Equilibrium and kinetic folding studies of two YchN-like proteins: the Tm0979 dimer and the Mth1491 trimer

by

Céline Galvagnion

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, 2007
©Céline Galvagnion, 2007

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

Proper folding of a protein to its native state is critical for the protein to be fully functional under biological conditions. Understanding protein folding and protein folding evolution within the same structural family are key to understand which processes assist or hinder protein folding and how to prevent misfolding. Tm0979 from *Thermotoga maritima*, Mth1491 from *Methanobacterium thermoautotrophicum* and YchN from *Escherichia coli* belong to the homologous superfamily of YchN-like proteins (SCOP and CATH: 3.40.1260.10). The structures of these proteins have been solved as part of structural proteomics projects, which consist of solving protein structures on a genome wide scale. In solution, Tm0979 forms a homodimer whereas Mth1491 folds as a trimer and YchN is a homohexamer. The structures of the individual monomeric subunits of these three proteins have high structural similarity, despite very low sequence similarity. The biological roles of these proteins are not yet well defined, but seem to be involved in catalysis of sulphur redox reactions. This thesis focuses on characterisation of the Tm0979 homodimer and the Mth1491 homotrimer, as well as the determination of the folding mechanisms of these two proteins. The folding mechanisms of the proteins are compared to each other and to the mechanisms of other dimeric and trimeric proteins. The evolution and basis of oligomeric structure within the YchN family are analyzed. Mutations of Tm0979 and Mth1491 are designed as a basis for future work to investigate processes responsible for switches in oligomeric protein quaternary structure.

Acknowledgements

Dedication

I would like to dedicate my Masters thesis to my parents, Sylvie and Christian Galvagnion, as well as my sisters, Aurélie and Amandine, for their love and support.

Table of Contents

List of Tables

List of Figures

xv

List of Abbreviations

3D: Three Dimension

3Dee: Database of Protein Domain Definitions

A_{280}: Absorbance at 280nm

ANS: 1-Anilino-8-Naphtalene Sulfonate

ASA: Accessible Surface Area

CATH: Class, Architecture, Topology, Homologous superfamily

COG: Clusters of Orthologous Groups

ΔG_U: Free energy of unfolding in water

CD: Circular Dichroism

CH_3COONa: Sodium Acetate

DTT: Dithiothreitol

E. coli: *Escherichia Coli*

EDTA: Ethylene Diamine Tetra Acetic Acid

f_x: Fraction of the species x

GdmCl: Guanidine hydrochloride

G_x: Free energy of the specie x

HCl: Hydrochloride acid

HOMSTRAD: Homologous Structure Alignment Database

I, I_2, I_3: Monomeric, dimeric and trimeric intermediate, respectively

Indels: Insertion and deletions of residues

IPTG: Isopropyl-β-D-thiogalactopyranoside

ITC: Isothermal Calorimetry

kDa: Kilo Dalton

Kd: Dissociation constant

K_U: Equilibrium constant of unfolding

k_f: Folding rate constant

k_u: Unfolding rate constant

M. thermoautotrophicum: *Methanobacterium thermoautotrophicum*

Mth1491: Conserved hypothetical protein from *Methanobacterium thermoautotrophicum*

m_U: Dependence of ΔG_U on denaturant concentration

N, N_2, N_3: Native monomer, native dimer, native trimer, respectively

$Na_3C_6H_5O_7$: Sodium Citrate

NaCl: Sodium Chloride

Na_2HPO_4: Sodium Phosphate

Na_2SO_4: Sodium Sulfate

NCBI: National Center for Biotechnology Information

OD: Optical Density

P: Total protein concentration

SCOP: Structural Classification of Proteins

SDS: Sodium Dodecyl Sulfate

T. maritima: *Thermotoga maritima*

Tm0979: Conserved hypothetical protein from *Thermotoga maritima*

TS‡: Transition state

U: Unfolded protein

YchN: Conserved hypothetical protein from *Escherichia coli*

1 Introduction

1.1 Relation between protein structure, function and evolution

Proteins can be classified into different structural classes depending on the arrangement of their secondary structure elements according to structural databases and classification schemes such as SCOP (Murzin *et al.*,1995), CATH (Orengo *et al.*,1997), DALI (Holm *et al.*,1996), 3Dee (Russell *et al.*, 1992) and HOMSTRAD (Mizuguchi *et al.*,1998). A large proportion of genes, up to 90% in eukaryotes, encodes for oligomeric proteins. The characteristics of families and superfamilies will be discussed in section 1.1.1, and then protein classification will be described by taking CATH classification as a reference in section 1.1.2. Finally, evolution of structure related proteins will be discussed in section 1.1.3.

1.1.1 Structure levels of proteins

The primary structure of a protein consists of its linear amino acids sequence. Each amino acid has a particular propensity to form secondary structure elements, such as α-helices, β-strands, irregular loops and turns. Upon folding, secondary structure elements are formed and there are clearly restraints on the ways in which secondary structures can be packed together to achieve optimal packing of hydrophobic residues in the core of the protein, which is central to the formation of its tertiary structure. The protein can then oligomerise, which represents its quaternary structure.

1.1.2 Protein structure classification

For most of the known classification methods, at the lowest level in a structural classification, proteins are grouped if they belong to the same class, i.e., if they have similar secondary structure compositions and packing (CATH: (Orengo *et al.*,1997), SCOP: (Murzin *et al.*,1995). There are three major classes: mainly α, mainly β, and α-β. Approximately 25% of proteins are mainly α, 25% are mainly β and almost half are α-β proteins (Martin *et*

al.,1998). Each class is divided in different categories where the proteins are classified based on the architecture of their structure, i.e. the relative orientations of the secondary structures in three-dimension (3D) and the order in which they are connected. Finally, those classes are divided into subcategories called families based on the hypothetical or known function of the proteins and will be described in more detail in the following sections.

1.1.3 Families and superfamilies

Most families include members with different quaternary structure (Orengo *et al.*,2005). However the molecular basis and evolution of quaternary structure are poorly understood. During the course of evolution, proteins derived from a common ancestral protein can change their sequences and diverge by mutations or substitutions of the residues and also by insertions and deletions of residues (indels), giving rise to families of homologous proteins (Orengo *et al.*,2005). Many protein family resources present a hierarchical classification whereby very close relatives, for example with high sequence similarity (e.g. >40% sequence identity), are grouped together into families. These close relatives frequently share common functional properties. More remote homologues that have lower sequence similarity (<30%) are grouped together into broader evolutionary families or superfamilies, as is the case for Tm0979, Mth1491 and YchN. It is difficult to recognize very divergent relatives by comparing their sequences alone, and as for Tm0979, Mth1491 and YchN, the remote homologues could only be detected by comparing their structures.

1.2   Description of protein-protein interactions

Most proteins fold as oligomers (79% in *Escherichia coli* (*E. coli*)) (Goodsell *et al.*,2000). Oligomers can be split into two categories: homomers are formed by association of *n* identical monomers; heteromers are formed by association of different monomers. Principles that govern association are complex. Gaining an understanding of these principles may help us to understand why proteins misfold and how proteins' oligomeric states evolve within the same structure family. Based on the characteristics of oligomer interfaces, hydrophobic interaction seem to stabilize the interface whereas ionic interactions and hydrogen-bonding seems to govern the selectivity of the interface (Chothia *et al.*,1975). Different amino acids in the interface may therefore make differing contributions to binding and amino acids not located in the interfaces can also affect binding. Protein-protein interfaces can also be characterized by their geometric characteristics (interface size, shape, atomic packing, planarity and complementarity) and will be reviewed in the following section (Ponstingl *et al.*,2005).

1.2.1   Nature of oligomer interfaces

The structural characteristics of protein-protein interfaces for homomer and heteromers have been analyzed in detail by a number of groups (Jones *et al.*,1995; Jones *et al.*,1996; Goodsell *et al.*,2000; Nooren *et al.*,2003; Ponstingl *et al.*,2005). Various trends have been observed and are outlined below. Terms for describing interfaces and illustrative examples of different types of interfaces are summarized in Table 1.1.

Table 1.1: Different types of protein-protein interactions.

Types of protein-protein interactions	Characteristics of the interface:	Protein example:
Symmetry of the interface: Isologous	Same surface on both monomer	Arc repressor, Tm0979
Heterologous	Different surface used for each monomer	Mth1491
Necessity of the oligomeric state: Obligate (protomers often expressed together)	Monomers are not found as stable on their own in vivo, they need to form the oligomer to be stable.	Arc repressor, human cathepsin D, cro repressor
Non-obligate (perform regulatory role)	Monomers can exist independently	Sperm lysin, RhoA-RhoGAP, bovine G protein
Life time of the complex: Transient	Weak transient: dynamic oligomeric equilibrium in solution, interactions are continuously made and broken	Tm0979
	Strong transient: require molecular trigger to shift the oligomeric equilibrium	Human papillomavirus E2, heterotrimeric G protein
Permanent	Very stable complex which only exists in its oligomeric form	Mth1491

1.2.1.1 Amino acid populations in interfaces

Solvation requirements lead to polar and charged functional groups being enriched on the exterior of the folded polypeptide whereas aliphatic groups tend to be hidden in the core of the fold away from contact with water molecules (Ponstingl *et al.,*2005). This hydrophobic effect is held as principal driving force for protein folding (Kauzmann,1959), (Dill,1990) as well as for protein-protein association (Chothia *et al.*,1975; Argos,1988; Janin *et al.*,1990; Young *et al.*,1994; Tsai *et al.*,1997). Interfaces of multimer tend to have a higher fraction of their area covered by carbon atoms (implying hydrophobic groups) than the rest of the surface (Ponstingl *et al.*,2005). Consistent with this, large, predominantly hydrophobic residues, like aromatic residues and the aliphatic residues as well as methionine and cysteines, have a tendency to occur in the interfaces rather than being in contact with solvent (Ponstingl *et al.*,2005). In contrast, polar, charged and smaller hydrophobic residues generally prefer to contact the solvent. Arginine is an exception, where a slight preference for inter-subunit contacts can be detected in spite of its positive charge (as seen for Mth1491 interface, section

1.4.1.2) (Ponstingl *et al.*,2005). Moreover, partly accessible residues tend to be more hydrophilic than the totally buried ones. Polar or charged residues also occur in interfaces, however, where they can make hydrogen bonds. Hydrogen bonds with their predominantly polar character and energy of 3-7 kcal.mol^{-1} are relatively short in range. Therefore, they are often held to confer specificity on protein-protein interactions (Ponstingl *et al.*,2005).

1.2.1.2 Size and shape

As a rule of thumb, a fraction of approximately 18% of the surface area of a subunit is involved in inter-subunit contacts. The fraction increases, however, with the multiplicity *n* of the subunit in the homomer (Ponstingl *et al.*,2005), thus it tends to be smaller for dimers and greater for trimers, tetramers and hexamers.

In homo-dimers with their two-fold symmetry, the part of the subunit contact is mapped onto itself by a 180° rotation (as seen for Tm0979 dimer, section 1.4.1.2) (Ponstingl *et al.*,2005). Such interfaces are referred to as *isologous* interfaces (Monod *et al.*,1965) (Table 1.1). Molecules with circular symmetry, like all trimers (as seen for Mth1491, section 1.4.1.2) (C_3) as well as some tetramers (C_4) and hexamers (C_6), usually have one dominant interface, which is made up of two distinct areas on the subunit surface (Ponstingl *et al.*,2005). Such interfaces are referred to as *heterologous* interfaces (Monod *et al.*,1965) (Table 1.1). The two surface areas cover approximately the same accessible surface area (ASA) of the subunit since they participate in one and the same contact (Ponstingl *et al.*,2005).

1.2.1.3 Planarity of the interfaces

The interfaces of protein-protein complexes tend to be flat (Argos,1988; Jones *et al.*,1996). There is a strong dependency of the planarity on the size of the interface (Ponstingl *et al.*,2005). This dependency can be explained by the fact that atoms in small interfaces are restrained in their scatter by belonging to the same or neighboring residues whereas in large interfaces, this restriction is relaxed, which enables formation of irregular shapes. Moreover,

the globularity of the subunit fold might imply that a very extensive interface area can only be achieved by increasing the 'ruggedness'. If one could approximate the two subunits forming an interface by two ellipsoids, for example, there would be a certain maximum interface area possible without distorting the ellipsoid shape. The only way to increase the contact area would be to introduce protrusions compensated by the shape of the respective partner. Apart from the contact, the original shape could be still kept (Ponstingl *et al.*,2005). Large interfaces tend thus to be flatter.

1.2.2 Comparison between permanent and transient complexes

Permanent complexes are very stable complexes that can only be detected in their oligomeric form whereas transient complexes are in equilibrium between their monomeric and oligomeric forms (Table 1.1). Obligate complexes tend to be characterised by unstable monomers and therefore only exist in their oligomeric form. In contrast, for non-obligate complexes, the monomeric protein is stable enough to survive on its own. In this section, the characteristics of the interface of permanent and obligate complexes will be compared with those of transient and non obligate complexes.

Permanent complexes are found to have protein-protein interfaces that are more packed but less planar and with less intersubunit hydrogen bonds than nonobligatory complexes (Jones *et al.*,1996). On the other hand, transient complexes contain more hydrophilic residues in their interface than permanent complexes (Jones *et al.*,1996). Moreover, strong transient dimers are characterized by larger, less planar and sometimes more hydrophobic interfaces, whereas weak homodimers tend to have smaller contact areas between protomers and interface are more planar and polar on average (Nooren *et al.*,2003).

Comparisons with trends observed for oligomers suggest that Tm0979 oligomeric characteristics are consistent with it forming a weak transient complex; experimentally

Tm0979 monomers have significant stability *in vitro* (B Cheyne, G. Meglei, K.A. Vassall, P.B. Stathopulos, E.M. Meiering, unpublished results). In constrast, Mth1491 seems more likely to be a permanent complex because the trimer interface is larger, more hydrophobic and more specific than the Tm0979 dimer interface. Characteristics of these two interfaces will be discussed in more detail in section 1.4.1.2.

1.3 Protein folding

In order to be fully active, proteins need to fold properly to their 3D native conformations. The native state is believed to correspond to the lowest Gibbs free energy state of the protein. Based on Anfinsen's experiments, the native state and the pathway for the protein to reach this state are encoded in the amino-acid sequence (Anfinsen,1973). This hypothesis is supported by the fact that there is an astronomical number of possible conformations that proteins can adopt so it would take too long for the unfolded protein to randomly search for its native structure (Levinthal,1968). Therefore, three types of mechanisms, have been proposed for protein folding. 1) The framework model consists of the formation of secondary structural elements from the unfolded protein independently of tertiary structure (Ptitsyn,1973; Kim *et al.*,1990). Then the formation of the native conformation occurs by diffusion of the secondary structure elements until they collide and coalesce to give the tertiary structure (Bashford *et al.*,1988; Karplus *et al.*,1994). 2) The nucleation model postulates that some neighboring residues in the primary sequence would form native secondary structure that would act as a nucleus. Native structure would then propagate in a stepwise manner and the tertiary structure would form as a necessary consequence of the secondary structure (Wetlaufer,1973; Wetlaufer,1990). 3) The hydrophobic-collapse model suggests that the protein collapses rapidly around its hydrophobic side chains and then rearranges from the restricted conformational space occupied by the collapses intermediate (Kuwajima,1989; Ptitsyn,1995). Both framework

model and hydrophobic-collapse involve the formation of intermediates upon folding whereas the nucleation model involves simultaneous formation of secondary and tertiary structure. The role of intermediates in folding is complex and not fully understood; this will be discussed further in the following sections (1.4.1, 1.4.2 and 1.4.3).

1.3.1 Monomeric protein folding

Early studies on protein folding were focussed on small monomeric proteins which do not contain prolines or form disulfide bonds in order to analyze simple mechanisms that govern protein folding (Jackson,1998). Many small proteins fold with a 2-state mechanism involving the formation of native monomer (N) from unfolded monomer (U), U $\leftrightarrow$ N. The folding of larger monomeric proteins, however, tends to involve the formation of intermediates (I), U $\leftrightarrow$ I $\leftrightarrow$ N. For monomeric proteins, intermediates tend to act as kinetic traps and slow the folding process. However, intermediates may also play a role in helping proteins fold, as in the case of larger oligomeric proteins which will be discussed next.

1.3.2 Dimeric protein folding

Oligomeric proteins represent approximately 15% of currently known protein structures. This percentage is biased, however, by the fact that monomeric proteins are easier to crystallize than oligomeric ones. Actually, a survey of *E. coli* proteins in SWISSPROT showed that most of the proteins (79%) folds as oligomers and 38.2% fold as a dimer (Goodsell *et al.*,2000). Moreover, from all oligomeric structures determined to date, more than half are dimers. Dimer is therefore the most common oligomeric state and dimeric protein folding has recently been investigated by studying, both, small dimeric proteins (Wendt *et al.*,1995), (Bowie *et al.*,1989; Milla *et al.*,1994; Zitzewitz *et al.*,1995; Jana *et al.*,1997; Rosengarth *et al.*,1999; Satumba *et al.*,2002; Jia *et al.*,2005; Maity *et al.*,2005) followed by studies on very large proteins such as chaperones (Doyle *et al.*,2000). Dimeric

proteins are found to fold through various mechanisms that may or may not involve the formation of intermediates. The different mechanisms and the role of these intermediates will be discussed in the following sections.

1.3.2.1 Dimer folding via 2 state mechanism

As observed for monomeric proteins, most small dimeric proteins fold via a 2-state mechanism, which consists of the formation of the native dimer directly from the unfolded monomers (Zitzewitz *et al.*,1995), (Topping *et al.*,2004), (Gloss *et al.*,2002; Placek *et al.*,2005), (Kim *et al.*,2000; Kim *et al.*,2001). The 2-state mechanism involves the formation of native dimer (N_2) from unfolded monomer (U): $2U \leftrightarrow N_2$. Two-state folding suggests that the burial of hydrophobic residues and the cooperative interactions within the dimer upon oligomerisation play an important role in the stability of small dimeric proteins and are strong enough to drive the folding of these proteins. Some DNA-binding proteins seem to fold through a 2-state mechanism as well but in this particular case, folding may be coupled to the physiological role of the protein, *in vivo*, which involves an equilibrium between the oligomeric state and the DNA bound state (Jana *et al.*,1997), (Bowie *et al.*,1989; Milla *et al.*,1994). This equilibrium suggests that protein concentration in the cell may be regulated by this coupling insofar as the protein binds to the DNA when the DNA-binding process is required at low protein concentration; otherwise, it is stored as a dimer.

1.3.2.2 Dimer folding via multiple states

In contrast, most of the large oligomeric proteins fold through a multiple states mechanism, involving the formation of monomeric (I) and/or dimeric intermediate(s) (I_2): $2U \leftrightarrow 2I \leftrightarrow N_2$ or $2U \leftrightarrow I_2 \leftrightarrow N_2$. Intermediates can play complex roles such as helping proteins to fold or facilitating aggregation. Some proteins related to misfolding diseases involving prion, amyloid formation and protein aggregation are proposed to fold via the formation of intermediate(s) which favour protein misfolding (Galani *et al.*,2002; Zhu *et al.*,2003),

9

(Svensson *et al.*,2006). However, other large proteins need the formation of these intermediates to fold properly. For example, the very large chaperone SecA is proposed to fold through the formation of different intermediates which help this large protein fold into the correct native structure and avoid the non-productive pathway that leads to aggregation (Doyle *et al.*,2000). Analogous mechanism is proposed for the large enzyme beta Galactosidase whose intermediates are supposed to be the substrates for chaperones increasing the efficiency of cellular protein folding (Nichtl *et al.*,1998).

To sum up, the role of folding intermediates is not yet well defined and requires further study. Intermediates can be monomeric or oligomeric, on or off pathway. The structure of these intermediates may play a role in the propensity for aggregate formation or proper folding. Characterization of oligomer interface formation may therefore help us to understand how a protein folds and if the formation of an intermediate is required or not.

1.3.3 Trimeric protein folding

Trimeric proteins represent only a small proportion of oligomeric proteins, 21% in *E. coli* (Goodsell *et al.*,2000). To date, very few studies have been conducted to investigate trimer folding. As for monomer and dimer, some of the first folding studies were performed on small proteins.

1.3.3.1 *Trimer folding mechanisms*

One of the earliest trimer folding studies was to investigate the mechanism of a coiled-coiled trimer (Marti *et al.*,2004). The designed three stranded coiled-coil, Lpp-56 (trimeric coiled-coil) (Bjelic *et al.*,2006), unfolds with a simple two state mechanism involving unfolded monomers (U) and native trimer (N_3). Few trimeric proteins have been reported to fold via the formation of intermediates. γ-carbonic anhydrase (Simler *et al.*,2004) is one of the rare examples of proteins that fold through the formation of monomer intermediates: $3U \leftrightarrow 3I$

↔ N$_3$. Monomeric, dimeric and trimeric (I$_3$) kinetic intermediates were observed upon Bactoriophage T4 fibritin (Guthe *et al.*,2004) folding and illustrate that folding of trimeric protein can be very complex, $3U \leftrightarrow 3I \leftrightarrow N_2 + I \leftrightarrow N_3$.

1.4 YchN-like proteins

Tm0979 from *Thermotoga maritima* (*T. maritima*), Mth1491 from *Methanobacterium thermoautotrophicum* (*M. thermoautotrophicum*) and YchN from *E. coli* belong to the homologous superfamily of YchN-like proteins (CATH: 3.40.1260.10, SCOP : superfamily dsrEFH-like). Tm0979 belongs to the dsrH family of conserved hypothetical proteins found in bacteria and archaea (Pfam 04077, (Bateman *et al.*,2004); COG 2168 (Tatusov *et al.*,2001)). The dsr locus encodes various proteins involved in sulphur metabolism (Pott *et al.*,1998). It has been shown that mutations in the dsrH gene in the phototrophic bacterium *Chromatium vinosum* completely abolish the ability of cells to oxidize intracellular sulphur; however, the molecular function of the protein is not known (Pott *et al.*,1998). Based on structural and sequence analysis, Tm0979 may play a role in intracellular sulphur oxidation (Gaspar *et al.*,2005).

Mth1491 is a conserved hypothetical protein from *Methanobacterium thermoautotrophicum* (PfamB 4177, COG 1416) (Christendat *et al.*,2002). Based on PSI-BLAST analysis, Mth1491 sequence share high sequence similarity with conserved hypothetical proteins which contain a "conserved cysteine-containing domain" proposed to function as a disulfide bond redox regulator (Christendat *et al.*,2002). Moreover, these proteins all have highly conserved cysteines that align with Cysteine 72 of Mth1491. DsrF, a small soluble protein in the metabolic pathway for the oxidation of sulphur in phototrophic bacteria, also carries a conserved cysteine-containing domain (Pott *et al.*,1998). Those

observations suggest that Mth1491 may play a role in sulphur oxidation, as proposed for Tm0979.

YchN from *E. coli* belongs to a family of conserved hypothetical proteins known as COG1553 in the National Center for Biotechnology Information (NCBI) database of Clusters of Orthologous Groups (Shin *et al.*,2002). All the members of the COG1553 are assumed to have an uncharacterized ancient conserved region involved in intracellular sulphur reduction. Moreover, YchN share sequence homology with DsrE protein from *Chromatium vinosum*, and is also another member of the gene cluster *dsrABEFHCMK*, products of which conduct intracellular oxidation of stored sulphur (Pott *et al.*,1998). Those observations suggest that YchN may be involved as well in intracellular sulphur metabolism, as suggested for Tm0979 and Mth1491. Moreover, the monomers of these three proteins exhibit high structural similarity, despite very low sequence homology. Since structure is much better conserved than sequence, evolutionary relationships among those three proteins were investigated herein. The tertiary and quaternary structure of those proteins will be described in section 1.4.1 and evolutionary relationships will be discussed in section 1.4.2.

1.4.1 Description of YchN-like protein fold

The tertiary structure of Tm0979, Mth1491 and YchN is very similar and consists of a central four or five stranded β-sheet flanked by four α-helices (Figure 1.1). As described in Figure 1.2, Tm0979, Mth1491 and YchN fold as a homodimer, homotrimer and homohexamer.
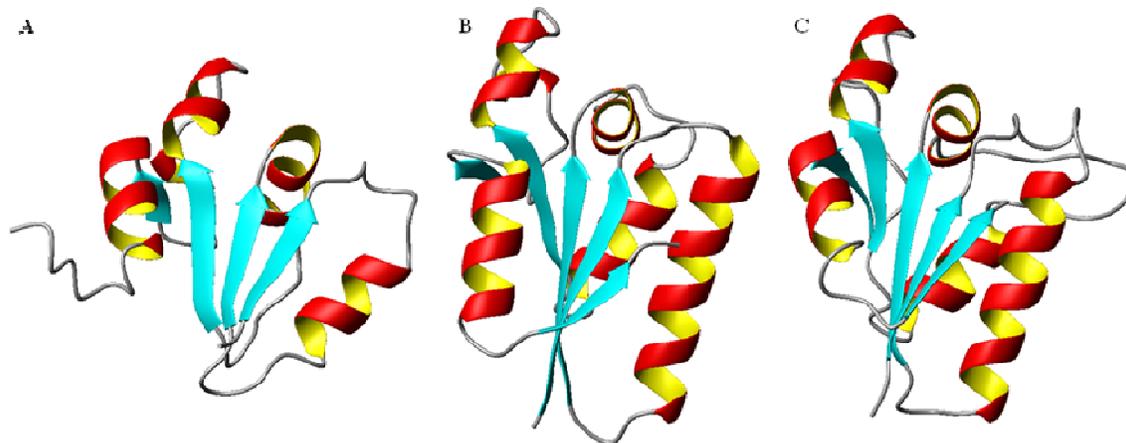
Figure 1.1: Monomer structures of Tm0979 (A), Mth1491 (B) and YchN (C).
The beta strands are coloured in cyan, alpha helices in red and yellow and the loops and turn in grey.
Ribbon diagrams were generated using MolMol (Koradi et al.,1996) using PDB accession codes: A:1x9a,
B:1l1s, C:1jx7.
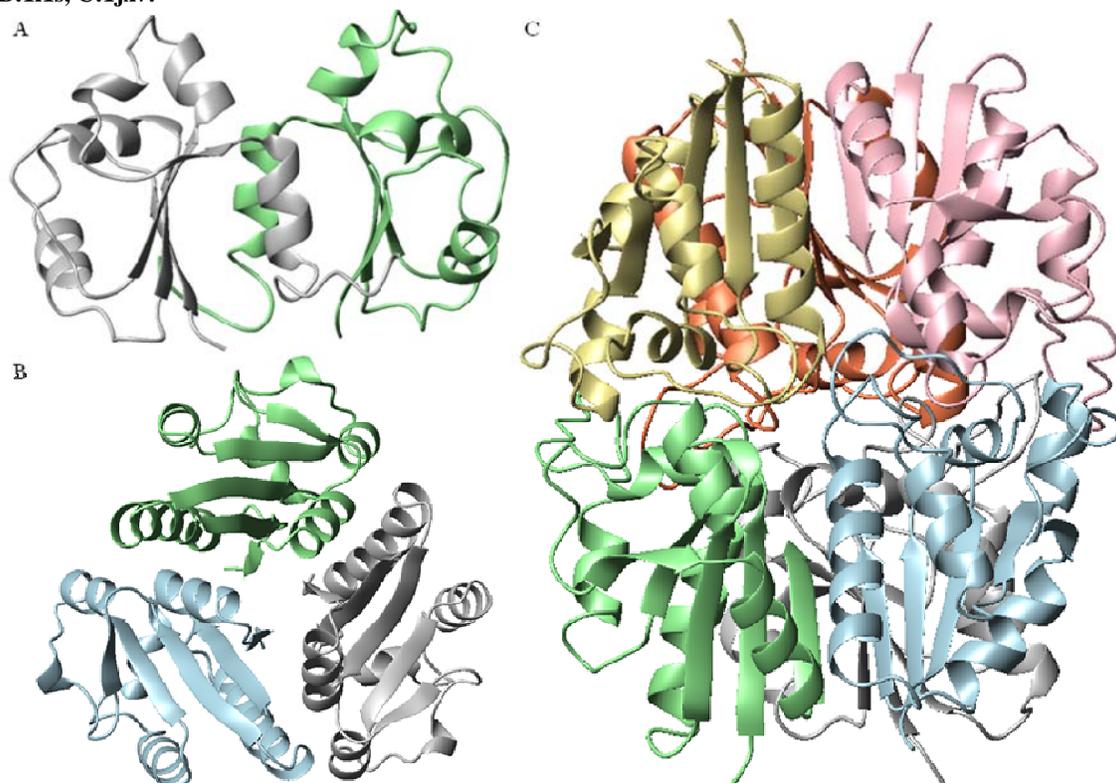
Figure 1.2: Three dimensional structure of Tm0979 dimer (A), Mth1491 trimer (B), YchN hexamer (C).
A, Each subunit of the Tm0979 homodimer is coloured in different colours (grey, pale green). B. Each
subunit of Mth1491 homotrimer is coloured using different colours (light grey, pale green and light blue).
C. Each subunit of YchN homohexamer is coloured with different colours (light grey, pale green, light
blue, khaki, pink and coral). Ribbon diagrams were generated using MolMol (Koradi et al.,1996) using
PDB accession codes: A:1x9a, B:1l1s, C:1jx7.

13

1.4.1.1 Comparison of the monomer structures of Tm0979 and Mth1491

Superposition of the Tm0979 and Mth1491 monomers, in Figure 1.3, shows the structural similarities between these two proteins. Actually, the first four beta strands and the four α-helices superpose very well meaning that the structure of the monomer is conserved within the YchN-like family (Figure 1.3A). The main differences between the structures of these two monomers occur in the length of the amino acid sequence of the proteins which results in longer helix-α_1 , a second helix-α_2, longer β-strands and the formation of a fifth β-strand for the Mth1491 monomer (Figure 1.3B). In fact, Tm0979 sequence is composed of 89 amino acids whereas Mth1491 sequence consists of 111 amino acids, which can explain why longer secondary structure elements are observed in the Mth1491 monomer. The formation of longer elements of secondary structure results in a more structured monomer for Mth1491 compared to Tm0979 and also a more packed oligomer. Interactions between the monomers are discussed in the following section.
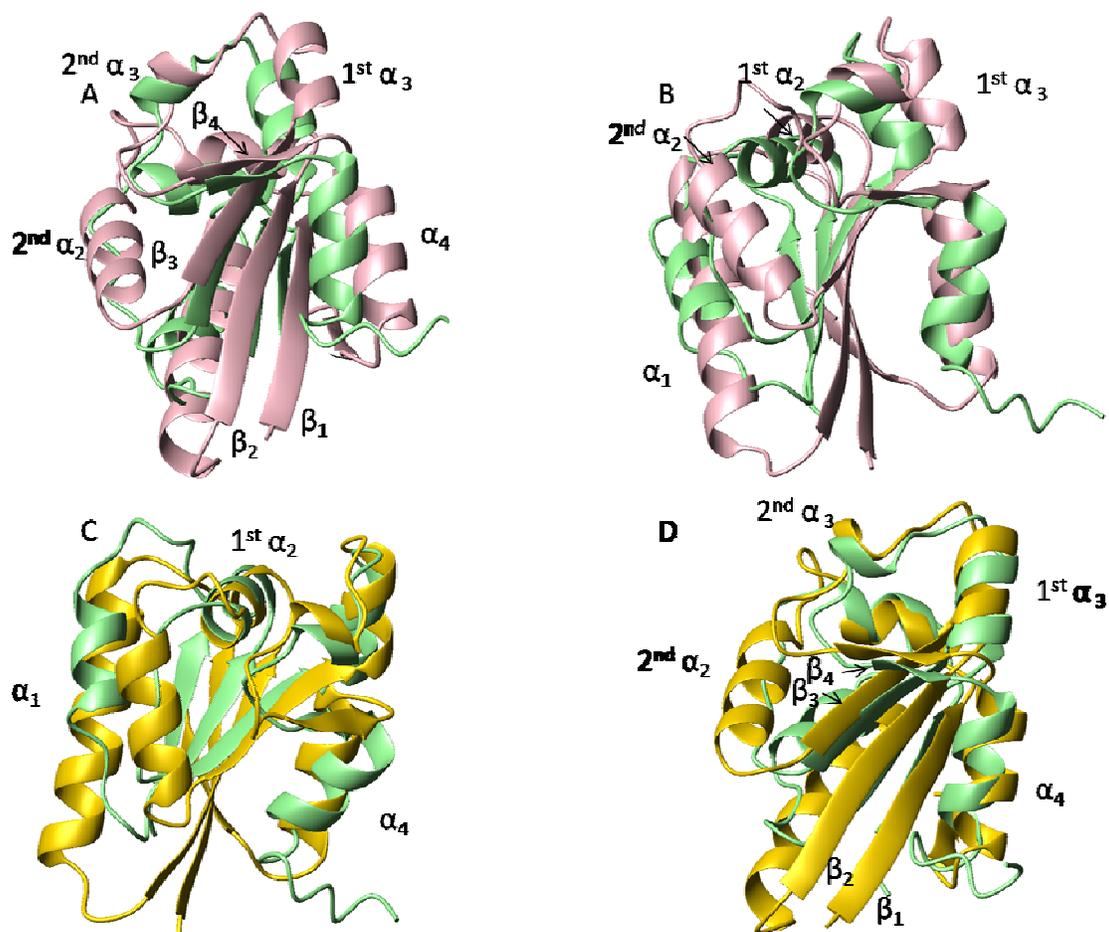
Figure 1.3: Superposition of Tm0979 and Mth1491 monomers.
A and B: Different views of the superposition of Tm0979 (green) and Mth1491 (pink) obtained from Protein structure comparison service SSM at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm), authored by E. Krissinel and K. Henric (Krissinel *et al.*,2004). C and D: Different views of the superposition of Tm0979 (green) and Mth1491 (yellow) obtained from Superpose (http://wishart.biology.ualberta.ca/SuperPose/) (Maiti *et al.*,2004). Ribbon diagrams were generated using MolMol (Koradi *et al.*,1996) using PDB accession codes: 1x9a and 1l1s.

1.4.1.2 Comparison of Tm0979, Mth1491 and YchN quaternary structure

In solution, Tm0979 folds as a dimer whereas Mth1491 forms a trimer and YchN a hexamer (Gaspar *et al.*,2005), (Christendat *et al.*,2002), (Shin *et al.*,2002). As mentioned above, the quaternary structure of Mth1491 is more packed and structured than that of Tm0979. This is illustrated in the difference in the interface composition for both proteins. On one hand, the Tm0979 interface is mainly formed by hydrophobic interactions between the fourth α-helix of each monomer, and between the fourth α-helix of one monomer and the first β-strand of the other monomer (Figure1.4).
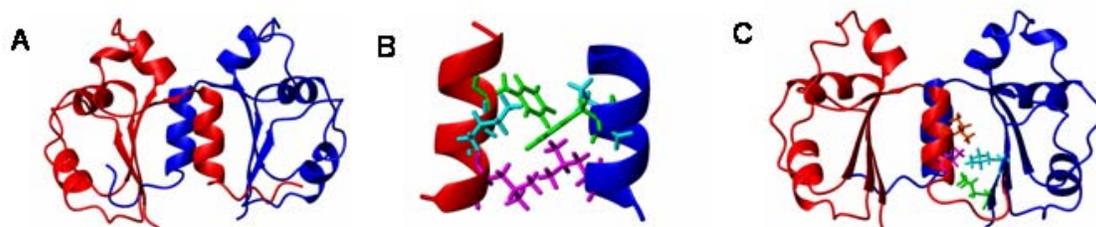
Figure 1.4: Tm0979 dimer interactions.
A. Tm0979 dimer. B. Tm0979 dimer interaction between the two α4-helices (one blue, the other one red) involving Leu79 (magenta), Ile76 (cyan) and Phe75 (green). C. Tm0979 dimer interaction between α4-helix of one subunit (Ile76 in magenta and Leu79 orange) and β1-strand of the other one (Leu3 in green and Leu5 in cyan). Ribbon diagrams were generated using MolMol (Koradi *et al.*,1996) using PDB accession code: 1x9a.

On the other hand, the Mth1491 trimer interface consists not only of hydrophobic interactions (Figure 1.5A) but also of ionic interactions, such as the salt bridges and hydrogen bonding (Figure 1.5B) (Christendat *et al.*,2002). Therefore, based on interface interaction, Mth1491 interface seems to be more specific than Tm0979 one.

Concerning YchN homohexamer, each homotrimer are stabilised by hydrophobic interactions and H-bond interactions mainly (Figure 1.6A). The hexamer is formed by association of the two homotrimers and form a dimer of trimer. The interface is mainly composed of hydrophobic interactions (Figure 1.6B).
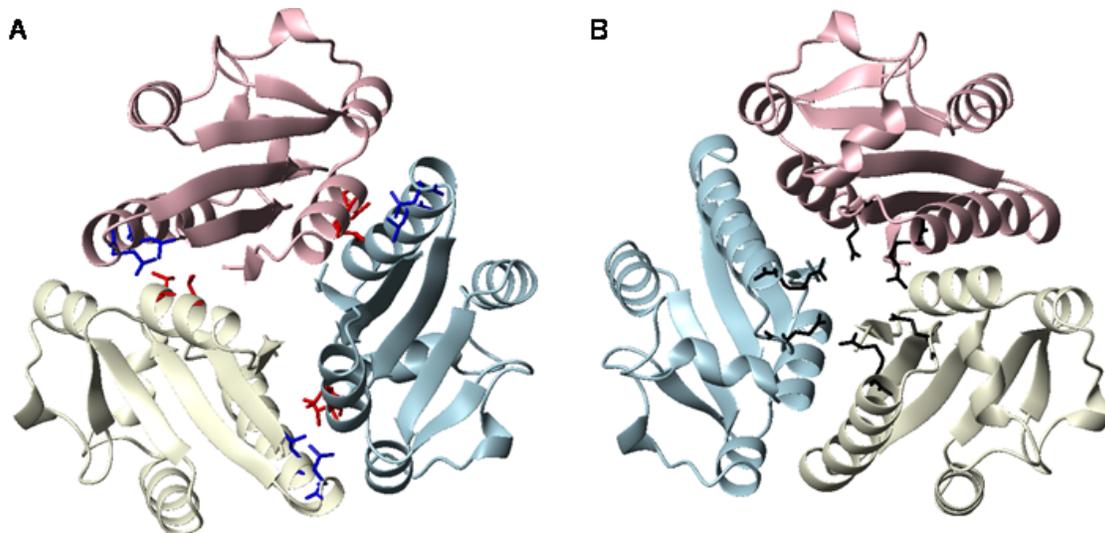
Figure 1.5: Mth1491 trimer interactions.
A, Mth1491 trimer hydrophobic interaction between helix-α$_1$ of one subunit (blue residues: Leu28 and Leu32) and helix-α$_4$ of the neighboring subunit (red residues: Val97 and Val101), B, ionic interaction between helices-α$_4$ (Arg17 and Glu12 are coloured in black). Ionic interactions also occur between helix-α$_1$ and helix-α$_4$ involving Arg103 and Asp31. Ribbon diagrams were generated using MolMol (Koradi *et al.*,1996) using PDB accession code: 1l1s.
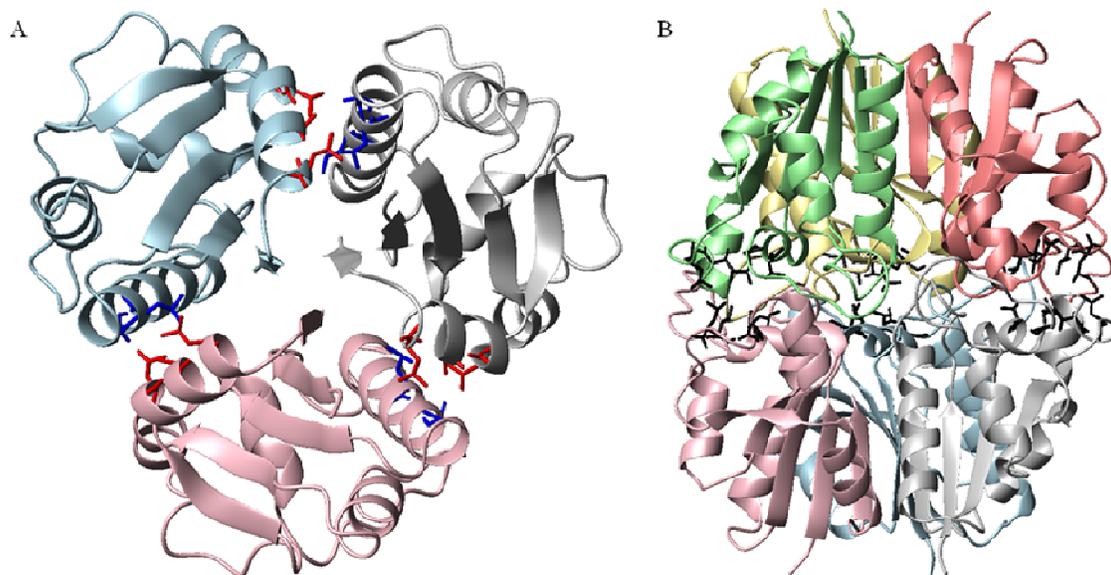
Figure 1.6: YchN trimer and hexamer interactions.
A, YchN trimer hydrophobic interactions between the helix-α$_1$ of one subunit (blue residues: Leu24, Ala27 and Leu28) and the helix-α$_4$ of the neighbouring subunit (red residues: Leu101, Ala105 and Leu109). B, YchN hexamer hydrophobic interactions. The trimer-trimer interaction on the equatorial interface is formed mainly by the inner L1–L1' and outer L3–L3' loop interactions. Residues Thr47, Leu50, Ile87, and Leu90 of each subunit make a hydrophobic patch with the same residues of a two-fold related subunit that seems to mediate stable trimer-trimer interaction. Ribbon diagrams were generated using MolMol (Koradi *et al.*,1996) using PDB accession code: 1jx7.

