

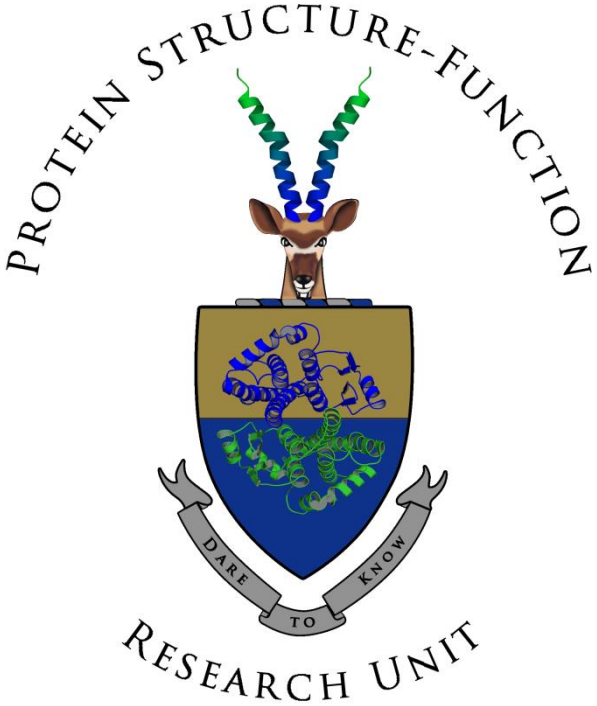
https://dspace.org/

Stability and kinetics of folding of the domain-swapped dimer of the FOXP3 forkhead domain

Perumal, Kershia

2018
https://hdl.handle.net/10539/28885

Downloaded from a DSpace repository. https://dspace.org/

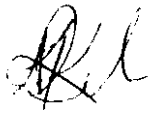


Stability and kinetics of folding of the domain-swapped dimer of the FOXP3 forkhead domain

Kershia Perumal

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before any degree or examination at any other university.



Kershia Perumal

13 August 2018

Research Outputs

2014 - 2016

Conference presentations

Name	Date	Location	Title	Authors
7th Cross Faculty Symposium	1 March 2016	University of the Witwatersrand, South Africa	Stability and folding of the FOXP3 FHD domain swapped dimer	K. Perumal S. Fanucchi
Molecular Biosciences Research Thrust	3 December 2015	University of the Witwatersrand, South Africa	Stability and folding of the FOXP3 FHD domain swapped dimer	K. Perumal S. Fanucchi

This work is dedicated to my soulmate Rowan Gengan, my mother, Shirley and Ma for their unwavering support, prayers and love throughout this degree and to God.

Always go with the choice that scares you the most, because that's the one that is going to help you grow.

Caroline Myss

Abstract

Crystal structures show that the forkhead domain (FHD) of FOX transcription factors is monomeric, except in the P-subfamily which exhibit formation of domain-swapped dimers. The crystal structure of the FOXP2 FHD shows a mixture of monomer and domain-swapped dimer (DSD) whereas the FOXP3 FHD is solely dimeric. The FOXP3 DSD is shown to be crucial for its suppressive function. Furthermore, mutations in the FOXP3 FHD are linked to a severe autoimmune disease known as the IPEX (immunodysregulation, polyendocrinopathy, enteropathy, X-linked) syndrome. Here, we explore the domain-swapping mechanism of the FOXP3 DSD using urea-induced equilibrium unfolding, stopped-flow kinetics, size-exclusion chromatography, circular dichroism, intrinsic and extrinsic (ANS) fluorescence spectroscopy. Our results show that the FOXP3 DSD is completely dimeric at micromolar concentrations as low as 4 µM which is in contrast to classical domain-swapped structures which exhibit a mixture of monomer and dimer. Interestingly, equilibrium unfolding of the FOXP3 FHD monitored by tryptophan fluorescence follows a three-state ($N_2 \leftrightarrow I_2 \leftrightarrow 2U$) folding mechanism whereby the dimer partially unfolds to form a dimeric intermediate which precedes complete unfolding. Urea-induced equilibrium unfolding monitored by circular dichroism shows a two-state folding mechanism with a free energy change of unfolding of 19.3 ± 0.97 kcal.mol^{-1}. Circular dichroism spectra shows that the intermediate strongly resembles the native state, whereas intrinsic and ANS fluorescence spectra as well as size-exclusion chromatography reveal that the structure of the intermediate is not as compact as the native state. Stopped-flow kinetics were monitored using tryptophan fluorescence above 320 nm. Unfolding kinetics indicate one kinetic phase except for the single jump from native to the intermediate state. Refolding kinetics are biphasic. A three-state unfolding mechanism (($N_2 \leftrightarrow I_2 \leftrightarrow 2U$) is proposed for the folding pathway of the FOXP3 DSD with the primary hydrophobic interface being disrupted upon formation less compact molten-globule intermediate which then dissociates into monomers (burst phase in I ↔ U transition) and subsequently unfolds completely.

Acknowledgments

Very special thanks to my supervisor, Dr. Sylvia Fanucchi for being a great source of support and encouragement throughout this challenging project and for being a fantastic role model throughout my studies.

I sincerely thank Prof. Heini Dirr for guiding me and for reminding me that scientific research is about asking the right questions.

Thank you to my colleagues at the PSFRU for their positivity, continuous support and motivation throughout my studies.

I would also like to acknowledge the University of the Witwatersrand and the National Research Foundation for financial assistance.

TABLE OF CONTENTS

Declaration.....i

Research outputs.....ii

Abstract.....iv

Acknowledgements.....v

List of abbreviations ix

List of figures.....xi

List of tables.....xiii

Chapter 1. Introduction..... 1

1.1 Transcription factors 1

1.2 Protein folding 4

1.2.1 The Levinthal paradox..... 4

1.2.2 Proposed folding models 4

1.2.3 Protein folding pathways 6

1.3 Protein-protein interactions 8

1.3.1 Stability of protein-protein interactions 9

1.3.2 Transcription factor oligomerisation 9

1.4 Domain swapping 11

1.4.1 Mechanism 13

1.4.2 Role of the hinge region and hydrophobic core in domain swapping of the FOXP3 DSD..... 15

1.4.3 Artefact of protein crystallisation or physiologically relevant?..... 16

1.5 Forkhead box family of transcription factors 18

1.6 FOXP subfamily 20

1.6.1 Structural similarities 20

1.6.2 The FOXP forkhead domain 21

1.6.3 FOXP2 22

1.7 FOXP3 ……. 24

1.7.1 Physiological role ……... 24

1.7.2 Structure …… 24

1.7.3 Comparison of the FOXP2 and FOXP3 domain-swapped dimers… 27

1.7.4 IPEX mutations disrupt domain swapping…….. 29

1.8 Problem identification ……. 31

1.8.1 Objectives…… 31

Chapter 2. Experimental …… 32

2.1 Materials …… 32

2.2 Methods…….. 32

2.2.1 Plasmid purification and verification …….. 32

2.2.2 FOXP3 FHD expression and purification ……. 33

2.2.3 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis …….. 34

2.2.4 Protein concentration determination…….. 35

2.2.5 Circular dichroism spectroscopy …….. 36

2.2.6 Fluorescence spectroscopy …….. 37

2.2.7 Size-exclusion chromatography …….. 38

2.2.8 DNA binding studies …….. 39

2.2.9 Reversibility and recovery …… 40

2.2.10 Urea-induced equilibrium unfolding…….. 41

2.2.11 Stopped-flow kinetics ……... 46

Chapter 3. Results…….. 50

3.1 Purification of wild-type FOXP3 FHD …….. 50

3.1.1 Plasmid verification ……... 50

3.1.2 FOXP3 FHD expression and purification ……... 50

3.1.3 Protein concentration determination ……... 51

3.2 Structural integrity of the FOXP3 FHD …… 55

3.2.1 Secondary structure …….. 55

3.2.2 Tertiary structure ……... 56

3.2.3 Quaternary structure …….. 60

3.3 DNA binding of the FOXP3 FHD…….. 62

3.4 Conformational stability of the FOXP3 domain-swapped dimer……. 63

3.4.1 Recovery and reversibility ……. 63

3.4.2 Urea-induced equilibrium unfolding…… 63

3.5 Kinetic studies ……. 68

3.5.1 Unfolding kinetics …… 70

3.5.2 Refolding kinetics …… 75

3.6 Properties of the equilibrium unfolding intermediate of the FOXP3 DSD 81

3.6.1 Secondary structure …….. 81

3.6.2 Tertiary structure …… 81

3.6.3 Quaternary structure …….. 83

3.6.4 ANS binding to the intermediate…….. 83

Chapter 4. Discussion …….. 85

4.1 Equilibrium unfolding of the FOXP3 FHD proceeds via a three-state model…... 85

4.2 The FOXP3 FHD intermediate ……. 86

4.3 Kinetic mechanism of (un)folding …… 92

4.4 Conclusion and future work …….. 94

References …… 97

List of Abbreviations

θ	Ellipticity
A_{280}	Absorbance at 280 nm
Å	Ångstroms
ÄKTA	Protein purification system
ASA	Accessible surface area
CD	Circular dichroism
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
DSD	Domain-swapped dimer
ds-DNA	Double-stranded DNA
DTT	Dithiothreitol
ε	Molar extinction coefficient
EMSA	Electromobility shift assay
Ets	E-twenty six DNA binding domain
FHD	Forkhead DNA binding domain
FOX	Forkhead box domain protein family
HNF-3	Hepatic nuclear factor 3
HLH	Helix-loop-helix motif
HTH	Helix-turn-helix motif
IMAC	Immobilised metal-ion affinity chromatography
IPEX	Immunodysregulation, polyendocrinopathy, enteropathy, X-linked syndrome
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_d	Dissociation constant
K_{eq}	Equilibrium constant

LB	Luria Bertani
Oct-1	Octamer 1
PDB	Protein data bank
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
POU	Derived from the following proteins containing this DNA binding domain: Pituitary-1, Octamer-1, Unc-86
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SgrAI	Type of restriction endonuclease
SEC	Size exclusion chromatography
SOC	Super optimal broth with catabolite repression
UV	Ultraviolet
WT	Wild-type

The IUPAC-IUBMB one- and three-letter abbreviations for the 20 standard amino acids were used

List of figures

Figure 1. Some of the most well-characterised DNA binding domains 3

Figure 2. Protein folding funnel 6

Figure 3. Schematic of domain swapping 12

Figure 4. Examples of proteins that exhibit domain-swapping..... 13

Figure 5. Structure of the 'winged' helix domain..... 20

Figure 6. Schematic of a typical FOXP protein..... 21

Figure 7. Sequence alignment of the FHD in FOX family members. 23

Figure 8. Structure of the FOXP3 domain-swapped dimer bound to DNA..... 26

Figure 9. Stabilising hydrophobic core residues in the FOXP3 DSD..... 26

Figure 10. Structural alignment of the FOXP2 FHD monomer with the FOXP2 FHD DSD..... 28

Figure 11. 15 % SDS-PAGE gel of small-scale induction trials..... 52

Figure 12 Elution profile of the FOXP3 FHD using the IMAC HisTrap™ column. 53

Figure 13. SDS-PAGE and calibration curve 54

Figure 14. Far-UV spectra of the wild-type FOXP3 FHD 57

Figure 15. Crystal structure of the FOXP3 domain-swapped dimer indicating locations of the tryptophan residues 58

Figure 16. Fluorescence spectra of the wild-type FOXP3 FHD 59

Figure 17. Size-exclusion chromatography of the FOXP3 FHD 61

Figure 18. Electrophoretic mobility shift assay of the FOXP3 FHD bound to Nelson DNA..... 62

