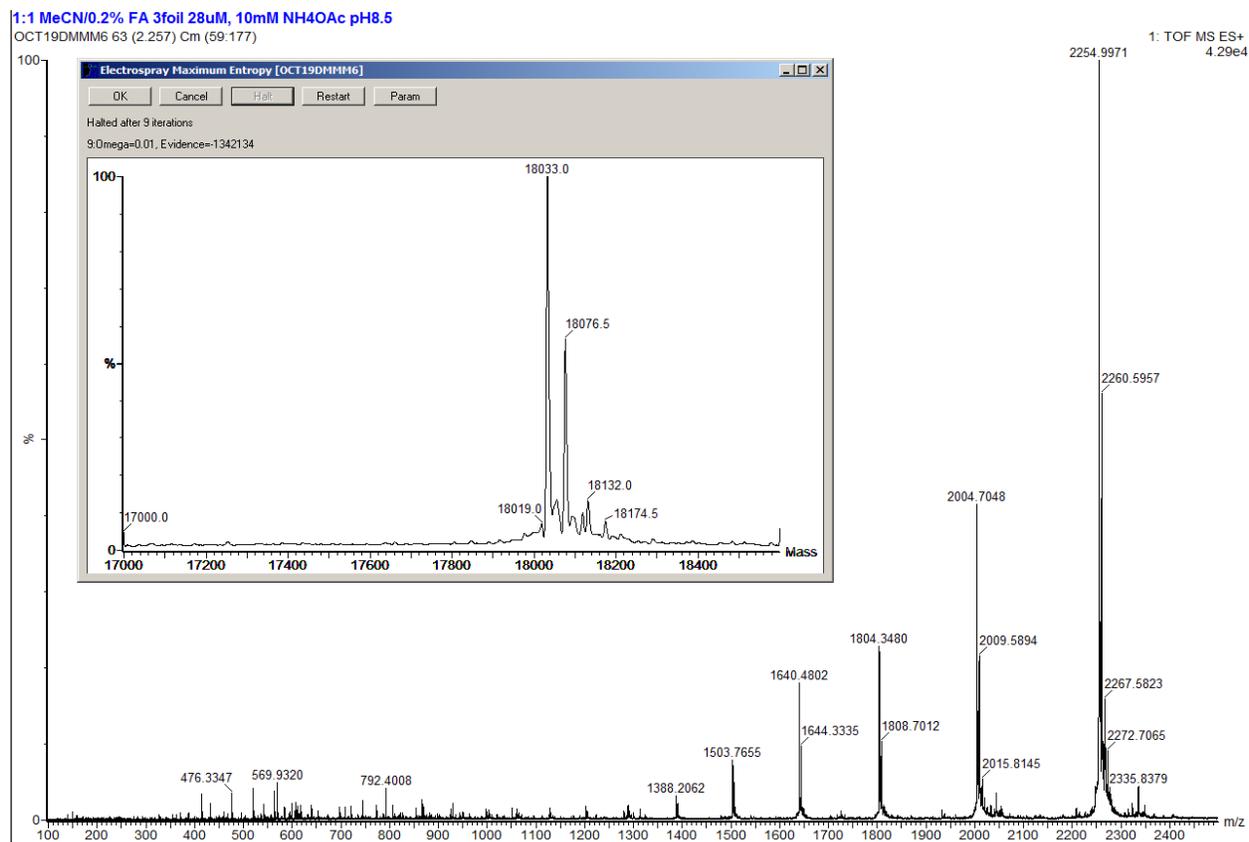


Appendix 5: Mass spectrometry

(A) Native MS spectrum of metal-free Threefoil. The calculated mass of Threefoil in the absence of metal (sodium) is 18034 a.m.u. versus the experimentally determined mass of 18033 a.m.u. (inset). (On-site mass spectrometry facility, dept. of Chemistry, The University of Waterloo)