17

1.4.2 YchN-like proteins hypothetical function and evolution

As suggested previously (Section 1.4), Tm0979, Mth1491 and YchN may play a role in sulfur metabolism. Moreover, as described in the previous section, Tm0979, Mth1491 and YchN share high structure similarity. As structure is better conserved than sequence, those three proteins probably evolved from a common ancestor. Actually, the gene cluster dsrABEFHCMK encodes, among others, three similar proteins dsrE, dsrF, and dsrH which have considerable structure homology with TusD, TusC and TusB from *E.coli*, respectively (Numata T, 2006). The TusBCD complex, a sulphur transfer mediator, is a heterohexamer composed of a dimer of the heterotrimers (Numata T, 2006). The structure similarity between the monomers of Tm0979 and TusB, and between the monomers of Mth1491, YchN and TusD suggests that Tm0979, Mth1491 and YchN monomers may form with or without other dsr proteins a more complex oligomer, functionally active *in vivo*. The functions of Tm0979, Mth1491 and YchN, however, remain unknown at this time. Therefore, the dimeric state of Tm0979 may be a regulatory way to store the inactive protein *in vivo*, as observed for some DNA-binding proteins (Section 1.3.2.1). The TusD monomer shares significant structure homology with the monomer of YchN and the monomer of Mth1491. Thus the YchN homohexamer may have evolved by differentiation to a more complex fold such as the TusBCD heterotrimer, probably due to a change of function or the addition of a new function. This heterotrimer may then have dimerised to form a hexameric complex. It should be noted that the order of events is not known, but various evolutionary mechanisms may link these proteins.

Evolution of Tm0979, Mth1491 and YchN structure, in particular quaternary structure, were investigated by characterizing the mechanism of folding of the Tm0979 dimer and the Mth1491 trimer. Those mechanisms were compared to folding mechanisms of other dimers

and trimers and are discussed in chapters 4 and 5. In the long term, engineering of Tm0979 and Mth1491 monomers will be considered in order to investigate the evolutionary pathway of YchN fold as well as the mutational process responsible for the switch of the oligomeric state. This work will be discussed in more detail in chapter 6 of this thesis.

2 Protein expression, purification and preparation

2.1 Introduction

Tm0979 is expressed by *T. maritima*, a hyperthermophilic organism which has a optimal growth temperature of 90°C (Adams,1994). On the other hand, Mth1491 is expressed by *M. thermoautotrophicum*, a thermophilic organism whose optimal growth temperature is 65°C (Smith *et al.*,1997). Proteins from thermophilic organisms tend to have higher stability against thermal denaturation, and this property was exploited in developing the protein purification protocol of those two proteins. The expression and purification of Tm0979 were previously studied and optimised in our laboratory by Joe Gaspar and Gabriela Meglei, and will be described in detail in the section 2.2.1.1 (Gaspar *et al.*,2005). Mth1491 expression and purification protocols were optimised by Christendat et al. (Christendat *et al.*,2000; Christendat *et al.*,2002). However, the initial purifications of Mth1491 resulted in very low yield and aggregation was observed throughout. Therefore, the expression and purification of Mth1491 required optimisation, described in section 2.3.

2.2 Materials and Methods

2.2.1 Tm0979 and Mth1491 expression

The gene sequence for Tm0979 was previously PCR-amplified from *T. maritima* genomic DNA and subcloned immediately after the N-terminal (His)6 tag and the thrombin cleavage site (with sequence MGSS(H)6SSGLVPRGSH) of the pET15b vector which contains an ampicillin and kanamycin resistance gene (Figure 2.1) (Gaspar *et al.*,2005). The recombinant Mth1491 expressing *E. coli* cells were obtained from Adelina Yee from the University of Toronto and used as described by Christendat and coworkers (Christendat *et al.*,2000). The protein was expressed in strain BL21(GoldλDE3) which contains T7 RNA polymerase under control of the β-lac promoter. Addition of isopropyl-β-D-thiogalactopyranoside (IPTG) induces expression of T7 RNA polymerase by binding the lac repressor and causing it to dissociate from the operator DNA so that T7 RNA polymerase can initiate the transcription from the T7 promoter, Tm0979 gene or Mth1491 gene. A starter culture of transformed BL21-GoldλDE3 cells was grown in 100 mL 2TY medium (1.6% (v/v) bacteriotryptone, 1% (v/v) yeast extract, 1% (v/v) sodium chloride (NaCl)) containing 100 µg/mL ampicillin and 50 µg/mL kanamycin at 37°C overnight. The starter culture was diluted, approximately 100-fold, into 1L of fresh 2TY medium containing 100µg/mL ampicillin and 50 µg/mL kanamycin, the following morning, and grown at 37°C to log phase (indicated by an optical density measured at 600 nm (OD_{600}) between 0.6-0.7), at which point Tm0979 expression was induced with 1 mM IPTG for 4 hours. Cells were then harvested bu centrifugation at 65000 g for 15 minutes at 4 °C. The supernatant was discarded and cell pellets were stored in 50 mL falcon tubes at-80°C. SDS-PAGE (30%) samples to follow the time course of protein expression were prepared by removing a 1 mL aliquot from the culture, before and every hour after induction, centrifuging at 14000 rpm for 2 minutes, removing the

supernatant and resuspending the pellet in 40 μL of 30% SDS-PAGE sample buffer (Figure

2.2).
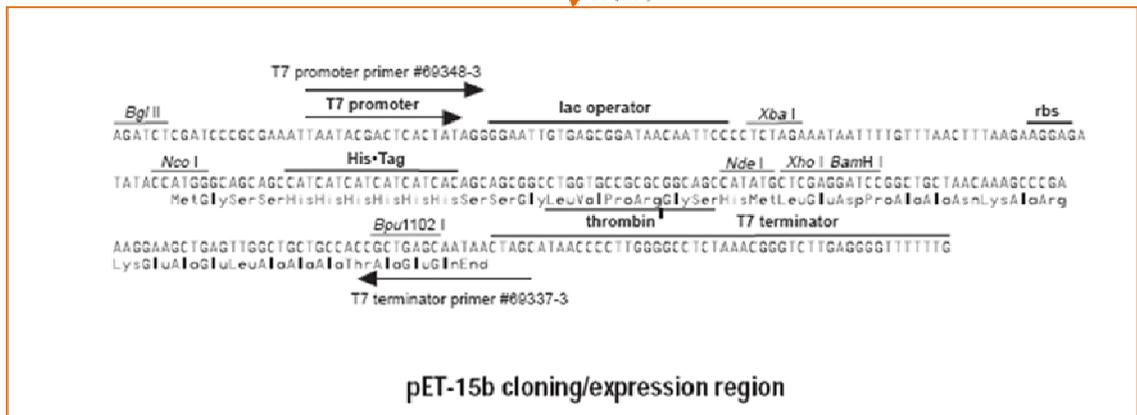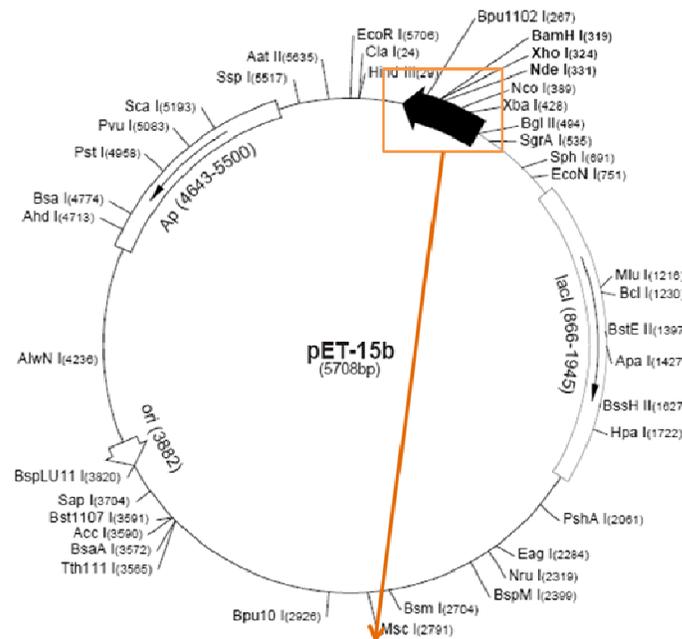
Figure 2.1: Circular map of the pET-15b vector (Cat. No. 69661-3). It carries an N-terminal His-Tag sequence followed by a thrombin site and three cloning sites. Unique sites are shown on the circle map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below in the zoomed window.
Figures from Novagen catalogue (http://depts.washington.edu/bakerpg/plasmid_maps/pet15bm.pdf).

23

A

M 1 2 3 4

M:marker

1:one hour after induction

2: two hours after induction

3: three hours after induction

4: four hours after induction

B

1 2 3 4 5 6 7

1: before induction

2: one hour after induction

3: two hours after induction

4: marker

5: three hours after induction

6: four hours after induction

7: Purified Mth
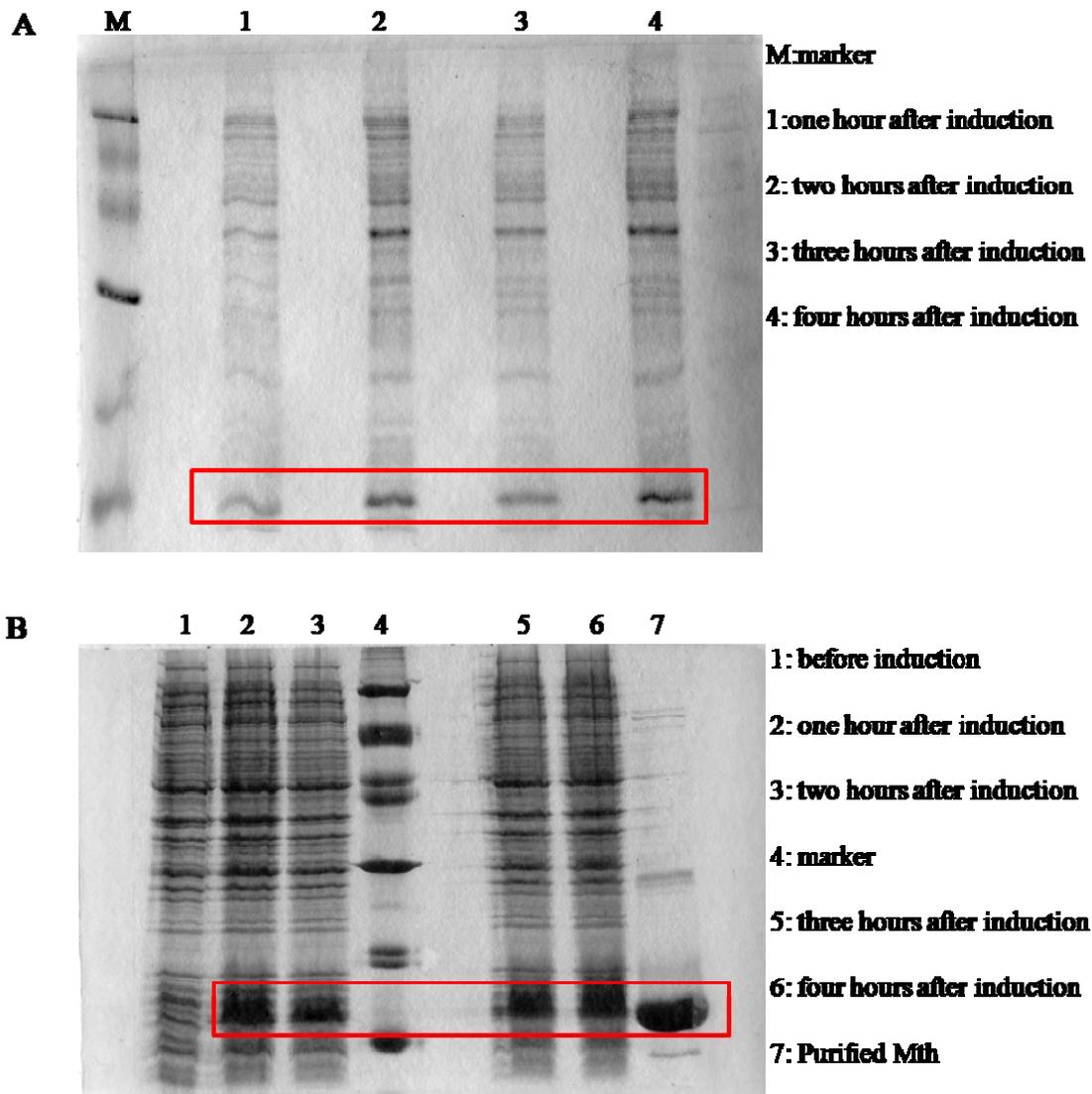
Figure 2.2: Expression gels (30% SDS) of Tm0979 and Mth1491.
A. Expression gel of Tm0979. Lines 1, 2, 3 and 4 correspond to the protein content of the cells after 1h, 2h, 3h and 4h, respectively, after induction. The position of Tm0979 on the gel is shown by the red rectangle. B. Expression gel of Mth1491. Line 1 corresponds to the protein content of the cells before induction. Lines 2, 3, 5, 6 correspond to the protein content of the cells 1h, 2h, 3h and 4h after induction, respectively. Line 4 corresponds to the marker. Line 7 corresponds to pure Mth1491 from previous purification. The position of Mth1491 on the gel is shown by the red rectangle. Samples were prepared by taking 50 µL of growing cells and diluting them in 50 µL loading buffer. Each sample is then boiled and centrifuged at 14000 rpm and the supernatant is loaded onto the gel.

2.2.2 Tm0979 and Mth1491 protein purification

All buffers used for Mth1491 purification were degassed at room temperature for 30 minutes prior to use during the purification. The cell pellets were resuspended in lysis buffer (50 mM sodium phosphate (Na_2HPO_4), 300 mM NaCl, 10 mM imidazole, pH 8). Cell suspension was run through an emulsifier (EmulsiFlex-C5 from Avestin) at approximately 17000 psi to lyse the cells. The resulting suspension was incubated at 60°C for 20 minutes in order to denature thermally labile *E. coli* proteins. More than half of the total protein is precipitated by the heat treatment (Figure 2.4 B). The heat-treated suspension was centrifuged at 45000 g for 20 minutes at 4°C and the pellet discarded. The supernatant was then loaded onto a Ni^{2+} affinity chromatography column (Qiagen) connected to a liquid chromatography system (BioLogic LC, Biorad) and eluted using an imidazole gradient from 20 mM to 500 mM imidazole in 50 mM Na_2HPO_4, 300 mM NaCl, pH 8 for Mth1491 purification and from 20 mM to 1 M imidazole in 50 mM Na_2HPO_4, 300 mM NaCl, pH 8, for Tm0979 purification. Tm0979 and Mth1491 both elute from the column at an imidazole concentration between approximately 300 mM and 400 mM, as shown in Figures 2.3A and 2.4A. For Mth1491 purification, ethylene diamine tetra acetic acid (EDTA) and dithiothreitol (DTT) were placed in each collecting tube, prior to Mth1491 elution, to give final concentrations of 1 mM and 10 mM, respectively, to minimise effect of any metal contamination and disulfide bond formation. After fraction collection, glycerol was added to eluted fractions containing Mth1491 to a final concentration of 10% (v/v), to stabilize the trimer and prevent aggregation. 30% SDS gel was prepared to check the protein content of the eluted fractions (Figures 2.3B and 2.4B). Samples were prepared by taking 20 μL of eluted fraction and diluting them in 50 μL loading buffer. Each sample was then boiled and centrifuged at 14000 rpm and the supernatant was loaded onto the gel. Fractions containing Tm0979 or Mth1491 were pooled and concentrated with an ultrafiltration system (200 mL Amicon cell / Millipore) using

membranes with a cut-off of 3 kiloDalton (kDa) (Millipore: YM3 ultrafiltration membranes made of regenerated cellulose, diameter: 63.5 mm).
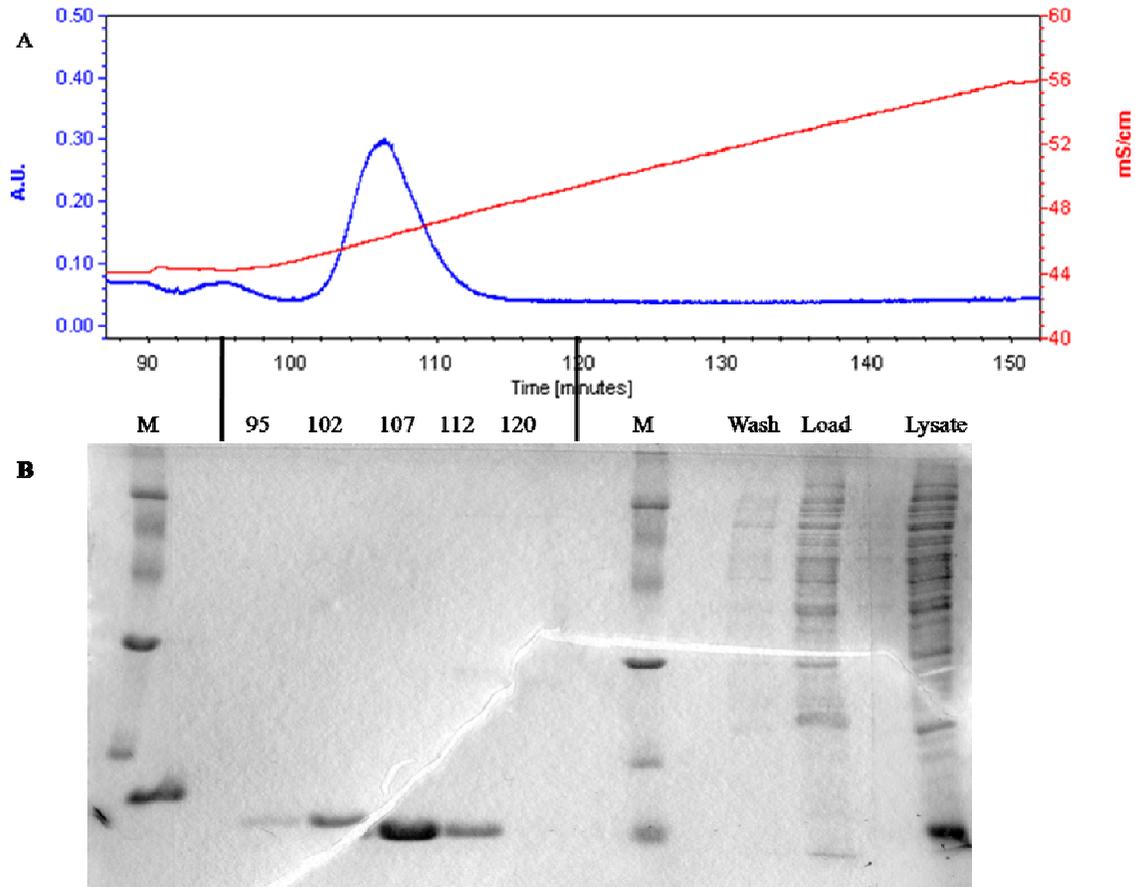
Figure 2.3: Ni²⁺ affinity chromatography column purification of Tm0979.
A. Elution profile of Tm0979 purification. This is the elution profile from Ni²⁺ affinity chromatography column. The blue line represents the change in absorbance at 280 nm (A₂₈₀). The red line represents the change in conductivity due to increasing concentration of imidazole in solution. The vertical black thick lines correspond to the alignment of the elution profile to the 30% SDS gel with time. Tm0979 elutes at 102-114 minutes, corresponding to 300-400 mM imidazole in 25 mM Na₂HPO₄, pH 8. B. 30 % SDS-PAGE gel showing the purity of Tm0979 during the purification process. From left to right: M: marker, 95, 102, 107, 112, 120: advancement time of the purification, Wash: composition of the eluent at the washing step, Load: composition of the eluent at the loading step, lystate: composition of the lysate after 60°C treatment. The imidazole ring of the histidines of Tm0979 His-tag interacts with Ni²⁺ and makes the proteins bind to the column while the lysate is loaded (Load). Then the column is washed to remove all the other proteins not bound to the column (Wash) and a gradient of imidazole is applied to the column. There is therefore a competition between the rings of the histidines of Tm0979 His-tag and of imidazole to bind Ni²⁺. The conductivity increases with increasing the concentration of imidazole. At 45 mS.cm⁻¹ more imidazole binds to the column and the protein can no longer bind and is eluted from the column.
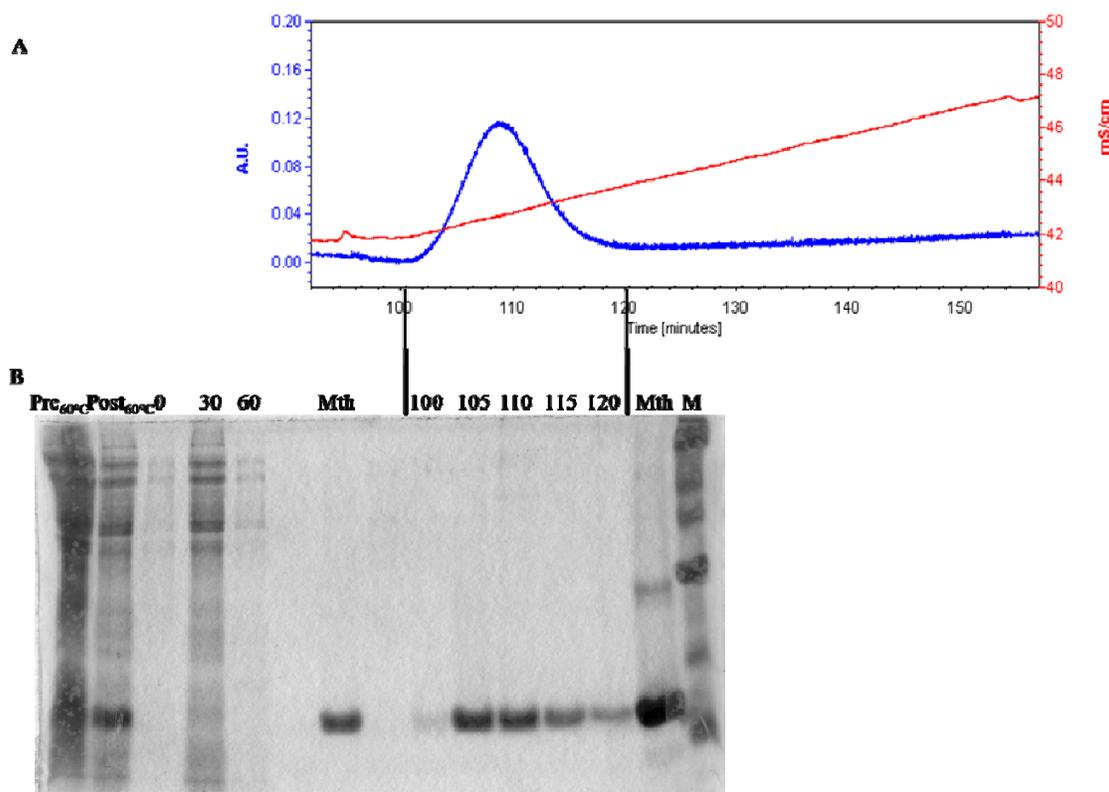
Figure 2.4: Ni^{2+} affinity chromotography column purification of Mth1491.
A. Elution profile of Mth1491 purification. This is the elution profile from Ni^{2+} affinity chromatography column. The blue line represents the change in absorbance at 280 nm (A$_{280}$). The red line represents the change in conductivity due to increasing concentration of imidazole in solution. The vertical black thick lines correspond to the alignment of the elution profile to the 30% SDS gel with time. Tm0979 elutes at 100-118minutes, corresponding to 300-400 mM imidazole in 25mM Na$_2$HPO$_4$, pH 8.
B. 30% SDS gel showing the purity of the Mth1491 during the purification process. From left to right: Pre$_{60°C}$: composition of the lysate before the 60°C water bath treatment, Post$_{60°C}$: composition of the lysate after the 60°C water bath treatment, 0, 30, 60, 100, 105, 110, 115, 120: advancement time of the purification, Mth: pure Mth1491 from previous purification, M: marker.
The imidazole ring of the histidines of Mth1491 His-tag interacts with Ni^{2+} and makes the proteins bind to the column while the lysate is loaded (0, 30 minutes). Then the column is washed (60 minutes) to remove all the other proteins not bound to the column and a gradient of imidazole is applied to the column. There is therefore a competition between the rings of the histidines of Mth1491 His-tag and of imidazole to bind Ni^{2+}. The conductivity increases with increasing the concentration of imidazole. At 45 mS.cm^{-1} more imidazole binds to the column and the protein can no longer bind and is eluted from the column (fraction 100-120).

The protein solution was exchanged into NMR buffer (25 mM Na_2HPO_4, 450 mM NaCl, pH 6.5), for Tm0979, and for Mth1491, after experimentation with various buffers into the most optimal buffer, to a citrate buffer (20 mM sodium citrate ($Na_3C_6H_5O_7$), 450 mM NaCl, 10% (v/v) glycerol, pH 6), by dilution/reconcentration by ultrafiltration. Protein concentration was determined by measuring the A_{280} (Absorbance of the solution at 280 nm) and using the experimentally determined extinction coefficient and theoretical molecular weight (His-tag included) of 11167.53 $M^{-1}.cm^{-1}$ and 12040 $g.mol^{-1}$, respectively, for Tm0979 (Gaspar *et al.*,2005) and 12058 $M^{-1}.cm^{-1}$ and 14720 $g.mol^{-1}$ for Mth1491.

2.3 Results and discussion

2.3.1 Mth1491 purification, optimisation and storage

Unfortunately, during initial purifications, Mth1491 aggregated during the concentration step and this resulted in very low yields of purified protein. Mth1491 is trimeric and has two cysteines per monomer (Figure 2.5). In oxidative conditions, these cysteines can form improper intermolecular disulfide bonds, which can result in aggregation of the protein. Mth1491 is an intracellular protein that functions in a reducing environment where disulfide bonds are not formed. Cysteine 72, in particular, is conserved within sequence related proteins as well as all of the proteins containing a conserved cysteine-containing domain and is thought to play a key role in sulfur oxidation (Christendat *et al.*,2002) (Section 1.4.2). In the case of Mth1491, improper disulfide bonds are made during the concentration step, within hours after purification, and drive protein aggregation. This process is enhanced at high protein concentration. Preventing aggregation was made possible by changing the conditions of Mth1491 purification and storage; the main changes are summarized in Table 1.1.
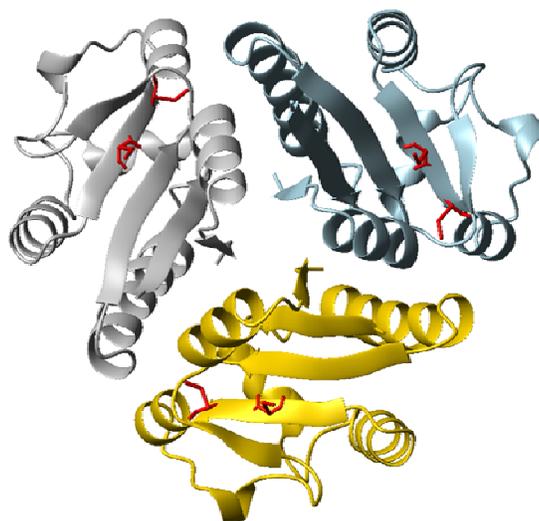
Figure 2.5: Representation of the six cysteines in Mth1491 trimer.
Each subunit is coloured in a different colour (yellow, light blue and light grey). Cysteines 70 and 72 are coloured in red. Ribbon diagrams were generated using MolMol (Koradi *et al.*,1996) using PDB accession code: 1l1s.

Table 2.1: Optimisation of the experimental conditions for stabilizing purified Mth1491.

Observation	Explanation	Solution
Initial conditions: 1. Mth1491 kept in NMR buffer. However Mth1491 aggregates during concentration.	Buffer does not keep the protein in a stable conformation.	Change buffer conditions: New buffer used corresponds to the one used for the determination of Mth1491 structure, i.e., 10 mM sodium acetate (CH_3COONa), 300 mM NaCl, 10 mM DTT, pH 5. Moreover, 1 M DTT is added into each collecting tubes to give a final concentration of 10 mM in eluted fractions.
2. Mth1491 no longer aggregates during the concentration step. However, the protein aggregates when samples are thawed from the -80°C freezer.	Mth1491 is a large protein. 10% glycerol is commonly used to stabilise larger protein. Also, metals can enhance the oxidation of disulfide bond.	Introduce glycerol to a final concentration of 10% (v/v) and EDTA to a final concentration of 1 mM in each collected fraction. In addition, all buffers used for Mth1491 were degassed before the purification.

The first studies conducted on Mth1491 in our lab were performed using NMR buffer and aggregation of the protein was observed under these conditions. Exchange of the protein into acetate buffer (10 mM CH_3COONa, 300 mM NaCl, 10 mM DTT, pH 5) used by Christendat and co-workers for the x-ray crystallographic determination of the Mth1491 structure results in dissolving the aggregates observed during the concentration process (Christendat *et al.*,2002). The protein was then frozen in liquid nitrogen and stored in a -80°C freezer. However, upon thawing of the frozen protein solution, some aggregates were observed, even after the protein was centrifuged or filtered with a 0.2 μm or 0.02 μm filter. Two parameters were then considered. First, Mth1491 needs to be stored in a long term reducing environment. Metal contamination needs as well to be eliminated because metals can catalyse disulfide bond formation, in particular, mixed-disulfide derivatives which lead to protein aggregation (Stadtman,1990; Kadokura *et al.*,2003). EDTA and DTT were thus introduced in each collecting tube during Mth1491 affinity column purification to give final concentrations of 1 mM and 10 mM, respectively. All buffers required for the use or the storage of Mth1491 after purification were degassed and contained 10 mM DTT and 1 mM EDTA. Secondly, other stabilising agents were added to stabilise the protein for long time

31

storage. As for other large protein, Mth1491 quaternary structure likely requires stabilisation in order to prevent the formation of larger oligomers or aggregates (Mallam *et al.*,2005). After the purification was completed, glycerol was therefore introduced in each collecting tubes which contain Mth1491 to a final concentration of 10 % (v/v). Glycerol was added as well in acetate buffer for storage of Mth1491.

In summary, Mth1491 is a large trimeric protein containing six cysteines per trimer. The protein environment needs therefore to be kept reducing and the native structure needs to be stabilised. The buffers used for Mth1491 storage were degassed and contained EDTA, DTT and glycerol. By comparing the concentration of the protein before and after storage at -80°C, Mth1491 no longer aggregates. Having found conditions to stabilize the purified protein, much better protein yields were obtained and protein folding experiments could then be undertaken.

3 Spectroscopic properties of Mth1491 and Tm0979 and optimization of Mth1491 refolding

3.1 Introduction

Characterising Tm0979 and Mth1491 equilibrium folding involves determining the free energy of folding, ΔG, and the m-value related to the protein folding. In order to do this, denaturation curves were measured at different protein concentrations and were fit to the appropriate model. These curves need to be at equilibrium when fitted in order to obtain accurate equilibrium constants, $i.e$, renaturation curves and denaturation curves at a particular protein concentration should superimpose when equilibrium has been reached. Previous experiments conducted on Tm0979 showed that the protein unfolds reversibly and denaturation curves reach equilibrium after 2 days (Bo Cheyne, Gabriela Meglei & Elizabeth Meiering, unpublished data). Therefore, denaturation curves samples were incubated for 2 days at room temperature before making measurements. Concerning Mth1491, no equilibrium curve analysis had been performed previously. In this chapter, the conditions required for reversible folding of Mth1491 trimer and the time required for the denaturation and renaturation curves to reach equilibrium will be described.

3.2 Materials and methods

3.2.1 Denaturation and renaturation curves sample preparation

3.2.1.1 Tm0979

Denaturation curve samples were prepared by 10-fold dilution of a stock solution of native Tm0979 into increasing concentrations of guanidinium chloride (GdmCl). Renaturation curve samples were prepared by 10-fold dilution of a stock solution of fully unfolded protein, incubated in 6 M GdmCl for 15 minutes, into decreasing concentrations of GdmCl. First, filtered ultrapure 8 M GdmCl (Sigma) and filtered MilliQ water were aliquoted in 1.5 mL Eppendorf tubes using micropipetman according to desired GdmCl concentration. Filtered NMR buffer was then introduced by diluting it 10-fold. Stock protein solution was added last using a Hamilton syringe (See Appendix A.1: Preparation of Tm0979 denaturation curve samples).

3.2.1.2 Mth1491

Denaturation curve samples were prepared by 10-fold dilution of a stock solution of native Mth1491, incubated in 10mM DTT and 1mM EDTA, into different concentrations of GdmCl. Renaturation curves samples were prepared by diluting a stock solution of fully unfolded Mth1491, incubated in 6M GdmCl, 10mM DTT, 1mM EDTA for 15 minutes, into decreasing concentrations of GdmCl. Filtered and degassed 8 M GdmCl and MilliQ water were used to obtain the different denaturant diluted samples, in 1.5 mL Eppendorf tubes, ranging from 0 to 6 M GdmCl final concentration. Filtered and degassed buffer was then added by diluting it 10-fold. DTT and EDTA were added to each sample to a final concentration of 10 mM and 1 mM, respectively, just before adding the protein using a Hamilton syringe and by 10-fold dilution of the denatured protein. Samples were kept under nitrogen at room temperature for desired equilibrium time (See appendix A.2: Preparation of Mth1491 renaturation curve samples).

3.2.2 Measurement of denaturation and renaturation curve samples

Equilibrium denaturation and renaturation curves were monitored by following the unfolding and refolding transition spectroscopically. Denaturation and renaturation of Tm0979 and Mth1491 were followed by circular dichroism (CD), using a Jasco 715 CD spectropolarimeter, at wavelength 215 nm (220 nm) for Tm0979 (Mth1491) and by steady state fluorescence, using a FL3-22 SPEX fluorolog, with an excitation wavelength of 280 nm (279 nm) and emission wavelength of 332 nm (321 nm) for Tm0979 (Mth1491). A 1 mm pathlength cuvette and a 1 cm pathlength 50μL cuvette were used for CD and fluorescence spectroscopy measurements, respectively. Samples were incubated in a 25°C water bath for 15 minutes prior to measurements and cuvettes were thermostatted to 25°C during the whole measurement. The change in fluorescence intensity and the change in ellipticity were plotted as a function of denaturant concentration for different protein concentrations. All equilibrium curves were remeasured over time to evaluate when equilibrium is reached.

3.3 Results

3.3.1 Spectroscopic properties of Tm0979 and Mth1491

Prior to denaturation curve measurements, the spectroscopic properties of the native and unfolded protein were recorded in order to determine the spectral conditions for best monitoring unfolding. A fluorescence excitation scan was performed for the native and denatured states of both Tm0979 dimer and Mth1491 trimer (data not shown). The maximum of emission was reached for an excitation wavelength of 279 and 280 nm for Tm0979 and Mth1491, respectively. Therefore, the emission scans were measured using these excitation wavelengths. Figure 3.1 shows the CD and fluorescence emission spectra of native and unfolded Tm0979. When monitored by CD, the maximum amplitude of signal is observed at 210 and 212 nm for the native and unfolded protein, respectively. When monitored by fluorescence, the maximum of emission is reached at 327 and 332 nm for the native and unfolded protein, respectively. The difference between the signal of the native and denatured protein is plotted as well and reaches a maximum at 332 and 210 nm when monitored by fluorescence and CD, respectively. Fluorescence-monitored denaturation and renaturation curves were measured at an emission wavelength of 332 nm in order to have the maximum amplitude, and hence sensitivity, for the curves. However, when denaturation curves were measured at 210 nm by CD, the signal to noise ratio was very poor and data were very scattered due to a high tension value above 600 volts. Therefore, the curves were measured at 215nm when monitored by CD. For Mth1491 (Figure 3.2), the difference between the native and unfolded protein reaches a maximum at 210 and 321 nm when measured by CD and fluorescence, respectively. As observed for Tm0979, the data were scattered when measured at an emission wavelength 210 nm by CD. Therefore renaturation and denaturation curves were recorded at an emission wavelength of 220 and 321 nm by CD and fluorescence spectroscopy, respectively.
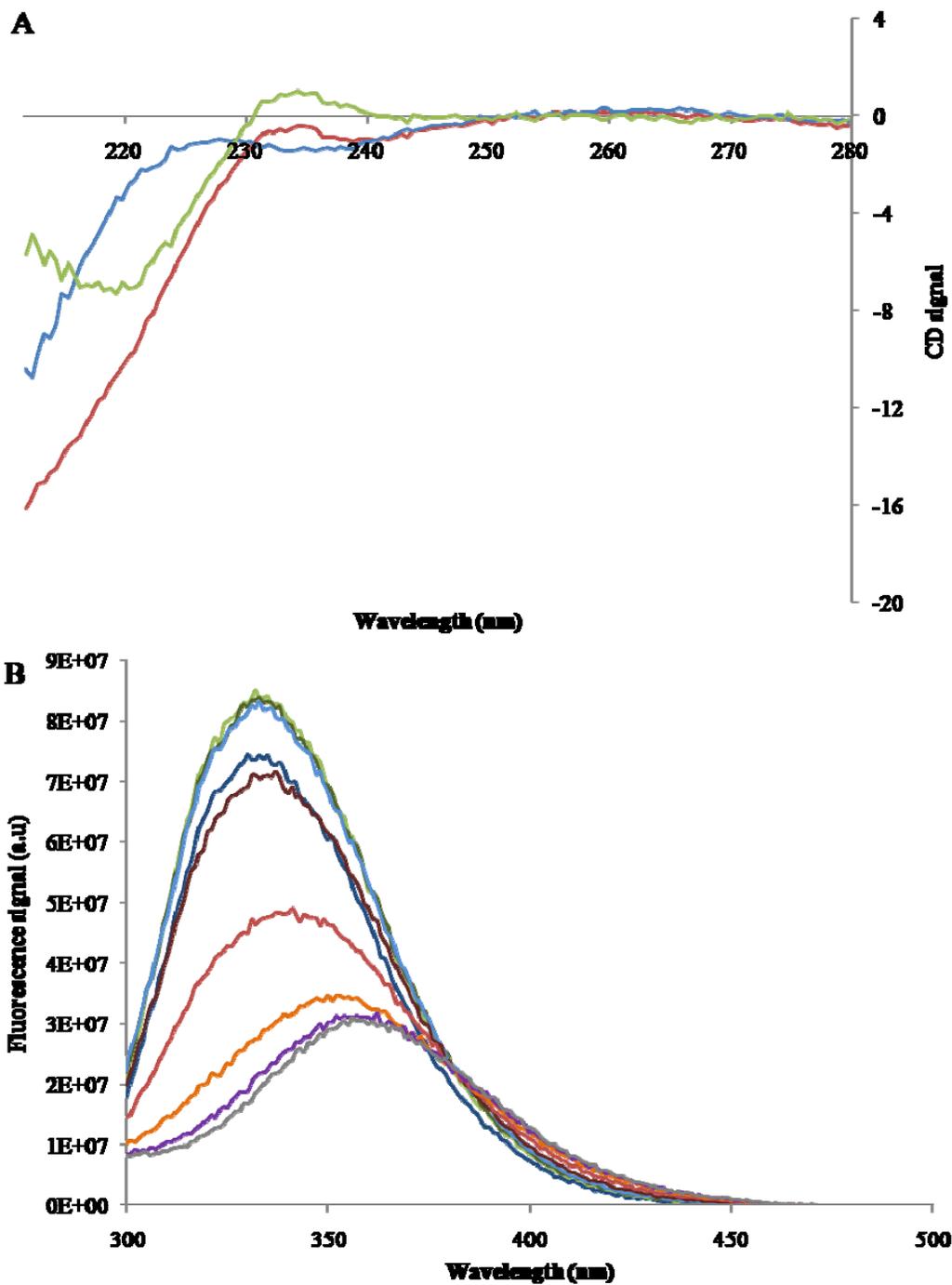
Figure 3.1: Tm0979 CD and fluorescence spectra as function of denaturant concentration.
A. The native and unfolded protein scans are plotted in red and blue lines, respectively. The denatured protein was incubated in 6 M GdmCl for 15 minutes prior to measuring the spectra. Tm0979 concentration was 12.5 μM and is expressed as dimer equivalent. The difference between the signal of the native and of the unfolded protein is plotted in green. The data recorded at lower wavelengths are not plotted because they correspond to high tension above 600volts and are thus not reliable. B. The fluorescence monitored scans were measured at 1.5 μM (dimer equivalent) in different GdmCl concentrations: 0M (dark blue line), 1 M (dark green line), 2 M (light green line), 2.4 M (light blue line), 2.8 M (dark red line), 3.2 M (light red line), 3.6 M (orange line), 4 M (purple line) and 5 M (grey line). All the scans were performed at 25°C in NMR buffer.
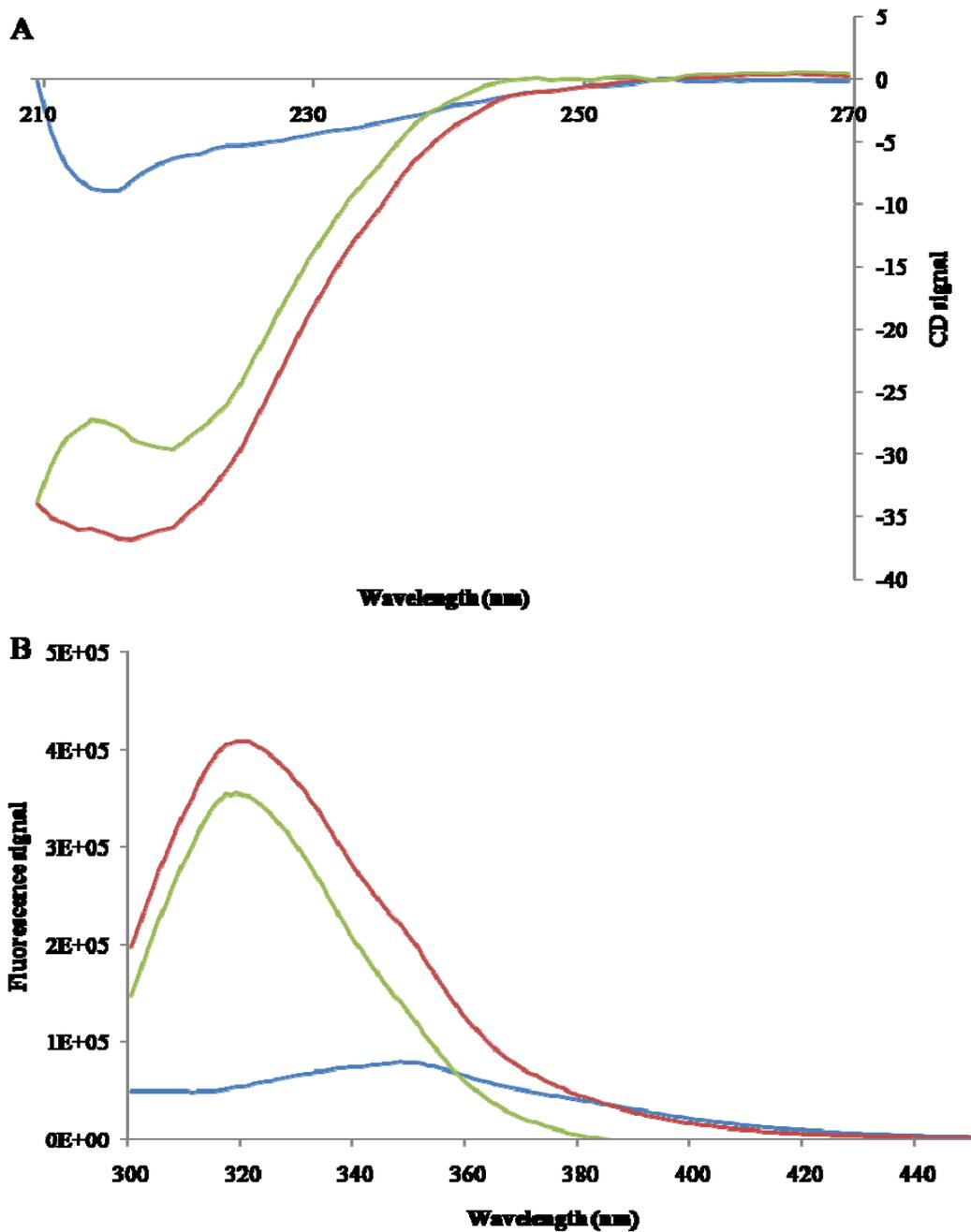
Figure 3.2: Mth1491 CD and fluorescence spectra as function of denaturant concentration.
A. CD scans of the native and unfolded trimer are plotted in red and blue lines, respectively. The difference between the signal of the native and the one of the unfolded protein is plot in green. The data recorded at low wavelengths are not plotted because they correspond to high tension above 600 volts and are thus not reliable. B. Fluorescence monitored scans of the native and unfolded protein are plotted in red and blue lines, respectively. The difference between the signal of the native and the one of the unfolded protein is plot in green.
All spectra were measured at 25°C and in citrate buffer at pH 6. Mth1491 concentration was 1 μM (trimer concentration). The denatured protein was incubated in 6M GdmCl for 15 minutes prior to measuring the spectra.