Figure 19. Recovery of the FOXP3 FHD 64

Figure 20. Unfolding and refolding curves of the FOXP3 FHD 66

Figure 21. Urea-induced equilibrium unfolding of FOXP3 FHD 67

Figure 22. Maximum emission wavelength of the FOXP3 FHD upon unfolding 69

Figure 23. Typical traces observed for unfolding monitored by stopped-flow kinetics 72

Figure 24. Urea dependence on the unfolding rate constant 74

Figure 25. Kinetic traces observed for refolding monitored by stopped-flow fluorescence 77

Figure 26. Effect of varying urea concentrations on refolding 80

Figure 27. Structural characterisation of the intermediate 82

Figure 28. Exposed hydrophobic surface of the FOXP3 DSD 89

Figure 29. The crystal structure of the FOXP3 FHD coloured according to B-factor 91

Figure 30. Proposed scheme of the (un)folding pathway of the FOXP3 DSD 96

List of tables

Table 1. Unfolding kinetic parameters obtained for the FOXP3 DSD 75

Table 2. Refolding kinetic parameters obtained for the FOXP3 DSD..... 79

CHAPTER 1: INTRODUCTION

1.1 Transcription factors

The regulation of gene expression is highly complex and involves a multitude of interactions between transcription factors and gene targets as well as between transcription factors and their regulators. These interactions are collectively known as the 'transcriptional regulatory network'. All cells in a multicellular organism contain the same genes and have similar processes that are regulated by 'house-keeping' transcription factors which are necessary for the functioning of every cell. In specific tissues, such as the lungs for example, other transcription factors have unique expression patterns and form part of distinct regulatory programmes. These distinct networks play a role in tissue specificity (Sonawane *et al.*, 2017) and can serve as biomarkers for disease (Scherzer *et al.*, 2008; Zhou *et al.*, 2008; Sykiotis and Bohmann, 2010). Transcription factors have crucial physiological roles, for example in metabolism, cell cycle regulation, reproduction, development and regulation of the immune system (Grandori *et al.*, 2000; Levy and Darnell, 2002; Amoutzias *et al.*, 2004; Desvergne *et al.*, 2006; Amoutzias *et al.*, 2007) and their structures usually contain specific domains that perform specific functions.

Transcription factors are generally classified according to their highly-conserved DNA binding domains (Harrison, 1991) which bind to specific DNA sequences (Schleif, 2013). The most well-studied of these DNA binding domains are homeobox, POU (derived from Pituitary-1, Octamer-1, Unc-86), paired box, cysteine-cysteine zinc finger, cysteine-histidine zinc finger, the basic element, the Ets (E-twenty six) domain as well as the helix-turn-helix motif (Pabo, 1992; Latchman, 1997) (Figure 1).

Transcription factors contain various domains within their structures such as dimerization, activation and repression domains which make them modular in nature. This modularity allows the activation and repression domains to be involved in up- or down-regulation of target genes, respectively (Frankel and Kim, 1991). This suggests that DNA binding alone may not be sufficient for activation of transcription of target genes (Latchman, 1997). A careful balance must be obtained between activated and

repressed target genes. This is achieved by regulating the ‘regulators’, in other words the regulation of transcription factors themselves (Latchman 1997). Transcription factors can be activated or repressed in response to specific stimuli, through ligand binding, due to changes in protein-protein interactions or due to post-translational modifications such as phosphorylation (Latchman 2005).

Transcription factors rarely work alone and in fact, many self-associate to form dimers and even higher-order oligomers. For some transcription factors, dimerisation is required prior to binding DNA and this is mediated through an associated dimerisation domain, for example a leucine zipper domain, or a DNA binding motif, such as the helix-loop-helix motif (Anthony-Cahill *et al.*, 1992; Latchman, 1997; Massari and Murre, 2000). Oligomerisation confers advantages such as increased stability and increased complexity, however, unwanted oligomerisation can lead to formation of pathogenic protein aggregates. A sophisticated example of a homodimeric DNA-binding protein is Type II restriction enzymes. Each subunit binds to half of a palindromic sequence (Pingoud and Jeltsch, 2001). Here, dimerisation is shown to significantly increase not only the binding affinity but also the binding specificity (Marianayagam *et al.*, 2004). Taken together, protein-protein interactions can be considered to be just as important in the gene regulation networks as protein-DNA interactions.

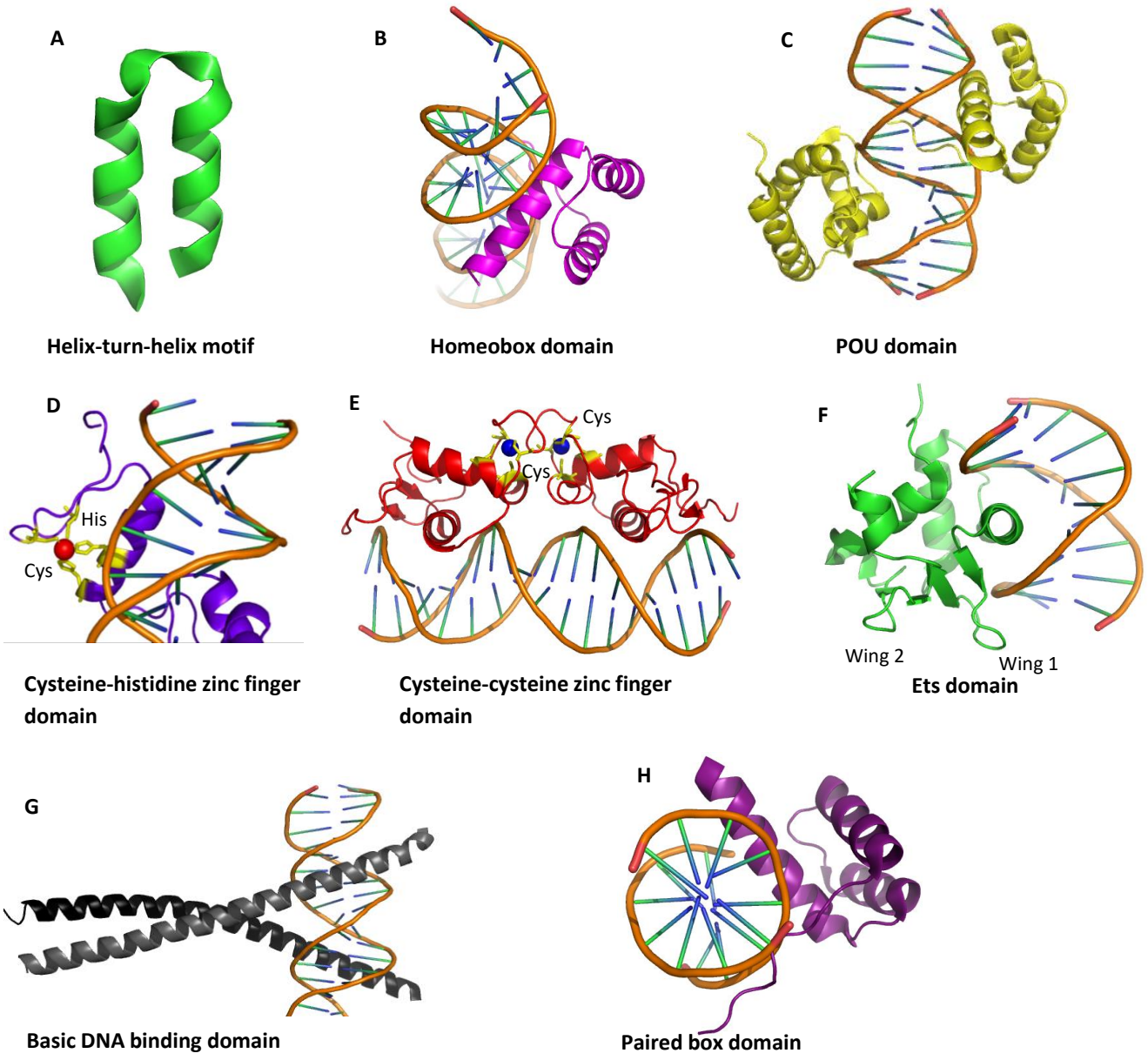


Figure 1. Some of the best-characterised DNA binding domains. A. The helix-turn-helix (HTH) motif consists of two helices separated by a beta turn. PDB code 1VRZ. B. Homeobox domain. PDB code 2LKX. C. POU domain. PDB code 3L1P. D. Cysteine-histidine zinc finger domain where two cysteines and two histidine residues coordinate a zinc ion (red sphere). PDB code 1CQT. E. Cysteine-Cysteine zinc finger consists of four cysteine residues that coordinate a zinc ion (blue) and binds DNA as a dimer. PDB code 5V3J. F. Ets domain belongs to the 'winged' helix-turn-helix superfamily. PDB code 1BC8. G. Basic DNA binding domain bound to DNA and is adjacent to the leucine zipper domain which assists dimerisation. PDB code 2DGC. H. Paired box domain of Pax3. PDB code 1FJL.Upon DNA binding in each example above (except for A), an alpha helix is inserted into the major groove of DNA and this is consistent throughout even though there are different proteins and different DNA sequences involved. Figure adapted from (Perumal, 2014).

1.2 Protein folding

1.2.1 The Levinthal paradox

While the stability of protein interactions has been studied in great detail, the accurate prediction of the three-dimensional native structure of a known polypeptide sequence remains elusive (Dobson, 2000), even though there has been considerable progress in this field with the development of *in silico* tools to predict secondary and tertiary structure based on protein models (Drozdetskiy *et al.*, 2015; Webb and Sali, 2016). It has been estimated that a protein consisting of 100 amino acids would take approximate 10^{30} years to fold to its correct conformation if it searched all possible conformations (Levinthal, 1969). However, proteins are known to fold into their functional native states extremely rapidly (in the order of seconds) (Dobson, 2000). This apparent contradiction is known as the Levinthal paradox and indicates that proteins do not fold via a random search of all possible conformations but instead, follow specific folding pathways (Levinthal, 1969).

Since protein function is dependent upon the fully-folded, native conformation of a protein, it is necessary to understand how a protein finds its way to its correctly folded structure in a reasonable period of time, without needing to access a myriad of incorrect alternatives.

1.2.2 Proposed folding models

Many proposed models have been suggested in order to explain the protein folding mechanism, such as the hydrophobic collapse model (Dill, 1985), the diffusion-collision model (Karplus and Weaver, 1994), the nucleation-condensation model (Fersht, 1997) and the jigsaw-puzzle model (Harrison and Durbin, 1985), none of which can fully explain or resolve the Levinthal paradox. The conformational space searched by a protein during folding however, can be reduced significantly through formation of specific folding pathways in the energy landscape (Onuchic *et al.*, 1999). This model is known as the landscape theory and is described with the use of folding funnels (Figure 2).

Initially, the unfolded protein exists in an species of unfolded speciess that is heterogeneous due to the relatively numerous conformational states that are available

to an unfolded polypeptide. As folding proceeds to a lower-energy state, stabilising native interactions (intramolecular interactions) are favoured over non-native interactions (intermolecular interactions) (Figure 2). The number of conformations accessible to a polypeptide decreases progressively down the funnel as more native, intramolecular contacts form and the protein becomes more compact. Hydrophobic groups interact with each other to expel water from the protein core and simultaneous exposure of polar amino acid side chains to the surface occurs (Dill *et al.*, 1995).