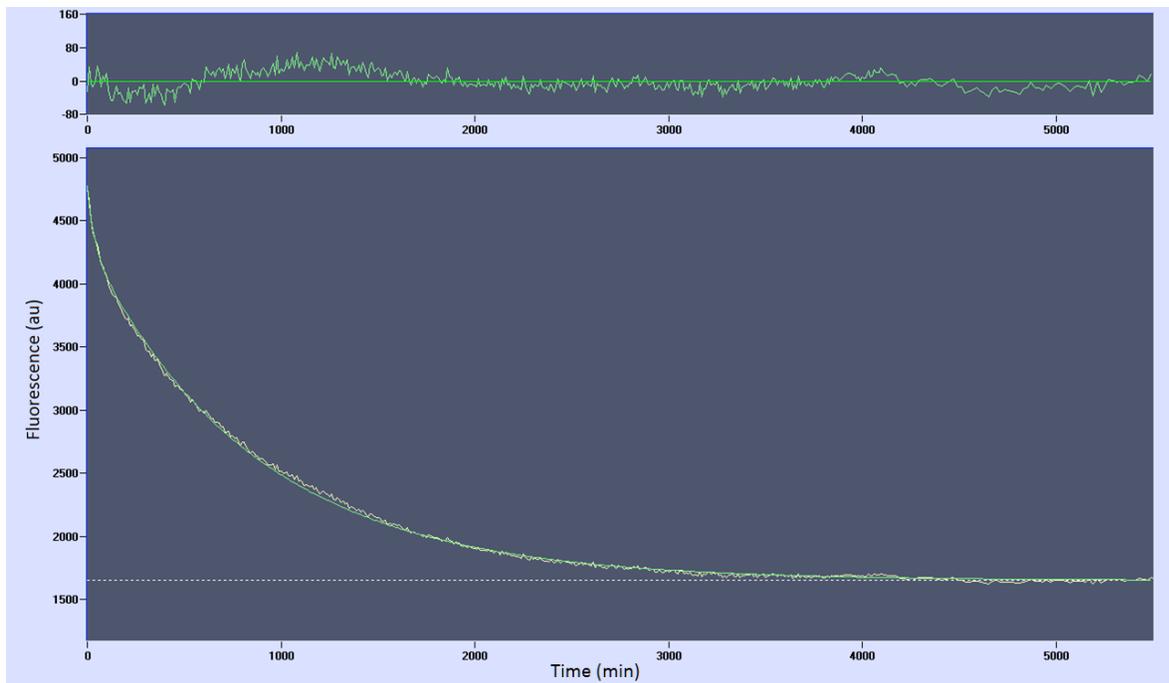


(B) Denatured MS spectrum of metal-free Threefoil in which native Threefoil signal is observed, with a mass of 18033 a.m.u. (inset). (On-site mass spectrometry facility, dept. of Chemistry, The University of Waterloo)