3.3.2 Mth1491 reversible folding

Mth1491 reversibility of GdmCl denaturation was investigated by measuring denaturation and renaturation curves prepared in the same buffer condition and at the same protein concentration. The curves were incubated for different time periods and measured until they were fully superposeable. However, reversibility of Mth1491 was not observed in the first studies. After trying different buffers, pH and salt conditions, Mth1491 was finally found to unfold reversibily in citrate buffer (20 mM $Na_3C_6H_5O_7$, 450 mM NaCl, 10 mM DTT, 1 mM EDTA, pH 6).

Figure 3.3 illustrates the initial attempt to measure a renaturation curve in acetate buffer (10 mM CH_3COONa, hydrochloric acid (HCl), 450 mM NaCl, 10 mM DTT, 1 mM EDTA, pH 5). This buffer was used by Christendat and co-workers to determine the structure of Mth1491 by x-ray crystallography (Christendat *et al.*,2002). Using these conditions, the signal of the renatured protein corresponded to approximately 41% of the signal of the native protein. This means that the protein does not fully refold and that these conditions do not sufficiently favour the formation of the native protein. Sodium sulphate (Na_2SO_4) is commonly used to stabilise proteins. However, stabilising the protein by adding sodium sulphate at pH 5 did not result in fully reversible unfolding (Figure 3.4.). However, upon switching the pH from 5 to 6 (Figure 3.5), the signal of the renatured protein reaches the signal of the native protein.
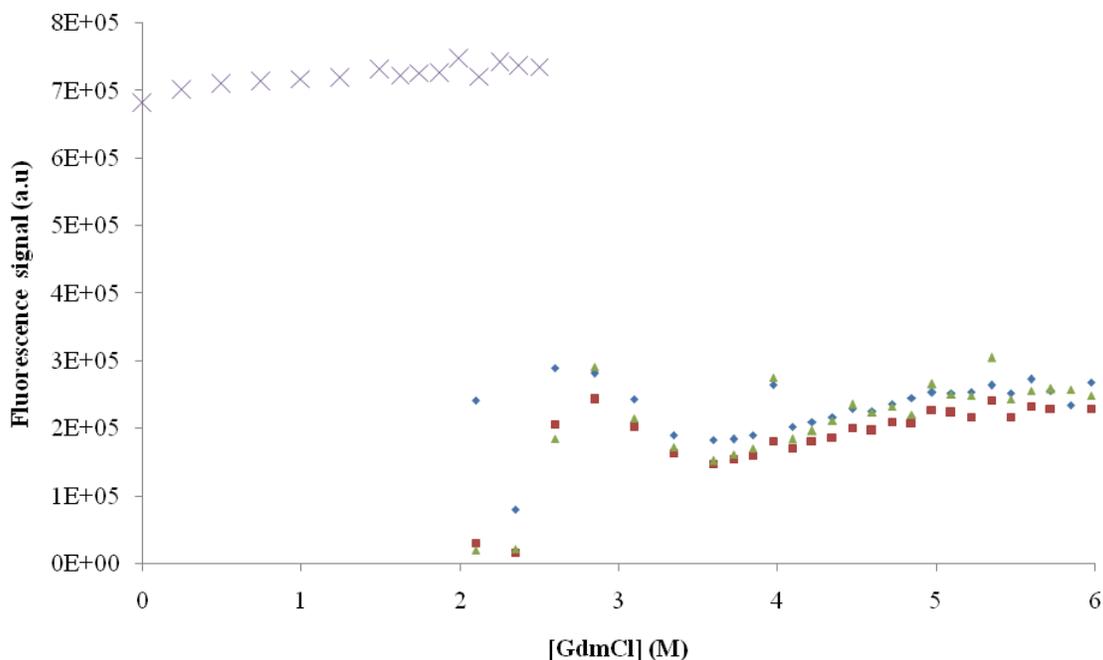
Figure 3.3: Initial attempt to measure the refolding of Mth1491 in acetate buffer.
Refolding curves of Mth1491 were monitored by fluorescence after incubation times of 2 days (blue diamonds), 4 days (red squares) and 1 week (green triangle). The purple crosses are the signal of the native protein, which samples renatured in less than 3 M GdmCl should reach. Tm0979 concentration was 1.5 µM and is expressed in dimer equivalent. The measurements were made at 25°C in acetate buffer (10 mM CH_3COONa, HCl, 450 mM NaCl, 10 mM DTT, 1 mM EDTA, pH 5).
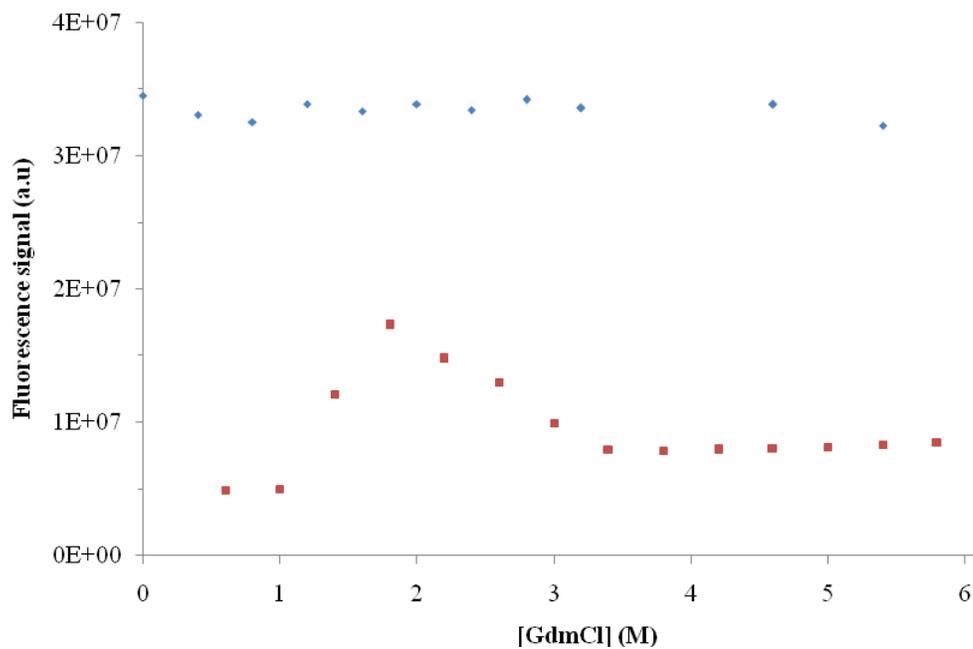
Figure 3.4: Mth1491 renaturation curve measurement attempt in stabilizing conditions at pH 5.
Renaturation (red squares) and denaturation (blue diamonds) curves were measured, after 1 day of incubation, at 1 µM (trimer equivalent) in acetate buffer and using 300 mM Na_2SO_4 as stabilizing agent, at 25°C.
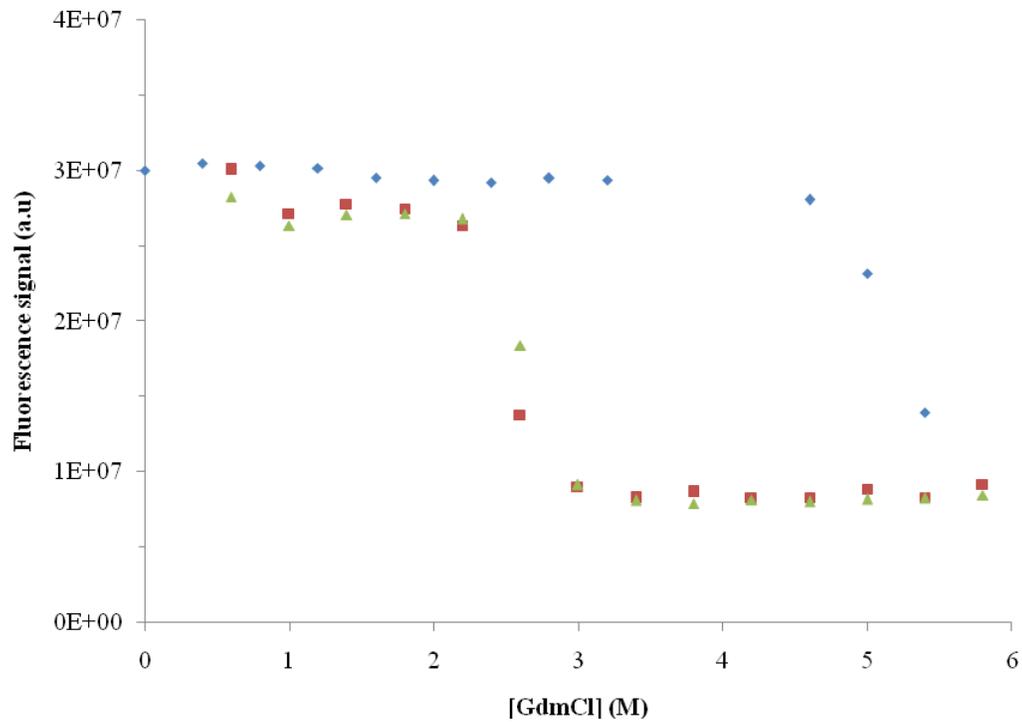
Figure 3.5: Mth1491 reversible folding in acetate buffer at pH 6.
Mth1491 renaturation curves were measured after 15 minutes (red squares) and 1 day (green diamonds), denaturation curves (blue diamonds) measured after 1 day of incubation.
Mth1491 concentration was 1µM (trimer equivalent). The measurements were done at 25°C, in acetate buffer at pH=6 and using Na_2SO_4 as stabilizing agent.

Reversible folding observed at pH 6 can be explained by considering the pI of Mth1491. The theoretical pI of Mth1491 was calculated using the web server Protein Calculator v3.3 (http://www.scripps.edu/~cdputnam/protcalc.html) and is 4.74. This pI estimate assumes all residues have pKa values that are equivalent to the isolated residues and gives a rough value of the experimental pI. At pH = pI, the net charge of a protein is 0 which favours aggregation, especially at increased protein concentration. Mth1491 renaturation curves are prepared by 10-fold dilution of a stock solution of unfolded protein. Mth1491 was incubated for 15minutes in 6 M GdmCl at pH 5 to obtain denatured protein. At pH 5, 10-fold concentrated Mth1491 is likely more prone to aggregation and this decreases reversibility of unfolding. This is the likely reason why refolding of unfolded monomer does not occur when measured at pH 5. At pH 6, Mth1491 overall charge is negative and thus the protein is more soluble in water. Refolding of the protein can thus be observed under these conditions.

3.3.3 Buffer condition optimisation

3.3.3.1 Buffer system

Although Mth1491 was found to fold reversibly at pH 6 in acetate buffer (Figure 3.5), acetic acid has a pka of 4.76, and so has very low buffering capacity at pH 6. The buffer system used for Mth1491 refolding condition should have similar chemical properties and structure as sodium acetate/acetic acid buffer system and should have a pKa close to 6. Succinic (pKa_1=4.16, pKa_2=5.61) and citric ($pKa1 = 3.14$, $pKa2 = 4.77$ and $pKa3 = 6.39$) acid are polyprotic acids and have one of their pKa values close to 6. Therefore denaturation and renaturation curves were prepared at pH 6 using these buffers (Figure 3.5). The main difference between the curves measured in citrate buffer and those measured in succinate buffer remains in the effect of these buffers on the flatness of the baselines. The baselines of the native and unfolded protein need to be well defined in order to fit accurately the curves. The equilibrium characteristics of Mth1491 folding were determined based on the fit of those

curves, so inaccurate baselines would result in inaccurate fit and thus in poorly defined characterisation of the protein folding. When measured in citrate buffer, the native and unfolded baselines of Mth1491 denaturation and renaturation curves were less scattered and therefore better defined. Citrate buffer were therefore used to prepare all subsequent denaturation and renaturation curves.

As shown in Figure 3.6, the midpoint of the renaturation curve measured in citrate buffer is far from the one of the denaturation curve, meaning that the equilibrium time may be long. In order to make this time shorter, the effect of a stabilizing salt on Mth1491 denaturation curve midpoint and equilibrium time was investigated as described in the next section.

3.3.3.2 Salt effect on midpoint

Figure 3.7 shows the effect of salt concentration on the denaturation curve midpoint for Mth1491. In Na_2SO_4 (stabilising conditions), the midpoint of Mth1491 denaturation curve is higher than in presence of NaCl. Moreover, as the concentration of Na_2SO_4 increases, the midpoint of the denaturation curve increases as well. An increase in the midpoint of the denaturation curve is usually characteristic of the stabilisation of the protein as more denaturant is required to make the protein unfold. Therefore, the stabilising effect of Na_2SO_4 is stronger than the one of NaCl and increases with salt concentration. This is consistent with expected relative effects of these salts based on the Hofmeister series. An increase in midpoint was also observed when the curves were monitored in 10% glycerol (data not shown), another stabilizing agent. The equilibrium time of denaturation curve should be as short as possible in order to prevent the formation of irreversible processes such as chemical modification and aggregation. The further apart are the denaturation and renaturation curves, the longer it will likely take for the curves to reach the equilibrium.
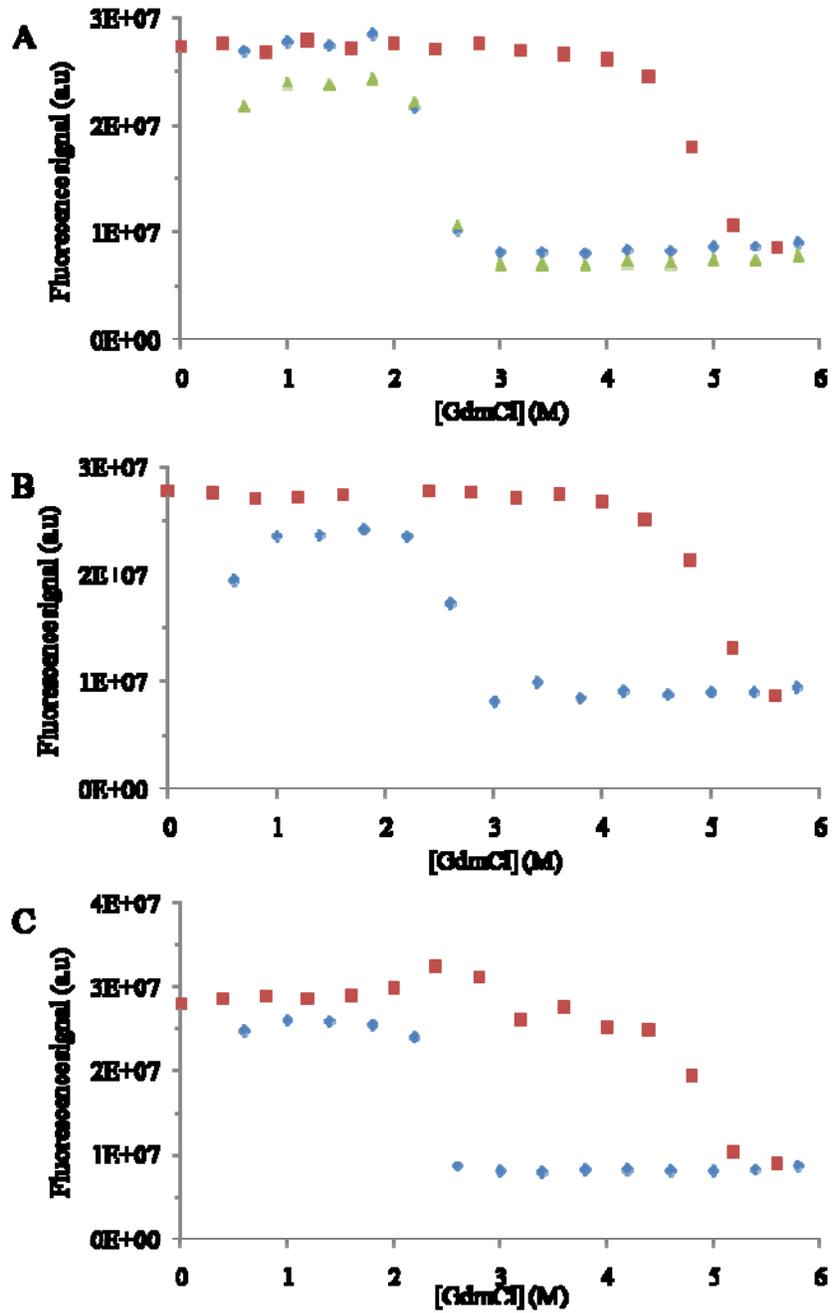
Figure 3.6: Mth1491 renaturation and denaturation measured at pH 6 using different buffer system.
A. Mth1491 renaturation curves were measured in acetate (10 mM) buffer at pH 6 after 15 minutes (blue diamonds) and 1 day (green triangle) of incubation. Mth1491 denaturation curve was measured after 1 day of incubation. B. Mth1491 renaturation (blue diamonds) and denaturation (red squares) curves of Mth1491 measured in citrate buffer (20 mM) after 1 day of incubation at pH 6. C. Mth1491 renaturation (blue diamonds) and denaturation (red squares) curves measured in succinate buffer (25mM) at pH 6 after 1 day of incubation. Samples contained 1 μM (trimer concentration) Mth1491, 10 mM DTT, 1 mM EDTA and 300 mM Na$_2$SO$_4$.

3.3.3.3

Since this difference is increased by these agents, Na_2SO_4 and glycerol were not used for the preparation of subsequent Mth1491 renaturation and denaturation curves. NaCl was therefore used as salt at a final concentration of 450mM.

In summary, Mth1491 folds reversibly at pH 6 in 20 mM citrate buffer with 10 mM DTT, 1mM EDTA and 450mM NaCl.
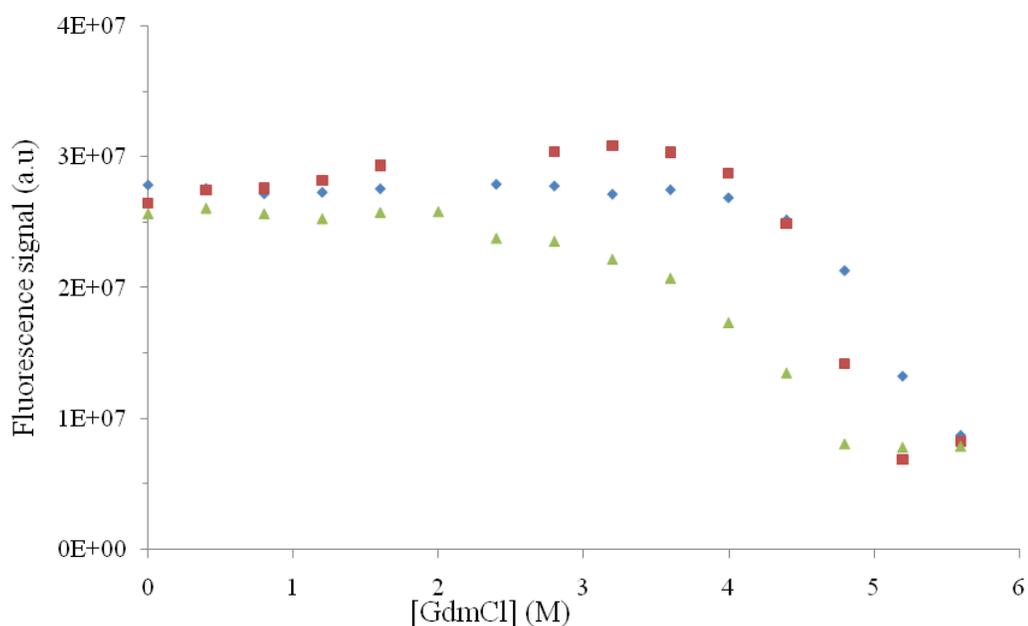
Figure 3.7: Effect of salt on Mth1491 stability.
Mth1491 denaturation curves measured after 1 day of incubation 20 mM citrate buffer with 300 mM Na_2SO_4 (blue diamonds), 200 mM Na_2SO_4 (red squares) and 450 mM (green triangle). Samples contained 1 μM (trimer concentration) Mth1491, 10 mM DTT, 1 mM EDTA and 20 mM sodium citrate at pH 6.

3.3.4 Equilibration time for Mth1491 GdmCl curves

Having explored optimal buffer conditions for maximizing reversibility and minimizing time required for equilibration, denaturation and renaturation curves were prepared and the equilibration time was investigated further. First, Mth1491 denaturation and renaturation curves were measured every day for one week (data not shown). The signal of points of Mth1491 renaturation and denaturation curves kept decreasing as the curves were remeasured. This decrease was probably caused by aggregation of Mth1491 due to the oxidation of the free cysteines (Figure 2.5). In order to avoid aggregation, samples were kept under N2 throughout the equilibration time and were discarded after measurements were made rather than remeasuring the same solution. Figure 3.8 shows Mth1491 denaturation and renaturation curves measured after various equilibration times. The first part of the transition, at a concentration of GdmCl lower than 2.5M, seems to reach equilibrium before 14 days of incubation. The difference between the equilibration time of the first and second part of the transition suggests that Mth1491 folding may not occurs via a simple 2 state mechanism but through a more complex mechanism, such as the formation of a trimeric intermediate (Figure 4.6). The denaturation and renaturation curves are only fully coincident, and hence clearly at equilibrium, after 23 days of sample incubation (Figures 3.8 C and D). However, equilibrium seems to be reached faster for renaturation curves than for denaturation curves. As shown in Figure 3.9, Mth1491 renaturation curves required 4 days of incubation time to reach the equilibrium. Therefore, renaturation curves measured at different protein concentration were used to further investigate Mth1491 folding mechanism (Chapter 4).
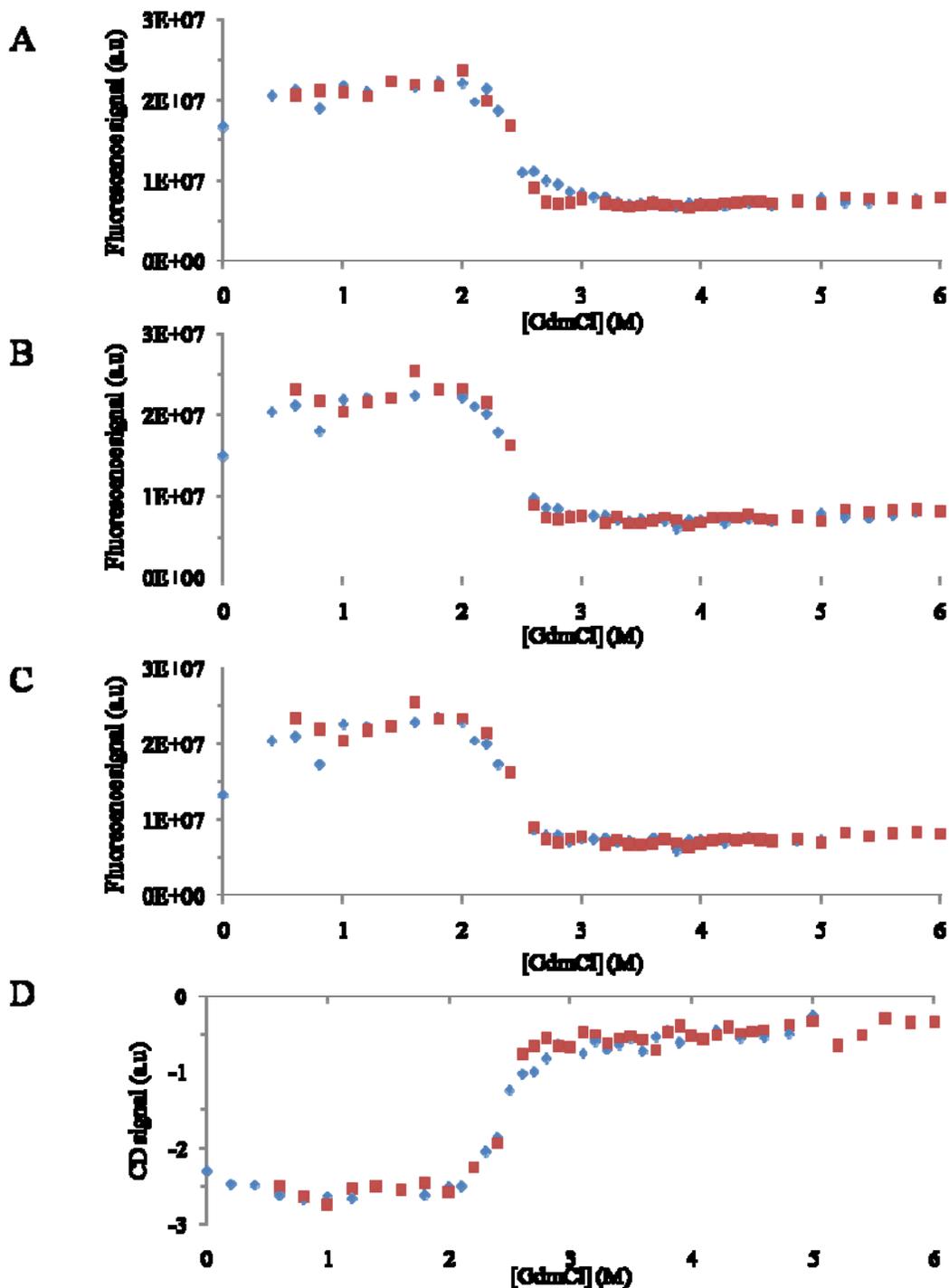
Figure 3.8: Reversible equilibrium curves of Mth1491.
Mth1491 renaturation (red squares) and denaturation (blue diamonds) curves measured by fluorescence after incubation time of 14 days (A), 19 days (B), 23 days (C) and by circular dichroism after an incubation time of 23 days (D). Samples contained 1 μM (trimer concentration) Mth1491, 20 mM sodium citrate, 450mM NaCl, 10mM DTT, and 1mM EDTA, pH 6.
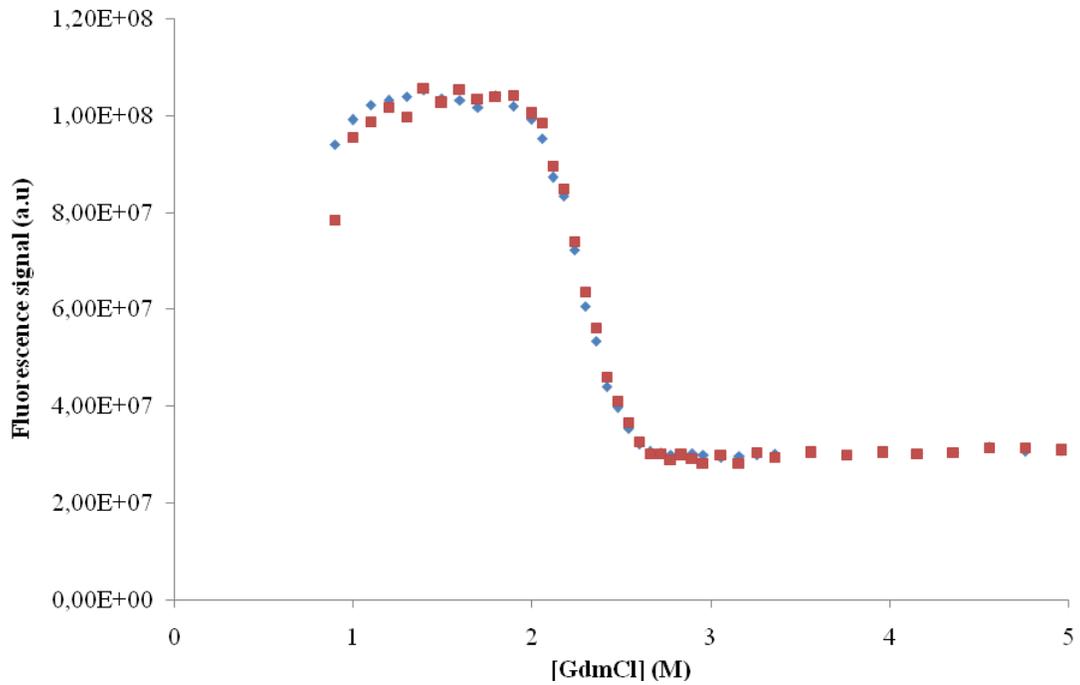
Figure 3.9: Equilibrium time for Mth1491 renaturation curve.
Mth1491 renaturation curves were measured by fluorescence spectroscopy after 3 (blue diamonds) and 4 days (red squares) of incubation. Samples contained 3 μM (trimer concentration) Mth1491, 20 mM sodium citrate, 450 mM NaCl, 10 mM DTT, and 1 mM EDTA, pH 6.

4 Equilibrium studies of Tm0979 and Mth1491 folding

4.1 Introduction

4.1.1 Protein folding

One of the great unsolved problems of science is the prediction of the three dimensional structure of a protein from its amino acid sequences: the 'folding problem' (Fersht,1999). At physiological conditions, proteins undergo reversible transition between their native (N) and unfolded (U) states. In its unfolded or denatured state, the protein makes many interactions with solvent (water). Upon folding, the protein exchanges those noncovalent interactions with others that it makes within itself. Its hydrophobic side chains pack with one another and its hydrogen bond donors and acceptors pair with each other to form hydrogen-bonded networks, (Fersht,1999). The formation of these interactions makes the native state marginally more stable than the unfolded state (Figure 4.1).
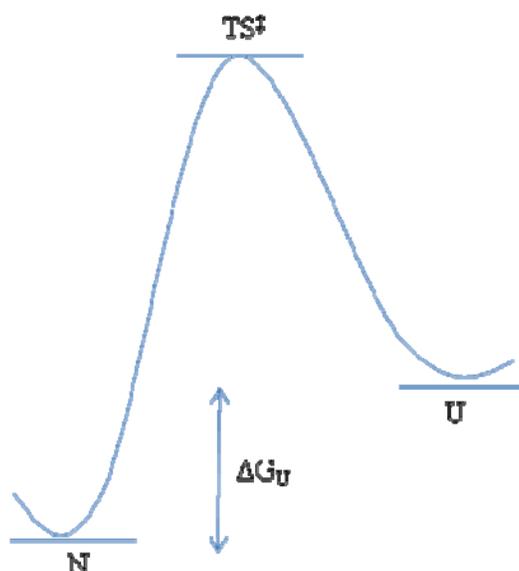
Figure 4.1: Reaction coordinate diagram of protein folding.
N, U and TS‡ are the native and unfolded state.
$\Delta G_U = G_U - G_N$ where ΔG_U represent the free energy of unfolding G_N is the free energy in the native state and G_U is the free energy in the unfolded state.

In this thesis, the stability of Tm0979 dimer and Mth1491 trimer were investigated by denaturant titration monitored by fluorescence spectroscopy and CD (spectroscopic details in section 3.2.2.1). The most commonly used denaturants are guanidinium chloride (GdmCl) and urea. The effects of these denaturants on protein stability are illustrated in Figure 4.2, and can be explained by energetic considerations. Denaturants solubilise all the constituent parts of the protein, i.e the polypeptide backbone as well as amino acid side chains (Fersht,1999). The free energy of transfer of the side chains and polypeptide backbone from water to solution of denaturant is linearly proportional to the concentration of denaturant (Tanford,1968; Tanford,1970). As the denatured state is more exposed to solvent than the native state, the denatured state is preferentially stabilized by denaturant. The free energy of unfolding at any GdmCl concentration $\Delta G_U{}^{GdmCl}$ is expressed by the following equation:

$$\Delta G_U{}^{GdmCl} = \Delta G_U - m_U[\text{GdmCl}] \qquad\qquad (4.1)$$

where $\Delta G_U{}^{GdmCl}$, is the free energy of unfolding at a specific GdmCl concentration, ΔG_U is the free energy of unfolding in water, m_U, is a constant of proportionality and is proportional to the change in accessible surface area upon unfolding. ΔG_U and m_U can be determined by fitting denaturation curves of a protein to the appropriate folding mechanism.
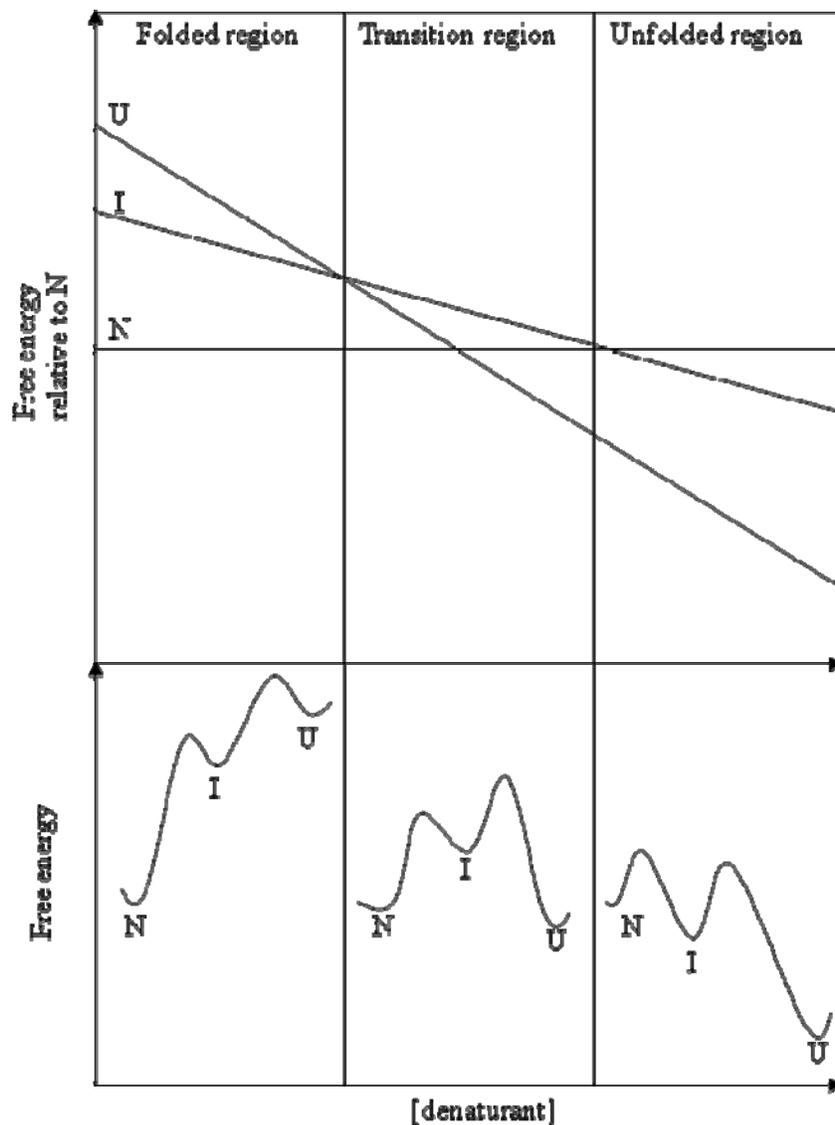
Figure 4.2: Gibbs free energy diagram of native (N), intermediate (I) and unfolded (U) states as a function of denaturant concentration.
As denaturant concentration is increased all states are thought to be stabilized, in proportion to their extent of exposed surface area, i.e. U is stabilized more than I, which is stabilized more than N, which is taken as the reference state here and so shown as having stability not altered by denaturant concentration. The figure is taken from (Fersht A, 1999).

4.1.2 Spectroscopic probes

As referred to the Section 4.1.1, the stability of the Tm0979 dimer and the Mth1491 trimer were investigated by measuring changes in fluorescence and CD signal upon unfolding and refolding. Changes in fluorescence in proteins are dominated by changes in tryptophan (Trp), tyrosine (Tyr) and/or phenylalanine (Phe) solvent exposure or hydrophobic environment whereas changes in CD signals are related to changes in secondary and tertiary structure. Figure 4.3 and 4.4 show the position of the Trp, Tyr and Phe in the Tm0979 dimer and Mth1491 trimer, respectively. Fluorescence is generally dominated by Trp, which typically has higher fluorescence than Phe and Tyr. For both proteins, the Trp residues are solvent exposed, however, the Trp residues in Tm0979 are further from the subunit interface than the ones in the Mth1491 trimer. Therefore, Tm0979 dimer dissociation would not be expected to be characterized by a significant change in fluorescence emission compared to the one due to monomer unfolding. However, as shown in Figure 4.4, Trp located on Mth1491 trimer are relatively close to the trimer interface and changes in Trp emission are more likely to occur for both trimer dissociation and monomer unfolding. The observed changes in Trp fluorescence will be discussed in more detail in sections 4.3 and 4.4.
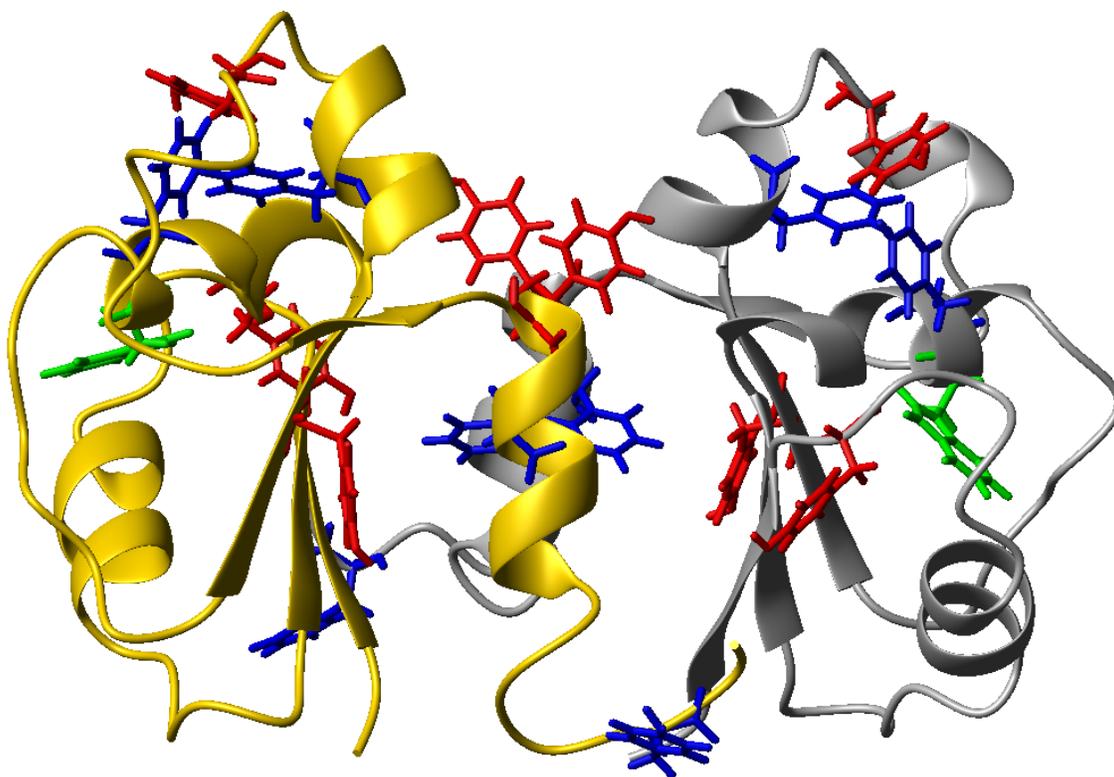
Figure 4.3: Location of tryptophan, phenylalanine and tyrosine residues in Tm0979 dimer.
Each of the Tm0979 subunits is coloured in different colour (gold and grey). The Trp residues are coloured in green lines, the Tyr in red and the Phe in blue. Ribbon diagrams were generated using MolMol (Koradi *et al.*,1996) using PDB accession code 1x9a.
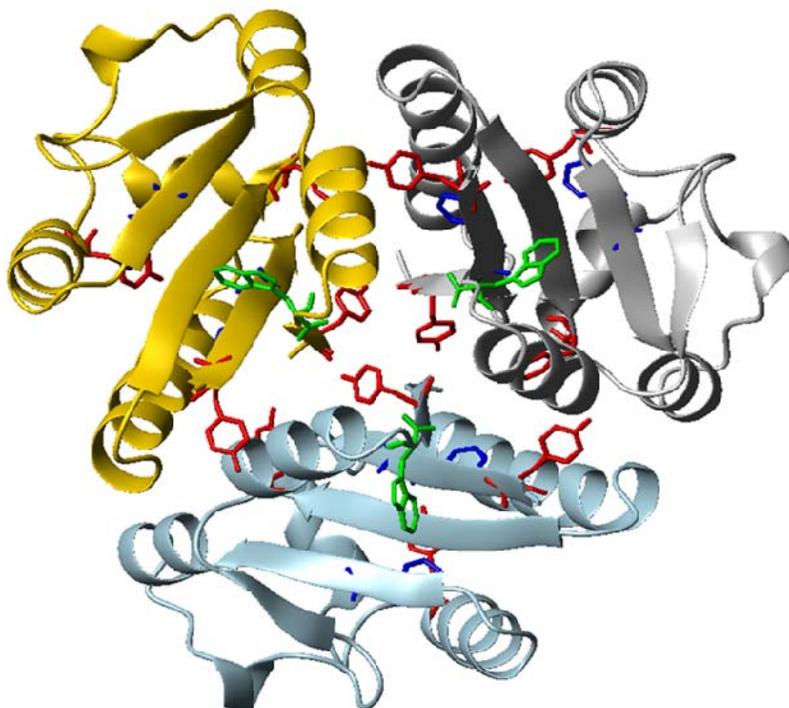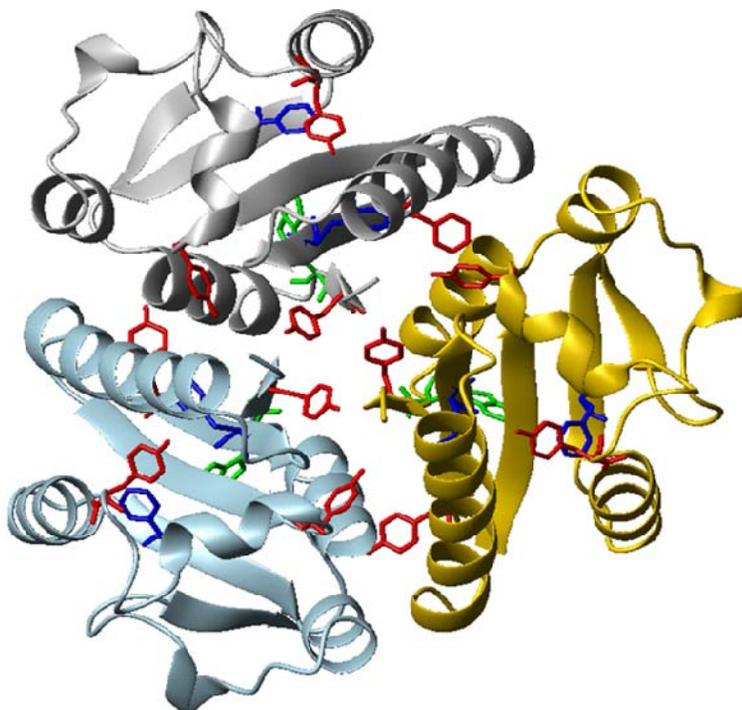
Figure 4.4: Location of tryptophan, phenylalanines and tyrosines residues in Mth1491 trimer.
A. View from one side of the C3 axis of the trimer. B. View from the opposite side of the C3 axis. Each of Mth1491 subunit is coloured in different colour (gold, grey and light blue). The Trp residues are coloured in green lines , the Tyr in red and the Phe in blue. Ribbon diagrams were generated using MolMol (Koradi *et al.*,1996) using PDB accession code 1l1s.