The folding process can occur via the formation of intermediates. These intermediates can either be on-pathway, that is, they can be found directly between the unfolded and folded states in the protein folding pathway or they can be off-pathway and result in misfolded or partially-folded, aggregates (Hartl and Hayer-Hartl, 2009). These aggregates can then form cytotoxic highly-ordered fibril structures that are associated with diseases such as cystic fibrosis (Bence *et al.*, 2001; Janowski *et al.*, 2001; Nilsson *et al.*, 2004), Alzheimer's (Murrell *et al.*, 1991) and Parkinson's diseases (Baba *et al.*, 1998). The origins of these diseases are due to mutations that result in changes in the stability or kinetic properties of the protein which leads to impaired function or disruption of protein folding control systems (Tan and Pepys, 1994; Bullock *et al.*, 1997). In the cell, proteins are in constant dynamic equilibrium between folded and unfolded speciess so folding assistance is often required by helper proteins called 'molecular chaperones' and folding catalysts (Gething and Sambrook, 1992). These biological catalysts can prevent aggregation of misfolded proteins and can speed up reactions in the folding mechanism that might otherwise be slow (Schiene and Fischer, 2000). Misfolded proteins are usually ubiquitinated and targeted for degradation when they are no longer needed, which provides an important step in regulation.

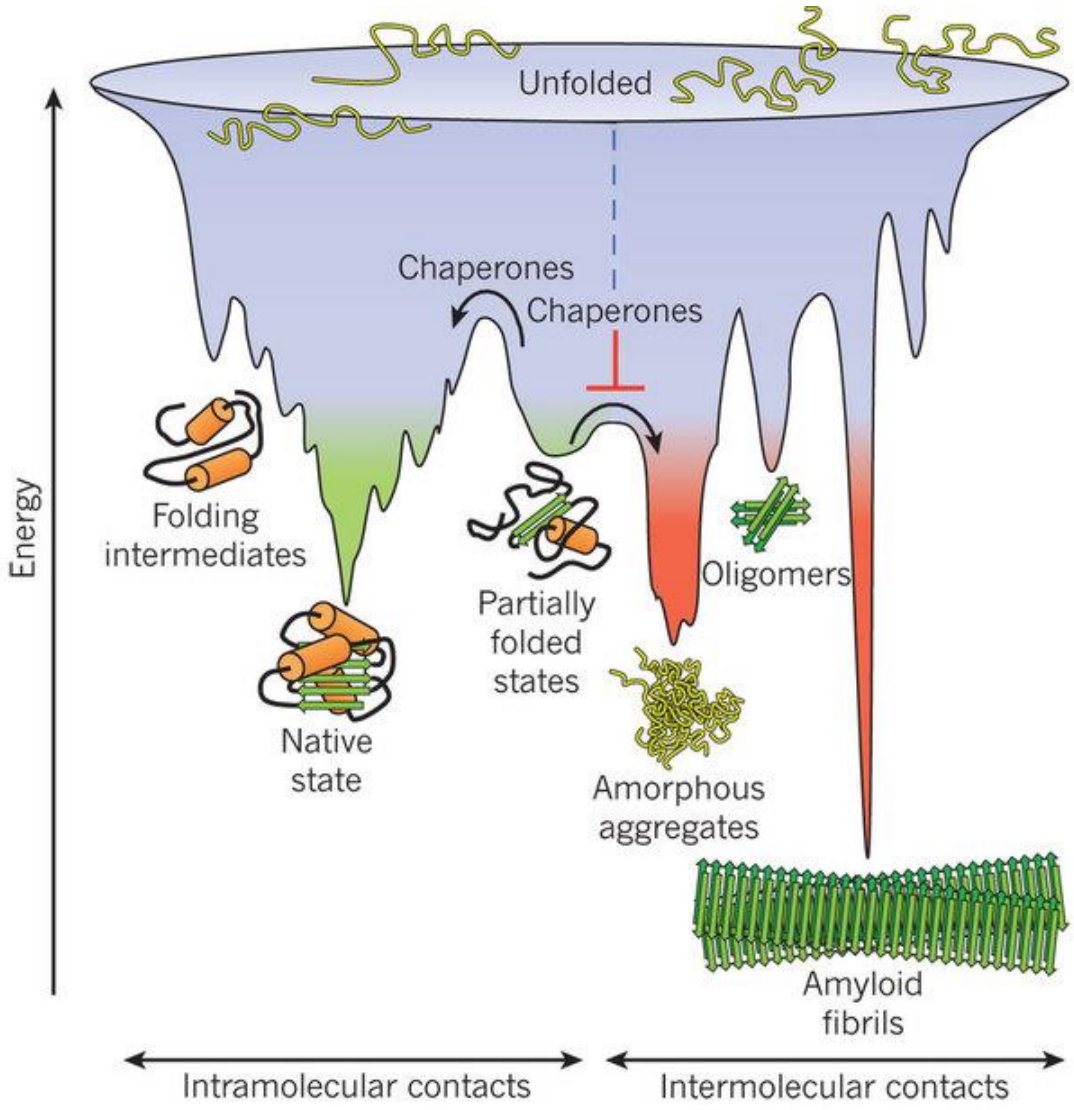


Figure 2. Protein folding funnel. The energy landscape of a protein as it folds to its native state is funnel-shaped. Intially, the unfolded species of protein molecules is highly disordered and is heterogenous. As folding proceeds, native interactions are favoured over non-native random interactions which may result in the formation of folding intermediates. These folding intermediates may serve as kinetic 'traps' as they are located in energy minima, however *in vivo*, molecular chaperones may assist in allowing these intermediates to fold into the lowest-energy native state. If mutations arise that promote misfolding, highly-ordered amyloid structures or amorphous aggregates may form and are associated with various diseases such as cystic fibrosis (Bence *et al.*, 2001), Alzheimer's (Murrell *et al.*, 1991) and Parkinson's disease (Baba *et al.*, 1998). Figure adapted from Harlt *et. al*. (2011).

1.2.3 Protein folding pathways and intermediates

While molecular chaperones are sometimes necessary to aid protein folding *in vivo*, numerous isolated unfolded proteins have been shown to refold spontaneously *in vitro*

to their fully-folded native states upon removal of denaturing conditions (Anfinsen, 1973). A set of elegant experiments by Anfinsen (1973) suggests that the protein sequence contains all the necessary information that is required for folding.

A folding pathway can be defined in terms of the speciess of conformational states that the protein encounters as it folds to form the native state. The folding pathway may be sequential or parallel, however all speciess of on-pathway intermediate states must converge to form the folded, native protein conformation (Walters, Milam et al. 2009) whereas off-pathway intermediates may form misfolded aggregates that are either degraded by the ubiquitin degradation pathway or may lead to disease.

Small proteins usually exhibit a two-state equilibrium unfolding mechanism where the molecules present at equilibrium consist of either fully folded (N) or fully unfolded (U) states. In a two-state model, the well-defined native and unfolded structures are separated by a high free-energy barrier. This is explained by the transition state theory. According to the transition state theory (Christensen and Pain, 1991), reacting molecules collide to form a high-energy, metastable transition state which then decomposes to form the products. The approaching reacting molecules, in this case the unfolded protein, must overcome a high-energy barrier, known as activation energy in order to form folded, native protein. The reorganisation of the transition state to the native conformation may be a rate-limiting step and in order to fully characterise the protein folding and refolding pathway, it is essential to characterise the structural and energetic properties of the transition states that form.

When proteins fold via multistate mechanisms, intermediate species will be detected either at equilibrium or transiently. The role of protein folding intermediates in the protein folding mechanism is controversial due to the variability of folding pathways. In some cases, intermediates are present in folding pathways while in others they are absent, which makes their significance in folding somewhat unclear (Roder and Colón, 1997). In complex, multidomain proteins, these intermediates may assist protein folding by reducing the conformational space that is available to the protein. Another view suggests that they may slow down the folding process which may be essential to ensure that the protein folds to the correct native conformation (Englander and Mayne, 2014).

1.3 Protein-protein interactions

1.3.1 Stability of protein-protein interactions

Proteins can interact with nucleic acids, membranes, small ligands and other proteins in order to perform distinct biological functions. An increase in the number of solved protein crystal structures has allowed detailed analysis into the principles that govern protein interactions.

Protein-protein interfaces are usually conserved (Valdar and Thornton, 2001; Moreira *et al.*, 2007) and are characterised according to: interface size; geometric complementarity; residue frequency; electrostatic interactions; hydrophobicity and the exposure of accessible surface area (ASA) all of which have been studied extensively (Chothia, 1974; Jones and Thornton, 1996; Sheinerman *et al.*, 2000; Glaser *et al.*, 2001; Moreira *et al.*, 2007). These studies allow the distinction between biologically relevant protein-protein interactions in complexes and those formed by crystalline contacts or packing in the crystal lattice in crystal structures (Valdar and Thornton, 2001; Moreira *et al.*, 2007).

The hydrophobic effect, hydrogen bonds, salt bridges, disulfide bonds, intrinsic secondary structure and packing interactions between amino acid residues have all been shown to be involved in stabilising protein-protein interactions as well as intramolecular interactions within the protein (Baldwin and Matthews, 1994; Young *et al.*, 1994; Pace *et al.*, 1996; Strickler *et al.*, 2006). Of these, hydrophobic interactions are shown to be the most important for association. This is due to the aqueous environment that causes the hydrophobic residues to collapse into the interior of a single protein molecule and also drives the association of many proteins via exposed hydrophobic surface patches (Chothia, 1974; Van den Burg *et al.*, 1994) as the burial of hydrophobic surface area contributes favourably to the binding energy (Chothia and Janin, 1975; Jones and Thornton, 1996). The burial of surface hydrophobic residues is an entropically-driven process thereby expelling disordered water molecules from the interface and allowing a water-tight seal around a critical set of energetically favourable interactions and is essential in the formation of high-affinity protein complexes (Bogan

and Thorn, 1998). It is proposed that clusters of hydrophobic residues on the protein surface interact with each other to form a network that stabilises the protein complex (Tisi and Evans, 1995) where clusters can range between 3 - 15 residues (Young *et al.*, 1994). The standard size of a protein-protein interface is approximately 500 - 2000 $Å^2$ (Moreira *et al.*, 2007) and a minimum of 381 $Å^2$ of buried surface area is necessary to form a stable protein-protein complex (Chen *et al.*, 2013).

In the event of oligomerisation, clusters of hydrophobic residues on the protein surface can thus interact with each other to form a network that stabilises the protein complex (Tisi and Evans, 1995). These clusters can range in size between 3 - 15 residues (Young *et al.*, 1994). The standard size of a protein-protein interface is approximately 500 - 2000 $Å^2$ (Moreira *et al.*, 2007) and a minimum of 381 $Å^2$ of buried surface area is necessary to form a stable protein-protein complex (Chen *et al.*, 2013).

Although hydrophobic residues are most important in stabilising protein-protein interfaces, polar residues exposed to the solvent are also involved in stabilising protein-protein complexes via the formation of salt bridges and hydrogen bonds (Robertson, 2002). Hydrogen bonds form between an electronegative atom (acceptor) and a hydrogen atom (donor). They are frequently observed at protein interfaces with an average of 1 bond per 100-200 $Å^2$ of surface area (Jones and Thornton, 1996). In addition, pairs of oppositely-charged residues can be found at the interface within 4 Å of each other (Barlow and Thornton, 1983). Whereas hydrophobic interactions are the driving forces that cause proteins to associate, electrostatic interactions play an important role in the binding specificity of protein-protein interactions and therefore are often involved in determining which proteins interact with each other (Fersht, 1987).