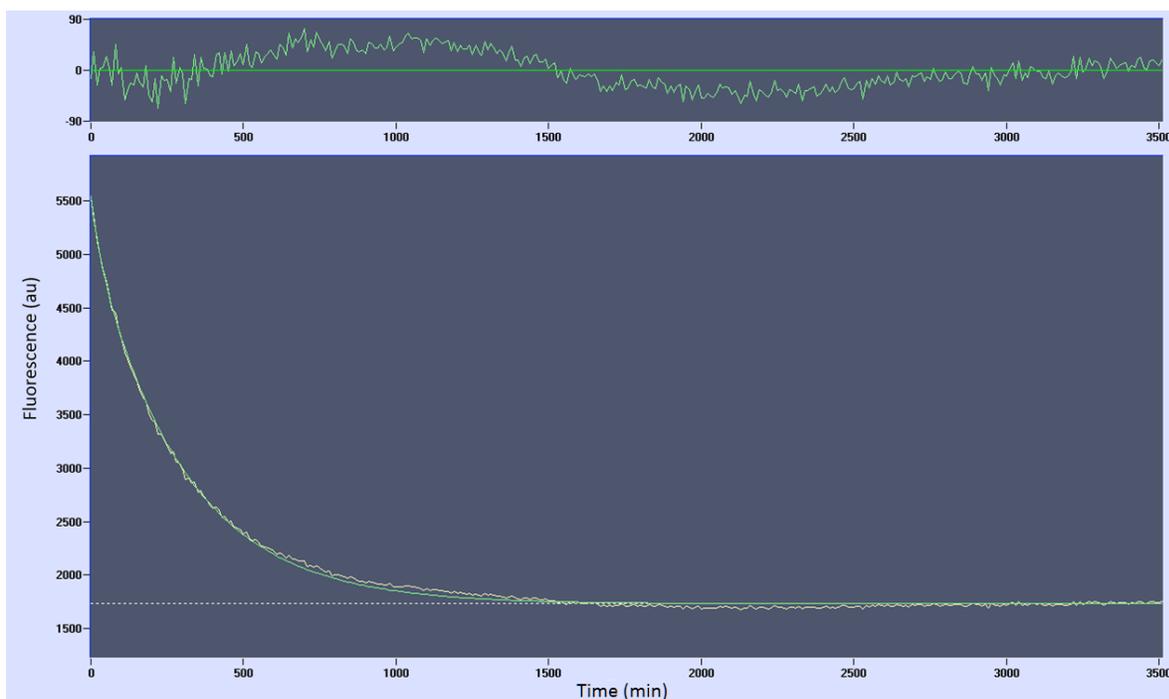


Appendix 6: Folding and unfolding kinetic traces of metal-free Threefoil, measured by plate reader (SpectraMax M5 fluorescence plate reader, Molecular Devices LLC)

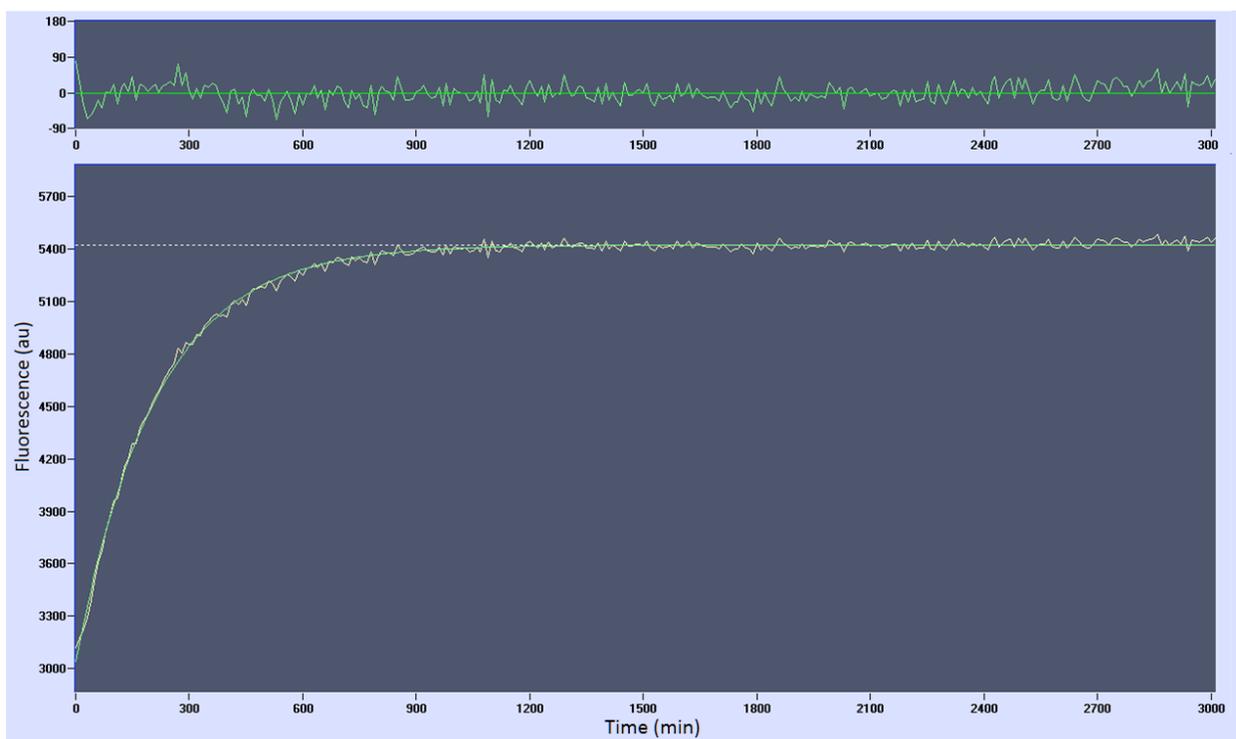
(A) Metal-free Threefoil unfolding in 1.1 M GuSCN, fit to a double exponential with no drift.