4.2 Methods

4.2.1 Denaturation and renaturation curve sample preparation

See section 3.2.1.

4.2.2 Measurment of denaturation and renaturation curves

See section 3.2.2.

4.2.3 Analysis of the curves

In order to determine the appropriate models for describing Tm0979 and Mth1491 unfolding, denaturation and renaturation curves were fit to different models corresponding to dimer and trimer dissociation without or with the formation of an intermediate. The dimer and trimer unfolding models are summarised in Tables 4.1 and 4.2 and Figures 4.5 and 4.6.

4.2.3.1 Two-state mechanism

For a dimeric or trimeric proteins, equilibrium denaturation curves are dependent upon protein concentration, as illustrated in Figures 4.5 B and 4.6 A. Unlike monomeric proteins which unfold by a two-state mechanism and have a symmetrical sigmoidal transition that is independent of protein concentration (Figure 4.5A), for dimeric proteins the transition is slightly asymmetric, being skewed towards lower denaturant concentration. The shape is unchanged with increasing protein concentration; however, as expected for an oligomeric system, due to mass action (Le Châtelier's principle), the midpoint shifts to higher denaturant concentration. As protein concentration increases, the ratio of dimer to unfolded monomer will increase, in order to satisfy the equilibrium constant. In contrast, for a monomeric protein, because the transition does not involve a change in molecularity, the ratio is independent of protein concentration.

Table 4.1: Dimer equilibrium unfolding mechanisms.

Folding mechanism:	Dimer 2 state (a)	Dimer 3-state Monomer intermediate (b)	Dimer 3-state Dimer intermediate (c)
	$N_2 \underset{K_U}{\rightleftharpoons} 2U$	$N_2 \underset{K_{U1}}{\rightleftharpoons} 2I \underset{K_{U2}}{\rightleftharpoons} 2U$	$N_2 \underset{K_{U1}}{\rightleftharpoons} I_2 \underset{K_{U2}}{\rightleftharpoons} 2U$
Free energy of unfolding	$\Delta G_U \quad 2G_U \quad G_{N2}$	$\Delta G_{U1} \quad 2G_I \quad G_{N2}$ $\Delta G_{U2} = G_U - G_I$ $\Delta G_U = \Delta G_{U1} + 2\Delta G_{U2}$	$\Delta G_{U1} \quad G_{I2} \quad G_{N2}$ $\Delta G_{U1} = 2G_U - G_{I2}$ $\Delta G_U = \Delta G_{U1} + \Delta G_{U2}$
	$\Delta G_U = -RT\ln K_U$	$\Delta G_{U1} = -RT\ln K_{U1}$ $\Delta G_{U2} = -RT\ln K_{U2}$	$\Delta G_{U1} = -RT\ln K_{U1}$ $\Delta G_{U2} = -RT\ln K_{U2}$
Equilibrium constant	$K_U = \dfrac{[U]^2}{[N_2]}$	$K_{U1} = \dfrac{[I]^2}{[N_2]}, \quad K_{U2} = \dfrac{[U]}{[I]}$	$K_{U1} = \dfrac{[I_2]}{[N_2]}, \quad K_{U2} = \dfrac{[U]^2}{[I_2]}$
Total protein concentration	$P = [U]/2 + [N_2]$	$P = [U]/2 + [I]/2 + [N_2]$	$P = [U]/2 + [I_2] + [N_2]$
Fraction protein	$f_u + f_{N2} = 1$ $f_u = [U]/2P$ $f_{N2} = [N_2]/P$	$f_u + f_I + f_{N2} = 1$ $f_u = [U]/2P$ $f_I = [I]/2P$ $f_{N2} = [N_2]/P$	$f_u + f_{I2} + f_{N2} = 1$ $f_u = [U]/2P$ $f_{I2} = [I_2]/P$ $f_{N2} = [N_2]/P$
Equations used for Origin fit	$Y = Y_{N2} + f_U(Y_U - Y_{N2})$	$Y = Y_{N2} + f_I(Y_I - Y_{N2} + K_2(Y_U - Y_{N2}))$	$Y = Y_U + f_{I2}/K_1(Y_{N2} - Y_U + K_1(Y_{I2} - Y_U))$
Equations used for Matlab fit	$4Pf_U^2 + K_U f_U - K = 0$	$4Pf_I^2 + K_{U1}(1+K_{U2})f_U - K_{U1} = 0$	$4P(1+K_{U1})f_U^2 + K_{U1}K_{U2}f_U - K_{U1}K_{U2} = 0$

a, Dimer 2-state mechanism: N_2, $[N_2]$ and f_{N2} are the native dimer, its concentration and its fraction, respectively, U, [U] and f_U are the unfolded monomer, its concentration and its fraction, ΔG_U is the free energy of unfolding, G_U and G_{N2} is the free energy in the unfolded and native state, R is the Gas constant (R= 0.00198 kcal.K^{-1}.mol^{-1}), T is the temperature (T = 298K), K_U is the equilibrium constant of unfolding. b, Dimer 3-state monomer intermediate mechanism: I, [I] and f_I are the monomer intermediate, its concentration and its fraction; ΔG_{U1} is the free energy of dissociation of the native dimer to monomeric intermediates, ΔG_{U2} is the free energy of unfolding of the monomer intermediate, G_I is the free energy of the monomeric intermediate, K_{U1} is the equilibrium constant of dissociation and K_{U2} equilibrium constant of monomer intermediate unfolding. c, Dimer 3-state dimer intermediate mechanism: I_2, $[I_2]$ and f_{I2} are the dimer intermediate, its concentration and its fraction; ΔG_{U1} is the free energy of unfolding of the native dimer to dimeric intermediate, ΔG_{U2} is the free energy of simultaneous dissociation and unfolding of the dimer intermediate, K_{U1} is the equilibrium constant of native dimer unfolding to dimer intermediate and K_{U2} equilibrium constant of dissociation and unfolding of dimer intermediate. All the concentrations are expressed in dimer equivalent. P, total protein concentration, is expressed as dimer equivalent, Y_{N2}, Y_U, Y_I, Y_{I2}, signal of the various states, f_{N2}, f_u, f_I, f_{I2}, the fraction of each species.
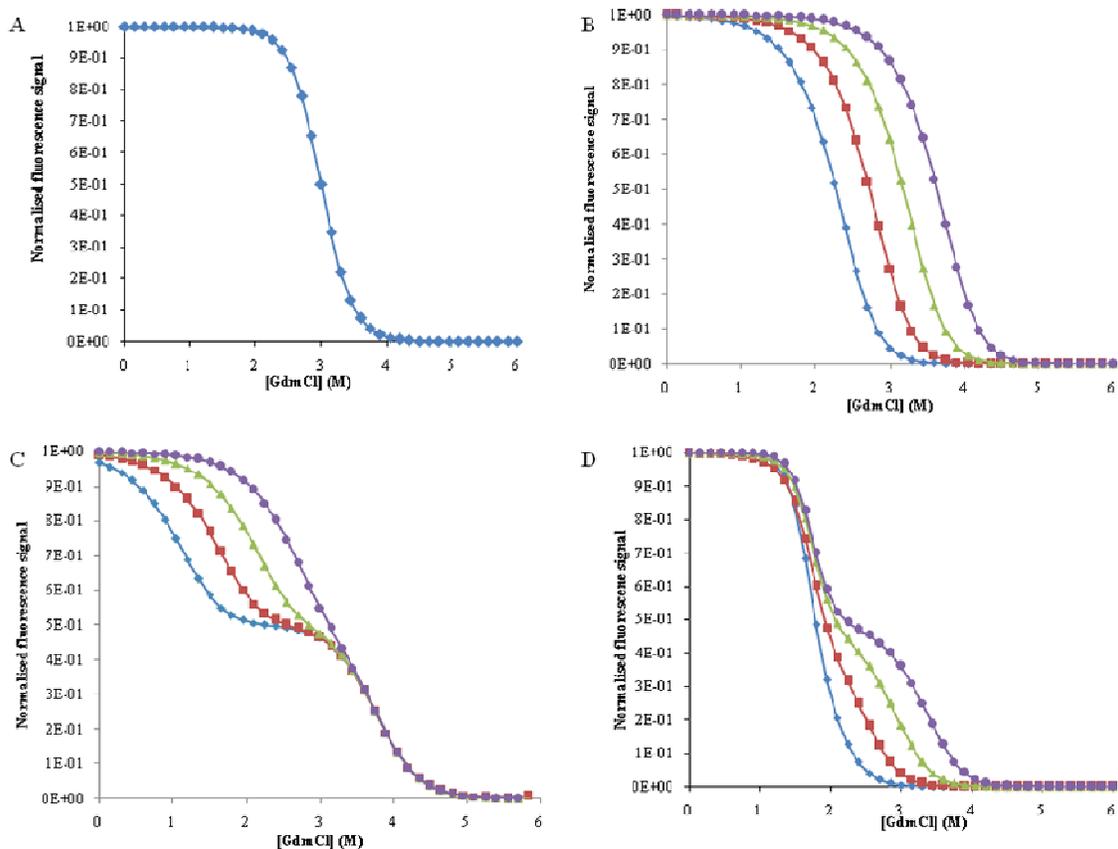
Figure 4.5: Simulation of a monomer 2-state, dimer 2-state and dimer 3-state folding.
A. Simulation of a monomer 2-state folding (ΔG_U = 7.5 kcal.mol^{-1}, m_U = 2.5 kcal.mol^{-1}). B. Simulation of a dimer 2-state folding (ΔG_U = 15 kcal.mol^{-1}, m_U = 3 kcal.mol^{-1}). C. Simulation of a dimer 3-state with monomer intermediate folding (ΔG_{UI} = 11kcal.mol^{-1}, ΔG_{U2} = 7.5kcal.mol^{-1}, m_{UI} = 2.7 kcal.mol^{-1} and m_{UI} = 2 kcal.mol^{-1}). D. Simulation of a dimer 3-state with dimer intermediate folding (ΔG_{UI} = 7 kcal.mol^{-1}, ΔG_{U2} = 14 kcal.mol^{-1}, m_{UI} = 4 kcal.mol^{-1} and m_{UI} = 3 kcal.mol^{-1}).
Signal of the native dimer, intermediate and unfolded protein are fixed to 1, 0.5 and 0, respectively. Curves are simulated at protein concentration of 0.5 µM (blue diamonds), 5 µM (red squares), 50 µM (green triangles) and 500 µM (purple circles). Protein concentration is expressed in dimer equivalent. Simulations calculated using equations of Table 4.1.

4.2.3.2 Three-state mechanism

A 3-state equilibrium denaturation of a dimer or trimer may involve either a monomeric, dimeric or trimeric intermediate (Tables 4.1 and 4.2). In general, an inflection in an equilibrium curve is an indication of the population of an intermediate, both for monomeric and dimeric proteins (Figures 4.5 and 4.6); however, as a general trend, for dimeric (trimeric) proteins a monomeric intermediate is populated more as protein concentration is decreased whereas a dimeric (trimeric) intermediate is populated more as protein concentration is increased. The increase of the corresponding intermediate can be observed by an increase in the inflection in the equilibrium curve as shown in Figures 4.5 C, D and 4.6 B, C. It should be noted, though, that depending on the nature of both transitions, for some protein concentrations only the native dimer and unfolded monomer are significantly populated and as a result equilibrium transitions will appear two-state. It is important for this reason to perform equilibrium curves at a range of protein concentrations in order to define the mechanism of unfolding. Note that when an intermediate is not significantly populated, fitting to a 2-state model will give the same overall ΔG_U and m_U as appropriate 3-state fitting when the intermediate is populated; however, in the former case no information is obtained for the intermediate.

Dimer or trimer unfolding without the formation of intermediates are characterised by ΔG_U and m_U (Equation 4.1). On the other hand, models describing dimer or trimer folding via the formation of intermediates are characterised by ΔG_{U1}, ΔG_{U2}, m_{U1} and m_{U2}. If the intermediate is a monomer, ΔG_{U1} is the free energy of dimer dissociation in water which leads to the formation of the monomeric intermediate, m_{U1} is proportional to the change in solvent exposed area upon this first step, ΔG_{U2} is the free energy of monomer intermediate unfolding in water and m_{U2} is proportional to the change solvent exposed area upon this second step.

Table 4.2: Trimer equilibrium unfolding mechanisms.

Folding mechanism:	Trimer 2 state (a)	Trimer 3-state Monomer intermediate (b)	Trimer 3-state Trimer intermediate (c)
	$N_3 \overset{K_U}{\underset{}{\rightleftarrows}} 3U$	$N_3 \overset{K_{U1}}{\underset{}{\rightleftarrows}} 3I \overset{K_{U2}}{\underset{}{\rightleftarrows}} 3U$	$N_3 \overset{K_{U1}}{\underset{}{\rightleftarrows}} I_3 \overset{K_{U2}}{\underset{}{\rightleftarrows}} 3U$
Free energy of unfolding	$\Delta G_U = 3G_U - G_{N3}$	$\Delta G_{U1} = 3G_I - G_{N3}$ $\Delta G_{U2} = G_U - G_I$ $\Delta G_U = \Delta G_{U1} + 3\Delta G_{U2}$	$\Delta G_{U1} = G_{I3} - G_{N3}$ $\Delta G_{U1} = 3G_U - G_{I3}$ $\Delta G_U = \Delta G_{U1} + \Delta G_{U2}$
	$\Delta G_U = -RT\ln K_U$	$\Delta G_{U1} = -RT\ln K_{U1}$ $\Delta G_{U2} = -RT\ln K_{U2}$	$\Delta G_{U1} = -RT\ln K_{U1}$ $\Delta G_{U2} = -RT\ln K_{U2}$
Equilibrium constant	$K_U = \dfrac{[U]^3}{[N_3]}$	$K_{U1} = \dfrac{[I]^3}{[N_3]}, \quad K_{U2} = \dfrac{[U]}{[I]}$	$K_{U1} = \dfrac{[I_3]}{[N_3]}, \quad K_{U2} = \dfrac{[U]^3}{[I_3]}$
Total protein concentration	$P = [U]/3 + [N_3]$	$P = [U]/3 + [I]/3 + [N_3]$	$P = [U]/3 + [I_3] + [N_3]$
Fraction protein	$f_u + f_{N3} = 1$ $f_u = [U]/3P$ $f_{N3} = [N_3]/P$	$f_u + f_I + f_{N3} = 1$ $f_u = [U]/3P$ $f_I = [I]/3P$ $f_{N3} = [N_3]/P$	$f_u + f_{I3} + f_{N3} = 1$ $f_u = [U]/3P$ $f_{I3} = [I_3]/P$ $f_{N3} = [N_3]/P$
Equations used for Matlab fit	$27P^2f_U^3 + f_U K_U - K_U = 0$	$27P^2f_U^3 + f_U K_{U1}(1 + K_{U2}) - K_{U1};$	$(1 + 1/K_{U1})^3 f_{I3}^3 - 3f_{I3}^2(1 + 1/K_{U1})^2 +$ $f_{I3}(3(1 + K_{U1})^2 + K_{U2}/27P^2) - 1 = 0$

a, Trimer 2-state mechanism: N_3, $[N_3]$ and f_{N3} are the native trimer, its concentration and its fraction, respectively, U, $[U]$ and f_U are the unfolded monomer, its concentration and its fraction, ΔG_U is the free energy of unfolding, G_U and G_{N3} is the free energy in the unfolded and native state, R is the Gas constant (R= 0.00198 kcal.K^{-1}.mol^{-1}), T is the temperature (T = 298K), K_U is the equilibrium constant of unfolding. b, Trimer 3-state monomer intermediate mechanism: I, $[I]$ and f_I are the monomer intermediate, its concentration and its fraction; ΔG_{U1} is the free energy of dissociation of the native trimer to monomeric intermediates, ΔG_{U2} is the free energy of unfolding of the monomer intermediate, G_I is the free energy of the monomeric intermediate, K_{U1} is the equilibrium constant of dissociation and K_{U2} equilibrium constant of monomer intermediate unfolding. c, Trimer 3-state trimer intermediate mechanism: I_3, $[I_3]$ and f_{I3} are the trimer intermediate, its concentration and its fraction; ΔG_{U1} is the free energy of unfolding of the native trimer to trimer intermediate, ΔG_{U2} is the free energy of simultaneous dissociation and unfolding of the trimer intermediate, K_{U1} is the equilibrium constant of native trimer unfolding to trimer intermediate and K_{U2} equilibrium constant of dissociation and unfolding of trimer intermediate. P is the total protein concentration. All the concentrations are expressed in trimer equivalent. Y_{N3}, Y_U, Y_I, Y_{I2}, signal of the various states, f_{N3}, f_U, f_I, f_{I2}, the fraction of each species.
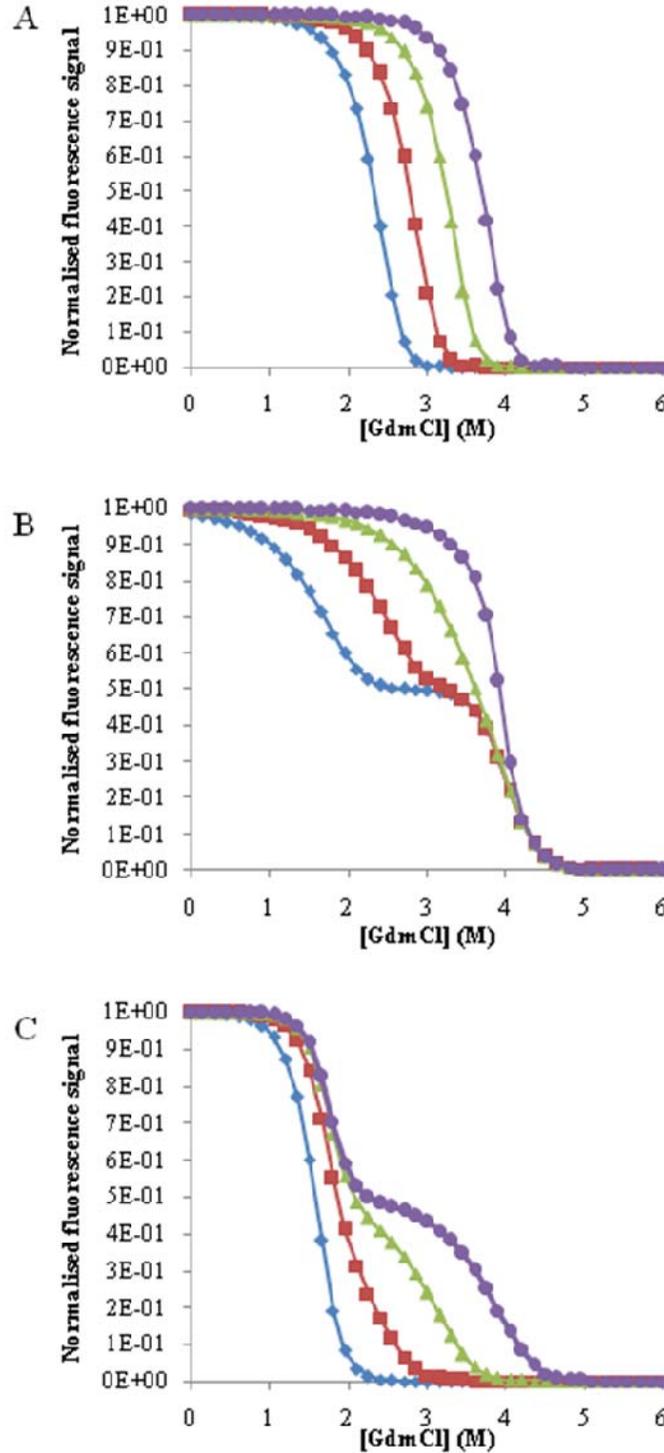
Figure 4.6: Simulation of trimer 2-state and trimer 3-state folding.
A. Simulation of trimer 2-state folding (ΔG_U = 30 kcal.mol^{-1}, m_U = 6 kcal.mol^{-1}). B. Simulation of trimer 3-state with monomer intermediate folding (ΔG_{U1} = 21.5 kcal.mol^{-1}, ΔG_{U2} = 12 kcal.mol^{-1}, m_{U1} = 3.5 kcal.mol^{-1} and m_{U1} = 3 kcal.mol^{-1}). C. Simulation of trimer 3-state with trimer intermediate folding (ΔG_{U1} = 7 kcal.mol^{-1}, ΔG_{U2} = 21 kcal.mol^{-1}, m_{U1} = 4 kcal.mol^{-1} and m_{U1} = 3.5 kcal.mol^{-1}). Signal of the native trimer, intermediate and unfolded protein are fixed to 1, 0.5 and 0, respectively. Curves are simulated at protein concentration of 0.5 µM (blue diamonds), 5 µM (red squares), 50 µM (green triangles) and 500 µM (purple circles). Protein concentration is expressed in trimer equivalent. Simulation calculated using equations of Table 4.2.

60

If the intermediate is an oligomer (dimer or trimer), ΔG_{U1} is the free energy of formation of the oligomeric intermediate in water, m_{U1} is proportional the change in solvent exposed area associated with this step, ΔG_{U2} is the free energy of dissociation of the oligomeric intermediate and the unfolding of the monomers and m_{U2} is proportional to the change in solvent exposed area associated with this second step.

4.2.4 Fitting of the curves

Fitting of the experimental curves was performed using OriginPro7.5 (OriginLab) and MATLAB R2007a according to different models (Tables 4.1 and 4.2). First, the denaturation or renaturation curves measured at different protein concentration were scaled by aligning the native protein and the unfolded protein signals. Then, the folded slope and intercepts (S_f and B_f, respectively) were determined by linear regression by selecting the data at low GdmCl concentration, *i.e.* before the transition region, where the measured signal corresponds to the native protein signal. The unfolded slope and baseline (S_u and B_u, respectively) were determined by linear regression by selecting the data at high GdmCl concentration, *i.e.* after the transition region, where the measured signal corresponds to the unfolded protein signal. The equilibrium curves were then globally fit by fixing B_u, S_u, B_f and S_f to the values determined previously by linear regression. The thermodynamic constants were shared and allowed to be varied. For the models involving an intermediate, its fluorescence was allowed to vary and its slope was fixed to 0.

4.3 Results

4.3.1 Tm0979 unfolding involves formation of a monomeric intermediate

4.3.1.1 Determination of Tm0979 folding mechanism

Tm0979 equilibrium chemical denaturation curves monitored by fluorescence spectroscopy as well as far UV CD were globally fit to various dimer folding models (Table 4.1) with two different fitting programs: OriginPro7.5 and MATLAB R2007a. Tm0979 denaturation curves at various protein concentrations measured by fluorescence and CD were initially fit globally (i.e. simultaneously) to the dimer 2-state mechanism, which describes the concurrent dissociation and unfolding of native dimer to unfolded monomers: $N_2 \leftrightarrow 2U$. (Table 4.3, Figures 4.7). ΔG_U and m_U were shared among the different denaturation curves. As shown in Figure 4.7, the dimer 2-state model does not fit the data properly and other models need to be considered.

Table 4.3: Thermodynamic parameters from fitting of GdmCl Curves of Tm0979.

	ΔG_{U1} (kcal.mol^{-1})	ΔG_{U2} (kcal.mol^{-1})	m_{U1} (kcal.mol.M^{-1})	m_{U2} (kcal.mol.M^{-1})	Signal of the intermediate	Slope of the intermediate
Dimer 2-state						
OriginPro 7.5						
Fluorescence	21.53 ± 1.05	-	4.66 ± 0.34	-	-	-
CD	15.42 ± 0.95	-	3.04 ± 0.31	-	-	-
MATLAB R2007a						
Fluorescence	21.15 ± 1.46	-	4.53 ± 0.46	-	-	-
CD	15.14 ± 1.36	-	2.93 ± 0.44	-	-	-
Dimer 3-state (I)						
MATLAB R2007a						
Fluorescence	7.2	6.2 ± 0.3	0.6 ± 0.2	2.0 ± 0.10	6.46 ± 0.18 x10^7	4.3 ± 1.1x10^6
CD	7.2	2.8 ± 0.7	0.2 ± 0.3	1.2 ± 0.2	-22.6 ± 2.7	-6.7 ± 3.2

For the parameters of the dimer 2-state mechanism, ΔG_{U1} represents the free energy of unfolding, m_{U1} is characteristic to the change in solvent exposed area upon unfolding. For the dimer 3-state mechanism, ΔG_{U1} represents the free energy of dimer dissociation and monomeric intermediate formation, ΔG_{U2} represents the free energy of the monomeric intermediate unfolding, m_{U1} is characteristic to the change in solvent exposed area upon dimer dissociation and monomeric intermediate formation and m_{U2} is characteristic to the change in solvent exposed area upon monomeric intermediate unfolding. The signal of the intermediate corresponds to the fluorescence and CD signal of the monomeric intermediate based on the denaturation curve fit. Note: for dimer-3state fit, $\Delta G_{U1}$was fixed to 7.2 kcal.mol-1 corresponding to Kd value for dimer in water as measured by SEC and DLS (Section 4.3.1.2).
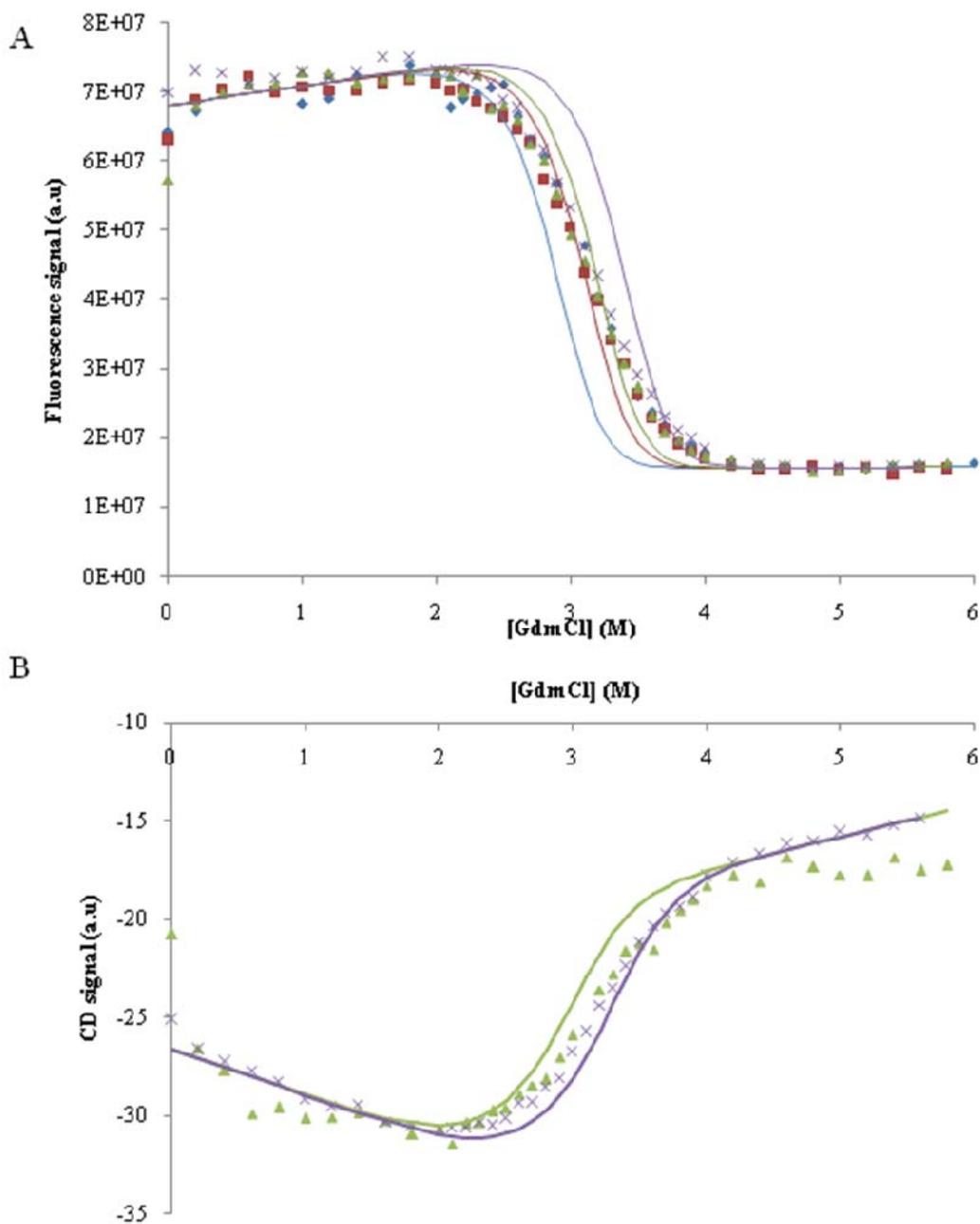
Figure 4.7: OriginPro 7.5 fit of Tm0979 fluorescence- and CD-monitored denaturation curves to dimer 2-state model.
A. Tm0979 denaturation curves measured by fluorescence spectroscopy at 0.5 (blue diamonds), 2.5 (red squares), 5 (green triangles) and 25 μM (violet crosses) protein. B. Tm0979 denaturation curves measured by circular dichroism at 5 (green triangles) and 25 μM (violet crosses) protein. Tm0979 concentrations are given in dimer equivalents. The lines of best fit are plotted in the same colour as the corresponding fitted data. The fitting was performed with OriginPro 7.5 and equations used are those describing the dimer 2-state folding mechanism (Table 4.1). Fitted values are summarized in Table 4.3.

The curves were then fit to the dimer 3-state model with monomer (I) or dimer (I_2) intermediate. Dimer 3-state model with dimer intermediate did not fit well the experimental denaturation curves. Actually, experimental denaturation curves shift slightly with protein concentration at GdmCl concentration lower than 2M, i.e, before the beginning of the transition observed by fluorescence or CD. Based on the simulation of 3-state folding, this trend is very different from those expected for a dimeric intermediate (Figure 4.5, Table 4.1). However, this trend is more likely due to the formation of monomeric intermediate. Actually, dimer 3-state mechanism with monomer intermediate well describes the experimental data. This mechanism consists of the dissociation of the native dimer to form a monomeric intermediate which then unfolds: $N_2 \leftrightarrow 2I \leftrightarrow 2U$.

4.3.1.2 Characterisation of the Tm0979 native dimer and monomeric intermediate

The first step of the mechanism of Tm0979 consists of the dissociation of dimer. Tm0979 dimer was previously characterized by size exclusion chromatography (SEC) and static light scattering experiments (SLS) (Figures 4.8 and 4.9) and dissociation constant (K_d) in absence of denaturant was calculated based on the results of these experiments. Figure 4.8 shows the results of SEC experiments (KA Vassal, E.M Meiering, unpublished experiments). The percentage of dimer was calculated and plotted as a function of protein concentration. The K_d determined based on the fit of this plot corresponds to 2.6 ± 0.5 μM. Figure 4.9 shows the Debye plot resulting from SLS measurements and corresponds to a K_d of 6 μM. Isothermal calorimetry (ITC) experiments were conducted on Tm0979 dimer and results in a K_d of 26 ± 10 μM. However, due to the high uncertainty of this value, K_d determined by ITC was not taken into account. An average value of 5 μM was used as K_d with corresponding value of ΔG_{UI} fixed to 7.2 kcal.mol^{-1}. ΔG_{U2}, m_{UI} and m_{U2} were allowed to vary. These equilibrium constants were shared among the different denaturation curves.
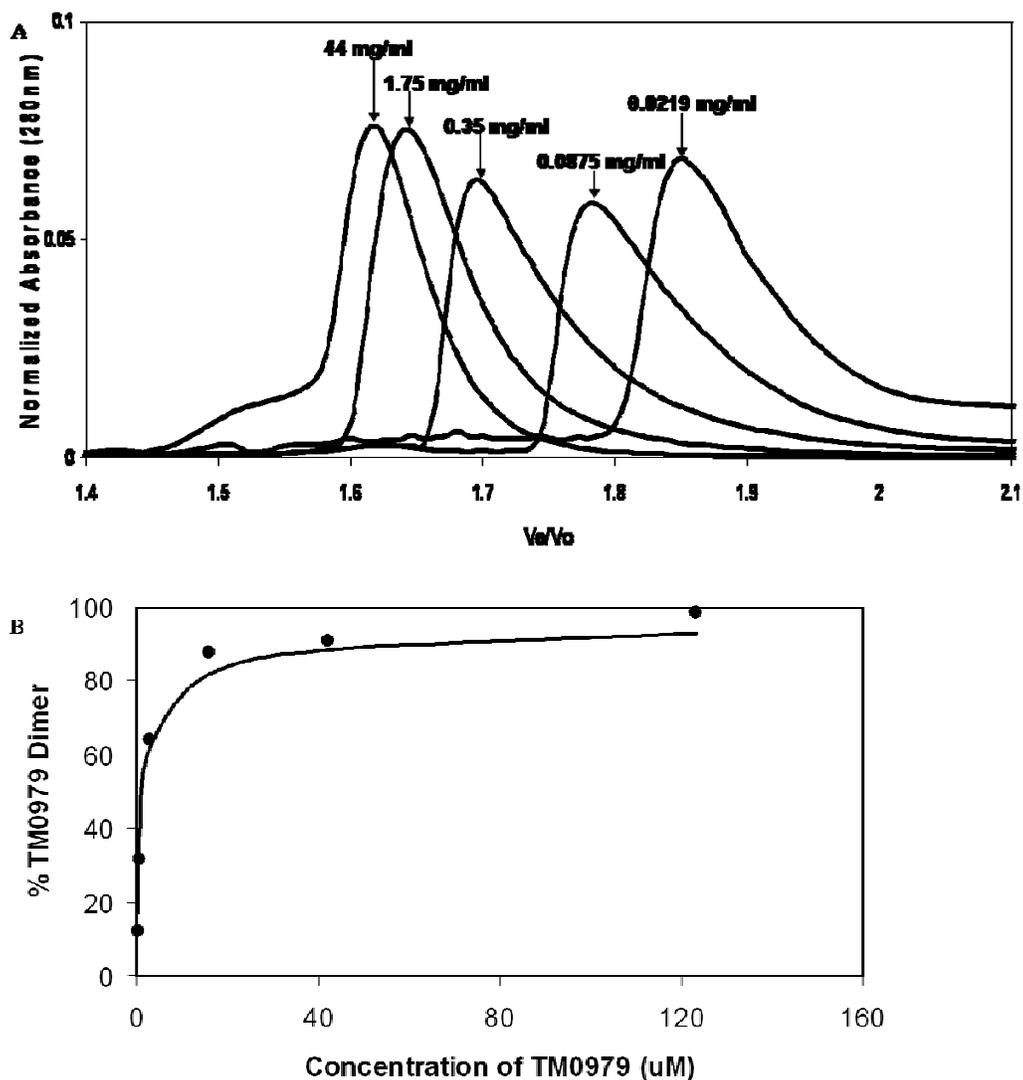
Figure 4.8: Characterisation of Tm0979 dimer by Size exclusion Chromatography (SEC) experiments.
A. Overlaid Elution profiles of Tm0979 at different concentrations. Various concentrations of Tm0979, expressed in the figure as monomer equivalents, were loaded onto a Superdex column. The absorbance values for the different elution profiles were normalized to better facilitate comparison. The position of elution is represented by Ve/Vo, where Ve is the elution volume of the peak and Vo is the void volume as determined by the elution volume of the totally excluded blue dextran. For Tm0979, samples of concentrations 0.0219, 0.0875, 0.35 and 1.75 mg/mL, 2mL were injected onto the column. For the higher concentration Tm0979 samples, 10 mg/mL, 30 mg/mL and 44 mg/mL, 1mL, 0.8 mL and 0.6 mL respectively was injected. All protein samples contained 450 mM NaCl and 25 mM Phosphate at pH 6.5. The samples were eluted with buffer containing 450 mM NaCl 25 mM phosphate at pH 6.5 at a flow rate of 2 ml/min. The elution positions of the various Tm0979 samples were compared to the standards Ovalbumin, beta-lactaglobulin, carbonic anhydrase, trypsin Inhibitor, cytochrome C. The column void volume was determined by the elution of the totally excluded solute blue dextran from the column. B. K_d value for Tm0979 as determined by gel filtration chromatography performed at different concentrations. The above shows a plot of % dimer vs. concentration in dimer equivalents. The data points represent actual data while the curve represents the fit of the data points to equation the following equation:
$\%D = ((8[D] + K_d) - (K_d^2 + 16\ K_d[D])^{1/2})/0.08[D]$, where [D] is the concentration of Tm0979 dimer (with column dilution) and Kd is the dissociation constant.

65
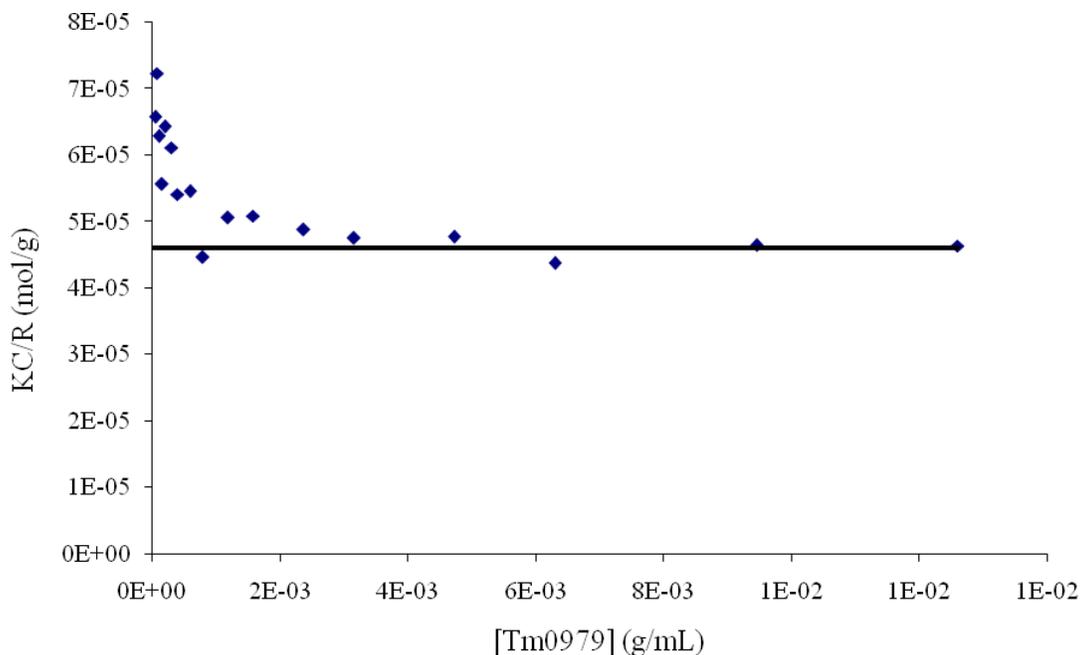
Figure 4.9: Debye Plot of Tm0979. KC/R is plot as a function of protein concentration.

By comparing the uncertainty of those values determined by fluorescence and by CD, the values were better determined by fluorescence experiment.

Actually, the amplitude of the transition measured by CD is quite small and the native and unfolded baselines were not well defined (Figure 4.10). This leads to less accurate fitted values as discussed previously in section 3.3.3.1 and will not be considered in the rest of the discussion. Based on the fit of the experimental data measured by fluorescence, ΔG_{U2} is 6.2 kcal.mol^{-1} and $\Delta G_{Utotal} = \Delta G_{U2} + 2\Delta G_{U2} = 19.6$ kcal.mol^{-1}.The values of m_{U1} and m_{U2} are 0.6 and 2.01 kcal.mol^{-1}.M^{-1}, respectively, which corresponds to a total m_{Utotal}, $m_{unfolding} = m_{U1} + 2m_{U2}$, of 4.6 kcal.mol^{-1}.M^{-1}. m_{U1} is very small compared to m_{U2} meaning that more surface area per monomer is exposed upon monomer unfolding than upon dimer dissociation. Theoretical m_U values were calculated using empirical equations (Myers *et al.*,1995) and are summarized in Table 4.4. In fact, Myers and co-workers gathered equilibrium constants determined for different proteins, monomeric and oligomeric, and observed a linear

66

correlation between the number of amino acids of a protein and the solvent exposed area (ΔASA) upon its folding (ΔASA = (-907 + 93(number of amino acids per protein)). Moreover, they observed a linear correlation between ΔASA and m_U determined by equilibrium studies conducted for those proteins. Different correlation were observed depending on the denaturant used to study the protein unfolding (for GdmCl, m_U = 859 + 0.22ΔASA).