1.3.2 Transcription factor oligomerisation

Proteins can oligomerise to form dimers, trimers, or higher order multimers. Protein dimerisation, in particular, involves the association of two identical monomers to form a homodimer while heterodimers form through the association of two different polypeptide chains (Spit *et al.*, 1998). Protein oligomers have several advantages over monomeric forms, namely regulation of protein activity via reversible oligomerisation, higher local protein concentration, larger binding surfaces (Bennett *et al.*, 1995) and

perhaps even novel biological activity (Changeux and Edelstein, 1998). Oligomerisation may confer additional advantages such as increased stability, regulation of binding site accessibility and increased complexity, allowing for the evolution of new functions (Marianayagam *et al.*, 2004; Wu *et al.*, 2014).

It is therefore not surprising that while some transcription factors are capable of interacting with DNA as monomers (Figures 1B-D and Figures 1F and 1H), transcription factor dimers are fairly common. In some cases transcription factors form dimers so as to bind symmetrically to DNA sequences (Abel and Maniatis, 1989) as in the case of the glucocorticoid receptor which contains a four-cysteine zinc finger domain (Luisi *et al.*, 1991) (Figure 1E) and the transcription factor C/EBP alpha, which contains a basic leucine zipper motif (bLZ) (Hu and Sauer, 1992) (Figure 1G). Dimerisation of some transcription factors such as the mammalian telomeric protein TRF1 (Bianchi *et al.*, 1997) can also be facilitated by structural motifs such as the helix-loop-helix (HLH) motif (Jones, 2004) and the helix-turn-helix (HTH) motif (Figure 1A) TRF1 is an example of a transcription factor that dimerises upon DNA binding, however not all transcription factors dimerise upon binding to DNA as in the case of steroid receptors (Luisi *et al.*, 1991). In fact, many transcription factors associate in solution with high affinity (in the nanomolar range) in the absence of DNA (Schleif, 1988; Neet and Timm, 1994; Foguel *et al.*, 1998) which suggests a free energy link between folding and DNA binding (von Hippel, 1994; Chen *et al.*, 2013).

In addition to transcription factors binding to DNA in their dimeric and monomeric oligomeric states, some transcription factors exhibit a unique type of oligomerisation, termed domain swapping, in order to carry out their physiological functions. Domain swapping is a unique oligomerisation event whereby specific regions of protein chains are exchanged to form an intertwined structure (Schlunegger *et al.*, 1997) The octamer-1 (Oct-1), a POU-member (Figure 1C) (Reményi *et al.*, 2001) and members of the ‘winged’ helix superfamily (Figure 1F), the forkhead box (FOX) P-subfamily (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011; Chu *et al.*, 2011) are examples of transcription factors that exhibit domain swapping.

1.4 Domain swapping

Domain swapping was first identified in the crystal structure of diphtheria toxin at an atomic resolution of 2.3 Å (Bennett and Eisenberg, 1994) and is a type of oligomerisation event in which specific regions of proteins are exchanged to form an intertwined dimer or higher order oligomer (Schlunegger *et al.*, 1997). The exchanged regions can vary in size from just one secondary structural element such as an α-helix or β-sheet (Khazanovich *et al.*, 1996) to an entire domain (Bennett *et al.*, 1995) or many domains (Liu *et al.*, 1998; Liu *et al.*, 2002). If the structure of a protein exhibits stable forms of both monomer and domain-swapped oligomer, it is considered to be a bona fide example of 3D domain swapping. In the case of two subunits exchanging structural elements, the intertwined structure is known as a domain-swapped dimer (DSD).

The structures of the monomer and the subunits of the DSD are identical except for the hinge-loop region connecting the two subunits (Ogihara *et al.*, 2001; Rousseau *et al.*, 2003) (Figure 3). The only other difference between monomer and the DSD may be in a secondary interface where new interactions may occur due to the close proximity of the subunits (Bandukwala *et al.*, 2011; Murawska *et al.*, 2017; Nandwani *et al.*, 2017).

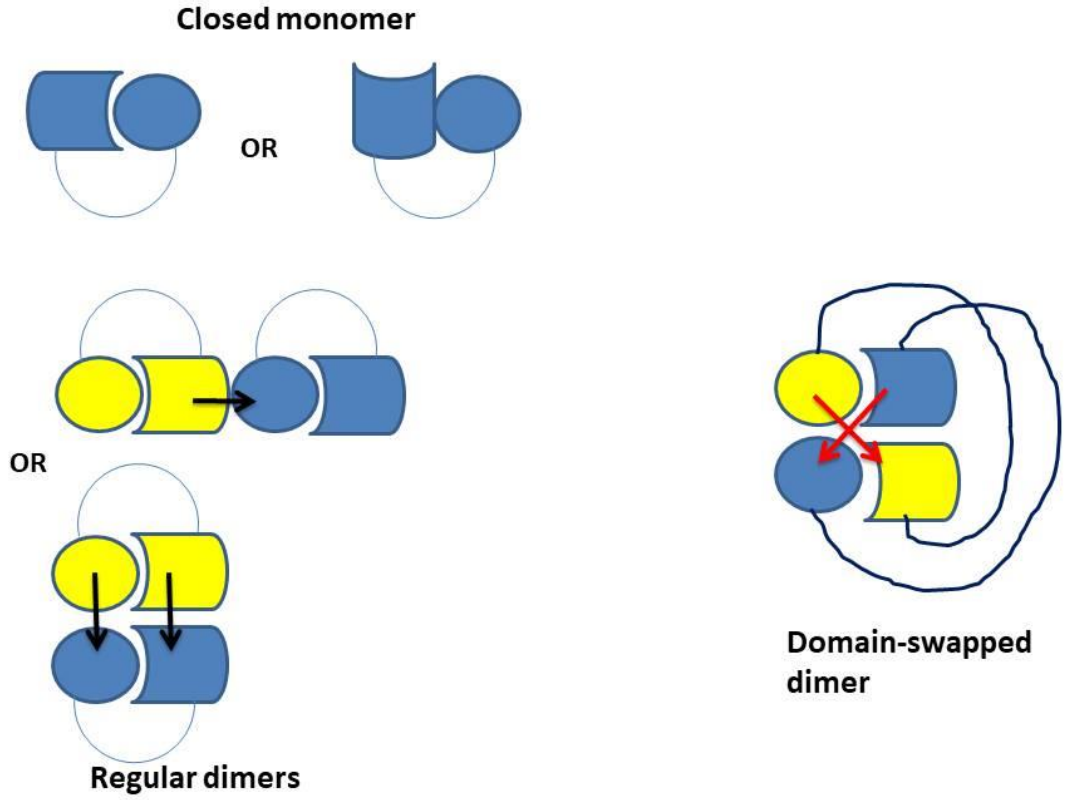


Figure 3. Schematic of domain swapping. The subunits in the domain-swapped dimer are usually identical. The hinge loop conformations are different in the closed monomer and open monomers in the domain-swapped dimer (Bennett *et al.*, 1995). The primary interface is shown by black arrows and the unique secondary interface that forms in domain-swapped dimers is indicated by the red arrows and is not present in regular dimers. Figure created using Microsoft PowerPoint 2010.

A study identified that many different proteins with a diverse range of functions can form domain-swapped dimers (Figure 4) (Huang *et al.*, 2012). For example, the immunoglobulin superfamily which has a diverse range of physiological functions in the immune system shows a diverse range of swapping modes where different structural elements are exchanged to varying degrees (Huang *et al.*, 2012). It is not clear how or why these domain-swapped structures form (Liu and Eisenberg, 2002). Even though domain swapping is frequent (approximately 10 % in protein folds and 5 % in protein families) , there is no clear pattern in terms of sequence or structure as to whether or not a protein will domain swap (Liu and Eisenberg, 2002) and little is known about the exact mechanism of domain swapping (Hayes *et al.*, 1999; Kuhlman *et al.*, 2001; Rousseau *et al.*, 2001; Schymkowitz *et al.*, 2001). This leads us to the following

question: how does a single polypeptide sequence gain access to two different conformational states?

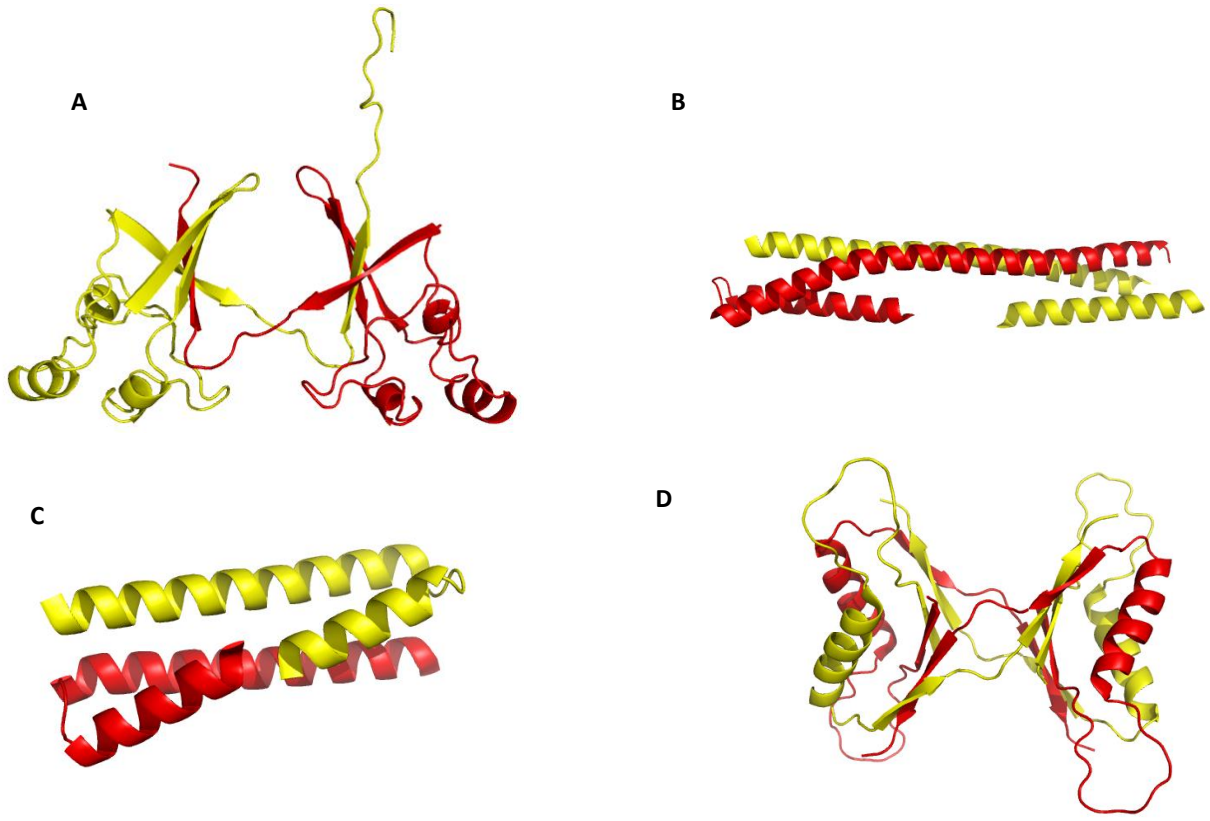


Figure 4. Examples of proteins that exhibit domain-swapping. These proteins all swap to different extents and have different physiological functions ranging from cell-cycle control to exocytosis of transport vesicles. A. Cell-cycle regulatory protein, $p13^{suc1}$ (Bourne *et al.*, 1995). PDB code 1PUC. B. T-SNARE (Antonin *et al.*, 2002). C. Designed helical bundle (Lovejoy *et al.*, 1993). PDB code 2C5J. D. B1 domain of streptococcal protein G (Frank *et al.*, 2002). PDB code 1MPE. One monomer is shown in red and the other monomer is shown in yellow. Figure was created using PyMOL (DeLano Scientific).