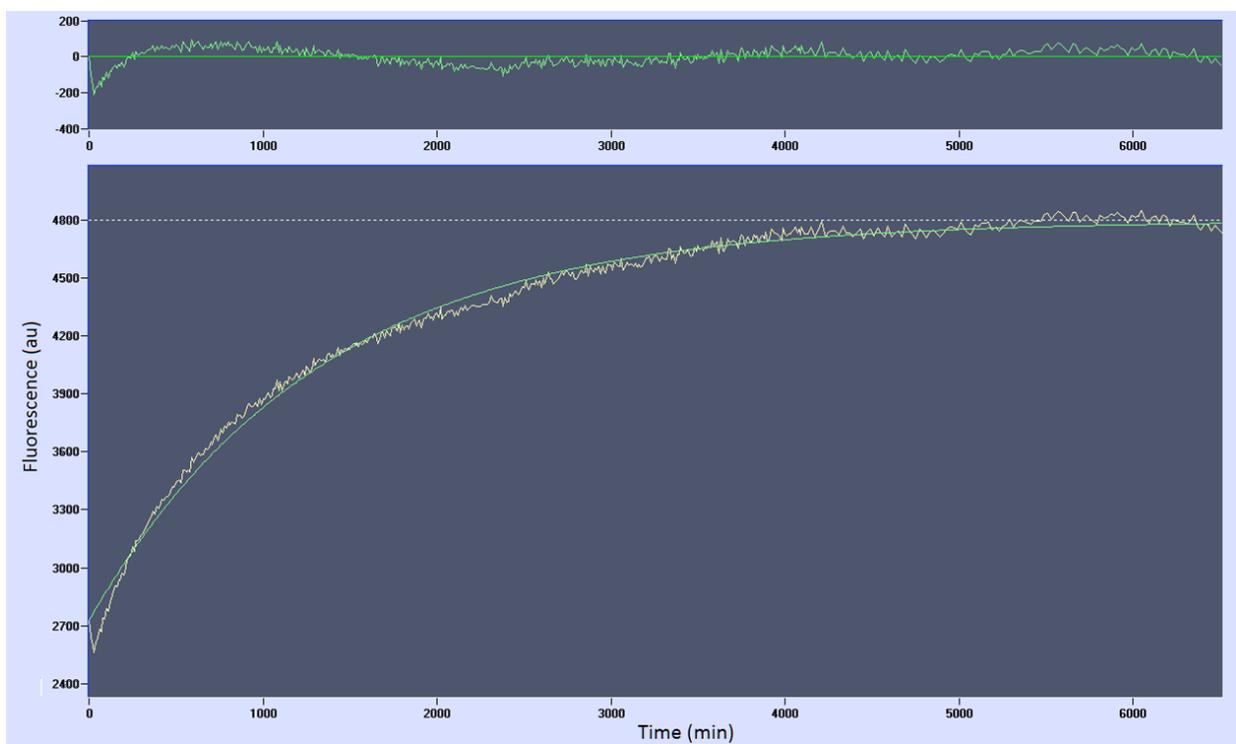
(B) Metal-free Threefoil unfolding in 1.4 M GuSCN, fit to a double exponential with no drift.



(C) Metal-free Threefoil folding to 0.04 M GuSCN, fit to a single exponential with no drift.