Based on the number of amino acids of Tm0979 dimer and Tm0979 monomer, theorical $m_{unfolding}$ and $m_{monomer\ unfolding}$ were calculated and correspond to 4.2 and 2.4 kcal.mol^{-1}.M^{-1} (Table 4.4), respectively and were compared to m_{Utotal} and m_{U2}. Theoretical m values are very similar to the ones determined from the curve fitting to the dimer 3-state model.

Table 4.4: Interface characteristics of Tm0979 and Mth1491.

Protein	Tm0979	Mth1491
Pdb code	1x9a	1l1s
Number of amino acids per subunit	87	111
$m_{unfolding}$ **(kcal.mol^{-1}.M^{-1})**[†]	4.2	7.5
Interface area (Å)[*]	1077	1438
$m_{dissociation}$ **(kcal.mol^{-1}.M^{-1})**[‡]	1.3	1.8
$m_{monomer\ unfolding}$ **(kcal.mol^{-1}.M^{-1})**[‡]	2.4	2.9
% of monomer buried in the interface[♦]	20	25
% polar residues in the interface per monomer[#]	25	30

[†]$m_{unfolding}$ **calculated based on the following equations (Myers *et al.*,1995):**
ΔASA = (-907 + 93(number of amino acid of the protein))
[*]**Calculation for interface characterisation made using Getarea (http://pauli.utmb.edu/cgi-bin/get_a_form.tcl):**
Interface area = (2 x surface area of monomer – surface area of dimer)/2
[‡]$m_{dissociation}$ **and** $m_{monomer\ unfolding}$ **calculated with the following equation (Myers *et al.*,1995):**
m = 859+0.22 x ΔASA
ΔASAdissociation = 2 x Interface area
ΔASAmonomer unfolding = (surface area of unfolded monomer –surface area of monomer)
Accessible surface area of unfolded monomer were estimated using the following website http://roselab.jhu.edu/utils/unfolded.html [♦]**% monomer buried is the % of surface area buried within the interface per monomer and is calculated as follow: %monomer buried = (Interface area) / (surface area of monomer)**
[#]**% polar residues is the amount of polar residues buried at the interface per monomer and is calculated as follow: % polar residues = (2 x polar surface area of monomer – polar surface area of dimer) / (interface area)**

Actually, m_{U2} for unfolding of the experimental intermediate (2.01 kcal.mol^{-1}.M^{-1}) (Table 4.3) is smaller than the predicted value of $m_{monomer\ unfolding}$ = 2.6 kcal.mol^{-1}.M^{-1}based on the crystal structure (Table 4.4). This also suggests that less buried surface is exposed upon monomeric intermediate unfolding than expected, due to the experimental intermediate having an expanded structure. An alternative explanation for the relatively low value of m_{U2} could be incomplete unfolding of Tm0979 monomeric intermediate due to the presence of residual structure in the unfolded monomers.

Considering the monomeric intermediate fluorescence (6.46 x10^7) and CD signal (-22.6), the values are clearly different from those of the native dimer (fluorescence: 7.3 x10^7,CD: -26.6), but are still closer to native dimer than to unfolded monomer (fluorescence: 1.5 x10^7,CD: -24.3) (Figure 4.10 and Table 4.3).

In summary, Tm0979 equilibrium unfolding is not a simple 2-state process. Actually, at GdmCl concentration below 3M, an inflection is observed and is more pronounced at lower protein concentrations. Moreover, the data are well described by a dimer 3-state mechanism and the formation of a monomeric intermediate is confirmed by previous experiments (SEC, SLS). The characteristics of this intermediate are discussed further in Section 4.4.
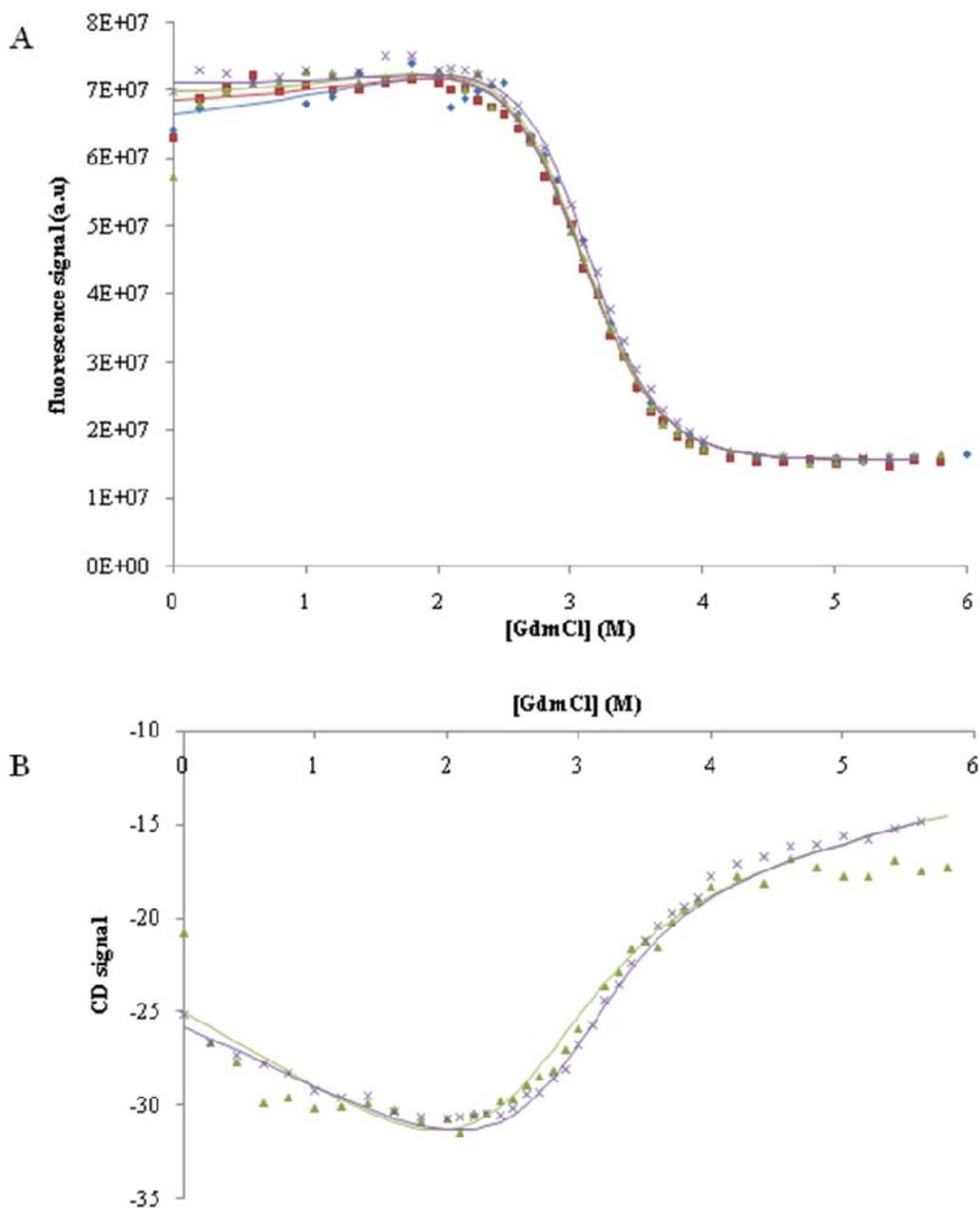
Figure 4.10: MATLAB fit of Tm0979 fluorescence and CD denaturation curves to dimer 3-state model with monomeric intermediate.
A. Tm079 denaturation curves measured by fluorescence spectroscopy at 0.5 (blue diamonds), 2.5 (red squares), 5 (green triangles) and 25 µM (violet crosses) protein. B. Tm0979 denaturation curves measured by CD at 5 (green triangles) and 25 µM (violet crosses) protein. Tm0979 concentration is expressed in dimer equivalent. The fit traces are plotted in the same colour as the data fitted. The fitting was performed with MATLAB R2007a and equations used are those describing the dimer 3 state folding mechanism via the formation of a monomeric intermediate (Table 3.1). Fitted constants are summarized in Table 4.3. The fitted values of the denaturant independent fluorescence and CD of the intermediate are $7.14 \pm 0.04 \times 10^7$ and -22.6, respectively. There are points of inflection at ~ 2.8-3 M GdmCl, indicating formation of monomeric intermediate.

4.3.2 Mth1491 may fold via the formation of a trimer intermediate

Mth1491 equilibrium chemical renaturation curves monitored by fluorescence as well as CD were globally fit to trimer folding model (Table 4.2) using MATLAB R2007a. Limitations in formulation of equations did not allow for data fitting using OriginPro 7.5. The Mth1491 renaturation curves were initially globally fit to the trimer 2-state mechanism, which consists of the concurrent dissociation and unfolding of native trimer to unfolded monomers: $N_3 \leftrightarrow 3U$ (Table 4.5, Figures 4.11). ΔG_U, the Gibbs free energy of unfolding for the trimer and m_U, the dependence of ΔG_U on denaturant concentration, were shared while simultaneously fitting the different denaturation curves. As shown in Figure 4.11, the trimer 2-state model fit the data quite well. However, there is a consistent deviation between the experimental data and the lines of best fit for the last part of the transition, for GdmCl concentration higher than 2.2M. Actually, the experimental data have a higher fluorescence signal than the one predicted by the fit. This deviation from the fit is more pronounced at high protein concentration.

Table 4.5: Thermodynamic parameters from GdmCl Curves of Mth1491.

	ΔG_{UI} (kcal.mol^{-1})	ΔG_{U2} (kcal.mol^{-1})	m_{UI} (kcal.mol.M^{-1})	m_{U2} (kcal.mol.M^{-1})	Signal of the intermediate
Trimer 2-state MATLAB R2007a					
Fluorescence	45.06 ± 1.10	-	13.51 ± 0.50	-	-
CD	37.89 ± 1.57	-	10.24 ± 0.68	-	-
Trimer 3-state (I$_3$) MATLAB R2007a					
Fluorescence	17.18 ± 0.58	26.36 ± 1.29	7.27 ± 0.22	5.45 ± 0.63	4.09 ± 0.17 x10^7
CD	17.21 ± 4.42	25.21 ± 5.59	7.25 ± 1.75	5.02 ± 2.17	-12

For the parameters of the trimer 2-state mechanism, ΔG_{U1} represents the free energy of unfolding, m_{U1} is related to the change in solvent exposed area upon unfolding. For the trimer 3-state mechanism, ΔG_{U1} represents the free energy of trimer unfolding to a trimeric intermediate, ΔG_{U2} represents the free energy of dissociation and unfolding of the trimeric intermediate to three unfolded monomers, m_{U1} is characteristic to the change in solvent exposed area upon the trimeric intermediate formation from Mth1491 native trimer and m_{U2} is related to the change in solvent exposed area upon the dissociation and unfolding of the trimeric intermediate. The signal of the intermediate corresponds to the fluorescence and CD signal of the trimeric intermediate based on the fit of the denaturation curves.
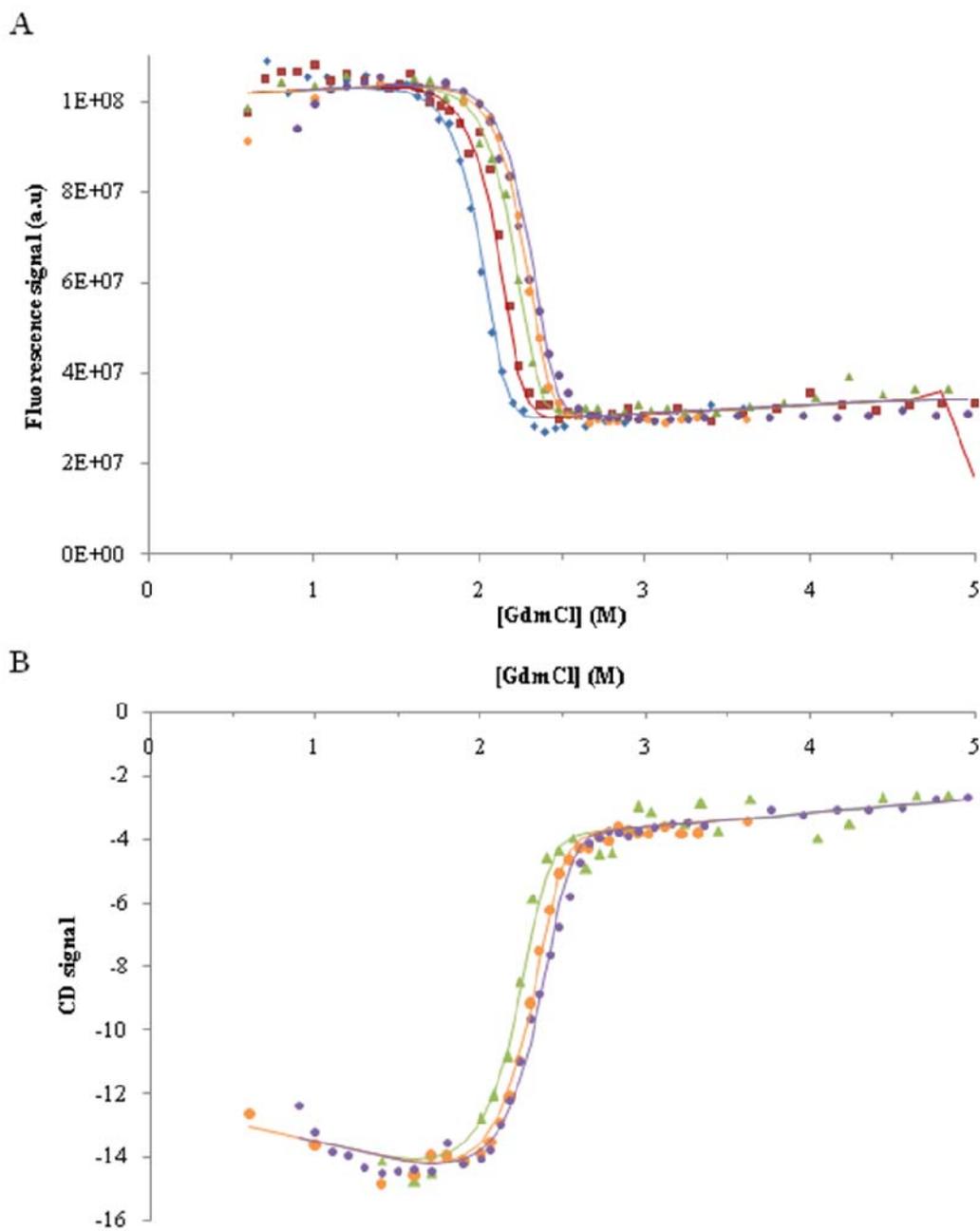
Figure 4.11: MATLAB R2007a fit of Mth1491 fluorescence and CD denaturation curves to a trimer 2-state model.
A. Mth1491 renaturation curves were measured by fluorescence spectroscopy at 0.1 μM (blue diamonds), 0.33 μM (red squares), 1 μM (green triangles), 2 μM (orange circles), 3 μM (violet circles) protein. B. Mth1491 renaturation curves were measured by circular dichroism at 1 μM (green triangles), 2 μM (orange circles), 3 μM (violet circles) protein. Mth1491 concentration is expressed in trimer equivalent. The fit traces are plotted in the same colour as the data fitted. The fitting was performed with MATLAB R2007a and equations used are those describing the trimer 2-state folding mechanism (Table 4.5).

By comparing this observation with the simulation of trimer folding (Table 4.2 and Figure 4.6), the inflection observed at GdmCl concentration higher than 2.2M suggests the formation of a trimeric intermediate. Another explanation would be that the renaturation curves are not quite at equilibrium and that the deviation is due more to an equilibration time problem than to the formation of a trimeric intermediate.

The curves were also fit to the trimer 3-state model with monomer (I) or trimer (I_3) intermediate. The trimer 3-state model with monomer intermediate fit the data to a comparable extent as the trimer 2-state model for the last part of the transition, i.e. there were still some systematic deviations between the data and the fits (fits not shown). However, the trimer 3-state with trimeric intermediate model fit the whole curves very well. This mechanism involves the formation of the trimeric intermediate from the native trimer, and subsequent concurrent dissociation and unfolding of the trimeric intermediate to form unfolded monomers: $N_3 \leftrightarrow I_3 \leftrightarrow 3U$ (Figures 4.12 and Table 4.5). ΔG_{U1} is the Gibbs free energy of unfolding of the native trimer into trimeric intermediate, ΔG_{U2}, the Gibbs free energy of trimeric intermediate dissociation and unfolding to form three unfolded monomers, m_{U1} and m_{U2} are the dependence of ΔG_{U1} and ΔG_{U2}, respectively, on denaturant concentration. The average fitted values of ΔG_{U1} and ΔG_{U2} are 17.20 and 25.79 kcal.mol^{-1}, respectively, which corresponds to an overall Gibbs free energy of unfolding of 43 kcal.mol^{-1} ($\Delta G_{Utotal} = \Delta G_1 + \Delta G_2$). The average values of m_{U1} and m_{U2} are 7.26 and 5.24 kcal.mol^{-1}.M^{-1}, corresponding to an overall m_{Utotal} of 12.5 kcal.mol^{-1}.M^{-1} ($m_{unfolding} = m_1 + m_2$). This m_{Utotal} value is very high compared to the theoretical one based on the empirical equation (7 kcal.mol^{-1}.M^{-1}). Moreover, the ΔG_{Utotal} is a very high value compared to the overall stability of other trimeric protein. Therefore, further experiments are required to confirm the formation of a trimeric intermediate, such as cross-linking experiment at different GdmCl concentrations and different protein concentrations as discussed in Section 4.4.
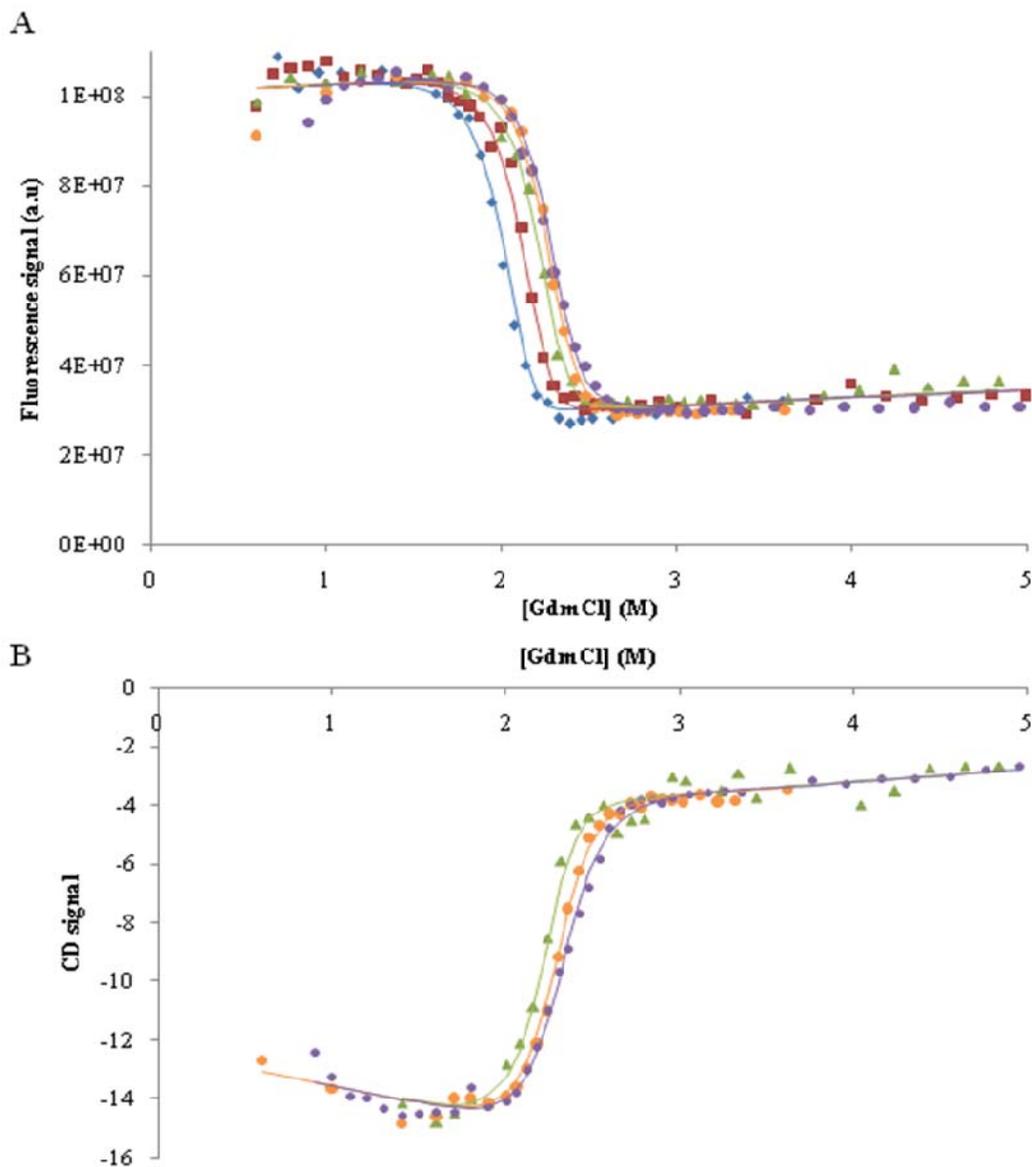
Figure 4.12: Matlab2007a fit of Mth1491 fluorescence and CD denaturation curves fitted to trimer 3 state model (trimeric intermediate).
A. Mth1491 renaturation curves were measured by fluorescence spectroscopy at 0.1 μM (blue diamonds), 0.33 μM (red squares), 1 μM (green triangles), 2 μM (orange circles), 3 μM (violet circles). Mth1491 concentration is expressed in trimer equivalent. The fit traces are plotted in the same colour as the data fitted. B. Mth1491 renaturation curves were measured by circular dichroism at 1 μM (green triangles), 2 μM (orange circles), 3 μM (violet circles). Mth1491 concentration is expressed in trimer equivalent. The fit traces are plotted in the same colour as the data fitted. The fitting was performed with MATLAB R2007a and equations used are those describing the trimer 3 state folding mechanism via the formation of a trimeric intermediate (Table 4.2). Fitted values are summarized in Table 4.5.
Intermediate signals:
Fluorescence: $4.09 \pm 0.17 \times 10^7$ this is really very low i.e. much closer to signal for U than for N_3, consistent with quite big ΔG_{Utotal} and $m_{U\,total}$, also with fluorophore being in the interface.
CD: -12, this signal is much closer to N_3 so the trimeric intermediate still retains most of secondary structure, consistent with it still having significant stability.

73

4.4 Discussion

4.4.1 Tm0979 folding mechanism

4.4.1.1 Tm0979 monomeric intermediate

Tm0979 folds from unfolded monomers to native dimer via the formation of monomeric intermediate. Formation of the monomeric intermediate is supported by previous studies in the lab (SEC, SLS, ITC, G. Meglei, KA Vassal, EM Meiering, unpublished results). Those experiments confirmed existence of an equilibrium between dimer and monomer that shifted with protein concentration, according to a K_d of ~5μM. These results show that Tm0979 has the capacity to form a stable monomer which is consistent with its 3-state folding mechanism

The Gibbs free energy of unfolding of the monomeric intermediate, ΔG_{U2}, is 6.2 kcal.mol^{-1}, with an m_{U2} of 2.01 kcal.mol^{-1} (Table 4.3). When compared to other monomeric protein stabilities (Jackson,1998), the Tm0979 monomer intermediate is quite stable. The equilibrium between the native dimer and the monomeric intermediate suggests that the monomeric intermediate should be stable enough to be detected in the equilibrium curves and that the stability of the monomer may be actually comparable to the stabilities of other monomeric proteins. The percentage of buried monomer at the interface is smaller for Tm0979 than for dimeric proteins that unfolds via a 2-state mechanism (Figure 4.13 A, Appendix A.3: Thermodynamic and kinetic parameters characteristic of chemical induced unfolding of dimeric proteins described by a two-state transition). Actually, monomeric species of these 2-state dimeric proteins are not stable on their own as the interface area exposed to the solvent upon dissociation is bigger; extensive exposure of hydrophobic residues to the solvent is highly destabilising.
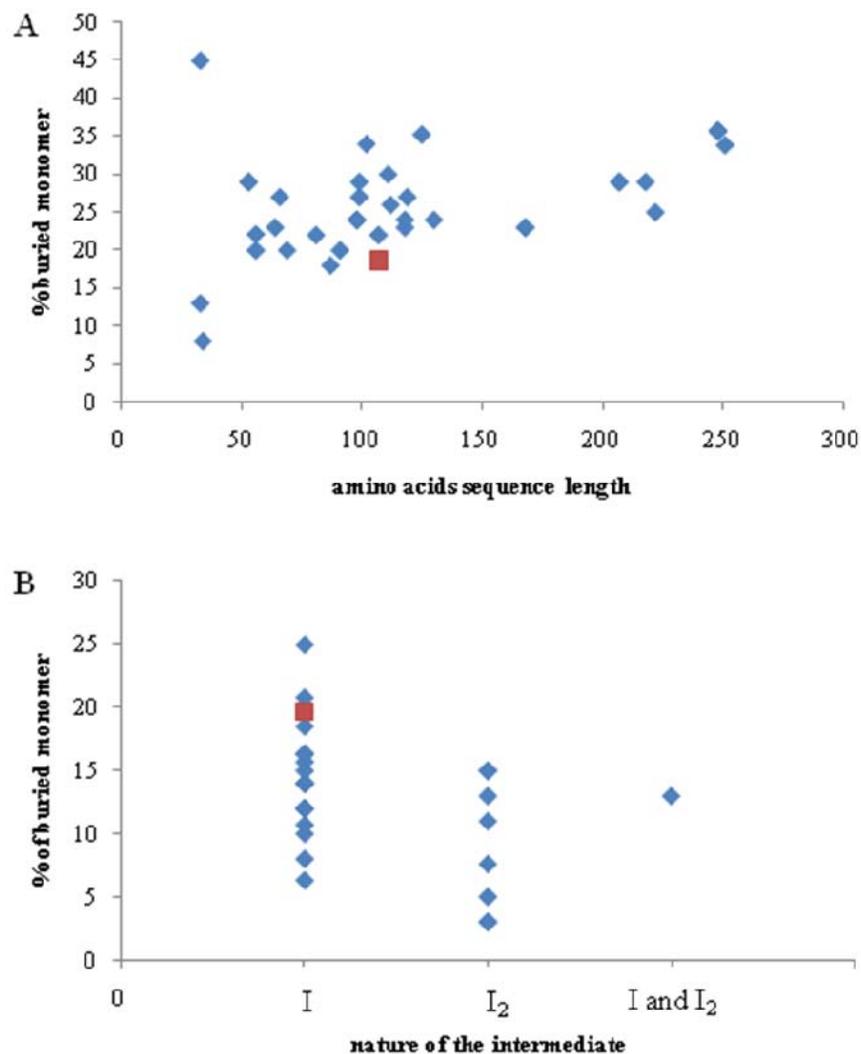
Figure 4.13: Comparison of the percentage of buried monomer surface area buried at the interface for Tm0979 and for dimeric proteins that unfold via a 2-state or 3-state mechanism.
A. % of buried monomer surface at the interface for dimeric proteins that fold via a 2-state mechanism plotted as a function of monomer chain length (blue diamonds). B. % of buried monomer at the interface for dimeric proteins which fold via a 3-state mechanism plotted as a function of the nature of the intermediate (blue diamonds). % buried monomer of Tm0979 % is represented by a red square. I and I_2 represent a monomeric and dimeric intermediate, respectively. % of buried monomer at the interface of dimeric proteins is gathered in Appendix 3 and 4.

These intermediates are thus not detectable upon unfolding. However, Tm0979 monomer is more stable in solution as less buried surface is exposed upon dimer dissociation and monomeric intermediate formation and can thus be detected upon unfolding. Moreover, the percentage of buried monomer at the dimer interface is similar to the percentage buried at the interface of dimeric proteins that unfold via the formation of a monomeric intermediate (Figure 4.13 B, Appendix A.4: Thermodynamic and kinetic parameters characteristic of chemical induced unfolding of dimeric proteins described by a three-state transition). This similarity is consistent with the observation that Tm0979 unfolds through a dimer 3-state mechanism with monomer intermediate.

4.4.2 Mth1491 folding mechanism

4.4.2.1 *Mth1491 folds via a 2-state mechanism or via the formation of a trimeric intermediate*

Mth1491 is proposed to fold via a 2-state or 3-state mechanism with a trimer intermediate. Further experiments need to be done in order to verify the formation of a trimeric intermediate and are discussed in Chapter 6. Whereas folding of many dimeric proteins has been analyzed, only a handful of studies have been performed for trimeric proteins. It is interesting to compare the results obtained for Mth1491 with those obtained for other trimers. The overall stability of Mth1491 is 43 kcal.mol^{-1} compared to 26 kcal.mol^{-1} for the trimer intermediate. The native trimer is quite stable compared to other studied trimers such as Lpp-56 (Bjelic *et al.*,2006) and SIV trimer (Marti *et al.*,2004) ($\Delta G_U \sim 20$kcal.mol^{-1}), which contain 56 and 70 amino acids per monomer, respectively, compared with 111 amino acids for Mth1491. However, the stability of Mth1491 trimer intermediate and those trimeric proteins are quite similar ($\Delta G_{U2} = 25.79$kcal.mol^{-1}). This suggests that due to its relatively larger size, Mth1491 may not be able to fold directly from unfolded monomer to native compact trimer.

4.4.2.2 *Necessity of a trimer intermediate formation for the protein to fold properly*

The formation of a quite stable trimer intermediate may be a necessary step to organise Mth1491 trimeric interface. Each monomer may rearrange its tertiary structure upon the second step of the folding mechanism to form the native trimer. The folding mechanism of oligomeric proteins tends to involve one or more intermediates upon unfolding as the length of their amino acids chain increase and as their oligomeric state increases (Rumfeldt, JA, dimer review paper to be published). *In vivo*, chaperones are often involved in facilitating the folding of big complexes. The formation of oligomeric intermediates usually helps the protein to fold in its proper native state *in vitro*, as for large dimeric proteins, such as histones, bacterial luciferase, organophosphorus hydrolase and SecA (Doyle *et al.*,2000) (Appendix A.4: Thermodynamic and kinetic parameters characteristic of chemical induced unfolding of dimeric proteins described by a three-state transition). Mth1491 proper folding may require the formation of a trimeric intermediate to drive the formation of the more complex and stable native trimer.

5 Kinetic studies of Tm0979 and Mth1491 folding

5.1 Introduction

In pioneering studies, Christian Anfinsen showed that the small proteins ribonuclease A and staphylococcal nuclease could be reversibly denatured *in vitro* (Anfinsen,1973). Upon removal of denaturant, such as GdmCl, many proteins spontaneously refold to their folded native structures after denaturation (Fersht,1999). The amino acid sequences of proteins encode the information on how to attain the final folded structure.

The folding kinetics of monomeric proteins have the potential to be extremely complex due to processes such as proline and disulfide isomerization, intra- and inter-molecular misfolding as well as on and off pathway intermediates. Oligomer folding can involve these phenomena as well as the additional complexity of the protein concentration dependent bimolecular association step. For Tm0979 and Mth1491, we can assume that there are no complications due to disulfide bond formation since Tm0979 contains no cysteines, while Mth1491 has no native disulfide bonds and the folding experiments are performed under reducing conditions. Both proteins contain only transprolines (two prolines per monomer for Tm0979 and one proline per monomer for Mth1491); in such cases proline isomerization reactions, which are generally much slower than the major folding reaction, can be neglected. Having characterized the reversible equilibrium unfolding of both proteins in Chapter 3, in this chapter, the kinetics of folding and unfolding of Tm0979 and Mth1491 folding are described. The results can be interpreted in a relatively simple way that takes into account protein concentration dependence.

In the following sections of the introduction, the terminology, equations and observed results for the kinetics of folding and unfolding of monomeric, dimeric and trimeric proteins are reviewed. Relevant schemes and equations are summarized in Table 5.1.

Table 5.1: kinetic models for monomer, dimer and trimer 2-state folding.

Folding model:	Monomer folding	Dimer folding	Trimer folding
	$N \underset{k_f}{\overset{k_u}{\rightleftharpoons}} U$	$N_2 \underset{k_f}{\overset{k_u}{\rightleftharpoons}} 2U$	$N_3 \underset{k_f}{\overset{k_u}{\rightleftharpoons}} 3U$
Rate of unfolding	$\dfrac{dU}{dt} = k_u[N] - k_f[U]$	$\dfrac{dU}{dt} = k_u[N_2] - k_f[U]^2$	$\dfrac{dU}{dt} = k_u[N_3] - k_f[U]^3$
Rate of folding	$\dfrac{dN}{dt} = k_f[U] - k_u[N]$	$\dfrac{dN_2}{dt} = k_f[U]^2 - k_u[N_2]$	$\dfrac{dN_3}{dt} = k_f[U]^3 - k_u[N_3]$
If the protein goes from unfolded conditions to strongly native condition, i.e., $[U]_0 \gg [N]_0$ and $k_f \gg k_u$	Change in unfolded protein concentration with time: $[U] = [U]_0 \exp(-k_f t)$ (5.1)	Change in unfolded protein concentration with time: $[U] = \dfrac{[U]_0}{(1+2k_f[U]_0 t)}$ (5.2)	Change in unfolded protein concentration with time: $[U] = \dfrac{[U]_0}{\sqrt{(1+2k_f[U]_0 t)}}$ (5.3)
If the protein goes from native conditions to strongly denatured condition, i.e., $[N]_0 \gg [U]_0$ and $k_u \gg k_f$	Change in native monomer concentration with time: $[N] = [N]_0 \exp(-k_u t)$ (5.4)	Change in native dimer concentration with time: $[N_2] = [N_2]_0 \exp(-k_u t)$ (5.5)	Change in native trimer concentration with time: $[N_3] = [N_3]_0 \exp(-k_u t)$ (5.6)

U, N, N_2, N_3 represent the unfolded protein, native monomer, dimer and trimer, respectively. ku and kf are the rate constant of folding and unfolding, respectively. [U], [N], $[N_2]$, $[N_3]$ are the concentration of unfolded protein, native monomer, native dimer and native trimer at a time t, $[U]_0$, $[N]_0$, $[N_2]_0$, $[N_3]_0$ are the concentration of unfolded protein, native monomer, native dimer and native trimer at t=0

5.1.1 Protein folding

In reversible protein folding, the protein travels back and forth between the native (N) and unfolded (U) states over a free energy barrier, the highest point corresponding to the transition state ($TS^{\ddagger}$) as described in an energy diagram (Figure 5.1). In the transition state, noncovalent interactions are in the process of being made and broken making this state very unstable and not directly detectable. For two-state folding, the rate of refolding and unfolding is proportional to the free energy difference between U and $TS^{\ddagger}$ and between N and $TS^{\ddagger}$, respectively. Intermediates may also form; these are generally difficult to trap because of their unstable nature, but they may be detected during kinetics studies when intermediates can become transiently populated (see section 5.1.3).
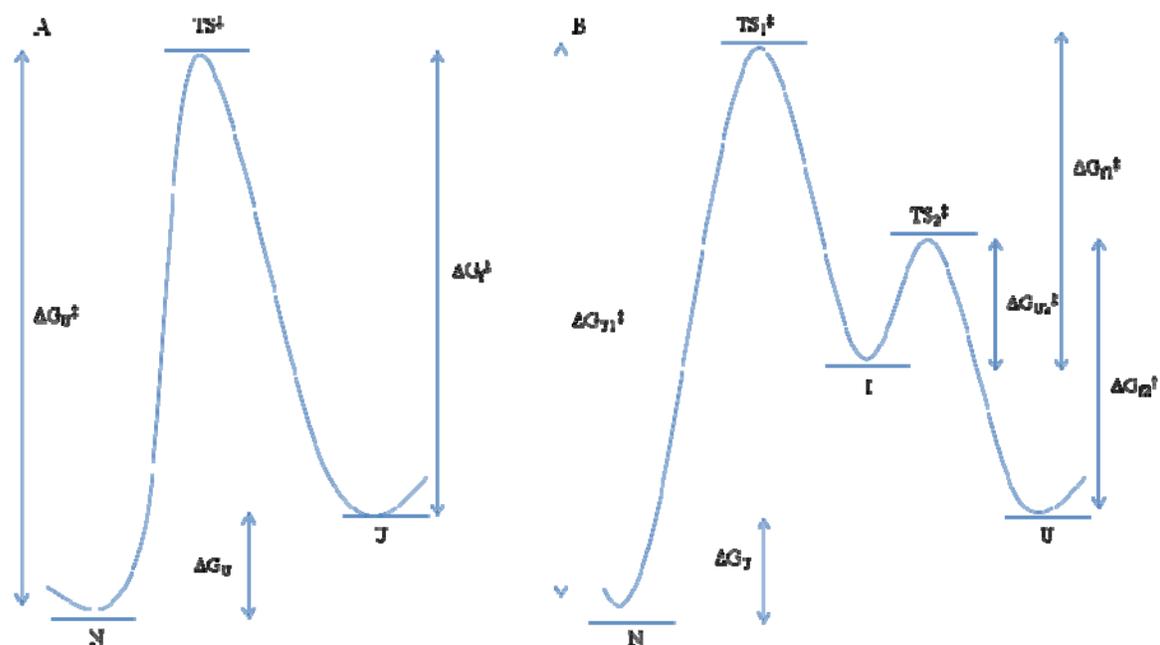
Figure 5.1: Reaction coordinate diagram for folding of a monomeric protein.
A. Two state folding of a monomeric protein. ΔG_U is the free energy of unfolding, $\Delta G_U^{\ddagger}$ is the activation free energy for unfolding and $\Delta G_f^{\ddagger}$ is the activation free energy for folding. N, U and $TS^{\ddagger}$ represent the native, unfolded monomer and the transition state, respectively. B. Three state folding of a monomeric protein. $\Delta G_{U1}^{\ddagger}$ is the activation free energy for unfolding of the native monomer to a monomeric intermediate, $\Delta G_{U2}^{\ddagger}$ is the free activation energy for unfolding of the monomeric intermediate, $\Delta G_{f1}^{\ddagger}$ is the activation free energy for folding of the native monomer from monomeric intermediate and $\Delta G_{f2}^{\ddagger}$ is the activation free energy for folding of the monomeric intermediate. I, $T_{S1\ddagger}$ and $T_{S2\ddagger}$ represent the monomeric intermediate, the first and second transition state.
$$\Delta G_U = G_U - G_f, \Delta G_U^{\ddagger} = G_{TS\ddagger} - G_N, \Delta G_f^{\ddagger} = G_{TS\ddagger} - G_U$$
$$\Delta G_{U1}^{\ddagger} = G_{TS1\ddagger} - G_N, \Delta G_{U2}^{\ddagger} = G_{TS2\ddagger} - G_I, \Delta G_{f1}^{\ddagger} = G_{TS1\ddagger} - G_I, \Delta G_{f2}^{\ddagger} = G_{TS2\ddagger} - G_U$$

Characterizing the folding of a protein involves determining the sequence of structural conversions, including any populated intermediates, the rates of interconversion of different species, and characterization of their structures and energetics (Fersht,1999).