1.4.1 Mechanism of domain swapping

Very few studies have looked at the mechanism of domain swapping (Nagarkar *et al.*, 2010), although there has been significant recent developments (Medina *et al.*, 2016; Fagagnini *et al.*, 2017; Medina *et al.*, 2017; Murawska *et al.*, 2017; Nandwani *et al.*, 2017; Huang *et al.*, 2018). Some hypotheses have been proposed, the first of which used the monomeric and dimeric structures of diphtheria toxin (Bennett *et al.*, 1995). In

most cases, the first step in domain swapping may involve breaking intramolecular interactions that stabilise the monomeric form, followed by formation of an 'open' monomer *via* alteration of the hinge-loop conformation. According to Bennett *et al*. (1995), the closed monomers are disrupted to form open monomers which then associate to form DSDs. The large kinetic barrier between the closed monomer and open monomer is associated with very slow interconversion between the monomer and dimer. However, this large kinetic barrier can be overcome by a change in pH, temperature, changes in denaturant concentration (Rousseau *et al.*, 2001), binding of a ligand or mutation in the protein. (Bennett *et al.*, 1995). The best-studied domain-swapped protein to date is the $p13^{suc1}$ cell-cycle regulatory protein. It is found in the yeast *Schizosaccharomyces pombe* and functions by binding to and suppressing cdc2 (a cyclin-dependent kinase). The $p13^{suc1}$ protein belongs to the CDC28 kinase-specific class of proteins that are essential for cell cycle progression (Endicott and Nurse, 1995). However, it is not clear whether the domain-swapped dimer is essential for its function.

The domain swapping mechanism has been outlined for $p13^{suc1}$ (Khazanovich *et al.*, 1996; Rousseau *et al.*, 1998; Rousseau *et al.*, 2001), and cyanovirin-N, an HIV-inactivating protein (Liu *et al.*, 2012). In both these cases, the monomeric form must completely unfold and then refold to form the domain-swapped dimer. This process is thought to be followed by rebuilding of intermolecular connections. The intermolecular connections formed should be the same as the intramolecular connections that were broken in the monomer and thus should contribute little to the overall free energy change in the monomer-dimer transition (Heringa and Taylor, 1997; Robertson, 2002). In another domain-swapping protein, barnase, a ribonuclease from bacteria, the regions involved in structural exchange can only fold partially in an independent manner which suggests that domain swapping must occur at an intermediate stage of folding (Zegers *et al.*, 1999). Together, these studies imply that the domain swapping mechanism is unique for each protein and is dependent on the folding pathway of the particular protein (Rousseau *et al.*, 2003).

Indeed, to date, a unifying mechanism of domain swapping remains elusive. However, it seems that in order to gain access to the domain-swapped conformational state, the monomer has to partially or fully unfold first as both computational and experimental work shows (Rousseau *et al.*, 2001; Yang *et al.*, 2004; Liu *et al.*, 2012). This implies that if the folding pathway was initiated by the folded monomer, intra-chain bonds are

required to be broken. The intra-chain bonds are then replaced by identical inter-chain bonds with another monomer in addition to hinge-loop backbone conformational changes. This process requires a considerable amount of energy, called 'activation energy' as there usually is a partitioning of the two conformations with respect to their free energy difference (Liu *et al.*, 2012). Amino acid residues at the hinge-loop region as well as at the new interface created by domain swapping play a key role in the domain swapping mechanism (Kuhlman *et al.*, 2001; Rousseau *et al.*, 2001; Schymkowitz *et al.*, 2001). Furthermore, Medina *et al.*, 2016 shows that domain swapping leads to reduced flexibility in the hinge-loop region.

1.4.2 Role of the hinge region and hydrophobic core in domain swapping of the FOXP DSD

Since the structures of the monomer and DSD are almost identical except for the hinge-loop region and the dimerisation interface, any differences in these regions (Nandwani *et al.*, 2017), for example mutations, can either favour or disfavour domain swapping (Rousseau *et al.*, 2003). Much focus on the mechanism behind domain swapping has, therefore been on the hinge-loop region. The length of the hinge-loop region, either by shortening or lengthening (Kortt *et al.*, 1994; Perisic *et al.*, 1994; Green *et al.*, 1995; Murray *et al.*, 1995) and flexibility of the hinge-loop region (Rousseau *et al.*, 2001; Staniforth *et al.*, 2001; Newcomer, 2002) has been shown to either promote (Ogihara *et al.*, 2001) or disrupt domain swapping as in the case of $p13^{suc1}$ (Rousseau *et al.*, 2003) and a non-domain swapping protein, single chain monellin (MNEI) (Nandwani *et al.*, 2017).

While most studies on the hinge-loop and domain swapping propensity have focused on the strain and loop length, the link between amino acid composition at or near the hinge-loop region has received little attention (Nandwani *et al.*, 2017). Domain-swapped dimers have been frequently observed in cystatins (Janowski *et al.*, 2001; Staniforth *et al.*, 2001; John-Baptiste *et al.*, 2012; Murawska *et al.*, 2017) and on average, the cystatin hinge-loop is six amino acid residues shorter than MNEI. A recent study of the effect of hinge-loop amino acid residue composition after deletion upon domain swapping indicates that hydrophobicity at the hinge region in addition to strain introduced favours domain swapping (Nandwani *et al.*, 2017).

Strain in the hinge-loop, introduced by proline residues in this region, can also control domain-swapping in proteins such as $p13^{suc1}$ (Rousseau *et al.*, 2001) and ribonuclease A (Miller *et al.*, 2010). Indeed, proline residues are frequently observed in hinge-loop regions in domain-swapping proteins (Bergdoll *et al.*, 1997). In fact, more than 40 % of domain-swapped proteins contain proline residues in the hinge-loop region. These proline-rich hinge regions form a more extended conformation that may be beneficial to domain swapping as they may assist formation of the "open" structures to facilitate exchange (Huang *et al.*, 2018).

For example, in $p13^{suc1}$, there are two proline residues located in the hinge region at positions 90 and 92. Mutation of either proline to alanine substantially shifts the monomer-dimer equilibrium. Pro90 is unfavourable in the monomeric form and the mutation P90A shifts the equilibrium towards the monomer, whereas Pro92 is unfavourable in the dimeric form and the P92A mutation shifts the equilibrium towards the dimer (Rousseau *et al.*, 2001). Rousseau *et al*. (2001) suggest that the strain introduced by proline at a particular position in the hinge region can act as a molecular 'loaded spring' that relieves tension by changing its conformation. This, of course, depends on whether the hinge-loop conformation is more strained in the monomer or the DSD (Rousseau *et al.*, 2001).

In addition to the hinge-loop, residues in the hydrophobic core have also been shown to play important roles in stabilising the domain-swapped dimer (Byeon *et al.*, 2003; Byeon *et al.*, 2004). This is exemplified by a series of hydrophobic core mutations in the B1 domain of Protein L from *Peptostrepococcus magnus* that were shown either to favour or disfavour domain swapping (O'Neill *et al.*, 2001). It has been suggested that these hydrophobic core mutations introduce enough strain to overcome the entropic cost of dimer formation thereby favouring domain swapping (O'Neill *et al.*, 2001). A computational study of the B1 domain of the immunoglobulin G binding protein (GB1) also shows that hydrophobic core residues are involved in domain swapping (Sirota *et al.*, 2008). Mutation of Ala34 to a phenylalanine residue results in the destabilisation of the monomer and simultaneous stabilisation of the domain-swapped structure. Computational analysis showed that Phe34 side chains from both monomers are packed tightly against one another at the dimer interface, thus stabilising the domain-swapped dimer. (Sirota *et al.*, 2008).

1.4.3 Artefact of protein crystallisation or physiologically relevant?

There are over 100 proteins that have been shown to domain swap (Berman *et al.*, 2000), some domain-swapped structures are thought to be merely artefacts of protein crystallisation due to the high concentrations used, while other domain-swapped proteins have been shown to be physiologically relevant. In the case of bovine pancreatic ribonuclease (RNase A), DSDs form under non-native conditions during reconstitution of lyophilised protein in acetic acid (Liu *et al.*, 2001). Further studies show that RNase A can adopt two different quaternary structures by swapping one of two different domains, the N-terminal α-helix or the C-terminal β-strand. Appreciable amounts of RNase A domain-swapped dimers are shown to form only under non-native, acidic conditions, at very high protein concentrations (~ 200 mg/ml) (Libonati *et al.*, 1996; Gotte *et al.*, 1999) and formation of domain-swapped dimers may even require incubation at 60 ^{o}C (Gotte *et al.*, 2003).

Domain swapping has been associated with amyloid or fibril formation in neurodegenerative diseases (Janowski *et al.*, 2001; Knaus *et al.*, 2001; Staniforth *et al.*, 2001), aggregation (Rousseau *et al.*, 2001), the evolution of oligomeric proteins (D'Alessio, 1999) as well as the regulation of protein activity (Gotte *et al.*, 1999; Park and Raines, 2000; Liu and Eisenberg, 2002; Newcomer, 2002). Protein concentration can determine whether the monomeric or dimeric conformation is favoured. Thus regulation of the concentration of certain proteins can regulate when dimers form which could be significant in regulating the activity of that particular protein. Since domain swapping is highly important in protein interactions (Sicheri *et al.*, 1997; Rousseau *et al.*, 2001; Cowan-Jacob *et al.*, 2005) and has been shown to regulate the function of numerous proteins (Remenyi *et al.*, 2001; Park *et al.*, 2010), it is therefore implied that the DSD could have a function different to that of its monomeric counterpart (Changeux and Edelstein, 1998). Macromolecular crowding in the cell could increase the local concentration of a particular protein (Cole and Ralston, 1994; Lindner and Ralston, 1995; Rivas *et al.*, 1999), thus increasing its propensity to domain swap and carry out its specific function.

Some DNA binding proteins such as transcription factor Oct-1 and SgrAI (restriction endonuclease) have been found to domain swap which seems to be important for their particular functions and regulation of their activity (Remenyi *et al.*, 2001; Park *et al.*,

2010). Oct-1 function is regulated by the formation of different assemblies of the DSD in response to binding different DNA sequences (Remenyi *et al.*, 2001). SgrAI forms dimers that bind to palindromic DNA sequences and cleaves DNA slowly. A study shows that these dimers can form domain-swapped tetramers (Park *et al.*, 2010). These tetramers form active enzyme, which is shown to cleave at an even higher rate than the dimeric form, upon DNA binding (Park *et al.*, 2010). This implies that domain-swapping regulates activity of the restriction endonuclease SgrAI. Two members of the FOX (forkhead box) P-subfamily of transcription factors (FOXP2 and FOXP3) have been shown to form DSDs via their forkhead domains (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011). The FOXP3 DSD has been shown to bind two distant strands of antiparallel DNA and bring them together within close proximity which is believed to be necessary for its function (Bandukwala *et al.*, 2011).