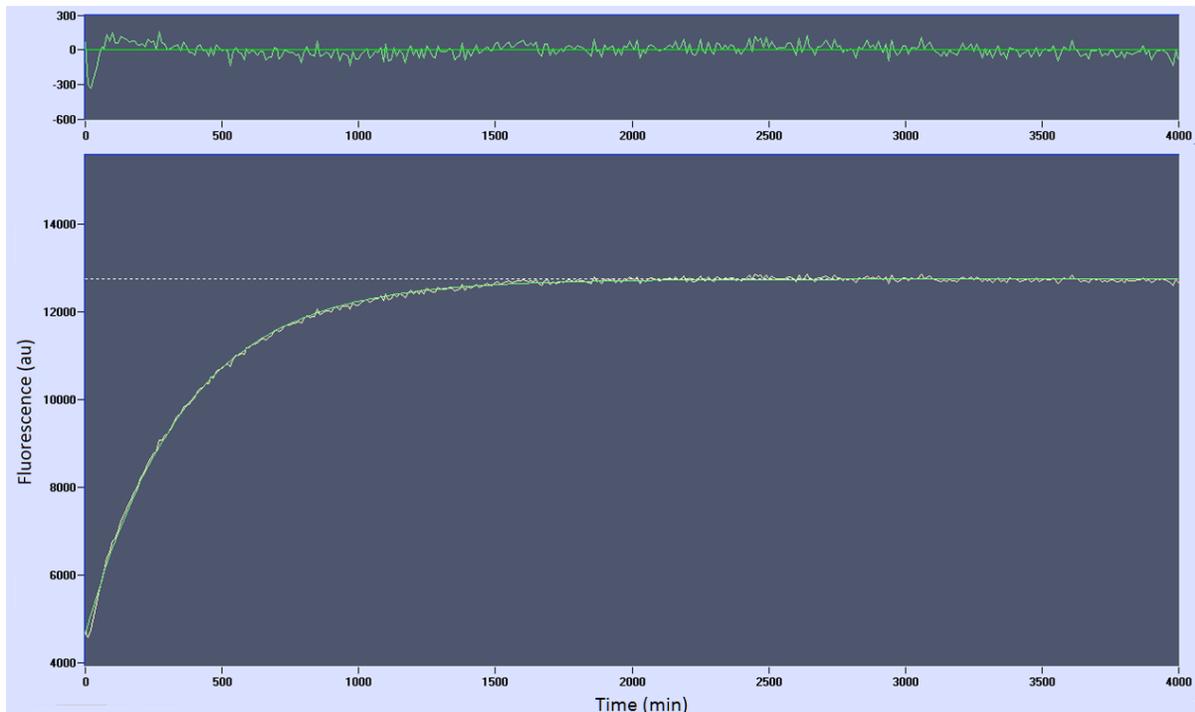


(D) Metal-free Threefoil folding to 0.16 M GuSCN, fit to a single exponential with no drift.

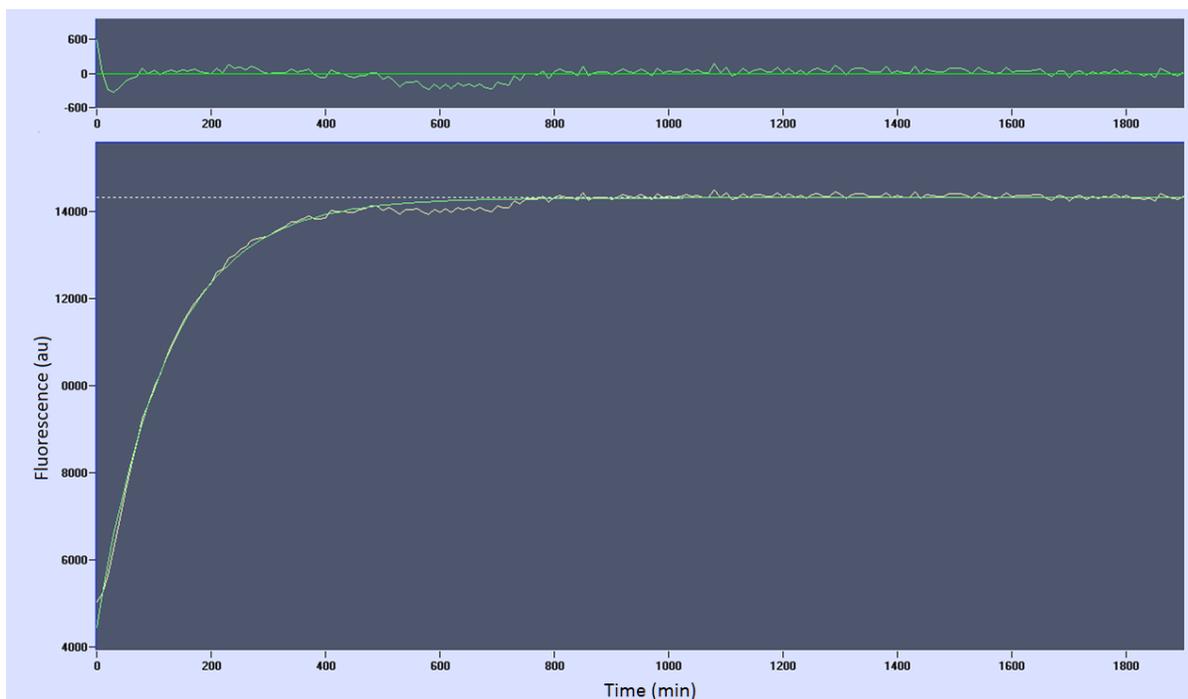


Appendix 7: Folding and unfolding kinetic traces in the presence and absence of carbohydrates, measured by plate reader (SpectraMax M5 fluorescence plate reader, Molecular Devices LLC)