5.1.2 Kinetic intermediates

Many small, monomeric proteins fold with simple 2-state kinetics and show wide variation in folding rates, from microseconds to seconds (Jackson,1998). The same trend is observed for small dimeric proteins whereas larger dimers tend to fold through the formation of monomeric or/and dimeric intermediates. The roles of these intermediates remain controversial. Kinetic intermediates have been proposed in some cases to help the protein fold more efficiently and rapidly. For example, SecA needs molecular chaperones to fold properly, *in vivo*. It has been proposed that SecA requires the formation of monomeric or dimeric kinetic intermediate to fold properly *in vitro* (Doyle *et al.*,2000). Moreover, recent studies performed on archael histones have shown the formation of kinetic intermediates makes the protein fold faster (Topping *et al.*,2004). For this specific case, the role of the kinetic intermediate can be explained by the fact that the formation of one or several intermediates divides the energy barrier between the unfolded and native protein into several smaller one, which may be easier and faster to traverse as illustrated in Figure 5.2.
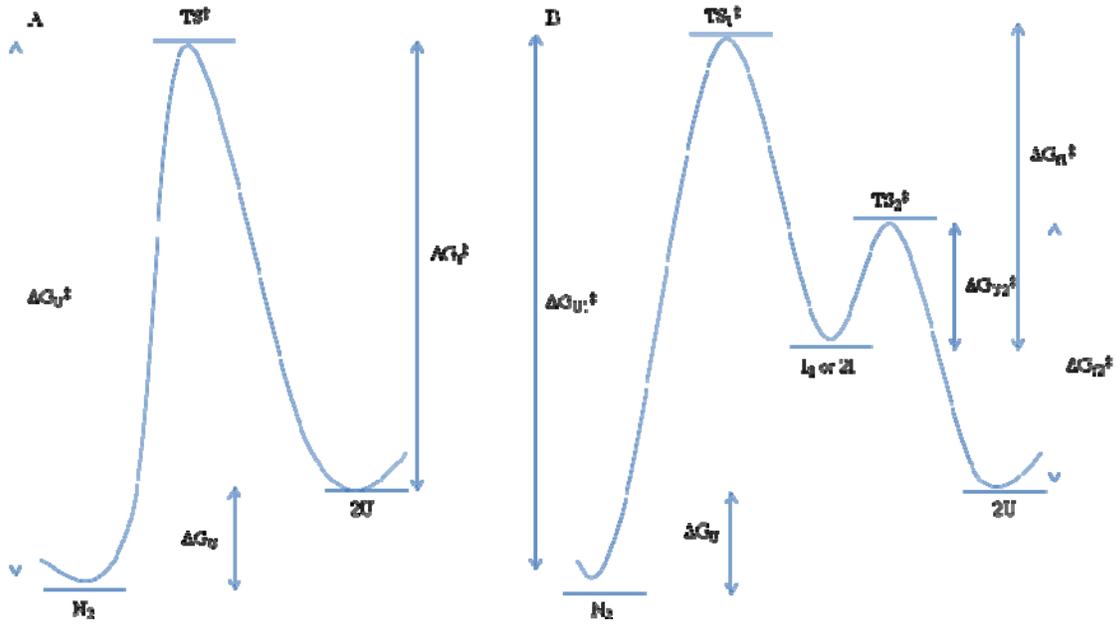
Figure 5.2: Reaction coordinate diagram for folding of a dimeric protein.
A. Two state folding of a dimeric protein. ΔG_U is the free energy of unfolding, $\Delta G_U^{\ddagger}$ is the activation free energy for unfolding and $\Delta G_f^{\ddagger}$ is the activation free energy for folding. N, U and $TS^{\ddagger}$ represent the native, unfolded monomer and the transition state, respectively. B. Three state folding of a dimeric protein. $\Delta G_{U1}^{\ddagger}$ is the activation free energy for unfolding of the native dimer to intermediate, $\Delta G_{U2}^{\ddagger}$ is the free activation energy for unfolding of the intermediate, $\Delta G_{f1}^{\ddagger}$ is the activation free energy for folding of the native monomer from intermediate and $\Delta G_{f2}^{\ddagger}$ is the activation free energy for folding of the intermediate. I, I_2 $T_{S1\ddagger}$ and $T_{S2\ddagger}$ represent the monomeric or dimeric intermediate, the first and second transition state.
$\Delta G_U = 2G_U - G_f$, $\Delta G_U^{\ddagger} = G_{TS\ddagger} - G_N$, $\Delta G_f^{\ddagger} = G_{TS\ddagger} - 2G_U$
$\Delta G_{U1}^{\ddagger} = G_{TS1\ddagger} - G_N$, $\Delta G_{U2}^{\ddagger} = G_{TS2\ddagger} - 2G_I$ or $\Delta G_{U2}^{\ddagger} = G_{TS2\ddagger} - G_{I2}$, $\Delta G_{f1}^{\ddagger} = G_{TS1\ddagger} - 2G_I$ or $\Delta G_{f1}^{\ddagger} = G_{TS1\ddagger} - G_{I2}$,
$\Delta G_{f2}^{\ddagger} = G_{TS2\ddagger} - 2G_U$

5.2 Methods

5.2.1 Unfolding kinetic measurements of Tm0979 and Mth1491

Unfolding rates were sufficiently slow that they could be measured by manual mixing experiments. Unfolding was monitored by fluorescence using a Fluorolog-3 (HORIBA; JOBIN YVON-SPEX) with excitation and emission wavelengths of 280 and 332 nm, respectively, for Tm0979 and 279 and 321 nm for Mth1491. Unfolding of the native protein was performed by diluting native protein stock solution 10-fold in different concentrations of GdmCl buffered with 20 mM citrate (10 mM DTT, 1 mM EDTA) for Mth1491 or 25 mM phosphate for Tm0979. Filtered 8 M GdmCl, milliQ water, and 10-fold buffer were used to prepare the different solutions of GdmCl, which are referred to as unfolding buffer. 90 μL of unfolding buffer was placed in an Eppendorf tube, and then 10 μL of native protein was added and mixed by pipetting. The unfolding protein was then introduced into a 50 μL quartz cuvette (pathlength of 3 mm) and the unfolding signal was monitored by fluorescence. To minimize temperature artifacts, both protein and GdmCl solutions were pre-equilibrated at 25°C prior to initiation of unfolding. The time required for the mixing before the introduction of the protein in the cuvette was measured using a stop-watch and is referred to as dead time (average dead time = 12s). Different unfolding buffers were prepared with a concentration of GdmCl ranging from 5.6 to 7 M for Mth1491 and from 4 to 6.2 M for Tm0979. At the end of each unfolding measurement, the concentration of GdmCl of the protein in the cuvette was checked by measuring its refractive index.

5.2.2 Refolding kinetics measurements

5.2.2.1 Tm0979 refolding kinetics

Refolding kinetics at moderate GdmCl concentrations were sufficiently slow that they could be measured using manual mixing techniques and monitored by fluorescence using a Fluorolog-3 (HORIBA; JOBIN YVON-SPEX) and by CD using a J715 CD

spectropolarimeter (Jasco, Easton, MD) (emission at 215 nm), by diluting a stock solution of unfolded protein 10-fold in buffer so that the final concentration of buffer is 25 mM phosphate. The stock solution of unfolded protein was prepared by incubating native protein in 6 M GdmCl for 15 minutes, which is sufficient time for the protein to become very highly unfolded (>90%). The unfolded protein was then diluted in a refolding buffer, with final GdmCl concentration varying from 1.6 to 3.2 M. As for the unfolding kinetic, 90 μL of refolding buffer was placed in an Eppendorf tube, and then 10 μL of denatured protein was added and mixed by pipetting. The refolding protein was then placed in a 50 μL quartz cuvette (pathlength = 3 mm) (500 μL quartz cuvette with 3 mm pathmength when recorded by CD) and the refolding signal was recorded optically in real time. To minimize temperature artifacts, both protein and GdmCl solutions were pre-equilibrated at 25°C prior to initiation of unfolding. The dead time was measured using a stop-watch and was on average 12 seconds. At the end of each refolding measurement, the concentration of GdmCl of the protein in the cuvette was checked by measuring its refractive index.

At low GdmCl concentration, Tm0979 refolding is too fast to be recorded by manual mixing experiment. Accordingly, for 0.6M to 2.04 M GdmCl, refolding of Tm0979 was monitored by stopped flow, using an SFM4/Q instrument (Molecular Kinetics, Pullman, WA) interfaced to the Fluorolog 3 (Figure 5.3). Temperature was maintained at 25°C using a circulating water bath connected to the stopped-flow instrument. 8 M GdmCl buffered in 25 mM phosphate was loaded in the syringe 2, buffer was loaded in syringe 3 by pumping in the solution at 100 μL.s^{-1}. Air bubbles were eliminated by pumping these solutions out at 300 μL.s^{-1}. Finally, unfolded protein was loaded into syringe 4 by pumping in the solution at 22 μL.s^{-1}; air bubbles were eliminated from the syringe by pumping out the solution at 66 μL.s^{-1} (Figure 5.3).
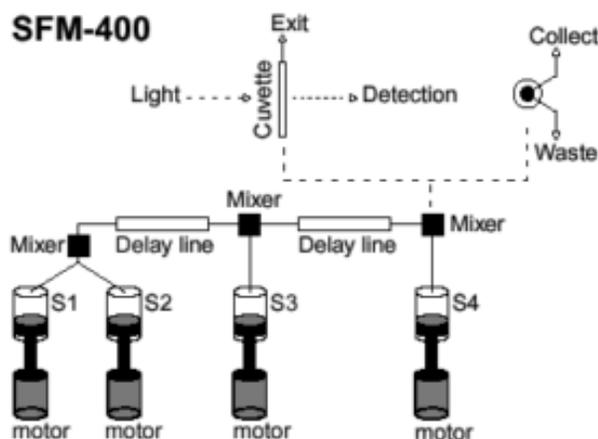
Figure 5.3: SFM-400 diagrams for Stopped-Flow.
S1, S2, S3 and S4 are the syringes 1, 2, 3 and 4 respectively. The motors push the desired amount of solution out of the corresponding syringe. The mixers mix the solutions out of the syringes before reaching the cuvette.

5.2.2.2 Mth1491refolding kinetics

Mth1491 refolding was found to be too fast to be monitored by manual mixing experiment. Refolding of Mth1491 was therefore measured by stopped-flow, as described for Tm0979, from 0.6M to 2.2M GdmCl. All solution were degassed for 30minutes prior to loading into syringes and were kept in reducing condition by adding DTT and EDTA so that the final concentrations were 10mM and 1mM, respectively.

5.2.3 Analysis of kinetic folding

5.2.3.1 Two-state kinetics

A two-state kinetic mechanism for a monomer involves single exponential unfolding and refolding traces at all denaturant concentrations (and are independent of protein concentration) (Table 5.1, Equations 5.1 and 5.4). In contrast, for dimers and trimers, due to the change in molecularity associated with the transition, under most conditions the observed kinetics are not exponential and are dependent on protein concentration. In addition, the equations for refolding are different from that for unfolding (Table 5.1 Equation 5.2, 5.5 for dimer, Equations 5.3, 5.6 for trimers).

The folding kinetics of a dimer or trimer 2-state kinetic mechanism are in theory always non-exponential (Table 5.1). Under strongly native conditions (the folded baseline region), the kinetics of folding can be described by the Equations 5.3 and 5.4 in Table 5.1 for dimer and trimer, respectively.

A plot of the natural logarithm of the observed unfolding and refolding rate constants for a monomer results in a V-shaped profile often referred to as a chevron plot (Fersht A,1999). For a dimer or a trimer, a smooth transition will not be observed for this type of plot since the observed rate will not be a simple sum of the unfolding and refolding rate constants as it is for monomers.

5.2.3.2 Three-state kinetics

A three-state equilibrium mechanism must at least involve a three state kinetic mechanism. Depending on the number and nature of intermediates, as well as the relative forward and backward rates controlling each step, the observed unfolding kinetics may be single exponential, double exponential, or complex (Nölting,2006). As a result, a plot of the natural logarithm of the observed kinetic rate constants may deviate quite significantly from the typical chevron plot observed for monomers and the two-state dimer discussed above.

In the case of monomeric intermediate formation, as observed upon equilibrium unfolding of Tm0979, at high denaturant concentrations, the apparent rate of unfolding is slower than expected based on linear extrapolation of the rates at low denaturant resulting in curvature of the unfolding arm of the chevron-like plot. The curvature is due to the association rate playing a part in the observed relaxation time.

As stated above, the observed folding kinetics depend on the relative forward and backward rates of all processes in the reaction. For a three state mechanism involving a monomeric intermediate, if the association rate constant is very fast, all that will be observed

is a single exponential according to the rate of monomer folding. If on the other hand monomer refolding is very fast one would see only see a second or third order reaction in case of dimer or trimer refolding, respectively. It is also possible to observe both the first order and second order processes.

5.2.4 Fitting of the kinetic data

5.2.4.1 *Fitting of the unfolding and refolding traces*

Unfolding and refolding kinetics were fit to integrated equations for first, second or third order kinetic processes, as appropriate (Table 5.2), using Biokine32 (Biologic) and OriginPro7.5 (OriginLab) software. For processes involving no change in molecularity (e.g. monomer folding or unfolding), the time course should follow a simple exponential, with a small linear drift being observed in some cases due to experimental factors such as diffusion of reagents or lamp instability. For reaction mechanisms involving intermediates, more than 1 exponential process may be observed. Unfolding and refolding kinetic traces could be conveniently fit to a single or double exponential equation with a linear drift using Biokine32 software (Biologic). Refolding kinetic traces were also fit to second order or third order equation using OriginPro7.5 (OriginLab) in order to determine if the rate limiting step was dimer or trimer association, respectively.

Table 5.2: Equations used to fit the refolding and unfolding kinetic traces of Tm0979 and Mth1491.

Fitting program	Biokine	Origin
Unfolding transition:		
One step process:	$Y = ax + b + c\exp(-k_u t)$	
Two steps process:	$Y = ax + b + c_1\exp(-k_{u1} t) + c_2\exp(-k_{u2} t)$	
Refolding transition:		
No change in molecularity:		
One step process	$Y = ax + b + c\exp(-k_f t)$	
Two steps process	$Y = ax + b + c_1\exp(-k_{f1} t) + c_2\exp(-k_{f2} t)$	
Dimer association		$Y = Y_{N2} + (Y_U - Y_{N2})*f_U$
Trimer association		$Y = Y_{N3} + (Y_U - Y_{N3})*f_U$

Unfolding transition: a, b and c correspond to the change in the unfolded protein signal with time, the extrapolated value of the unfolded signal at t=0 and the total amplitude of the unfolding transition, respectively. k_u is the rate constant of unfolding. If the unfolding occurs in 2 steps, c_1 and c_2 correspond to the amplitude of each step, k_{u1} and k_{u2} are the rates of unfolding of the corresponding steps.
Refolding transition: a, b and c correspond to the change in signal of the native protein with time, the extrapolated value of the signal of the native protein at t=0 and the total amplitude of the refolding transition. If the refolding occurs in 2 steps, c_1 and c_2 correspond to the amplitudes of each step, k_{u1} and k_{u2} are the rates of unfolding of the corresponding steps. Y_U, Y_{N2} and Y_{N3} are the signals of the unfolded monomers, the native dimer and the native trimer, respectively. f_U is the fraction of the unfolded monomer and it varies with time as described in Appendices 3 and 4.

5.2.4.2 Chevron plot analysis

Tm0979 observed kinetics correspond to a 2-state monomer folding. A plot of the natural logarithm of the observed unfolding and refolding rate constants for a monomer results in a V-shaped profile often referred to as a chevron plot. Upon monomer 2-state unfolding and refolding, the observed or measured rate constant corresponds to k_{obs} and is the sum of the folding and unfolding rates constants (Fersht,1999),(Table 5.1, Equation 5.1, 5.4):

$$k_{obs} = k_u + k_f \tag{5.7}$$

It has been observed experimentally for many small proteins that the logarithm of the rate constants of unfolding and folding are linearly proportional to the concentration of GdmCl and can be described by the following equations (Tanford C, 1968, Tanford C, 1970):

$$\ln k_u = \ln k_{uH2O} + m_u[GdmCl] \tag{5.8}$$

$$\ln k_f = \ln k_{fH2O} - m_f[GdmCl] \tag{5.9}$$

where k_f and k_{fH20}, are the rate constants of folding in GdmCl and in water, respectively, k_u and k_{uH20} are the rate constants of unfolding in GdmCl and in water, respectively, m_u and m_f are the dependence of lnk_u and lnk_f on [GdmCl] and are proportional to the difference in accessible surface area between the transition state and the native protein and between the unfolded state and the transition state, respectively.

The natural logarithm of the observed kinetic rate constants of a monomer can be plotted together in a chevron plot (Fersht,1999). By rearranging Equations 5.7, 5.8 and 5.9, the corresponding kinetic constants can be determined by fitting the chevron plot using the following equation:

$$lnk_{obs} = ln(k_f^{H2O}exp(m_f[GdmCl]/RT)+k_u^{H2O}exp(m_u[GdmCl]/RT)) \qquad (5.10)$$

with OriginPro7.5 (OriginLab). Moreover, equilibrium constants determined in the previous chapter (Chapter 4) can be compared with the kinetic constants determined in this chapter. Actually, the equilibrium constant of unfolding (K_U) can be linked to k_f and k_u, for monomeric and oligomeric protein, by the following equations:

$$\Delta G_U = G_N - G_U$$

$$= G_N - G_{TS\ddagger} - (G_U - G_{TS\ddagger})$$

$$= \Delta G_U^{\ddagger} - \Delta G_f^{\ddagger}$$

$$\Delta G_U = -RTlnK_U \qquad (5.11)$$

$$K_U = k_u/k_f \qquad (5.12)$$

$$m = m_u + m_f \qquad (5.13)$$

where ΔG_U is the free energy of unfolding, $\Delta G_U^{\ddagger}$ is the activation free energy for unfolding, $\Delta G_f^{\ddagger}$ is the activation free energy for folding, and G_N, G_U and $G_{TS\ddagger}$ are the free energy of the native, unfolded protein and the transition state, respectively.

5.3 Results

5.3.1 Tm0979 folding pathway

5.3.1.1 Unfolding kinetics

Kinetic unfolding traces monitored by fluorescence are shown for Tm0979 between 4 and 6.4 M GdmCl in Figure 5.4. The kinetic traces are well fit by a single exponential equation. Under highly unfolding conditions, all unfolding reactions are single exponential (Table 5.1), and so additional considerations are required to interpret the unfolding results. There are two situations that may be envisaged as being likely occur during unfolding. The transition monitored by fluorescence may correspond to the simultaneous dissociation and unfolding of the native dimer. Alternatively, it could correspond to unfolding of the native dimer to form an intermediate, as observed for the equilibrium data, with the formation of the monomeric intermediate or the unfolding of the intermediate being potentially rate determining . The amplitudes of the kinetic traces for the unfolding traces correspond to 93% of the signal observed by equilibrium curve measurment. This suggests that the transition observed by kinetics may not correspond to the entire unfolding transition. Moreover, the fluorescence of the intermediate corresponds to 90% of the one of the native protein, approximately. This observation suggests that the rate limiting step observed by kinetics measurement is the unfolding of the monomeric intermediate, the dissociation being too fast to be observed.

The natural logarithm of the unfolding rates varies linearly with GdmCl concentration (Figure 5.5). The unfolding rate in water corresponds to 7.3×10^{-4} s^{-1} with a $m_u^{\ddagger}$ of 0.53 kcal.mol^{-1}.K^{-1}. The magnitude of $m_u^{\ddagger}$ is similar to those observed for unfolding of other monomers of similar size (Jackson,1998), and this reinforces the idea that Tm0979 monomer unfolding may be the rate limiting step.

Figure 5.4: Unfolding kinetics of Tm0979.
A. Fluorescence-monitored unfolding kinetics of Tm0979 in 4 M GdmCl fit to a single exponential equation with a linear drift (equation 5.6 in Table 5.1). Shown inset is the residual of the single exponential fit with k_u = 3.4×10^{-2} s^{-1}. B. Fluorescence-monitored unfolding kinetics of Tm0979 in 5 M GdmCl fit to a single exponential equation. Shown inset is the residual of the single exponential fit with k_u = 8.3×10^{-2} s^{-1}. C. Fluorescence-monitored unfolding kinetics of Tm0979 in 6 M GdmCl fit to a single exponential equation. Shown inset are the residuals of the single exponential fit with k_u = 1.9×10^{-1} s^{-1}. Residuals indicate the kinetics are well described by a single exponential. Total protein concentration is 1µM (dimer equivalent).

Figure 5.5: Rate constants of unfolding of Tm0979.
Dependence of the natural logarithm of the observed unfolding rate constant $k_{u\text{-obs}}$ on denaturation concentration. $k_{u\text{-obs}} = 7.3 \times 10^{-4}\, s^{-1}$, $m_u^{\ddagger} = 0.53\ kcal.mol^{-1}.K^{-1}$.

5.3.1.2 Refolding kinetics

Kinetic refolding traces were monitored by manual mixing fluorescence between 1.4 and 3.2M GdmCl, by stopped flow fluorescence between 0.6 M and 2.04 M and by manual mixing CD between 1.8 and 3.2 M for Tm0979 at 1.5μM (expressed in dimer equivalent). The kinetic refolding traces were well fit by a single equation, suggesting that the transition observed corresponds to monomer refolding. In order to verify this hypothesis, the kinetic refolding traces were also fit to a second order equation describing the simultaneous folding and association of monomers to form the native dimer (Table 5.2, Appendix A.5: Kinetic model of dimer two-state folding). A systematic deviation between the experimental traces and the fit was observed, suggesting that the refolding process observed is not the association step. In order to test this further, kinetic refolding traces were monitored at 3μM in order to determine the dependencies of the observed refolding rates upon protein concentration. Figure 5.7 shows the dependence of the natural logarithm of the observed refolding rate constant $k_{f\text{-}obs}$ determined at 1.5 and 3μM with GdmCl concentration. The natural logarithms of the rates vary linearly and similarly with GdmCl concentration for stopped-flow and manual mixing fluorescence and for manual mixing CD experiments. However, a deviation from linearity is observed for the rates measured by stopped-flow fluorescence below 1.08 M and for rates measured by manual mixing fluorescence and CD below 2.2 M at both protein concentrations. The amplitude of the refolding transitions monitored by manual mixing fluorescence and CD at GdmCl concentration below 2.2 M is very small compared to the expected value. Thus, the reaction is becoming too fast to measure accurately as most of the reaction is completed in the dead time. This results in less accurate refolding rate constants, which can explain why those rates do not follow the same trends as the other measured rates. These slow rates may be inaccurate, or may correspond to small amplitude, slow rates due to proline isomerization (Fersht,1999) and so were thus not taken into account for further analysis.

The kinetic refolding rate constants measured at 1.5µM do not change significantly compared to those measured at 3µM. This again suggests that the observed refolding rate is characteristic of a process involving no change in molecularity, which reinforces the hypothesis that the transition observed is the monomer folding. As both folding and unfolding of Tm0979 may correspond to a monomer folding process, the dependence of the natural logarithm of the observed unfolding and refolding rate constants were plotted and fit together as described in the following section.
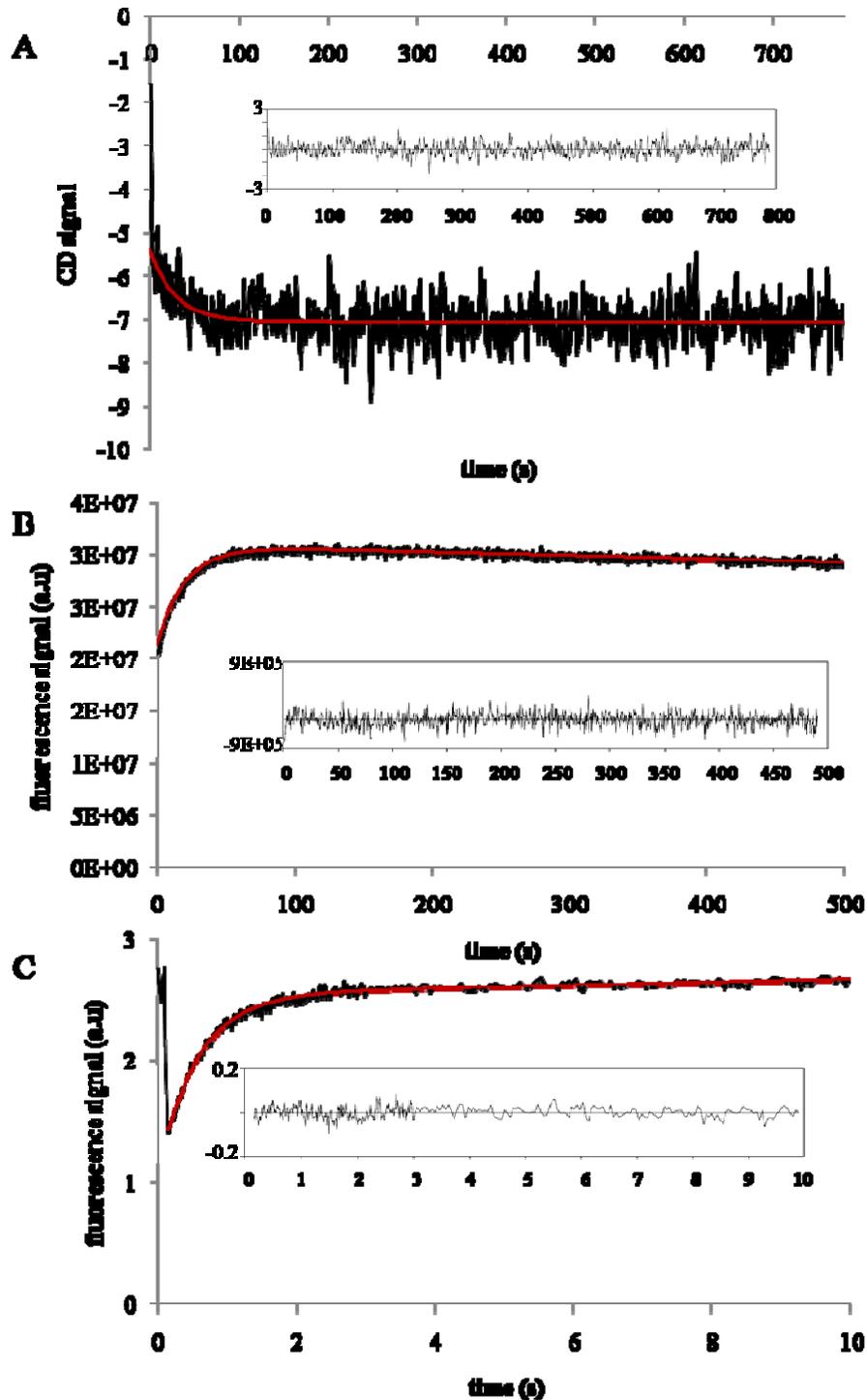
Figure 5.6: Refolding kinetics of Tm0979.
A. CD-monitored refolding kinetics of Tm0979 in 2.8 M GdmCl fit to a single exponential equation. Shown inset are the residuals of the single exponential fit with $k_f = 3.5 \times 10^{-2} s^{-1}$. B. Fluorescence-monitored refolding kinetics of Tm0979 in 2.4 M GdmCl fit to a single exponential equation. Shown inset are the residuals of the single exponential fit with $k_f = 4.9 \times 10^{-2} s^{-1}$. C. Stop flow fluorescence-monitored refolding kinetics of Tm0979 in 0.6 M GdmCl fit to a single exponential equation. Shown inset are the residuals of the single exponential fit with $k_f = 1.7 s^{-1}$. Residuals indicate the kinetics are well described by a single exponential. Total protein concentration is 1.5μM (dimer equivalent).
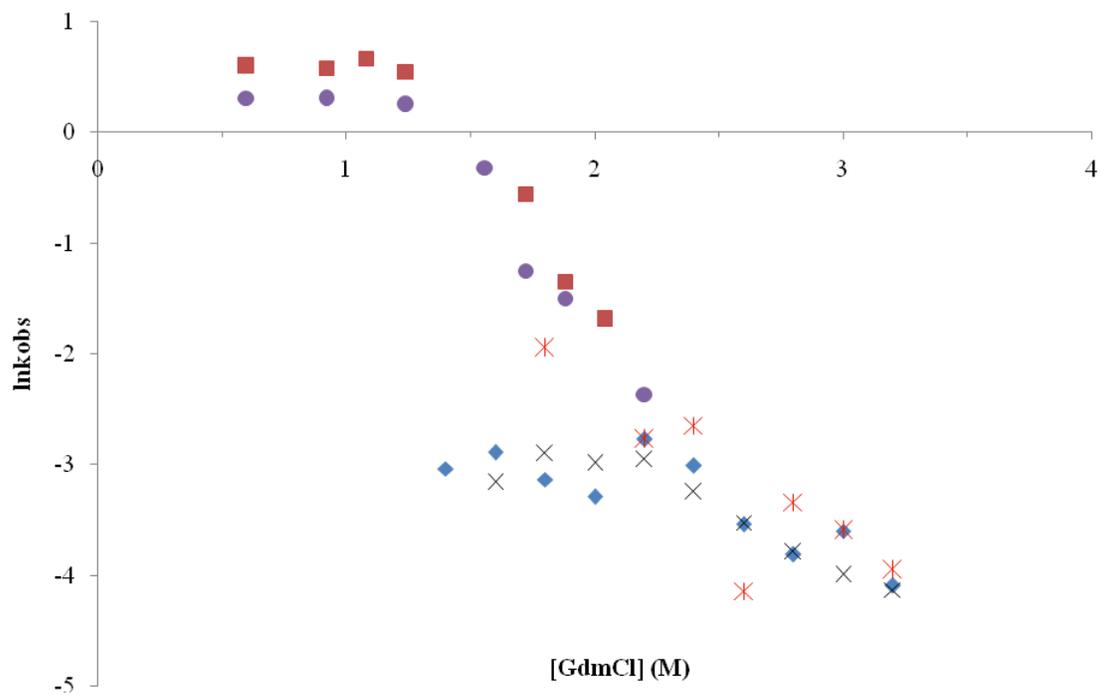
Figure 5.7: Rates of refolding of Tm0979.
Dependence of the natural logarithm of the observed refolding rate constant $k_{f\text{-obs}}$ on denaturation concentration measured at 1.5 and 3 µM by stopped-flow fluorescence (1.5 µM: red squares / 3 µM: purple circle), manual mixing fluorescence (1.5 µM: blue diamonds / 3 µM: black crosses) and manual mixing CD (1.5 µM: red crosses). Total protein concentration expressed in dimer equivalents.

5.3.1.3 Chevron plot

The dependence of the natural logarithm of the observed unfolding and refolding rate constants of Tm0979 with GdmCl concentration are plotted in Figure 5.8. The kinetic constants determined from the fitting of the chevron plot are summarized in Table 5.3. Equilibrium constants were calculated from the kinetic rate constants using Equations 5.11, 5.12 and 5.13 and compared to the equilibrium values determined in Chapter 4. The ΔG_U and m values calculated from the kinetic experiments are very similar to the values for monomer unfolding, ΔG_{U2} and m_{U2}, obtained from equilibrium measurements. This observation suggests that the monomeric species observed by kinetic studies corresponds to the monomeric intermediate observed by stability experiment, and that folding and unfolding of the monomer are rate limiting for Tm0979 kinetics.
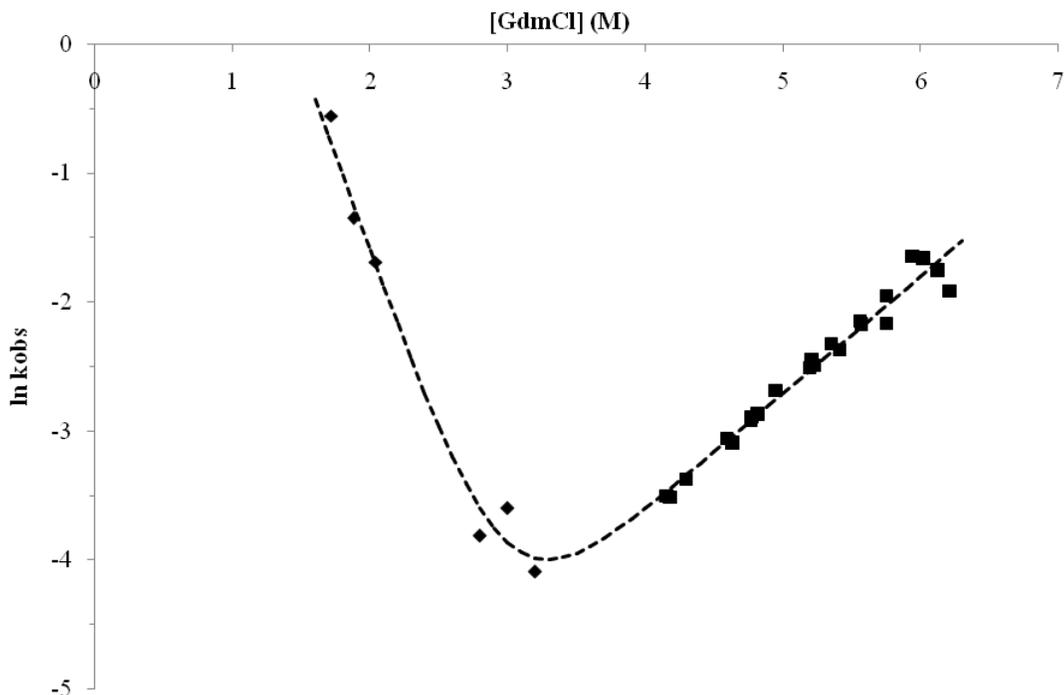
Figure 5.8: Rates of unfolding and refolding of Tm0979 monomer.
Dependence of the natural logarithm of the observed rate constant $k_{obs} = k_u + k_f$ on denaturation concentration measured in GdmCl concentration ranging from 1.7 to 6.4 M. Refolding and unfolding rate constants are plotted as diamonds and squares, respectively. Samples measured in 1.5 µM (dimer concentration).

Table 5.3: Kinetics and equilibrium folding constants of Tm0979 monomer.

Kinetic folding rate constants	Calculated equilibrium constants	Experimental equilibrium constants
$k_u = 7 \times 10^{-4} s^{-1}$	$K_U = 9.2 \times 10^{-6} mol$	$K_{U2} = 5 \times 10^{-6} mol$
$k_f = 7.6 \times 10^{1} mol^{-1}.s^{-1}$	$\Delta G_U = 6.9 kcal.mol^{-1}$	$\Delta G_{U2} = 6.2 kcal.mol^{-1}$
$m_u = 0.54 kcal.mol^{-1}.M^{-1}$	$m_U = 2.3 kcal.mol^{-1}.M^{-1}$	$m_{U2} = 2.01 kcal.mol^{-1}.M^{-1}$
$m_f = 1.76 kcal.mol^{-1}.M^{-1}$		

Kinetic folding rate constants were determined by fitting the chevron plot to Equation 5.10. Calculated equilibrium constant were determined using Equations 5.11, 5.12 and 5.13. Experimental equilibrium constants were determined in the previous chapter and are characteristic of the monomer folding.

5.3.2 Mth1491 folding pathway

5.3.2.1 Unfolding kinetics

Kinetic unfolding traces were monitored by both fluorescence and CD and are shown for GdmCl concentration of 5.5, 5.8 and 6.9 M in Figure 5.9. The kinetic traces are well fit by a single exponential equation for both optical probes. The natural logarithm of the unfolding rates measured by fluorescence and CD varies linearly with GdmCl concentration (Figure 5.10). The unfolding rate constants in water are 6.45×10^{-13} and 7.04×10^{-12} s^{-1} with corresponding $m_u^{\ddagger}$ values of 2.3 and 2.1 $kcal.mol^{-1}.M^{-1}$, when measured by fluorescence and CD, respectively. The kinetic unfolding rate constants and the corresponding $m_u^{\ddagger}$ values determined by the two probes are very similar, thus the transition observed by fluorescence likely corresponds to the one measured by CD. The amplitude of the kinetic unfolding traces measured by fluorescence and CD corresponds to almost all of the expected amplitude based on equilibrium curves measured at the same protein concentration. This suggests that the unfolding transition observed by both fluorescence and CD may be Mth1491 complete unfolding.

In comparing the experimental $m_u^{\ddagger}$ for Mth1491 with values of $m_u^{\ddagger}$ for other monomeric proteins, it does not seem that the refolding transition of Mth1491 corresponds to that expected for unfolding of a monomeric subunit of the protein. Actually, the $m_u^{\ddagger}$ determined for Mth1491 is considerably larger than would be expected for a monomer of 112 amino acids (the subunit size for Mth1491) (Jackson,1998). Comparisons with data for other dimeric or trimeric proteins is also consistent with this. Thus, it seems likely that the observed unfolding kinetics for Mth1491 corresponds to trimer or perhaps dimer unfolding.

In summary, the unfolding transition observed by both fluorescence and CD manual mixing experiments seems likely to be the complete unfolding of Mth1491 trimer or perhaps

the unfolding of a dimeric or trimeric intermediate. Refolding kinetics were then also investigated and are described in the following section.
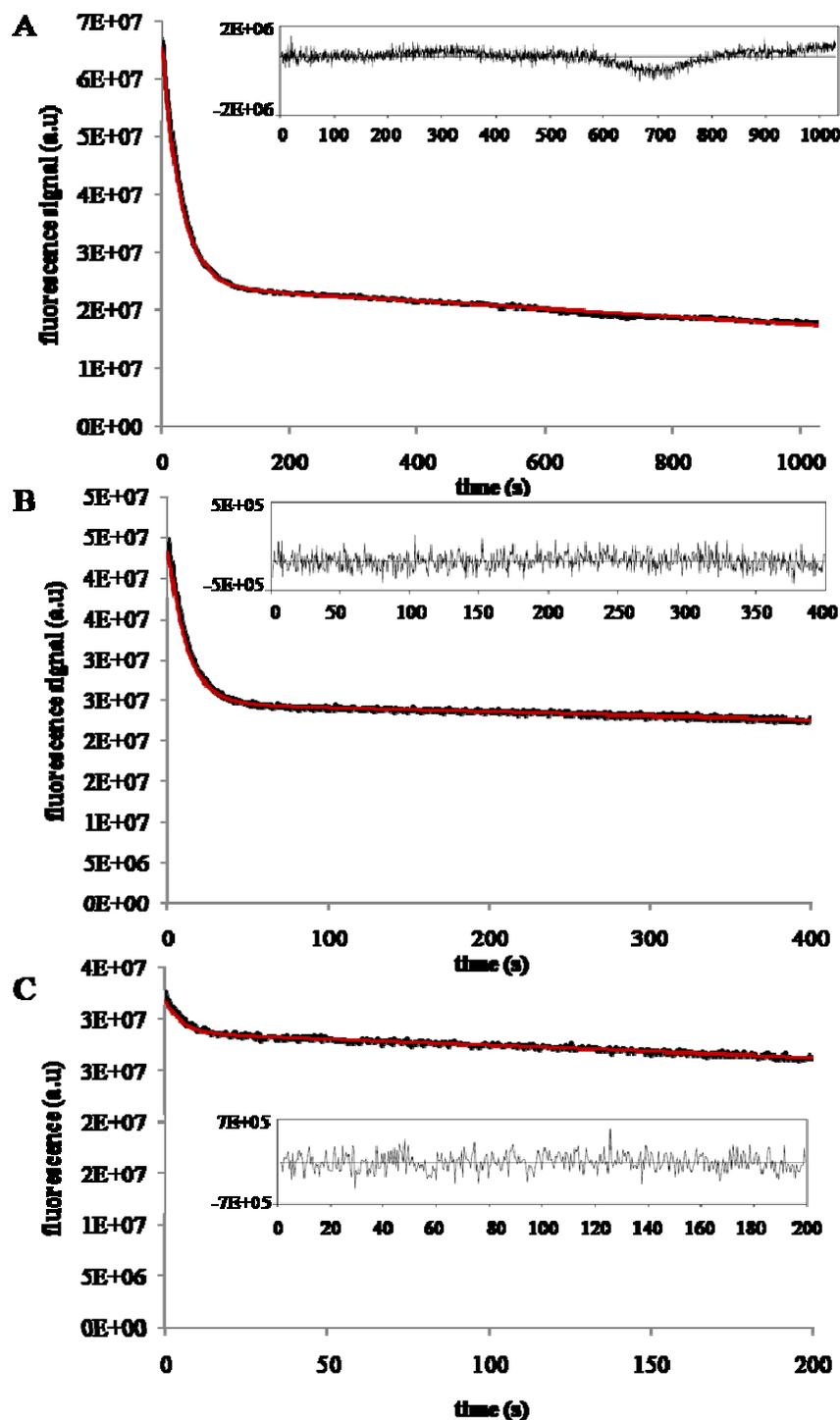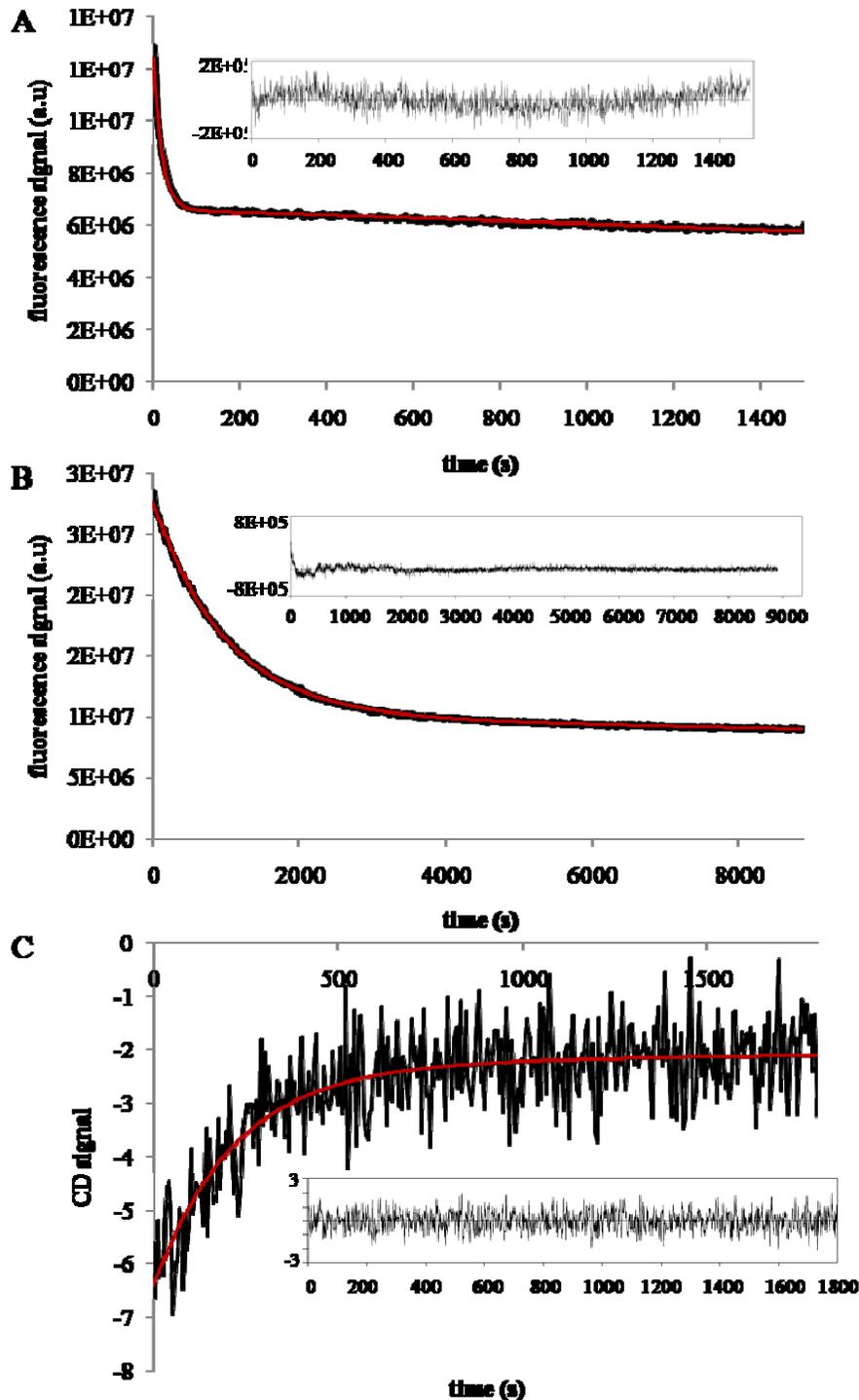
Figure 5.9: Unfolding kinetics of Mth1491.
A. Fluorescence-monitored unfolding kinetics of Mth1491 in 6.4 M GdmCl fit to a single exponential equation. Shown inset are the residuals of the single exponential with $k_u = 4.9 \times 10^{-2}$ s^{-1} which indicate the kinetics are well described by a single exponential. B. Fluorescence-monitored unfolding kinetics of Mth1491 in 5.5 M GdmCl fit to a single exponential equation. Shown inset are the residuals of the single exponential with $k_u = 1 \times 10^{-3}$ s^{-1} which indicate the kinetics are well described by a single exponential. C. CD-monitored unfolding kinetics of Tm0979 in 5.8 M GdmCl fit to a single exponential equation. Shown inset are the residuals of the single exponential fit with $k_u = 1.9 \times 10^{-1}$ s^{-1}. Residuals indicate the kinetics are well described by a single exponential. Total protein concentration is 1 μM (trimer equivalent).