1.5 Forkhead box family of transcription factors

The forkhead box (FOX) superfamily of transcription factors consists of 50 genes, to date, arranged into 19 subclasses, FOX A to S, that share homology only in their DNA binding, 'winged' helix forkhead domains (FHDs) (Kaufmann and Knochel, 1996; Mazet *et al.*, 2003). The 'winged' helix motif (Figure 1A) was first discovered in the DNA binding domain of the murine HNF-3 (hepatic nuclear factor-3) protein (Clark *et al.*, 1993). The murine HNF-3 protein shared a high degree of amino acid sequence similarity to the *Drosophila* forkhead gene and was later named FOXA1 (forkhead box A1) (Kaufmann and Knochel, 1996). This motif contains three α-helices (H1, H2 and H3) instead of the classical two α-helices in the HTH motif (Pabo and Sauer, 1984), an additional three β-strands (S1, S2 and S3) and two flexible loop regions that form the wings (W1 and W2) and is therefore known as the 'winged' helix motif. Strand S1 is situated directly after helix H1 and along with the other two strands, S2 and S3, forms antiparallel beta-sheets. There are slight structural differences between the 'winged' helix motifs among the FOX proteins. For example, the β-turn in the classical HTH motif is replaced either by a loop/coil conformation as in FOXM1 (Littler *et al.*, 2010) and FOXO3a (Tsai *et al.*, 2007) or a 3_{10}-helix as in FOXP2 (Stroud *et al.*, 2006) or FOXK1a (Tsai, 2006). In addition, the wing regions are truncated in FOXP2 compared to other FOX proteins (Figure 5) (Stroud *et al.*, 2006).

Even though the structure of the forkhead domain between these proteins is virtually unchanged, the FOX proteins have largely diverse yet important functions including regulation of the immune system (Su, 2003; Chae *et al.*, 2006), metabolism (Friedman and Kaestner, 2006), regulation of cellular proliferation and differentiation (Clark *et al.*, 1993; Kaufmann and Knochel, 1996), ageing (Partridge and Bruning, 2008) and embryonic development (Lu *et al.*, 2002; Rausa *et al.*, 2003). Mutations in the forkhead domain have been linked to a wide variety of human diseases such as tumorigenesis (FOXP1) (Banham *et al.*, 2001), alopecia (FOXN1) (Mecklenburg *et al.*, 2001), thyroid agenesis (FOXE1) (Castanet *et al.*, 2002), Axenfeld-Rieger syndrome (FOXC1) (Saleem *et al.*, 2003; Saleem *et al.*, 2004), a speech deficit disorder (FOXP2) and the IPEX (immune dysregulation, polyendocrinopathy, enteropathy and X-linked) syndrome (FOXP3), to name a few.

In order for FOX proteins to carry out their functions, the highly conserved forkhead domain must bind to DNA so as to either activate or repress transcriptional targets (Kaufmann and Knochel, 1996). The functional diversity of FOX proteins surely then relies on the differences between the amino acid sequences of the FHDs which would account for the differences in affinity and specificity for DNA sequences (Coffer and Burgering, 2004). The FHD of the FOX proteins is predominantly monomeric (Li *et al.*, 2004) and this is reflected in the solved crystal structures of the FHDs of FOXA3 (Clark *et al.*, 1993), FOXK1a (Tsai, 2006), FOXO3a (Tsai *et al.*, 2007), FOXO4 (Boura *et al.*, 2010), FOXM1 (Littler *et al.*, 2010) and FOXO1 (Brent *et al.*, 2008) in the presence of DNA. The only exception identified to date is the P-subfamily where FOXP2 and FOXP3 exhibit formation of domain-swapped dimers in their FHDs (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011).

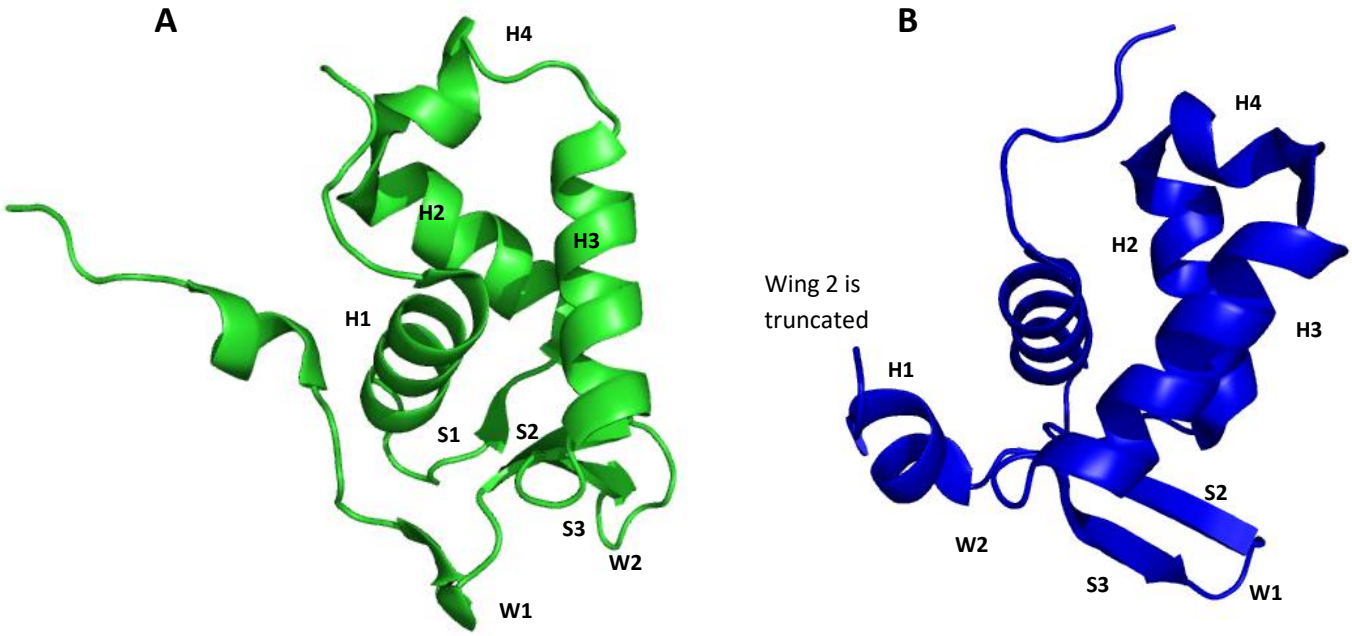


Figure 5. Structural differences between the 'winged' helix motif among FOX proteins. A. FOXM1 and B. FOXP2 FHD monomer transcription factors. The 'winged' helix motif exhibits three alpha-helices (H1, H2 and H3), a loop region (H4), three beta-strands (S1, S2 and S3) as well as two wing regions (W1 and W2). Even though the 'winged' helix motif is highly conserved, there are slight structural differences between the 'winged' helix motifs among the forkhead box (FOX) proteins. For example, the loop region in FOXM1 is replaced by a 3_{10}-helix in FOXP2 and the wings are truncated in FOXP2 compared to FOXM1. PDB codes 3G73 (FOXM1) and 2AO7 (FOXP2). Figure was created using PyMOL (Delano Scientific) (Perumal, 2014).

1.6 FOXP subfamily

1.6.1 Structural similarities

The FOXP subfamily consists of four members namely FOXP1, FOXP2, FOXP3 and FOXP4. These proteins contain structural domains unique to the FOXP-subfamily including an amino-terminal polyglutamine repression region (replaced by polyproline

in FOXP3), C_2H_2 zinc finger domain and leucine zipper domain. They also contain the highly conserved forkhead domain (FHD) characteristic of the FOX superfamily (Figure 5) (Bettelli *et al.*, 2005; Chae *et al.*, 2006; Lopes *et al.*, 2006). In the FOXP subfamily, the FHD is located near the C-terminus of the protein (Figure 6) whereas in most other FOX proteins, the FHD is located near the N-terminus (Clark *et al.*, 1993; Carson *et al.*, 2006).

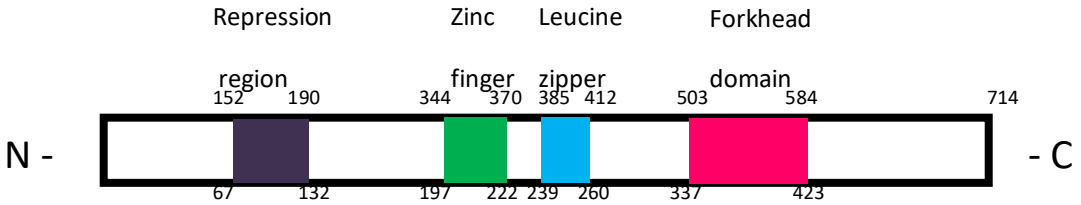


Figure 6. **Schematic of a typical FOXP protein.** Specialised motifs are highlighted. FOXP2 numbering is shown above and FOXP3 numbering is shown below. The polyglutamine repression region (dark purple) in FOXP1, FOXP2 and FOXP4 is replaced by a polyproline region in FOXP3 and these regions are involved in transcriptional repression (Chamberlain *et al.*, 1994) whereas the leucine zipper (blue) is involved in homo- or hetero-dimerisation between the FOXP proteins (Lai *et al.*, 2001), and the forkhead domain (pink) is involved in recognising and binding to specific DNA sequences which is thought to be stabilised by the zinc finger domain (Wang *et al.*, 2003) (green). Figure adapted from Perumal (2014).

Each structural domain is proposed to play a crucial functional role. The polyglutamine stretch is thought to confer repression or could be involved in mediating protein-protein interactions, (Chamberlain *et al.*, 1994), whereas the zinc finger domain can potentially bind DNA and thus stabilise the protein-DNA interaction (Wang *et al.*, 2003). The leucine zipper has been shown to bring two FOXP proteins together and is involved in homo- and hetero-dimerisation events (Li *et al.*, 2004). The FHD is involved in DNA binding specificity (Webb *et al.*, 2017) and mutations in this region have been associated with disease (Lai *et al.*, 2001). The FOXP FHD can form a DSD and FOXP dimers are thus thought to have two separate dimer interfaces, one at the FHD and one at the leucine zipper (Stroud *et al.*, 2006).

1.6.2 The FOXP forkhead domain

The FHD is a ~ 90 amino acid long region that is highly conserved in the FOX family and even more so within the FOXP subfamily (Figure 7). The FOXP FHD forms a 'winged helix' structure consisting of 3 α-helices, 3 β-sheets and 2 loop conformations or 'wings'. The FOXP FHD differs from the conventional winged helix motif of other FOX FHDs in that the turn between helix H2 and H3 is a 3_{10} helix known as helix H4; wing region W1 is truncated and forms a type I turn; and W2 is also shorter (Figure 5), making fewer contacts with DNA (Stroud *et al.*, 2006). As with all 'winged' helix motifs, the FOXP FHD binds to DNA via the insertion of helix H3 into the major groove of DNA (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011). Furthermore, the FHD is capable of forming a DSD in the absence of the rest of the protein (Stroud *et al.*, 2006).

The first FOXP FHD structure to be solved was that of FOXP2, which was solved by X-ray diffraction in the presence of DNA by Stroud *et al.,* (2006). Since then, the structures of the FHD of FOXP3 (Bandukwala *et al.*, 2011) and FOXP1 (Chu *et al.*, 2011) have also been solved; FOXP3 by X-ray crystallography in the presence of DNA and FOXP1 by nuclear magnetic resonance (NMR) spectroscopy in the absence of DNA. Both structures are similar which implies that DNA binding does not induce significant conformational changes in the FHD. The FHD is able to fold autonomously and bind DNA in the absence of the rest of the protein (Stroud *et al.*, 2006).