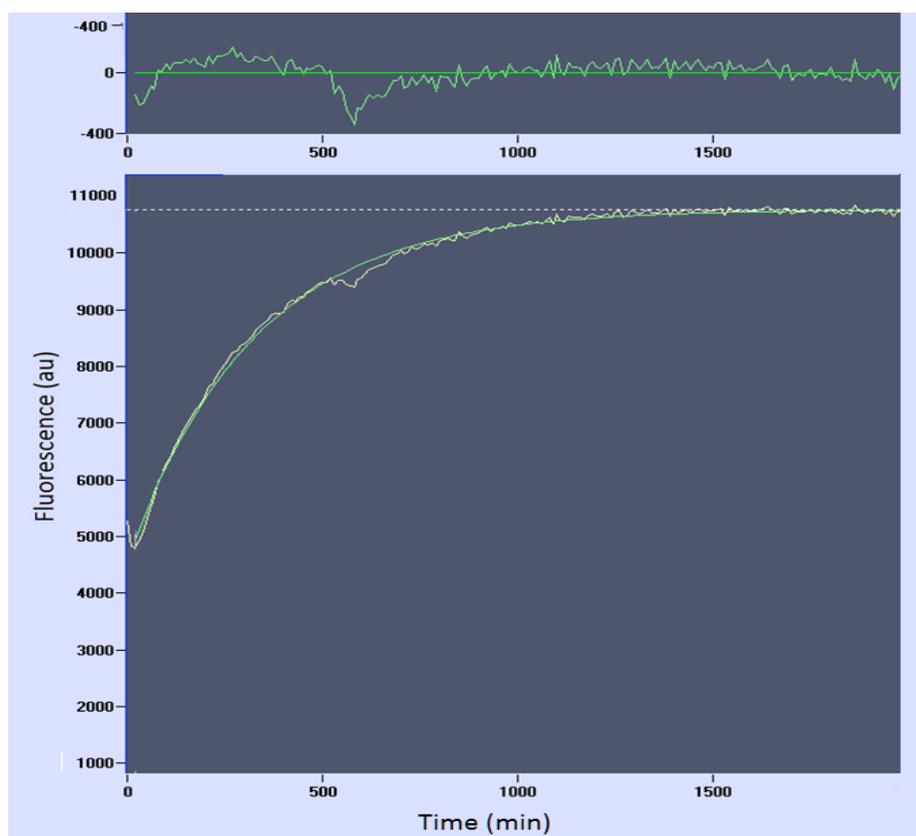
(A) Threefoil folding to 0.14 M GuSCN, 5 mM lactose, fit to a single exponential with no drift.



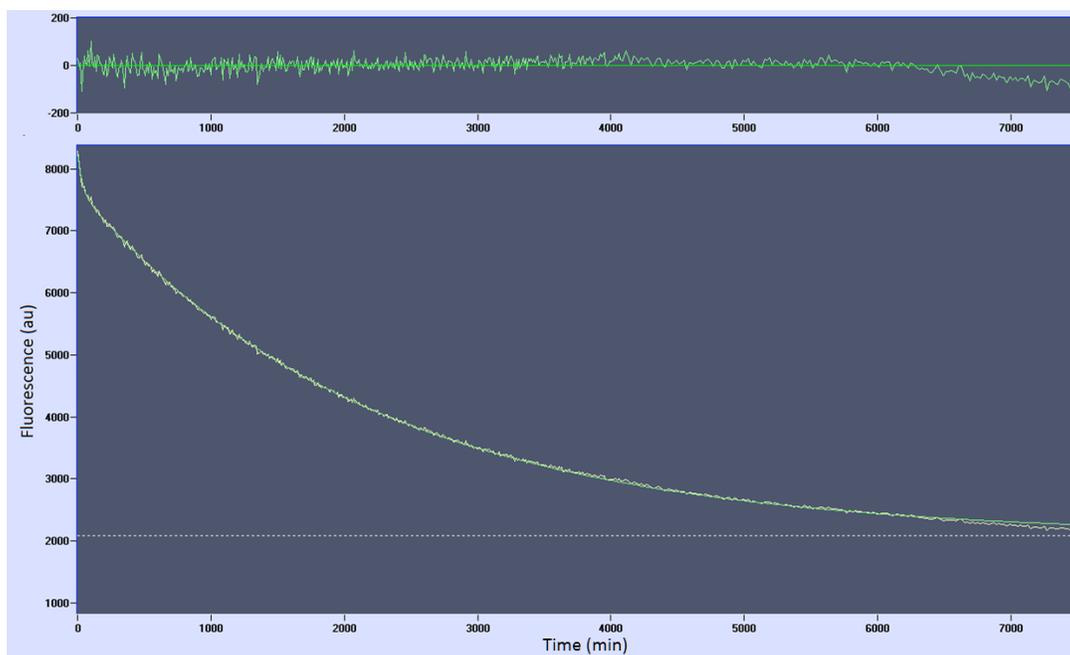
(B) Threefoil folding to 0.14 M GuSCN, 50 mM lactose, fit to a single exponential with no drift.