102

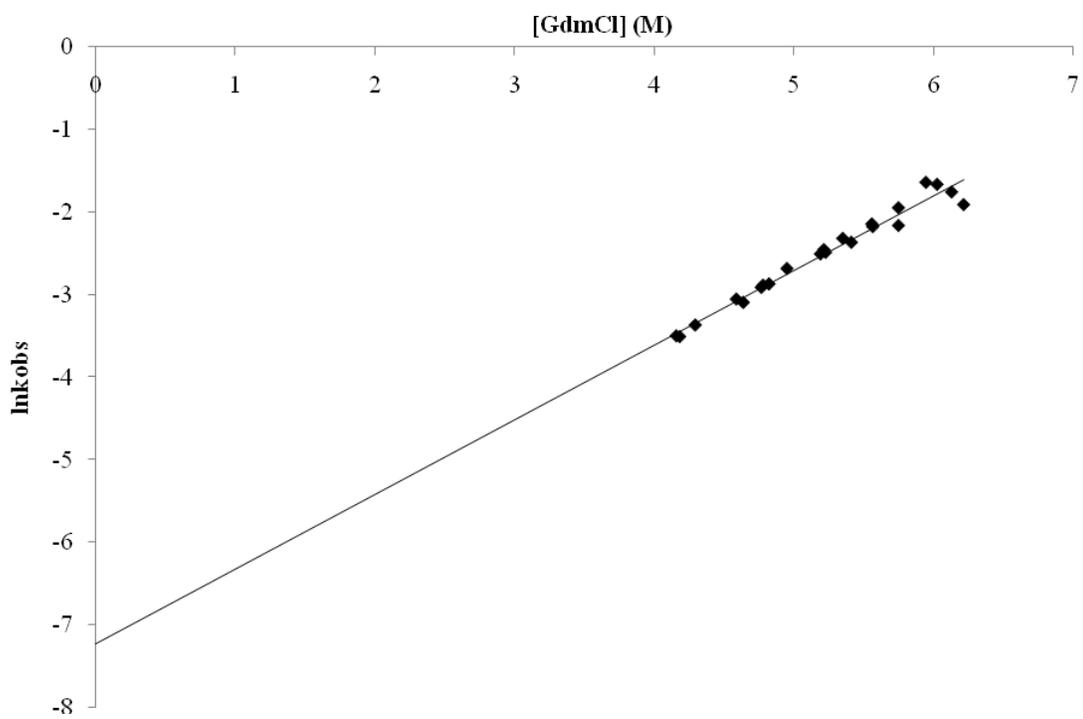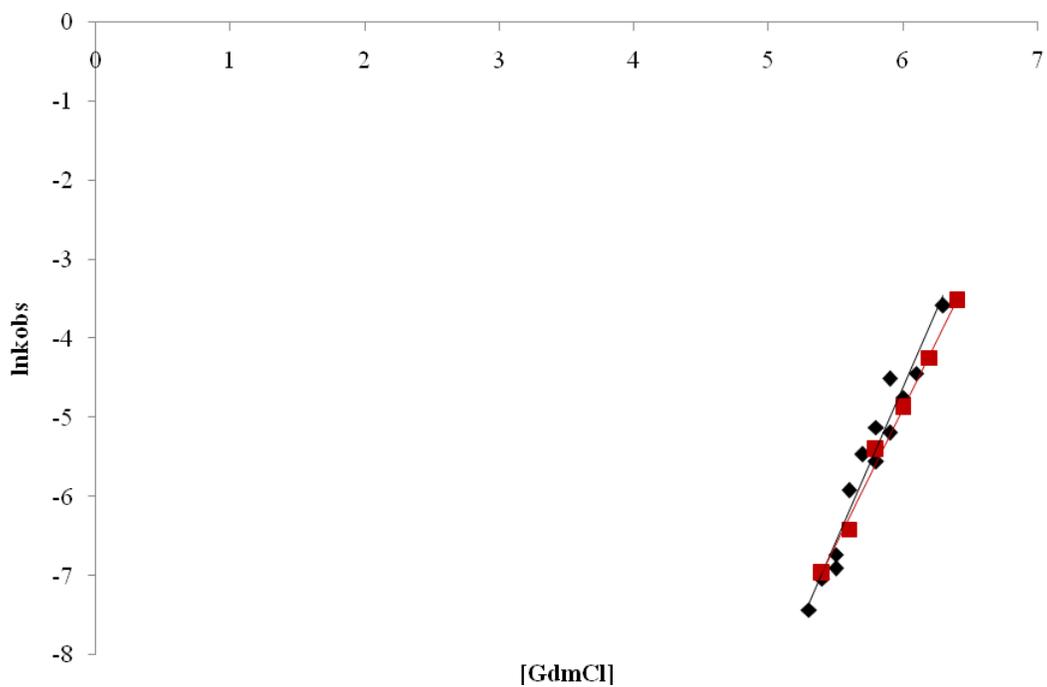**Figure 5.10: Rates of unfolding of Mth1491.**
Dependence of the natural logarithm of the observed unfolding rate constant $k_{u\text{-}obs}$ measured by fluorescence (black diamonds) and by CD (red squares) on denaturant concentration. In water, $k_{u\text{-}obs}$ = 6.45×10^{-13} s^{-1}, $m_u^{\ddagger}$ = 2.3 kcal.mol^{-1}.K^{-1} for fluorescence measurement, k_u = 7.04×10^{-12} s^{-1}, $m_u^{\ddagger}$ = 2.1 kcal.mol^{-1}.K^{-1} for CD measurement.

5.3.2.2 Refolding kinetics

Kinetic refolding traces of Mth1491 were monitored at 1 and 2µM protein by stopped flow fluorescence at GdmCl concentration ranging from 0.6 to 2.2 M. The refolding traces are well described by a single exponential equation for both protein concentrations, as shown in Figure 5.11. Single exponential refolding kinetics suggest that the observed transition corresponds to a process that does not involve a change in molecularity such as monomer folding (U → I) or perhaps native trimer formation from a trimeric intermediate (I_3 → N_3). In addition, the kinetic refolding traces were fit to a third order equation, and the fitted lines for individual traces passed well through the data points (Table 5.2, Appendix A.6: kinetic model of trimer two-state folding). However, the fitted rate constants are inconsistent. Thus, the observed kinetics do not appear to correspond to a trimer association reaction.

The natural logarithm of the refolding rate constants, obtained from single exponential equation fits, varies linearly with GdmCl concentration, as shown in figure 5.12. The denaturant dependence is similar for 1 and 2µM protein. The observed kinetic refolding rates in water are 6.2 and 0.2 s^{-1} for 1 and 2 µM trimer, respectively; the $m_f^{\ddagger}$ values are 1.3 and 1.1kcal.mol^{-1}.M^{-1}for 1 and 2 µM, respectively. The refolding rate constant of the observed process is thus not protein concentration dependent. This observation reinforces the idea that the process does not involve a change in molecularity. By comparing monomeric proteins (Jackson SE, 1998), the determined $m_f^{\ddagger}$ is quite close to what would be expected for folding of a protein the size of the Mth1491 monomer.
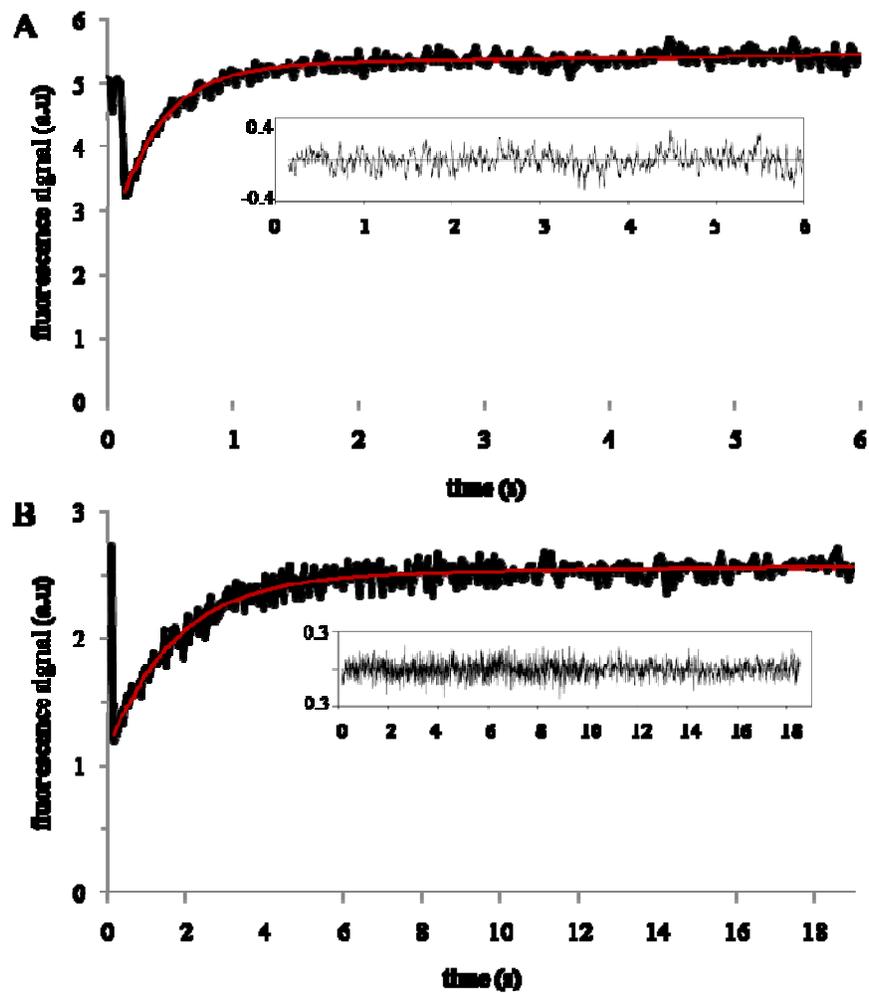
Figure 5.11: Refolding kinetics of Mth1491.
A. Stopped flow fluorescence-monitored refolding kinetics of Mth1491 at 1µM in 0.6 M GdmCl fit to a single exponential equation. Shown inset are the residuals of the single exponential with k_{f-obs} = 2.7 s^{-1}, which indicates the kinetics are well described by a single exponential. B. Stopped flow fluorescence-monitored refolding kinetics of Mth1491 at 2 µM in 1.88 M GdmCl fit to a single exponential equation. Shown inset are the residuals of the single exponential with k_{f-obs} = 0.58 s^{-1} which indicate the kinetics are well described by a single exponential.
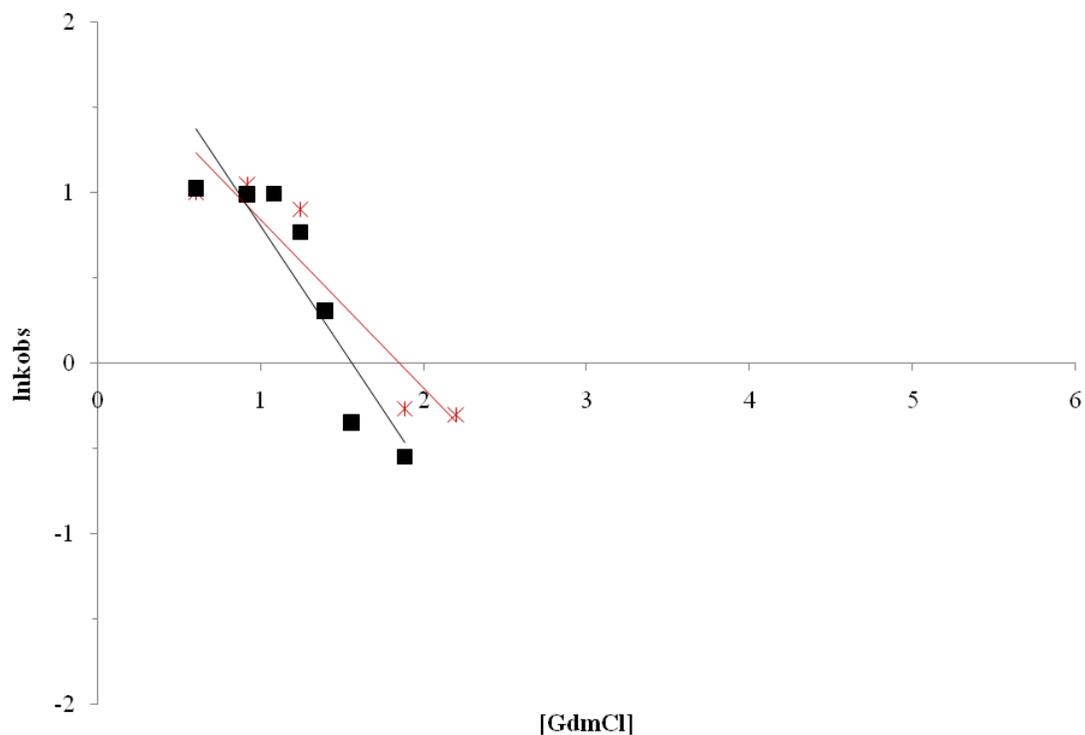
Figure 5.12: Rates of refolding of Mth1491.
Dependence of the natural logarithm of the observed refolding rate constant k_{f-obs} on denaturation concentration measured at 1 (black squares) and 2 μM (red crosses) by stopped flow fluorescence. Total protein concentration expressed in trimer equivalent. In water, the kinetic refolding rate constant is 6.2 s^{-1} and 0.2 s^{-1} at 1 and 2 μM, respectively, with $m_f^{\ddagger}$ of 1.1 and 1.3 kcal.mol^{-1}.M^{-1}, respectively.

5.4 Discussion

5.4.1 Tm0979 folding pathway

The Tm0979 folding pathway was investigated by measuring the unfolding and refolding of the dimer by manual mixing fluorescence and CD as well as by stopped flow fluorescence. Analysis of the kinetic unfolding and refolding kinetic traces suggests that Tm0979 folds through the formation of a monomeric intermediate, which is also observed by equilibrium studies. The folding and unfolding of this intermediate is proposed to be the rate limiting step of Tm0979 dimer folding, with association and dissociation being too fast to be measured. The same folding mechanism was observed for other dimeric proteins, such as SOD.

In addition to the kinetic data that support this mechanism, the results are also consistent with the fact that dimerization of Tm0979 is relative weak, and the monomer has significant stability. A careful quantitative analysis of the amplitudes and initial and final signals in the kinetic experiments would also be useful to validate the proposed mechanism.

5.4.2 Mth1491 folding pathway

The Mth1491 folding pathway was investigated by following the unfolding and refolding of the trimer by manual mixing fluorescence and CD as well as stopped flow fluorescence. Mth1491 folding mechanism is less well defined and may be more complex than the one observed for Tm0979. Actually, the unfolding transition of Mth1491 corresponds to a 2-state process. The process can then correspond to the fully unfolding of the native trimer to unfolded monomer or the unfolding of an intermediate. While observed refolding kinetics certainly do not correspond to trimer association, they may be monomer folding or another transition involving intermediates that does not involve a change in molecularity. Formation of monomeric and oligomeric intermediates was been observed for other trimeric

proteins, such as bacteriophage T4 Fibritin (Guthe *et al.*,2004) ($3U \leftrightarrow 3I \leftrightarrow I_2 + I \leftrightarrow I_3 \leftrightarrow N_3$), HIV and SIV gp41 six-helix bundles, 3 stranded coiled-coil (Marti *et al.*,2004) ($3U \leftrightarrow I_2 + I \leftrightarrow N_3$), Tumor necrosis factor TNFα (Hlodan *et al.*,1995) ($U \leftrightarrow I' \leftrightarrow I'' \leftrightarrow I''' \leftrightarrow I'''_2 \leftrightarrow I_3 \leftrightarrow N_3$), which indicates that the folding of trimer may be very complex. In theses studies, the formation of intermediates was proposed to help protein folding. Actually, it seems unlikely that unfolded monomers fold and associate simultaneously. Actually, unfolded monomers may first fold and then associate to form a kinetic dimer or trimer intermediate. The intermediate may then rearrange and form the native trimer.

Mth1491 folding mechanism is thus more complex than Tm0979 and may involve the formation of monomeric and oligomeric intermediates. Double jump experiments may help us to understand better the trimer folding.

6 Summary and future work

6.1 Characterization of Tm0979 and Mth1491 folding

Tm0979 and Mth1491 equilibrium and kinetic folding were investigated by measuring their structural transitions by fluorescence spectroscopy and CD. For Tm0979 the mechanism for the structural transition is relatively simple and well defined, whereas for Mth1491 the mechanism is more complex and less well defined.

Considerable equilibrium and kinetic data as well as dimer dissociation measurements by size exclusion chromatrography and dynamic light scattering indicate that Tm0979 unfolds through the formation of a monomeric intermediate. Furthermore, the formation of this intermediate is the rate limiting step during folding, the association step being too fast to be observed. Some characteristics of this intermediate should be further investigated and proposed experiments are described in section 6.1.1. Many other dimeric proteins have also been reported to fold via monomeric intermediates, and the results obtained for Tm0979 are in line with studies on other proteins.

In contrast, for Mth1491, equilibrium studies suggest that the protein unfolds through a 2-state mechanism or through the formation of a trimeric intermediate. The Mth1491 folding kinetics are complex and may involve the formation of a monomeric and/or oligomeric intermediate(s). The complexity of folding of Mth1491 is not surprising because other trimeric proteins were found to exhibit complex kinetics transitions involving the formation of monomeric, dimeric and trimeric intermediates (Guthe *et al.*,2004). Experiments proposed to further define the presence or absence of intermediate on the pathway of Mth1491 folding are described in more detail in section 6.1.1-same section as for Tm0979.

6.1.1 Further validation of folding mechanisms and intermediate formation

In fitting of the equilibrium curves, there is still some uncertainty regarding the spectral properties and energetics for the intermediates of both Tm0979 and Mth1491. In order to address these uncertainties, it would be useful to further test some other global data fitting procedures. In particular, it would be worth exploring which fitted parameters should be restricted (e.g. ΔG_{U1} for dissociation of Tm0979, based on independent measurements of dissocation) or shared (e.g. only m_{U1}, ΔG_{U2} and m_{U2}, or also fluorescence values of different species).

Global fitting may not be sufficient to definitively define when intermediates are populated for Tm0979 and Mth1491. Additional experiments should be considered to define the extent of intermediate formation as a function of GdmCl concentration. Chemical cross-linking experiments may be particularly useful in this regard. For example, crosslinking using glutaradehyde, as a function of GdmCl and protein concentration was very useful in defining the formation of monomeric intermediate by dimeric human superoxide dismutase as a function of GdmCl concentration. Analogous experiment should be conducted for Tm0979 and Mth1491. For Tm0979, this could clarify, for example, whether dimer dissociation is complete before the main optically monitored transition, as is suggested by the current fitted parameters. For Mth1491, it will be key to determine if a trimeric or dimeric form of the protein persists well into the optically monitored transition, and this would prove the formation of an oligomeric intermediate.

Another possibility to better define equilibrium monomer population is to to perform size exclusion chromatography analysis for Tm0979 at different GdmCl concentration. ANS binding experiment at different GdmCl concentration may be considered to detect the formation of the monomer.

Such experiments would address the remaining uncertainties in the global fitting of the GdmCl curves regarding monomer formation.

6.2 YchN-like protein design

One of the long term goals of research on Tm0979 and Mth1491 is rational design of quaternary structure. Accordingly at the start of this thesis mutations were designed to switch Tm0979 from dimer to trimer and to switch Mth1491 from trimer to dimer. The rationale for these designs is summarized below as a basis for future studies. Further details of the design process and rationale are given in my Chem794 research proposal.

6.2.1 Evolution of structure and function in protein families

Different members of protein families often differ in quaternary structure (Orengo *et al.*,2005), however, the evolution of the quaternary structure change and the molecular basis for the change are not well understood. Very few studies have addressed the rational design of switching quaternary structure. Such studies would provide fundamental tests of and insights into the principles governing protein association reactions. With the development of new computational tools for designing protein structures and interfaces, the YchN-like family is an attractive system for conducting such design experiments. In particular, it may be attractive to design Tm0979 to switch from dimer to trimer because the dimer association constant is relatively weak (~5 µM) and so may be relatively easy to disrupt, while the trimer association may be relatively strong which may facilitate creation of binding.

As described in Chapter 1, Tm0979, Mth1491 and YchN share high structure similarity and are proposed to be involved in catalyzing redox reactions involving sulfur metabolism. As is observed in general for different protein families, within this protein family (the YchN-like proteins), structure and function appear to be more conserved than sequence

It is interesting to consider the possible evolutionary relationships between the YchN-like proteins. One possibility is that these three proteins evolved from a more ancient and simple fold, or they could have evolved from each other with an accompanying switch in quaternary structure. Tm0979 is a relatively weak dimer and the dimeric state of the protein may correspond to a way to store the protein inactive. As seen for some DNA binding protein, Tm0979 may switch its oligomeric state from dimer to a more complex to be functionally active. The functionally active Tm0979 may be a homomer formed by association of Tm0979 monomers, a homotrimer as is observed for Mth1491 or perhaps a homohexamer as is observed for YchN. Possible point mutations that could affect a switch in Tm0979 from dimer to trimer were investigated and are described in section 6.2.2.

Another possibility would be the formation of a heterocomplex by association of Tm0979 monomer and other dsr proteins monomers as observed for TusBCD heterohexamer. Possible proteins susceptible of binding Tm0979 were investigated and are described in section 6.2.3.

6.2.2 Design of Tm0979

Sequence alignments of each of Tm0979, Mth1491 and YchN with sequence related proteins suggest that residues buried at the interface are more conserved than those located on the surface (appendices). This is especially true for residues involved in important interactions at the interface, such as Tyr72, Phe75 and Ile76 in helix-α_4 of Tm0979 and Arg103, as described in Chapter 1, section 4. In order to investigate possible point mutations that could bring about a switch in Tm0979 oligomeric state from dimer to trimer, the sequence of Tm0979 and Mth1491, Tm0979 and YchN were aligned based on structural alignment.

6.2.2.1 Mutation design based on sequence and structure comparison

Conserved residues of each of the three proteins were determined by using sequence alignments of proteins belonging to each corresponding sequence families. Related protein sequences were found using Pfam (Bateman *et al.*,2004) and Psiblast (Altschul *et al.*,1990) servers. The sequences were then aligned using Clustal (Aiyar,2000) and Muscle (Edgar, *et al.*, 2004) from EBI/EMBL. The structures of the three proteins were then aligned in order to identify similarities between them. The structure alignments (Table 6.1) were determined using Dali (Holm *et al.*,1993) and SSM (Krissinel *et al.*,2004) servers in addition to a new server using a method based on combinatorial extension of an alignment path defined by aligned fragment pairs (CE) (Shindyalov *et al.*,1998). Mutations were designed to destabilize the wild-type oligomer and stabilize the redesigned oligomer. The mutations to be made are summarized in Figure 6.1 and described further below.

Table 6.1: Structure alignment of primary sequences for Tm0979, Mth1491 and YchN.

Secondary structure	β1		α1		β2	α2
Tm0979	----MA*LVLV*KYGTDHPVEKLKIRSAKA---------ED*KIVL*IQNGVFWA------L---					
Mth1491	MVDY*RVVFHI*DED-DESRVLLLISNVRNLMADLES*VRIEVVA*YSMGVNVL-RRDSEY----					
YchN	--Q*KIVIVA*NGAPYGSESLFNSLRLAIALREQESNL*DLRLFL*MSDAVTAGLRGQKPGEGYN					

Secondary structure	α2′	β3	α3	α3′	β4	α4	β5
Tm0979	-EE-L--ET-----PA*KVYAIKD*DFLARGYSE--EDSKVP*LI*TYSE--FIDLLEGEEKFIG						
Mth1491	-SG-DV--SE*L*TGQGV*RFCAC*SNTLRASGMDGDDLLEGV*DVV*SSGVGHIVRRQTEGW*AYI*RP						
YchN	IQQMLEILTAQ---NVP*VKL*CKTCTDGRGISTLPLIDG*VEIGTL*--VELAQWTLSAD*KVLT*F						

Secondary elements corresponding to the amino acid sequence are indicated by underlined amino acid for α-helices and by underlined and italic amino acid for β-strands. The colours correspond to the conservation of the amino acid among the corresponding family: red, hydrophobic residues, pink, basic residue, green, small and/or polar residue, blue, acidic residue.

Figure 6.1: Proposed designed mutations of the Tm0979 sequence.
A. Mutations proposed on Tm0979 sequence based on Tm0979 and YchN structure alignment consist of the substitutions S21L, Y72L, I76V and G87TF, the insertion of L between A24 and E25. B. Mutations proposed on Tm0979 sequence based on Tm0979 and Mth1491 structure alignment consist of the insertion of LMADLESVR between A24 and E25 and the substitutions I74V, L78R and G81RP.

Based on the alignment of the Tm0979 and YchN structures, Tm0979 amino acids were mutated into residues that should destabilize the actual dimer and favor the formation of a new trimer oligomeric interface (Table 6.2). The same approach was used to design a mutant based on the alignment of Tm0979 and Mth1491 structure. The proposed mutations are summarized in Table 6.3. The secondary and tertiary structures of the mutants were predicted and results in a longer helix-α_4 and the formation of a fifth β-strand. Quaternary structure predictions results in the formation of a mutant trimer with the same interface as the native Mth1491 and YchN trimers.

Tm0979 dimer interface is less complex than Mth1491 trimer interface. Therefore, the evolution of YchN like protein would be conducted by studying the effect of mutation on Tm0979 sequence, first. Two mutants are proposed based on the comparison between the structure and sequence of Tm0979 and YchN monomers and between Tm0979 and Mth1491. These two mutants consist of different amino acids insertions and substitutions. The mutant A involves single point mutation only whereas the mutant B involves insertion of a piece of 9 amino acids. The studies of these two mutants would thus allow us to investigate the nature of the mutation process responsible of the switch of the oligomeric state.

Table 6.2: Proposed mutations on Tm0979 sequence designed to cause a switch of the quaternary structure based on Tm0979 and YchN structure alignment.

YchN hydrophobic residues involved in trimer stabilization	Tm0979 residue structurally aligned to YchN ones.	Proposed mutations	Explanations and consequences.
L24	S21	S21L	Favour the formation of a trimer interface
A27	A24	No mutation	
L28	Gap in Tm0979 sequence	Insertion of Leu beside A24	Destabilize dimer interface by changing conformation of the first α-helix, favour formation of trimer interface
L101	Y72	Y72L	Changes in Tm0979 structure because Y72 conserved among dsrH family, favour formation of trimer interface
A105	I76	I76A	Despite the fact that Ile76 is hydrophobic, I76A substitution make the side chain of this residue shorter destabilising the dimer and favouring formation of trimer interface
L109	E80	No mutation	L78 and L79 in Tm0979 can make hydrophobic interactions to stabilize an eventual trimer interface by a little distortion of the fourth α-helix, so no mutation is designed.
YchN residues involved in trimer stabilization by H-bonding	Tm0979 residue structurally aligned to YchN ones.	Proposed mutations	Explanations and consequences.
T108	L79	L79T	Destabilization of Tm0979 dimer interface and favouring of the formation of a trimer interface
K113	K84	No mutation	
T116	G87	G87TF	Insertion of the last two amino acid of YchN sequence in order to favour the formation of a fifth β-strand and consequently the trimer interface.

The color red indicates a hydrophobic residue, pink is a charged residue and green is a small polar residue. Residues can be compared using Table 6.1.

Table 6.3: Proposed mutations on Tm0979 sequence designed to cause a switch of the quaternary structure based on Tm0979 and Mth1491 structure alignment.

Mth1491 hydrophobic residues involved in trimer stabilization	Tm0979 residue structurally aligned to Mth1491 ones.	Proposed mutations	Explanations and consequences.
L28	-	Insertion of the last amino acids of helix-α_1	Destabilise the dimer interface and favour the formation of a trimer interface by formation of a longer helix-α_1
L32	-		
V97	E74	E74V	Favour formation of trimer interface
V101	I76	I76V	Despite the fact that Ile76 is hydrophobic, I76V substitution make the side chain of this residue shorter destabilising the dimer and favouring formation of trimer interface

Mth1491 residues involved in trimer stabilization by ionic interactions	Tm0979 residue structurally aligned to Mth1491ones	Proposed mutations	Explanations and consequences.
D31	-	Insertion of amino acids of end of helix-α_1	Destabilization of Tm0979 dimer interface and favouring of the formation of a trimer interface by the formation of a longer helix-$\alpha1$
R102	L78	L78R	Destabilisation of Tm0979 dimer interface by introducing a charged residue and favouring the formation of a trimer interface by specific interaction.
R112	**G87**	**G87RP**	Insertion of the last two amino acid of YchN sequence in order to favour the formation of a fifth β-strand and consequently the trimer interface.

Amino acids bolded are conserved in the corresponding protein family. The color red indicates a hydrophobic residue, pink is a charged residue and blue is a negatively charged residue. Residues can be compared using Table 6.1.

References

Adams, M. W. (1994). Biochemical diversity among sulfur-dependent, hyperthermophilic microorganisms. *FEMS Microbiol Rev* 15(2-3): 261-77.

Aiyar, A. (2000). The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment. *Methods Mol Biol.* 132: 221-41.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). Basic local alignment search tool. *J Mol Biol.* 215(3): 403-10.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* 181(96): 223-30.

Apiyo, D., K. Jones, J. Guidry and P. Wittung-Stafshede (2001). Equilibrium unfolding of dimeric desulfoferrodoxin involves a monomeric intermediate: iron cofactors dissociate after polypeptide unfolding. *Biochemistry.* 40(16): 4940-8.

Argos, P. (1988). An investigation of protein subunit and domain interfaces. *Protein Eng* 2(2): 101-13.

Barbar, E., B. Kleinman, D. Imhoff, M. Li, T. S. Hays and M. Hare (2001). Dimerization and folding of LC8, a highly conserved light chain of cytoplasmic dynein. *Biochemistry.* 40(6): 1596-605.

Bashford, D., F. E. Cohen, M. Karplus, I. D. Kuntz and D. L. Weaver (1988). Diffusion-collision model for the folding kinetics of myoglobin. *Proteins* 4(3): 211-27.

Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats and S. R. Eddy (2004). The Pfam protein families database. *Nucleic Acids Res.* 32(Database issue): D138-41.

Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats and S. R. Eddy (2004). The Pfam protein families database. *Nucleic Acids Res* 32(Database issue): D138-41.

Bjelic, S., A. Karshikoff and I. Jelesarov (2006). Stability and folding/unfolding kinetics of the homotrimeric coiled coil Lpp-56. *Biochemistry* 45(29): 8931-9.

Bose, K. and A. C. Clark (2001). Dimeric procaspase-3 unfolds via a four-state equilibrium process. *Biochemistry.* 40(47): 14236-42.

Bowie, J. U. and R. T. Sauer (1989). Equilibrium dissociation and unfolding of the Arc repressor dimer. *Biochemistry.* 28(18): 7139-43.

Chanez-Cardenas, M. E., G. Perez-Hernandez, B. G. Sanchez-Rebollar, M. Costas and E. Vazquez-Contreras (2005). Reversible equilibrium unfolding of triosephosphate isomerase from Trypanosoma cruzi in guanidinium hydrochloride involves stable dimeric and monomeric intermediates. *Biochemistry* 44(32): 10883-92.

Cheng, X., M. L. Gonzalez and J. C. Lee (1993). Energetics of intersubunit and intrasubunit interactions of Escherichia coli adenosine cyclic 3',5'-phosphate receptor protein. *Biochemistry.* 32(32): 8130-9.

Chothia, C. and J. Janin (1975). Principles of protein-protein recognition. *Nature* 256(5520): 705-8.

Christendat, D., V. Saridakis, Y. Kim, P. A. Kumar, X. Xu, A. Semesi, A. Joachimiak, C. H. Arrowsmith and A. M. Edwards (2002). The crystal structure of hypothetical protein MTH1491 from Methanobacterium thermoautotrophicum. *Protein Sci* 11(6): 1409-14.

Christendat, D., A. Yee, A. Dharamsi, Y. Kluger, A. Savchenko, J. R. Cort, V. Booth, C. D. Mackereth, V. Saridakis, I. Ekiel, G. Kozlov, K. L. Maxwell, N. Wu, L. P. McIntosh, K. Gehring, M. A. Kennedy, A. R. Davidson, E. F. Pai, M. Gerstein, A. M. Edwards and C. H. Arrowsmith (2000). Structural proteomics of an archaeon. *Nat Struct Biol* 7(10): 903-9.

Clark, A. C., S. W. Raso, J. F. Sinclair, M. M. Ziegler, A. F. Chaffotte and T. O. Baldwin (1997). Kinetic mechanism of luciferase subunit folding and assembly. *Biochemistry.* 36(7): 1891-9.

Dams, T. and R. Jaenicke (1999). Stability and folding of dihydrofolate reductase from the hyperthermophilic bacterium Thermotoga maritima. *Biochemistry.* 38(28): 9169-78.

De Francesco, R., A. Pastore, G. Vecchio and R. Cortese (1991). Circular dichroism study on the conformational stability of the dimerization domain of transcription factor LFB1. *Biochemistry.* 30(1): 143-7.

de Prat-Gay, G., A. D. Nadra, F. J. Corrales-Izquierdo, L. G. Alonso, D. U. Ferreiro and Y. K. Mok (2005). The folding mechanism of a dimeric beta-barrel domain. *J Mol Biol.* 351(3): 672-82.

Deu, E. and J. F. Kirsch (2007). The unfolding pathway for Apo Escherichia coli aspartate aminotransferase is dependent on the choice of denaturant. *Biochemistry* 46(19): 5810-8.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry* 29(31): 7133-55.

Dirr, H. W. and P. Reinemer (1991). Equilibrium unfolding of class pi glutathione S-transferase. *Biochem Biophys Res Commun.* 180(1): 294-300.

Doyle, S. M., E. H. Braswell and C. M. Teschke (2000). SecA folds via a dimeric intermediate. *Biochemistry.* 39(38): 11667-76.

Ferguson, P. L. and G. S. Shaw (2002). Role of the N-terminal helix I for dimerization and stability of the calcium-binding protein S100B. *Biochemistry* 41(11): 3637-46.

Fersht, A. (1999). Enzyme kinetics. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding.* New York, W.H. Freeman.

Galani, D., A. R. Fersht and S. Perrett (2002). Folding of the yeast prion protein Ure2: kinetic evidence for folding and unfolding intermediates. *J Mol Biol.* 315(2): 213-27.

Galani, D., A. R. Fersht and S. Perrett (2002). Folding of the yeast prion protein Ure2: kinetic evidence for folding and unfolding intermediates. *J Mol Biol.* 315(2): 213-27.

Gaspar, J. A., C. Liu, K. A. Vassall, G. Meglei, R. Stephen, P. B. Stathopulos, A. Pineda-Lucena, B. Wu, A. Yee, C. H. Arrowsmith and E. M. Meiering (2005). A novel member of the YchN-like fold: solution structure of the hypothetical protein Tm0979 from Thermotoga maritima. *Protein Sci* 14(1): 216-23.

Gasset, M., C. Alfonso, J. L. Neira, G. Rivas and M. A. Pajares (2002). Equilibrium unfolding studies of the rat liver methionine adenosyltransferase III, a dimeric enzyme with intersubunit active sites. *Biochem J* 361(Pt 2): 307-15.

Gittelman, M. S. and C. R. Matthews (1990). Folding and stability of trp aporepressor from Escherichia coli. *Biochemistry* 29(30): 7011-20.

Gloss, L. M. and C. R. Matthews (1997). Urea and thermal equilibrium denaturation studies on the dimerization domain of Escherichia coli Trp repressor. *Biochemistry* 36(19): 5612-23.

Gloss, L. M. and C. R. Matthews (1998). The barriers in the bimolecular and unimolecular folding reactions of the dimeric core domain of Escherichia coli Trp repressor are dominated by enthalpic contributions. *Biochemistry.* 37(45): 16000-10.

Gloss, L. M. and C. R. Matthews (1998). Mechanism of folding of the dimeric core domain of Escherichia coli trp repressor: a nearly diffusion-limited reaction leads to the formation of an on-pathway dimeric intermediate. *Biochemistry.* 37(45): 15990-9.

Gloss, L. M. and B. J. Placek (2002). The effect of salts on the stability of the H2A-H2B histone dimer. *Biochemistry* 41(50): 14951-9.

Gloss, L. M., B. R. Simler and C. R. Matthews (2001). Rough energy landscapes in protein folding: dimeric E. coli Trp repressor folds through three parallel channels. *J Mol Biol* 312(5): 1121-34.

Goodsell, D. S. and A. J. Olson (2000). Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 29: 105-53.

Grant, S. K., I. C. Deckman, J. S. Culp, M. D. Minnich, I. S. Brooks, P. Hensley, C. Debouck and T. D. Meek (1992). Use of protein unfolding studies to determine the conformational and dimeric stabilities of HIV-1 and SIV proteases. *Biochemistry.* 31(39): 9491-501.

Grimsley, J. K., J. M. Scholtz, C. N. Pace and J. R. Wild (1997). Organophosphorus hydrolase is a remarkably stable enzyme that unfolds through a homodimeric intermediate. *Biochemistry* 36(47): 14366-74.

Guthe, S., L. Kapinos, A. Moglich, S. Meier, S. Grzesiek and T. Kiefhaber (2004). Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *J Mol Biol* 337(4): 905-15.

Hlodan, R. and R. H. Pain (1995). The folding and assembly pathway of tumour necrosis factor TNF alpha, a globular trimeric protein. *Eur J Biochem* 231(2): 381-7.

Ho, S. P. and W. F. DeGrado (1987). The design of a four-helix bundle protein: Synthesis of peptides which self-associate into a helical protein. *J Am Chem Soc* 109(22): 6751-8.

Holm, L. and C. Sander (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol.* 233(1): 123-38.

Holm, L. and C. Sander (1996). Mapping the protein universe. *Science* 273(5275): 595-603.

Hornby, J. A., J. K. Luo, J. M. Stevens, L. A. Wallace, W. Kaplan, R. N. Armstrong and H. W. Dirr (2000). Equilibrium folding of dimeric class mu glutathione transferases involves a stable monomeric intermediate. *Biochemistry.* 39(40): 12336-44.

Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold Des* 3(4): R81-91.

Jana, R., T. R. Hazbun, A. K. Mollah and M. C. Mossing (1997). A folded monomeric intermediate in the formation of lambda Cro dimer-DNA complexes. *J Mol Biol.* 273(2): 402-16.

Janin, J. and C. Chothia (1990). The structure of protein-protein recognition sites. *J Biol Chem* 265(27): 16027-30.

Jia, H., W. J. Satumba, G. L. Bidwell, 3rd and M. C. Mossing (2005). Slow assembly and disassembly of lambda Cro repressor dimers. *J Mol Biol* 350(5): 919-29.

Jones, S. and J. M. Thornton (1995). Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol* 63(1): 31-65.

Jones, S. and J. M. Thornton (1996). Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 93(1): 13-20.

Kadokura, H., F. Katzen and J. Beckwith (2003). Protein disulfide bond formation in prokaryotes. *Annu Rev Biochem* 72: 111-35.

Kaplan, W., P. Husler, H. Klump, J. Erhardt, N. Sluis-Cremer and H. Dirr (1997). Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag. *Protein Sci* 6(2): 399-406.

Karplus, M. and D. L. Weaver (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* 3(4): 650-68.

Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14: 1-63.

Kim, D. H., D. S. Jang, G. H. Nam and K. Y. Choi (2001). Folding mechanism of ketosteroid isomerase from Comamonas testosteroni. *Biochemistry* 40(16): 5011-7.

Kim, D. H., D. S. Jang, G. H. Nam, S. Yun, J. H. Cho, G. Choi, H. C. Lee and K. Y. Choi (2000). Equilibrium and kinetic analysis of folding of ketosteroid isomerase from Comamonas testosteroni. *Biochemistry.* 39(42): 13084-92.

Kim, P. S. and R. L. Baldwin (1990). Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* 59: 631-60.

Koradi, R., M. Billeter and K. Wuthrich (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 14(1): 51-5, 29-32.

Kretschmar, M. and R. Jaenicke (1999). Stability of a homo-dimeric Ca(2+)-binding member of the beta gamma-crystallin superfamily: DSC measurements on spherulin 3a from Physarum polycephalum. *J Mol Biol* 291(5): 1147-53.

Kretschmar, M., E. M. Mayr and R. Jaenicke (1999). Kinetic and thermodynamic stabilization of the betagamma-crystallin homolog spherulin 3a from Physarum polycephalum by calcium binding. *J Mol Biol* 289(4): 701-5.

Krissinel, E. and K. Henrick (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D* 60(12): 2256-2268.

Krissinel, E. and K. Henrick (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1): 2256-68.

Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* 6(2): 87-103.

Lambeir, A. M., J. Backmann, J. Ruiz-Sanz, V. Filimonov, J. E. Nielsen, I. Kursula, B. V. Norledge and R. K. Wierenga (2000). The ionization of a buried glutamic acid is thermodynamically linked to the stability of Leishmania mexicana triose phosphate isomerase. *Eur J Biochem* 267(9): 2516-24.

Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique* 65(1): 44.

Liang, H. and T. C. Terwilliger (1991). Reversible denaturation of the gene V protein of bacteriophage f1. *Biochemistry.* 30(11): 2772-82.

Mainfroid, V., S. C. Mande, W. G. Hol, J. A. Martial and K. Goraj (1996). Stabilization of human triosephosphate isomerase by improvement of the stability of individual alpha-helices in dimeric as well as monomeric forms of the protein. *Biochemistry* 35(13): 4110-7.

Maiti, R., G. H. Van Domselaar, H. Zhang and D. S. Wishart (2004). SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res* 32(Web Server issue): W590-4.

Maity, H., M. C. Mossing and M. R. Eftink (2005). Equilibrium unfolding of dimeric and engineered monomeric forms of lambda Cro (F58W) repressor and the effect of added salts: evidence for the formation of folded monomer induced by sodium perchlorate. *Arch Biochem Biophys* 434(1): 93-107.

Malecki, J. and Z. Wasylewski (1997). Stability and kinetics of unfolding and refolding of cAMP receptor protein from Escherichia coli. *Eur J Biochem.* 243(3): 660-9.

Mallam, A. L. and S. E. Jackson (2005). Folding studies on a knotted protein. *J Mol Biol.* 346(5): 1409-21. Epub 2005 Jan 28.

Mallam, A. L. and S. E. Jackson (2005). Folding studies on a knotted protein. *J Mol Biol* 346(5): 1409-21.