1.6.3 FOXP2

FOXP2 functions as a transcriptional repressor (Li *et al.*, 2004). The FOXP2 protein is highly conserved and plays distinct roles throughout the vertebrates (Mazet *et al.*, 2003). It is involved in signal transduction pathways, developmental pathways and sensorimotor pathways (Kurt *et al.*, 2012). In fact, FOXP2 has been dubbed the "language gene" (Lai *et al.*, 2001; Fisher and Scharff, 2009) and was catapulted into the spotlight by the British KE family (Lai *et al.*, 2001). Approximately half of the members in the KE family are heterozygous for a FOXP2 missense mutation (R553H) that causes a severe speech and language disorder by impairing orofacial movements that are important for word formation (Lai *et al.*, 2001). In humans specifically, FOXP2 mutations are linked to a variety of diseases such as language defects (Lai *et al.*,

2001), autism and schizophrenia (Gong *et al.*, 2004; Tolosa *et al.*, 2010; Spaniel *et al.*, 2011) as well as cancer (Campbell *et al.*, 2010).

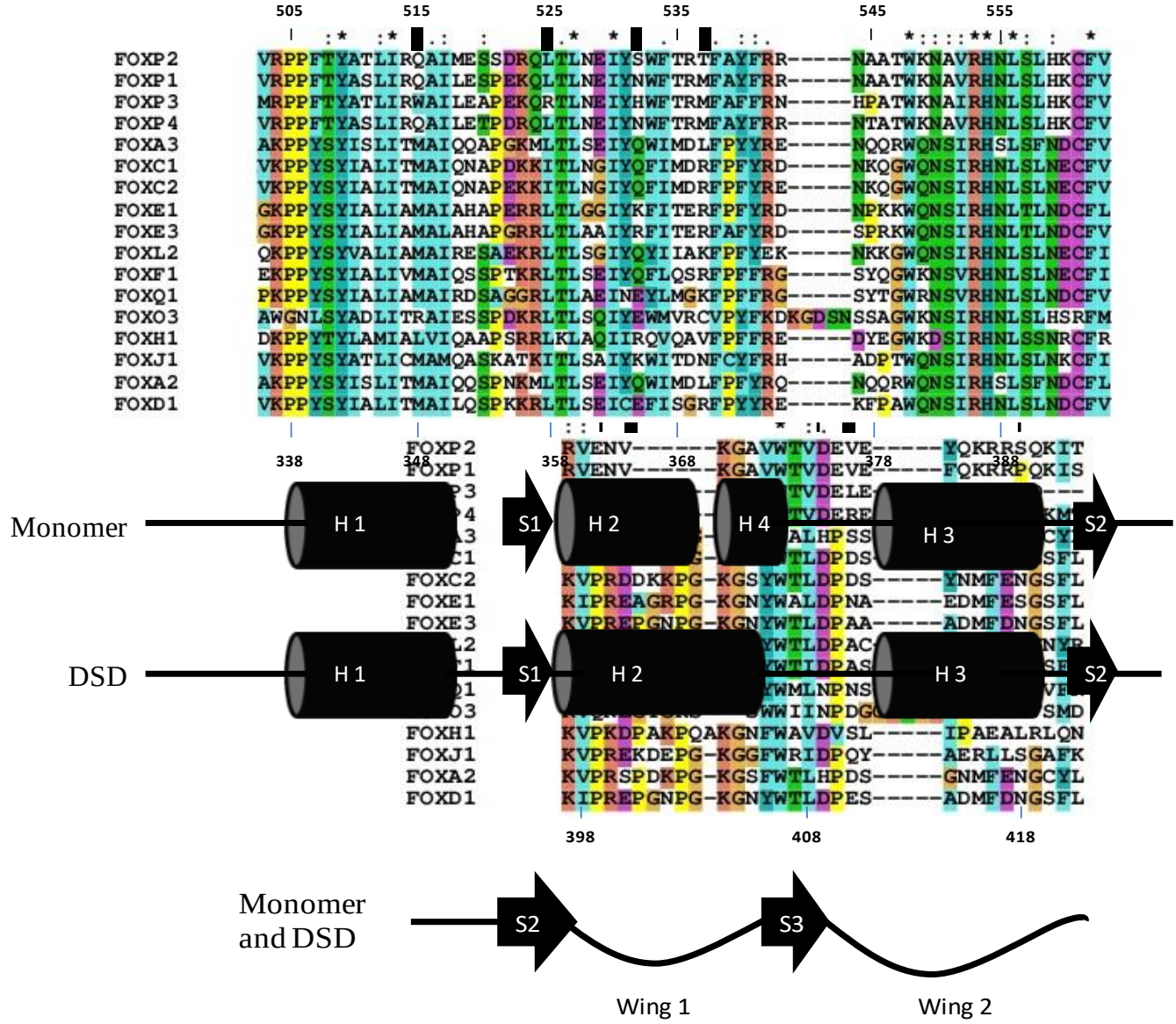


Figure 7. Sequence alignment of the FHD in FOX family members. The alignment was performed using ClustalX2 (Thompson *et al.*, 2002). Non-conserved residues are indicated by black blocks. The '*' indicates fully conserved residues and the ':' indicates similarity. The residues are numbered with respect to FOXP2 (top) and FOXP3 (bottom). The secondary structure is indicated below the alignment where DSD refers to the domain-swapped dimer. The recognition helix (H3) is shown to be the most highly conserved region in the FOX proteins whereas the hinge-loop region (H4), wing 1 and wing 2 are shown to be the least conserved in terms of the amino acid sequence (Perumal, 2014).

1.7 FOXP3

1.7.1 Physiological role

FOXP3 is regarded as the 'black sheep' of the P-subfamily as it has the ability to function as both a transcriptional repressor and an activator as opposed to the other FOXP members which are repressors under normal physiological conditions (Chamberlain *et al.*, 1994; Li *et al.*, 2004; Koh *et al.*, 2009; Bandukwala *et al.*, 2011). Unlike the other FOXP members that are expressed in a number of tissues including the brain, heart and lungs, FOXP3 is only expressed by a particular group of cells known as CD4+ T cells or regulatory T cells (T_{reg} cells) (Coffer and Burgering, 2004; Ziegler, 2006). FOXP3 functions to repress and activate target genes that are involved in effector T cell proliferation (Sakaguchi, 2005) and regulates homeostasis in the immune system by doing so (Koh *et al.*, 2009).

1.7.2 Structure

The crystal structure of the forkhead domain of FOXP3 shows formation of a stable DSD bound to DNA (Bandukwala *et al.*, 2011) where helix H1, strand S1 and helix H2 of one monomer fold over helix H3, strand S2 and S3 of the other monomer to create the domain-swapped dimer conformation (Figure 8). Although the forkhead domains of FOXP1 (Chu *et al.*, 2011), FOXP2 (Stroud *et al.*, 2006) and FOXP3 all dimerise, only the FOXP3 FHD exists solely as a stable dimer at physiological concentrations and does not show formation of monomers in the micromolar concentration range (Bandukwala *et al.*, 2011). Furthermore, even though the FOXP2 FHD is shown to dimerise, work in our lab have shown that the FOXP2 FHD is almost entirely monomeric at physiological concentrations (Perumal *et al.*, 2015; Morris and Fanucchi, 2016).

The crystal structure of the FOXP3 FHD was solved in the presence of DNA and its binding partner, another transcription factor protein, nuclear factor of activated T cells (NFAT1) (Bandukwala *et al.*, 2011). Furthermore, Bandukwala *et al*. (2011) also shows that the FOXP3 FHD requires the presence of a specific DNA sequence in order to interact with NFAT1 to form a ternary complex. The FOXP3 FHD dimer makes extensive contacts with NFAT1. In particular, residues Thr359, Asn361, His365, Glu399 and Glu401 are located at the centre of the FOXP3:NFAT1 protein-protein interface.

These key residues form hydrogen bonds, van der Waals interactions and electrostatic interactions with NFAT1.

The FOXP3 DSD is stabilised by a network of core hydrophobic residues: Tyr364, Trp366, Phe367, Phe371, Phe373, Phe374 and Trp381 (Figure 9) (Bandukwala *et al.*, 2011). These residues are highly conserved among the FOX proteins and a similar network is present in the FOXP2 DSD (Stroud *et al.*, 2006). Furthermore, an additional interface is formed between the H1 helices in the FOXP3 DSD whereas in the FOXP2 DSD, the H1 helices are pointing away from each other (Stroud *et al.*, 2006). This secondary interface is thought to provide more stability to the FOXP3 DSD (Bandukwala *et al.*, 2011).

In the full-length FOXP3 protein, the leucine zipper is involved in dimerisation (Lopes *et al.*, 2006) and the forkhead domain has also been shown to dimerise. Therefore, it is implied that there are two separate dimerisation interfaces: one at the leucine zipper domain and the other at the forkhead domain (Lopes *et al.*, 2006; Bandukwala *et al.*, 2011). Both dimerisation interfaces are shown to be important for the suppressor function. This implies that FOXP3 could form a dimer of dimers by forming clusters which may involve recruitment of other proteins to form large complexes to carry out its function *in vivo* (Bandukwala *et al.*, 2011). In support of this, the crystal structure of the murine FOXP3 leucine zipper domain was solved and shows formation of a two-stranded α-helical coiled-coil which can associate to form a dimer of dimers (Song *et al.*, 2012).

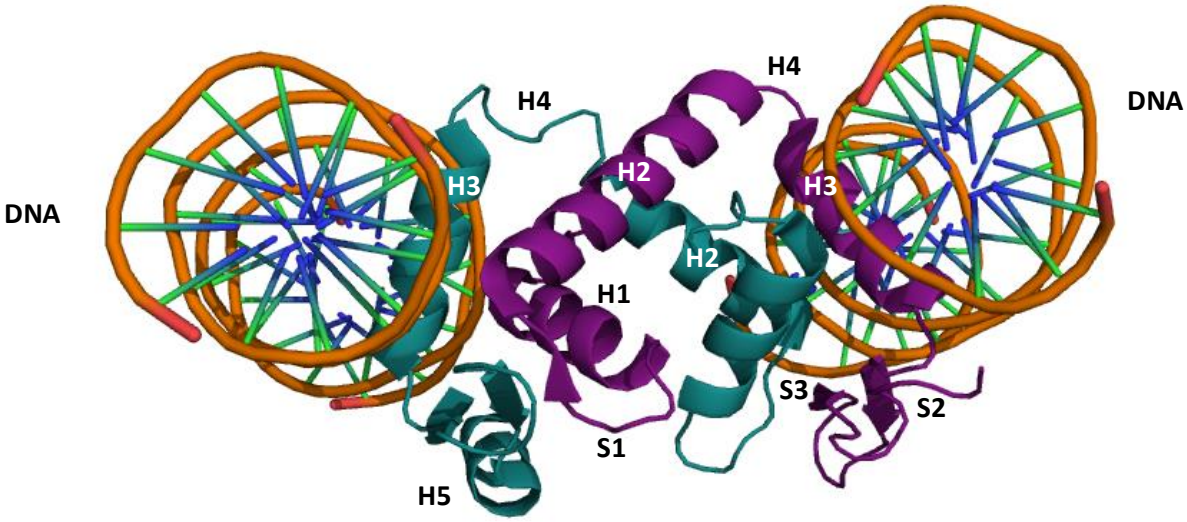


Figure 8. Structure of the FOXP3 domain-swapped dimer bound to DNA. The crystal structure of the FOXP3 FHD (PDB code 3QHR) shows formation of only a stable domain-swapped dimer (DSD). Helix H1, strand S1, and helix H2 of one monomer is exchanged with the same regions of the other monomer. This DSD is capable of binding two separate molecules of DNA where Helix H3, in both monomers, is inserted into the major groove of DNA. One subunit is shown in purple and the other in teal (Perumal, 2014).Figure was created using PyMOL.