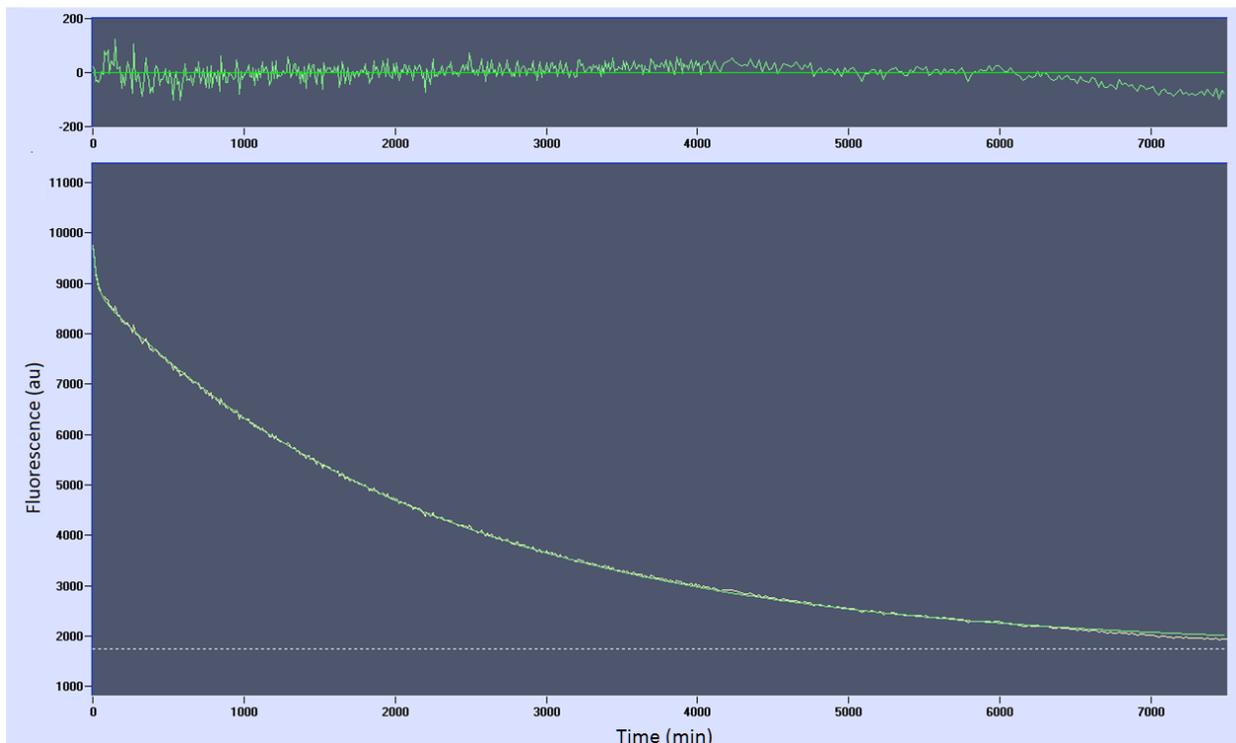
(C) Threefoil folding to 0.12 M GuSCN, 50 mM sucrose, fit to a single exponential with no drift.



(D) Threefoil unfolding in 2.27 M GuSCN, 5 mM lactose, fit to a double exponential with no drift.



(E) Threefoil unfolding in 3.13M GuSCN, 50 mM lactose, fit to a double exponential with no drift.



(F) Threefoil unfolding in 2.27 M GuSCN, 50 mM sucrose, fit to a double exponential with no drift.