Mallam, A. L. and S. E. Jackson (2006). Probing nature's knots: the folding pathway of a knotted homodimeric protein. *J Mol Biol.* 359(5): 1420-36. Epub 2006 May 2.

Mallam, A. L. and S. E. Jackson (2007). A Comparison of the Folding of Two Knotted Proteins: YbeA and YibK. *J Mol Biol.* 366(2): 650-65. Epub 2006 Nov 10.

Marti, D. N., S. Bjelic, M. Lu, H. R. Bosshard and I. Jelesarov (2004). Fast folding of the HIV-1 and SIV gp41 six-helix bundles. *J Mol Biol* 336(1): 1-8.

Martin, A. C., C. A. Orengo, E. G. Hutchinson, S. Jones, M. Karmirantzou, R. A. Laskowski, J. B. Mitchell, C. Taroni and J. M. Thornton (1998). Protein folds and functions. *Structure* 6(7): 875-84.

Mason, J. M., U. B. Hagemann and K. M. Arndt (2007). Improved stability of the Jun-Fos activator protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis. *J Biol Chem*.

Mateu, M. G. (2002). Conformational stability of dimeric and monomeric forms of the C-terminal domain of human immunodeficiency virus-1 capsid protein. *J Mol Biol.* 318(2): 519-31.

Mazzini, A., A. Maia, M. Parisi, R. T. Sorbi, R. Ramoni, S. Grolli and R. Favilla (2002). Reversible unfolding of bovine odorant binding protein induced by guanidinium hydrochloride at neutral pH. *Biochim Biophys Acta* 1599(1-2): 90-101.

Milla, M. E. and R. T. Sauer (1994). P22 Arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry* 33(5): 1125-33.

Mizuguchi, K., C. M. Deane, T. L. Blundell and J. P. Overington (1998). HOMSTRAD: a database of protein structure alignments for homologous families. *Protein Sci* 7(11): 2469-71.

Mok, Y. K., G. de Prat Gay, P. J. Butler and M. Bycroft (1996). Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain. *Protein Sci.* 5(2): 310-9.

Monera, O. D., G. S. Shaw, B. Y. Zhu, B. D. Sykes, C. M. Kay and R. S. Hodges (1992). Role of interchain alpha-helical hydrophobic interactions in Ca2+ affinity, formation, and stability of a two-site domain in troponin C. *Protein Sci* 1(7): 945-55.

Monod, J., J. Wyman and J. P. Changeux (1965). On the Nature of Allosteric Transitions: a Plausible Model. *J Mol Biol* 12: 88-118.

Murzin, A. G., S. E. Brenner, T. Hubbard and C. Chothia (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247(4): 536-40.

Myers, J. K., C. N. Pace and J. M. Scholtz (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 4(10): 2138-48.

Najera, H., M. Costas and D. A. Fernandez-Velasco (2003). Thermodynamic characterization of yeast triosephosphate isomerase refolding: insights into the interplay between function and stability as reasons for the oligomeric nature of the enzyme. *Biochem J* 370(Pt 3): 785-92.

Nichtl, A., J. Buchner, R. Jaenicke, R. Rudolph and T. Scheibel (1998). Folding and association of beta-Galactosidase. *J Mol Biol.* 282(5): 1083-91.

Nölting, B. (2006). Protein folding kinetics : biophysical methods. Berlin ; New York, Springer.

Nooren, I. M. and J. M. Thornton (2003). Diversity of protein-protein interactions. *Embo J* 22(14): 3486-92.

Orengo, C. A., A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells and J. M. Thornton (1997). CATH--a hierarchic classification of protein domain structures. *Structure* 5(8): 1093-108.

Orengo, C. A. and J. M. Thornton (2005). Protein families and their evolution-a structural perspective. *Annu Rev Biochem* 74: 867-900.

Pan, H., A. S. Raza and D. L. Smith (2004). Equilibrium and kinetic folding of rabbit muscle triosephosphate isomerase by hydrogen exchange mass spectrometry. *J Mol Biol* 336(5): 1251-63.

Placek, B. J. and L. M. Gloss (2005). Three-state kinetic folding mechanism of the H2A/H2B histone heterodimer: the N-terminal tails affect the transition state between a dimeric intermediate and the native dimer. *J Mol Biol.* 345(4): 827-36.

Ponstingl, H., T. Kabir, D. Gorse and J. M. Thornton (2005). Morphological aspects of oligomeric protein structures. *Prog Biophys Mol Biol* 89(1): 9-35.

Pott, A. S. and C. Dahl (1998). Sirohaem sulfite reductase and other proteins encoded by genes at the dsr locus of Chromatium vinosum are involved in the oxidation of intracellular sulfur. *Microbiology* 144 (Pt 7): 1881-94.

Ptitsyn, O. B. (1973). [Stages in the mechanism of self-organization of protein molecules]. *Dokl Akad Nauk SSSR* 210(5): 1213-5.

Ptitsyn, O. B. (1995). How the molten globule became. *Trends Biochem Sci* 20(9): 376-9.

Reece, L. J., R. Nichols, R. C. Ogden and E. E. Howell (1991). Construction of a synthetic gene for an R-plasmid-encoded dihydrofolate reductase and studies on the role of the N-terminus in the protein. *Biochemistry.* 30(45): 10895-904.

Rietveld, A. W. and S. T. Ferreira (1998). Kinetics and energetics of subunit dissociation/unfolding of TIM: the importance of oligomerization for conformational persistence and chemical stability of proteins. *Biochemistry* 37(3): 933-7.

Riley, P. W., H. Cheng, D. Samuel, H. Roder and P. N. Walsh (2007). Dimer dissociation and unfolding mechanism of coagulation factor XI apple 4 domain: spectroscopic and mutational analysis. *J Mol Biol* 367(2): 558-73.

Rosengarth, A., J. Rosgen and H. J. Hinz (1999). Slow unfolding and refolding kinetics of the mesophilic Rop wild-type protein in the transition range. *Eur J Biochem* 264(3): 989-95.

Rousseau, F., J. W. Schymkowitz, H. R. Wilkinson and L. S. Itzhaki (2002). The structure of the transition state for folding of domain-swapped dimeric p13suc1. *Structure.* 10(5): 649-57.

Rousseau, F., J. W. Schymkowitz, H. R. Wilkinson and L. S. Itzhaki (2004). Intermediates control domain swapping during folding of p13suc1. *J Biol Chem.* 279(9): 8368-77. Epub 2003 Dec 7.

Ruller, R., T. L. Ferreira, A. H. de Oliveira and R. J. Ward (2003). Chemical denaturation of a homodimeric lysine-49 phospholipase A2: a stable dimer interface and a native monomeric intermediate. *Arch Biochem Biophys* 411(1): 112-20.

Satumba, W. J. and M. C. Mossing (2002). Folding and assembly of lambda Cro repressor dimers are kinetically limited by proline isomerization. *Biochemistry* 41(48): 14216-24.

Shin, D. H., H. Yokota, R. Kim and S. H. Kim (2002). Crystal structure of a conserved hypothetical protein from Escherichia coli. *J Struct Funct Genomics* 2(1): 53-66.

Shindyalov, I. N. and P. E. Bourne (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* 11(9): 739-47.

Simler, B. R., B. L. Doyle and C. R. Matthews (2004). Zinc binding drives the folding and association of the homo-trimeric gamma-carbonic anhydrase from Methanosarcina thermophila. *Protein Eng Des Sel* 17(3): 285-91.

Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, J. N. Reeve and et al. (1997). Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics. *J Bacteriol* 179(22): 7135-55.

Stadtman, E. R. (1990). Metal ion-catalyzed oxidation of proteins: biochemical mechanism and biological consequences. *Free Radic Biol Med* 9(4): 315-25.

Svensson, A. K., O. Bilsel, E. Kondrashkina, J. A. Zitzewitz and C. R. Matthews (2006). Mapping the folding free energy surface for metal-free human Cu,Zn superoxide dismutase. *J Mol Biol.* 364(5): 1084-102. Epub 2006 Sep 7.

Talbott, M., M. Hare, A. Nyarko, T. S. Hays and E. Barbar (2006). Folding is coupled to dimerization of Tctex-1 dynein light chain. *Biochemistry.* 45(22): 6793-800.

Tanford, C. (1968). Protein denaturation. *Adv Protein Chem* 23: 121-282.

Tanford, C. (1970). Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24: 1-95.

Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova and E. V. Koonin (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 29(1): 22-8.

Timm, D. E., P. L. de Haseth and K. E. Neet (1994). Comparative equilibrium denaturation studies of the neurotrophins: nerve growth factor, brain-derived neurotrophic factor, neurotrophin 3, and neurotrophin 4/5. *Biochemistry* 33(15): 4667-76.

Timm, D. E. and K. E. Neet (1992). Equilibrium denaturation studies of mouse beta-nerve growth factor. *Protein Sci* 1(2): 236-44.

Topping, T. B. and L. M. Gloss (2004). Stability and folding mechanism of mesophilic, thermophilic and hyperthermophilic archael histones: the importance of folding intermediates. *J Mol Biol.* 342(1): 247-60.

Topping, T. B., D. A. Hoch and L. M. Gloss (2004). Folding mechanism of FIS, the intertwined, dimeric factor for inversion stimulation. *J Mol Biol.* 335(4): 1065-81.

Tsai, C. J., S. L. Lin, H. J. Wolfson and R. Nussinov (1997). Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci* 6(1): 53-64.

Wales, T. E., J. S. Richardson and M. C. Fitzgerald (2004). Facile chemical synthesis and equilibrium unfolding properties of CopG. *Protein Sci* 13(7): 1918-26.

Wallace, L. A. and H. W. Dirr (1999). Folding and assembly of dimeric human glutathione transferase A1-1. *Biochemistry* 38(50): 16686-94.

Wallace, L. A., N. Sluis-Cremer and H. W. Dirr (1998). Equilibrium and kinetic unfolding properties of dimeric human glutathione transferase A1-1. *Biochemistry.* 37(15): 5320-8.

Wendt, H., C. Berger, A. Baici, R. M. Thomas and H. R. Bosshard (1995). Kinetics of folding of leucine zipper domains. *Biochemistry* 34(12): 4097-107.

Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* 70(3): 697-701.

Wetlaufer, D. B. (1990). Nucleation in protein folding--confusion of structure and process. *Trends Biochem Sci* 15(11): 414-5.

Wojciak, P., A. Mazurkiewicz, A. Bakalova and R. Kuciel (2003). Equilibrium unfolding of dimeric human prostatic acid phosphatase involves an inactive monomeric intermediate. *Int J Biol Macromol* 32(1-2): 43-54.

Young, L., R. L. Jernigan and D. G. Covell (1994). A role for surface hydrophobicity in protein-protein recognition. *Protein Sci* 3(5): 717-29.

Zeeb, M., G. Lipps, H. Lilie and J. Balbach (2004). Folding and association of an extremely stable dimeric protein from Sulfolobus islandicus. *J Mol Biol* 336(1): 227-40.

Zhu, L., X. J. Zhang, L. Y. Wang, J. M. Zhou and S. Perrett (2003). Relationship between stability of folding intermediates and amyloid formation for the yeast prion Ure2p: a quantitative analysis of the effects of pH and buffer system. *J Mol Biol.* 328(1): 235-54.

Zitzewitz, J. A., O. Bilsel, J. Luo, B. E. Jones and C. R. Matthews (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry.* 34(39): 12812-9.

Zitzewitz, J. A., O. Bilsel, J. Luo, B. E. Jones and C. R. Matthews (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry* 34(39): 12812-9.

Appendix A.1: Preparation of Tm0979 denaturation curve samples.

Tm0979 denaturation curves @ 1 µM

Stock solution 10x buffer*. pH=6		
[phosphate] =	250	mM
[NaCl] =	4.5	M

*stored in the -20°C freezer as 1.5mL aliquots

Stock solution [Tm0979] = 10µM		
V total =	1700	µL
V protein =	68	µL
V buffer =	170	µL
V water =	1462	µL

Sample	[GdmCl]	V GdmCl	V protein	V buffer	V water	V total
1	0	0	40	40	320	400
2	0.2	10	40	40	310	400
3	0.4	20	40	40	300	400
4	0.6	30	40	40	290	400
5	0.8	40	40	40	280	400
6	1	50	40	40	270	400
7	1.2	60	40	40	260	400
8	1.4	70	40	40	250	400
9	1.6	80	40	40	240	400
10	1.8	90	40	40	230	400
11	2	100	40	40	220	400
12	2.1	105	40	40	215	400
13	2.2	110	40	40	210	400
14	2.3	115	40	40	205	400
15	2.4	120	40	40	200	400
16	2.5	125	40	40	195	400
17	2.6	130	40	40	190	400
18	2.7	135	40	40	185	400
19	2.8	140	40	40	180	400
20	2.9	145	40	40	175	400
21	3	150	40	40	170	400
22	3.1	155	40	40	165	400
23	3.2	160	40	40	160	400
24	3.3	165	40	40	155	400
25	3.4	170	40	40	150	400
26	3.5	175	40	40	145	400
27	3.6	180	40	40	140	400
28	3.7	185	40	40	135	400
29	3.8	190	40	40	130	400
30	3.9	195	40	40	125	400
31	4	200	40	40	120	400
32	4.2	210	40	40	110	400
33	4.4	220	40	40	100	400
34	4.6	230	40	40	90	400
35	4.8	240	40	40	80	400
36	5	250	40	40	70	400
37	5.2	260	40	40	60	400
38	5.4	270	40	40	50	400
39	5.6	280	40	40	40	400
40	5.8	290	40	40	30	400
41	6	300	40	40	20	400
sum =		6150	1640	1640	6970	16400

Appendix A.2: Preparation of Mth1491 renaturation curve samples.

Mth1491 renaturation curves @ 1 µM

Stock solution [Mth1491] = 10µM

V total =	1700	µL
V GdmCl =	1275	µL
V EDTA+DTT =	17	µL
V protein =	68	µL
V buffer =	170	µL
V water =	170	µL

Stock solution 10x buffer*. pH=6

[citrate] =	200	mM
[NaCl] =	4.5	M

*stored in the -20°C freezer as 1.5mL aliquots

Stock solution EDTA+DTT‡. pH=6

[EDTA] =	100	mM
[DTT] =	1	M

‡prepared using degassed milliQ water and stored in the -20°C freezer as 500uL aliquots

Sample	[GdmCl]	V GdmCl (8M)	V protein	V EDTA+DTT†	V 10xbuffer†	V water	V total
1	0.6	0	40	4	40	316	400
2	0.8	10	40	4	40	306	400
3	1	20	40	4	40	296	400
4	1.2	30	40	4	40	286	400
5	1.4	40	40	4	40	276	400
6	1.6	50	40	4	40	266	400
7	1.8	60	40	4	40	256	400
8	2	70	40	4	40	246	400
9	2.2	80	40	4	40	236	400
10	2.4	90	40	4	40	226	400
11	2.6	100	40	4	40	216	400
12	2.7	105	40	4	40	211	400
13	2.8	110	40	4	40	206	400
14	2.9	115	40	4	40	201	400
15	3	120	40	4	40	196	400
16	3.1	125	40	4	40	191	400
17	3.2	130	40	4	40	186	400
18	3.3	135	40	4	40	181	400
19	3.4	140	40	4	40	176	400
20	3.5	145	40	4	40	171	400
21	3.6	150	40	4	40	166	400
22	3.7	155	40	4	40	161	400
23	3.8	160	40	4	40	156	400
24	3.9	165	40	4	40	151	400
25	4	170	40	4	40	146	400
26	4.1	175	40	4	40	141	400
27	4.2	180	40	4	40	136	400
28	4.3	185	40	4	40	131	400
29	4.4	190	40	4	40	126	400
30	4.5	195	40	4	40	121	400
31	4.6	200	40	4	40	116	400
32	4.8	210	40	4	40	106	400
33	5	220	40	4	40	96	400
34	5.2	230	40	4	40	86	400
35	5.4	240	40	4	40	76	400
36	5.6	250	40	4	40	66	400
37	5.8	260	40	4	40	56	400
38	6	270	40	4	40	46	400
sum =		5280	1520	152	1520	6728	15200

†EDTA+DTT and 10x buffer added together

Appendix A.3: Thermodynamic and kinetic parameters characteristic of chemical induced unfolding of dimeric proteins described by a two-state transition

Protein	ref	Pdb Code	Chain length		Interface Characteristics			denaturant	$\Delta G\ (H_2O)$ (kcal/mol)	m (kcal Mol^{-1} M^{-1})	kinetic mechanism	$k_u(H_2O)$ (s-1)	m_u (kcal mol^{-1} M^{-1})	$k_f(H_2O)$ (M^{-1}s^{-1})	m_f (kcal mol^{-1} M^{-1})	β
					IA[a] (Å^2)	%MB[b]	%P[c]									
α-proteins:																
leucine zipper:																
LZ(12A16A)	(Wendt et al.,1995)		29					Urea	7.7	0.7	$N_2 \leftrightarrow 2U$	1.3x10^1	-	4.0x10^6	-	-
LZ(12A)	(Wendt et al.,1995)		29					Urea	9.1	0.9	$N_2 \leftrightarrow I_2$ $I_2 \leftrightarrow 2U$	1.3x10^0 2.3x10^0	-	- 6.2x10^6	-	-
GCN4-p1	(Zitzewitz et al.,1995)	2ZTA	33	α	817	25	13	GdmCl	10.2	1.8	$N_2 \leftrightarrow 2U$	3.3x10^{-3}	0.9	4.2x10^5	0.9	0.5
cFos-JunW	(Mason et al.,2007)		37					GdmCl	7.7	1.8	$N_2 \leftrightarrow I_2$ $I_2 \leftrightarrow 2U$	1.2x10^0 -	1.1 -	3.4x10^0(s^{-1}) 1.5x10^6	0.6 0.5	0.4 -
4-helix bundle																
LFB1	(De Francesco et al.,1991)	1JB6	33	α	1451	26	45	GdmCl	12.0	1.3						
TnC domain (LFIL)	(Monera et al.,1992)	1CTA	34	α	685	23	8	GdmCl	11.0	2.2						
α2(PRR) (artificial)	(Ho et al.,1987)		35	α				GdmCl	12.8	1.3						
ROP	(Rosengarth et al.,1999)	1ROP	63	α	1334	31	20	GdmCl	16.9	3.1	$N_2 \leftrightarrow 2U$	7.1x10^{-7}	0.5	1.8x10^6	2.6	0.8
ISSADD:																
hFoB	(Topping et al.,2004)	1BFM	67	α	-	-	-	GdmCl GdmCl*	5.7* 10.8	- 3.4	$N_2 \rightarrow 2U$ $N_2 \rightarrow 2U$	9.0x10^{-2} 3.6x10^{-3}	1.1 0.9	- 8.1x10^3	- 0.4	- 0.3
hMfB	(Topping et al.,2004)	1A7W	69	α	1619	31	20	GdmCl	14.1	3.8	$N_2 \leftrightarrow 2U$	9.0x10^{-4}	1.6	3.1x10^6	1.3	0.4
hPyA1	(Topping et al.,2004)	-	66	α	-	-	-	GdmCl	16.1	3.8	$N_2 \leftrightarrow 2M$ $M \leftarrow U$	3.6x10^{-5}	1.3	9.0x10^5 burst	1.7	0.6
FIS	(Topping et al.,2004)	1ETY	98	α	2270	36	24	Urea	15.2	2.9	$N_2 \leftrightarrow I_2$ $I_2 \leftarrow 2U$	6.8x10^{-3} -	0.7 -	3.3x10^1(s^{-1}) >1.0x10^7	0.6 -	0.7 -
								GdmCl	13.5	5.0	$N_2 \leftrightarrow I_2$ $I_2 \leftarrow 2U$	2.0x10^{-2} -	1.4 -	5.0x10^0(s^{-1}) >1.0x10^7	0.4 -	0.5 -
TR	(Gittelman et al.,1990; Gloss et al.,2001)	3WRP	108	α	2218	29	22	Urea	24.0	3.1		Complex	kinetics			
TR [2-66]$_2$	(Gloss et al.,1997; Gloss et al.,1998; Gloss et al.,1998)	3WRP (8-66)	65	α	1370	26	23	Urea	14.4	2.0	$N_2 \leftrightarrow I_2$ $I_2 \leftarrow 2U$	6.5x10^{-1} -	0.2 -	4.6x10^1(s^{-1}) 7.0x10^8	0.6 -	0.7 -

								Urea/GdmCl			Mechanism					
H2A/H2B	(Gloss et al.,2002; Placek et al.,2005)	1AOI (C,D)	119/122	α	2543	26/28	19	Urea	11.8	2.9	$N_2 \leftrightarrow I_2$ $I_2 \leftarrow 2U$	6.0×10^{-2} -	0.5 -	$5.7 \times 10^0 (s^{-1})$ $>1.0 \times 10^8$	1.1 -

other:

S100B	(Ferguson et al.,2002)	1UWO	91	α	1291	20	20	GdmCl	17.2	3.3					

α/β proteins:

Arc repressor	(Bowie et al.,1989; Milla et al.,1994)	1ARR	53	α/β	1964	37	29	Urea GdmCl	11.0 11.1	1.91 3.27	$N_2 \leftrightarrow 2U$	2.0×10^{-1} -	0.4* -	9.0×10^6 -	1.0 -	0.7 -
CopG	(Wales et al.,2004)	2CPG	56	α/β	1583	22	36									
ORF56	(Zeeb et al.,2004)	*						GdmCl	19.8	2.3	$N_2 \leftrightarrow 2U$	1.8×10^{-7}	1.3	7.0×10^7	2.4	0.6
λ Cro	(Jana et al.,1997)	5CRO	66	α/β	676	13	27	GdmCl	11.2	5.6*						
λ Cro (F58W)	(Satumba et al.,2002; Jia et al.,2005; Maity et al.,2005)	1D1L*	61					Urea GdmCl GdmCl	11.4 13.6 11.0	1.6 5.1* 3.4	$N_2 \leftrightarrow 2M$	3.0×10^{-1}	-	1.9×10^4	-	-
Tctex-1	(Talbott et al.,2006)	1YGT	111	α/β	1580	21	30	GdmCl	19.8	5.4						
p13^{suc1}: Δ8789	(Rousseau et al.,2002; Rousseau et al.,2004)	1PUC	112	α/β	2144	25	26	Urea	~20*	3.4	$N_2 \leftrightarrow I_2$ $I_2 \leftarrow 2U$	2.5×10^{-5} -	0.8 -	$2.8 \times 10^0 (s^{-1})$ -	-	-
KSI	(Kim et al.,2000; Kim et al.,2001)	8CHO	125	α/β	1067	15	35	Urea	22.0	4.0	$N_2 \leftarrow I_2$ $I_2 \leftarrow 2I$ $I \leftarrow U$	-	-	$1.7 \times 10^{-2} (s^{-1})$ 5.4×10^4 6.0×10^1	-	-
TmDHFR	(Dams et al.,1999)	1CZ3	168	α/β	1264	14	23	GdmCl Urea	33.9 34.4	9.3 4.7	na	4.6×10^{-12} -	1.3 -	$1.4 \times 10^{-1} (s^{-1})$ 2.0×10^{-2} -	1.0 0.9 -	- -
hGSTA1-1	(Wallace et al.,1998; Wallace et al.,1999)	1PKW	222	α/β	1495	13	25	Urea	26.8	4.2	$N_2 \rightarrow I_2$ $I_2 \rightarrow 2U$ $N_2 \leftarrow I_2$ $I_2 \leftarrow 2M$ $M \leftarrow U$	6.1×10^{-1} 2.9×10^{-5}	0.1 0.3	$6.0 \times 10^{-3} (s^{-1})$ 7.6×10^5 Burst*	0.1 0.3	
pGSTP1-1	(Dirr et al.,1991)	2GSR	207	α/β	1366	13	29	GdmCl Urea	23.3 27.2	11.1 4.6						
Sj26GST	(Kaplan et al.,1997)	1M9A	218	α/β	1342	12	29	Urea	26.0	4.5						
rTim	(Rietveld et al.,1998; Pan et al.,2004)	1R2T	248	α/β	3225	15	14	GdmCl	33.6	39.4*	$N_2 \rightarrow 2M$ $M \leftarrow U$	2.8×10^{-5}	1.9 -	$1.9 \times 10^2 (s^{-1})$	- 12.9	-
hTim	(Mainfroid et al.,1996)							Urea	19.3	1.7						

131

Protein	Reference	PDB		Fold	IA[a]	%MB[b]	%P[#]	Denaturant	ΔG	m	Mechanism	k_u	β	k_f	m	β
LmTim	(Lambeir et al.,2000)	1AMK						GdmCl	19.7	4.4						
HPV-16 E2	(Mok et al.,1996; de Prat-Gay et al.,2005)	1ZZF	81	α/β	887	15	22	Urea	11.0	3.1	$N_2 \leftarrow 2M$ $N_2 \leftarrow I_2$ $M \leftarrow U$	9.0×10^{-3} 5.0×10^{-1}	0.3 ~0	2.4×10^5 $2.0 \times 10^{-1} (s^{-1})$ $3.1 \times 10^1 (s^{-1})$	1.0 1.0 0.6	

β-proteins:

Protein	Reference	PDB		Fold	IA[a]	%MB[b]	%P[#]	Denaturant	ΔG	m	Mechanism	k_u	β	k_f	m	β
R67 DHFR*	(Reece et al.,1991)	-	-	β	-	-	-	GdmCl	13.2	3.5		-	-	-	-	-
Gene V protein	(Liang et al.,1991)	1VQB	87	β	870	13	18	GdmCl	16.3	3.6	$N_2 \leftrightarrow 2U$	3.0×10^{-4}	0.8	1.1×10^7	-3.1	0.8
HIV-1 protease	(Grant et al.,1992)	1Z1H	99	β	1874	28	27	Urea	14.2							
SIV-1 protease	(Grant et al.,1992)	1TCW	99	β	1760	26	29	Urea	13.3							
S3a	(Kretschmar et al.,1999; Kretschmar et al.,1999)	1HDF	102	β	152	3	34	GdmCl	19.4	8.2	na	6.1×10^{-6}	1.1			

neurotrophins:

Protein	Reference	PDB		Fold	IA[a]	%MB[b]	%P[#]	Denaturant	ΔG	m	Mechanism	k_u	β	k_f	m	β
mNGF	(Timm et al.,1992)	1BET	118	β	1411	20	23	GdmCl	19.3	4.8						
hNGF	(Timm et al.,1994)	1WWW (V,W)	118	β	1658	21	24	GdmCl	23.0	4.3						
BDNF	(Timm et al.,1994)		119	β				GdmCl	26.4	5.3						
NT-3	(Timm et al.,1994)	1B8K	119	β	1293	20	27	GdmCl	22.7	4.5						
NT-4/5	(Timm et al.,1994)	1HCF (A,B)	130	β	1752	21	24	GdmCl	20.8	5.1						

The free energy of unfolding, ΔG, is the energy difference between the unfolded and folded state protein, *m* value is related to changes in solvent exposure area during unfolding. k_u and k_f are the unfolding and folding rate respectively. β-value is a measure of the degree of compaction of the transition state, it is calculated as follows $m_f/(m_u+m_f)$

[a] IA= Interface Area per monomer ($Å^2$), calculation for interface characterisation made using Getarea (http://pauli.utmb.edu/cgi-bin/get_a_form.tcl): total interface area per dimer and is calculated as follows: Interface area = (2 x surface area of monomer – surface area of dimer)/2.

[b] % MB= percent monomer buried, the % of surface area buried within the interface per monomer and is calculated as follow: %monomer buried = (Interface Area) / (surface area of monomer).

[#] %P= percent polar, is the amount of polar residues buried at the interface per monomer and is calculated as follow: % polar residues = (2 x polar surface area of monomer – polar surface area of dimer) / (interface area).

na=the kinetic mechanism has not been determined

Appendix A.4: Thermodynamic and kinetic parameters characteristic of chemical induced unfolding of dimeric proteins described by a three-state transition

Protein		Ref	Pdb code	Chain length		I^a (Å²)	%M[b]	%P[c]	denaturant	ΔG (kcal/mol)	m (kcal mol^{-1} M^{-1})	Kinetic Mechanism	k_u (H₂O) (s-1)	m_u (kcal mol^{-1} M^{-1})	k_f(H₂O) (M^{-1}s-1)	m_f (kcal mol^{-1} M^{-1})	β
α-helical																	
CA-C	N₂↔2M	(Mateu,2002)	1A43	85	α	927	19	34	GdmCl	12.1	3.0						
of HIV-1	M↔U									4.5	1.8						
	N₂↔2U									**21.1**	**6.6**						
GR-LBD	N₂↔I₂	(Ferguson et	1M2Z	257	α	672	5	48	GdmCl	-	-						
	I₂↔2U	al.,2002)								19.5	2.3						
	N₂↔2U									-	-						
BthTx-1	N₂↔2M	(Ruller et	1QLL	121	α	458	6.3	40	GdmCl	10.1	1.4						
	M↔U	al.,2003)								7.2	1.9						
	N₂↔2U									**24.5**	**5.2**						
α & β proteins																	
LC8	N₂↔2M	(Barbar et	1CMI	85	α/β	293	5.4	44	GdmCl	8.4	1.8						
	M↔U	al.,2001)								7.5	2.0						
	N₂↔2U									**23.4**	**5.8**						
FXI	N₂↔2M	(Riley et	2F83	91	α/β	898	16	43	GdmCl	9.5	1.2						
a4 C321S	M↔U	al.,2007)	(F272-							2.6	0.6						
	N₂↔2U		E362							**14.7**	**2.4**						
CRP	N₂↔2M	(Cheng et	1I5Z	209	α/β	1267	11	11	GdmCl	12	2.4	N₂↔2M	1.2×10^1	-	-	-	-
	M↔U	al.,1993;								7.2	2.5	M↔U	9.4×10^{-5}	1.4	$1.2\times10^2(s^{-1})$	1.4	0.5
	N₂↔2U	Malecki et								**26.4**	**7.4**						
		al.,1997)															
DsbC	N₂↔2M	(Bjelic et	1EEJ	216	α/β	930	8	31	GdmCl	11.7	3.6	N₂→2M	6.8×10^0	-	-	-	
	M↔U	al.,2006)								13.9	6.0	na	6.3×10^{-1}	-	$4.8\times10^{-1}(s^{-1})$	-	
	N₂↔2U									**39.5**	**15.6**	na.	4×10^{-2}	-	$2.7\times10^{-2}(s^{-1})$	-	
GSTM1-1	N₂↔2M	(Hornby et	1GSU	219	α/β	1293	12	38	Urea	10.8	1.0						
	M↔U	al.,2000)								16.5	3.4						
	N₂↔2U									**43.8**	**7.8**						
	N₂↔2M								GdmCl	9.2	2.2						
	M↔U									16.3	5.1						
	N₂↔2U									**41.8**	**12.4**						
GSTM2-2	N₂↔2M	(Hornby et	1HNB	217	α/β	1697	14	32	Urea	12.4	1.8						
	M↔U	al.,2000)								14.8	3.1						
	N₂↔2U									**42.0**	**8.0**						
	N₂↔2M								GdmCl	9.8	3.3						
	M↔U									15.8	6.0						
	N₂↔2U									**41.4**	**15.3**						
Procaspase-3	N₂↔I₂	(Bose et	-	277	α/β	-	-	-	Urea	8.3	2.8						
	I₂↔2M	al.,2001)								10.5	0.5						
	M↔U									3.5	0.6						
	N₂↔2U									**25.8**	**4.5**						
TyrRS	N₂↔2M	(Bose et	4TS1	319	α/β	1642	12	17	Urea	13.8	0.9						
	M↔U	al.,2001)								13.9	2.5						
	N₂↔2U									**41.7**	**5.9**						

										ΔG	m						
	M↔U	al.,2002)								3.5	1.2						
	N₂↔2U									**15.6**	**7.1**						
eAATase	N₂↔I₂	(Deu et al.,2007)	1ASL	396	α/β				Urea	12.0	4.8						
	I₂↔2U									24.4	3.4						
	N₂↔2U									**36.4**	**8.2**						
Ure2p	N₂↔I₂	(Galani et al.,2002; Zhu et al.,2003)	1G6Y	354	α/β	1602	13	20	GdmCl	8.0	4.2	N₂↔I₂	10^{-8}- 10^{-9}	-	7×10^{-2} (s^{-1})	0.5	
	I₂↔2U									36.0	9.0	I₂↔2M	$\leq8\times10^{-12}$	2.8	10^{4}- 10^{11}	-	
	N₂↔2U									**44.0**	**13.2**	M←U			burst	-	
hPAP	N₂↔2M	(Wojciak et al.,2003)	1CVI	354	α/β	1469	9	25	GdmCl	4.2	3.4	N₂↔I₂	2.7×10^{0}				
	M↔U									6.4	2.0	I₂↔2M	5.4×10^{-2}				
	N₂↔2U									**17.0**	**7.4**	M←U	-				
SecA	N₂↔I₂	(Doyle et al.,2000)	2FSF	853	α/β	2768	7.6	30	Urea	8.4	4.1	N₂↔I^{a}₂	2.2×10^{-2}	-	2.6×10^{-2} (s^{-1})	-	
	I₂↔2U									14.1	1.5	I^{a}₂↔I^{b}₂	~5	-	5.0×10^{0} (s^{-1})	-	
	N₂↔2U									**22.5**	**5.6**	I^{b}₂←2M^{a}		-	~3×10^{9}		
												M^{a}←M^{b}		-	7.7×10^{0} (s^{-1})		
												M^{b}←U		-	7.1×10^{3} (s^{-1})		

knotted proteins

Protein	Transition	Reference	PDB	Res	Fold				Denat	ΔG	m	Kinetic					
YbeA	N₂↔2M	(Mallam et al.,2007)	1NS5	155	α/β	1355	15	29	Urea	13.3	1.6	N₂↔2M	4.3×10^{-4}	0.6	3.9×10^{4}	0.8	0.6
	M↔U									2.8	1.5	M↔U	8.1×10^{-4}	0.5	1.6×10^{-1} (s^{-1})	0.8	0.6
	N₂↔2U									**18.9**	**4.6**						
YibK	N₂↔2M	(Mallam et al.,2005; Mallam et al.,2006)	1J85	160	α/β	1989	22	51	Urea	18.9	1.8	N₂↔2M^{a}	4.9×10^{-7}	0.7	1.8×10^{4}	0.6	0.5
	M↔U									6.5	1.5	M^{a}↔M$^{b/c}$	9.0×10^{-5}	0.4	7.7×10^{-2} (s^{-1})	0.5	0.6
	N₂↔2U									**31.9**	**4.9**	M^{b}↔U^{b}	3.0×10^{-1}	0.3	1.3×10^{2} (s^{-1})	0.9	0.8
												M^{c}↔U^{c}	1.5×10^{-2}	0.3	1.5×10^{1} (s^{-1})	0.7	0.7

α/β-barrels

Protein	Transition	Reference	PDB	Res	Fold				Denat	ΔG	m	Kinetic					
yTIM	N₂↔2M	(Najera et al.,2003)	1YPI	247					Urea	16.8							
	M↔U									4.0							
	N₂↔2U									**24.7**							
TcTIM	N₂↔I₂	(Chanez-Cardenas et al.,2005)	1TCD	249					GdmCl	1.1							
	I₂↔2M									15.7							
	M↔U									4.2							
	N₂↔2U									**25.3**							
Luciferase	αβ$_N$↔αβ$_I$	(Clark et al.,1997)	1LUC	355/324	α/β	2202	15	36	Urea	4.5	2.38	αβ$_N$↔αβ$_I$		-	2.7×10^{-4} (s^{-1})	-	-
	αβ$_I$↔α$_U$+β$_U$									19.7	3.99	αβ$_I$↔α$_M$+β$_M$		-	2.4×10^{3}		
	αβ$_N$↔α$_U$+β$_U$									**24.2**	**7.27**	α$_M$↔α$_U$		-	2×10^{-3} (s^{-1})		
												β$_M$↔β$_U$		-	6×10^{-3} (s^{-1})		
OPH	N₂↔I₂	(Grimsley et al.,1997)	1PTA	362	α/β	1561	11	24	Urea	4.3	1.0						
	I₂↔2U									36.1	4.3						
	N₂↔2U									**40.4**	**5.3**						

β-proteins:

Protein	Transition	Reference	PDB	Res	Fold				Denat	ΔG	m	Kinetic					
dfx	N₂↔2M	(Apiyo et al.,2001)	1DFX	125	β	1746	24	69	GdmCl	5.5	-						
	M↔U									11.8							
	N₂↔2U									**34.6**							
apoSOD	N₂↔2M	(Svensson et al.,2006)	1HL4	153	β	725	10	20	GdmCl	12.4	2.7	N₂↔2M	2.4×10^{-4}	0.1	2×10^{9}	0.8	0.9
	M↔U									1.8	2.8	M↔U	9.5×10^{-4}	0.5	8×10^{-2} (s^{-1})	1.1	0.7
	N₂↔2U									**16.0**	**7.3**						
bOBP	N₂↔2M	(Mazzini et	1OBP	159	β	2728	25	30	GdmCl	-	-						

	M↔U	al.,2002)					5.0	1.9
	N₂↔2U						**-**	**-**
βB1-crystallin	N₂↔I₂	(Mateu,2002)	1OKI	251	β	GdmCl	4.4	5.4
	I₂↔2U						16.0	3.2
	N₂↔2U						**20.4**	**8.6**
AAO	N₂↔I₂	(Ferguson et al.,2002)	1AOZ	552	β	Urea	3.5	1.7
	I₂↔2U						13.6	1.2
	N₂↔2U						**17.1**	**2.9**
	N₂↔I₂					GdmCl	3.3	3.1
	I₂↔2U						12.3	1.7
	N₂↔2U						**15.6**	**4.8**

The free energy of unfolding, ΔG, is the energy difference between the unfolded and folded state protein, m value is related to changes in solvent exposure area during unfolding. ku and kf are the unfolding and folding rate respectively. β-value is a measure of the degree of compaction of the transition state, it is calculated as follows mf/(mu+mf)

a IA= Interface Area per monomer (Å2), calculation for interface characterisation made using Getarea (http://pauli.utmb.edu/cgi-bin/get_a_form.tcl): total interface area per dimer and is calculated as follows: Interface area = (2 x surface area of monomer – surface area of dimer)/2.

b% MB= percent monomer buried, the % of surface area buried within the interface per monomer and is calculated as follow: %monomer buried = (Interface Area) / (surface area of monomer).

 # %P= percent polar, is the amount of polar residues buried at the interface per monomer and is calculated as follow: % polar residues = (2 x polar surface area of monomer – polar surface area of dimer) / (interface area).

na=the kinetic mechanism has not been determined.

Appendix A.5: Kinetic model of a dimer 2-state folding.

$$N_2 \underset{k_f}{\overset{k_u}{\rightleftarrows}} 2U$$
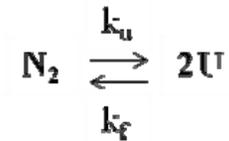
Unfolding rate:	Refolding rate:
$\dfrac{dN_2}{dt} = k_f[U]^2 - k_u[N_2]$	$\dfrac{dU}{dt} = k_u[N_2] - k_f[U]^2$

Conditions: Unfolding far from the transition	Conditions: Refolding far from the transition

$$\frac{dN_2}{dt} = -k_u[N_2]$$

$$f_{N2} = \frac{[N_2]}{P}$$

$$P\,\frac{df_u}{dt} = -Pk_u f_{N2}$$

$$\frac{df_u}{dt} = -k_u f_{N2}$$

$$\frac{df_u}{f_{N2}} = -k_u t$$

$$\ln f_{N2} - \ln f_{N20} = -k_u t$$

$$f_{N20} = 1 \text{ at } t=0 \text{ so :}$$

$$\boxed{f_{N2} = \exp(-k_u t)}$$

$$\frac{dU}{dt} = -k_f[U]^2$$

$$f_u = \frac{[U]}{2P}$$

$$2P\,\frac{df_u}{dt} = -4P^2 k_f f_u^2$$

$$\frac{df_u}{dt} = -2Pk_f f_u^2$$

$$\frac{df_u}{f_u^2} = -2Pk_f dt$$

$$\frac{1}{f_{u0}} - \frac{1}{f_u} = -2Pk_f t$$

$$f_{u0} = 1 \text{ at } t=0 \text{ so :}$$

$$\boxed{f_u = \frac{1}{1+2Pk_f t}}$$

N_2, $[N_2]$, f_{N2} represent native dimer, its concentration and its fraction. U, [U], f_U represent the unfolded monomer, its concentration and its fraction. ku and kf are the rate of unfolding and refolding, respectively. T is time and is expressed in seconds. P is the total concentration and is expressed in dimer equivalent.

Appendix A.6: Kinetic model of a trimer 2-state folding.

$$N_3 \underset{k_f}{\overset{k_u}{\rightleftharpoons}} 3U$$
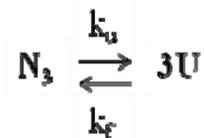
Unfolding rate:	Refolding rate:
$\dfrac{dN_3}{dt} = k_f[U]^3 - k_u[N_3]$	$\dfrac{dU}{dt} = k_u[N_3] - k_f[U]^3$

Conditions: Unfolding far from the transition	Conditions: Refolding far from the transition
$\dfrac{dN_3}{dt} = -k_u[N_3]$	$\dfrac{dU}{dt} = -k_f[U]^3$
$f_{N3} = \dfrac{[N_3]}{P}$	$f_u = \dfrac{[U]}{3P}$
$P\dfrac{df_u}{dt} = -Pk_u f_{N3}$	$3P\dfrac{df_u}{dt} = -27P^3 k_f f_u^3$
$\dfrac{df_u}{dt} = -k_u f_{N3}$	$\dfrac{df_u}{dt} = -9P^2 k_f f_u^3$
$\dfrac{df_u}{f_{N3}} = -k_u dt$	$\dfrac{df_u}{f_u^3} = -9P^2 k_f dt$
$\ln f_{N3} - \ln f_{N30} = -k_u t$	$\dfrac{1}{2f_{u0}^2} - \dfrac{1}{2f_u^2} = -9P^2 k_f t$
$f_{N20} = 1$ at t=0 so :	$f_{u0} = 1$ at t=0 so :
$\boxed{f_{N3} = \exp(-k_u t)}$	$\boxed{f_u = \dfrac{1}{\sqrt{1+18P^2 k_f t}}}$

N_3, $[N_3]$ and f_{N3} represent the native trimer, its concentration and its fraction. U, $[U]$ and f_U represent the unfolded monomer, its concentration and its fraction. ku and kf are the unfolding and refolding rate, respectively. T is the time and is expressed in second. P is the total concentration and is expressed in trimer equivalent.