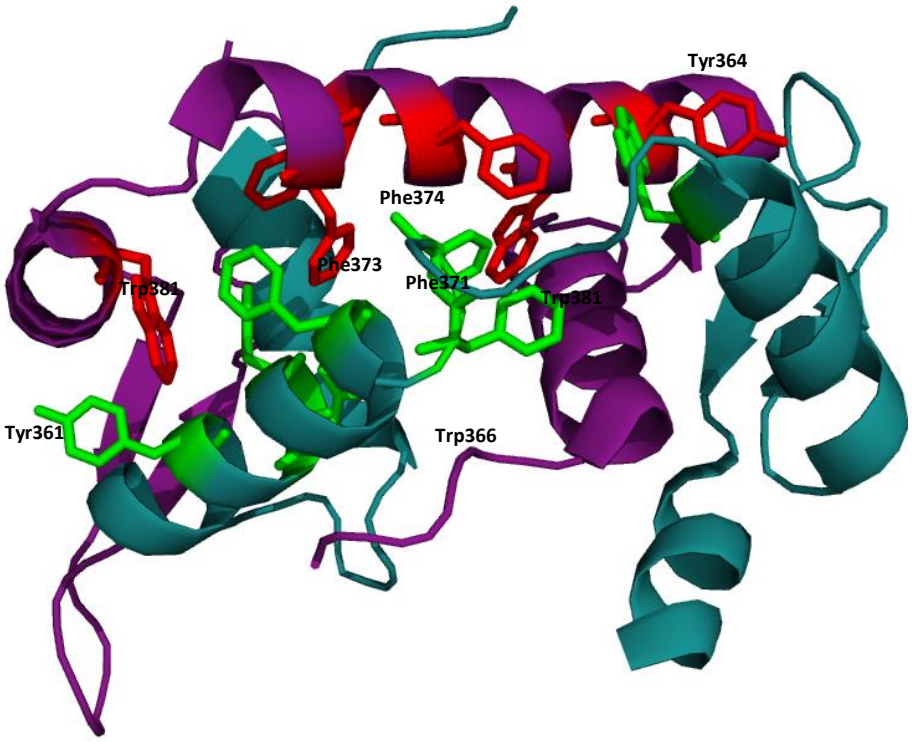


Figure 9. Stabilising hydrophobic core residues in the FOXP3 DSD. Domain swapping is mediated by a hydrophobic network of residues: Tyr364, Trp366, Phe367, Phe374 and Trp381. These residues (green and red) stabilise the DSD interface between one monomer (teal) and the other monomer (purple). PDB code 3QRF (Bandukwala *et al.*, 2011). Figure was created using PyMOL.

The interface formed between the coiled-coil regions is mainly stabilised by electrostatic interactions, although the molecular surface is highly hydrophobic (Song *et al.*, 2012). This hydrophobic region, on the coiled-coil surface, is shown to facilitate bundling of α-helices from separate dimers to form clusters. Furthermore, Song *et. al.*, 2012 showed using cross-linking and gel filtration analyses, that these clusters are not artefacts of crystallisation and indeed they do form in addition to monomer and dimer species Furthermore, the FOXP3 DSD is shown to be able to bind and bridge two separate molecules of DNA, bringing two distant DNA molecules together into close proximity (Figure 8). It is therefore suggested that formation of the DSD may facilitate higher-order complexes such as those formed during transcription (Bandukwala *et al.*, 2011).

1.7.3 Comparison of the FOXP2 and FOXP3 domain-swapped dimers

Despite the high sequence conservation between the FHD of the four FOXP family members (Figure 7), the FOXP3 DSD is considerably more stable than the others (Bandukwala, Wu *et al*. 2011). A comparison between the structures of the FOXP2 and FOXP3 FHD domain-swapped dimers might give clues as to what causes the FOXP3 DSD to be more stable. Indeed, there are noticeable structural differences between the FOXP2 DSD and the FOXP3 DSD (Figure 10). The FOXP3 DSD is more arched and more compact than the FOXP2 DSD and the hinge loop region is relatively longer (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011). There is a network of hydrophobic core residues (Tyr364, Trp366, Phe367, Phe373, Phe374 and Trp381) that stabilises the DSD . This hydrophobic network is similar in both the FOXP2 (Stroud *et al.*, 2006) and FOXP3 (Bandukwala *et al.*, 2011) FHDs however, any sequence difference in this network may account for the difference in stabilities of their DSDs.

Inspection of the FOXP FHD sequence reveals that it is highly conserved with only six residues that are non-homologous (Figure 7) across all FOXP members. Two of the six residues (His365 and Glu401 in FOXP3) are found at the NFAT: FOXP3 interface, Arg358 is located in strand S1 (between H1 and H2), Leu411 is situated in helix H5, and the two remaining residues (Trp348 and Met370) are found at the secondary domain swap interface. It is interesting that Trp348 is conserved across different species in FOXP3, but is absent in the other FOXP members (Bandukwala *et al.*, 2011). These

two hydrophobic and non-conserved residues, located in the secondary interface form interactions that are critical to stabilising the DSD. This is demonstrated by showing that mutations of these residues decrease the stability of the FOXP3 dimer and increase the proportion of monomer present (Bandukwala *et al.*, 2011).

The hinge-loop region of the FOXP FHD, as with all DSDs, plays a significant role in domain swapping. The difference in the FOXP3 hinge-loop region compared to the other FOXP members is that there is a proline residue located at position 378 in the hinge-loop region just before helix H3 (Figure 7). This Pro378 residue is replaced by alanine in all the other FOXP members and is a non-conserved residue in other FOX proteins. In previous studies on domain swapping it has been shown that proline residues are frequently found at the hinge-loop region and are key players in controlling domain swapping (Bennett *et al.*, 1995; Rousseau *et al.*, 2001) and thus Pro378 is likely to contribute significantly to formation and stability of the FOXP3 DSD.

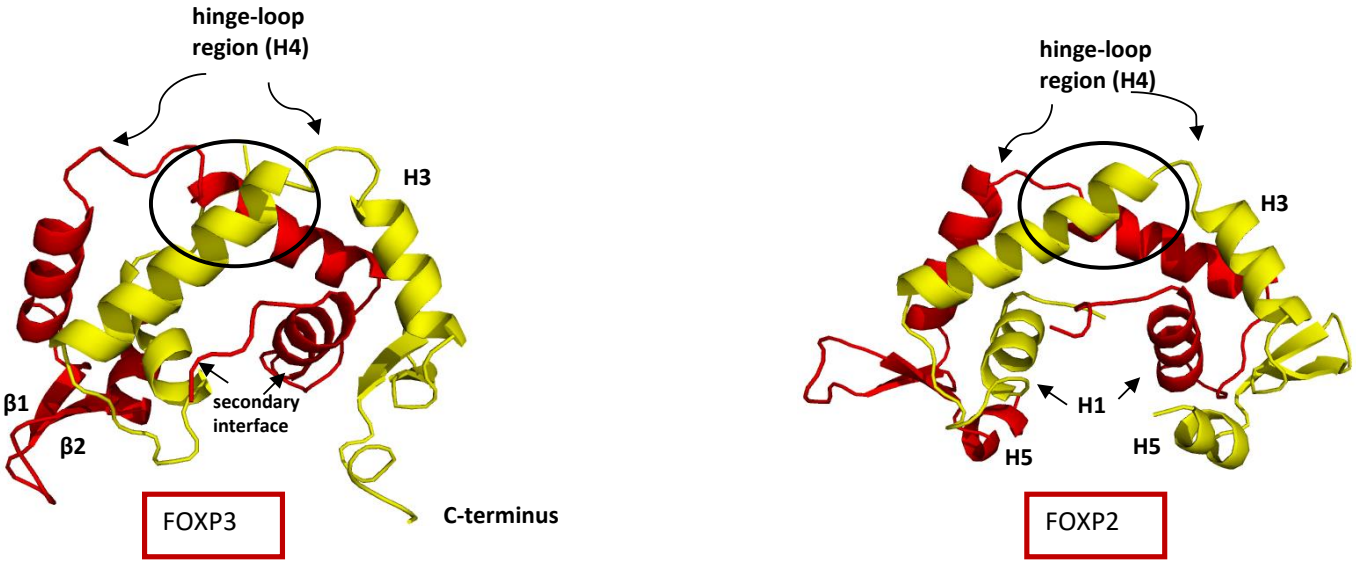


Figure 10. **Structural differences between the FOXP3 and FOXP2 domain-swapped dimers.** The FOXP3 DSD (PDB code 3QRH) (left) (Bandukwala *et al.*, 2011) is slightly more arched than the FOXP2 DSD (PDB entry 2A07) (Stroud *et al.*, 2006) (right). It contains a longer hinge (H4) region than FOXP2 and it contains an additional stabilising interface between the H1 helices whereas in FOXP2, the H1 helices are pointing away from each other.

The hydrophobic dimer interface is circled. (Bandukwala *et al.*, 2011). Image generated using PyMOL (DeLano, 2002)

As a corollary to this, Ala372 in FOXP3 is conserved as an alanine in the FOXP subfamily but is a proline in all other FOX members (Figure 7). Replacing the alanine residue with a proline in FOXP2 (A539P) strongly stabilises the monomeric form over the DSD (Stroud *et al.*, 2006). It is thought that the rigid proline residue prevents extension of the hinge region (H4) which is required for DSD formation (Stroud *et al.*, 2006). Since the A539P mutation results in an exclusively monomeric species in FOXP2 and the W348Q/M370T double-mutant destabilises the DSD in FOXP3, a W348Q/M370T/A372P triple-mutant was created in an attempt to destabilise domain swapping in FOXP3 by disrupting both the hinge-loop region and the secondary interface (Bandukwala *et al.*, 2011). However, this triple-mutant did not result in a completely monomeric species in FOXP3 as expected. The W348Q/M370T/A372P triple-mutant eluted as a single polydisperse peak from size-exclusion chromatography using multi-angle light scattering which suggested that it remained as a mixture of species (Bandukwala *et al.*, 2011). The additional stabilising secondary interface and different key residues in the hinge-region of the FOXP3 DSD may explain why it exists whereas as a more stable dimer than the FOXP2 FHD.

Residues other than Ala372 in the hinge region are less conserved among the FOXP proteins, for example His377 and Phe373 in FOXP3 are replaced by asparagine and tyrosine respectively in other FOXP members. Phe373 appears to have a direct influence on the dimer form as work performed in our lab on FOXP2 has shown that replacing Tyr540 (equivalent to Phe373 in FOXP3) with phenylalanine stabilises the DSD (Perumal *et al.*, 2015). These sequence differences between the FOXP2 and FOXP3 FHDs may give rise to differences in the stabilities of the DSDs formed by each. In turn, these sequence differences may explain why the FOXP3 FHD forms only a stable DSD which is necessary for its function.

1.7.4 IPEX mutations disrupt domain swapping

Mutations in FOXP3 lead to IPEX (immune dysregulation polyendocrinopathy enteropathy X-linked) syndrome. The gene encoding FOXP3 is located on the X-

chromosome and as such, the IPEX syndrome is a fatal autoimmune disease that only affects male infants. Symptoms of IPEX include anaemia, insulin-diabetes and chronic diarrhoea (Levy-Lahad and Wildin, 2001).

Certain missense mutations in the FOXP3 FHD, namely R347H, A384T, F371C and F373A are directly implicated in the IPEX syndrome by either affecting DNA binding or disrupting domain swapping, depending on whether they are located in helix 3 or helix 4, respectively (Bennett *et al.*, 2001; Bandukwala *et al.*, 2011). Furthermore, disease causing mutations in the hydrophobic core of the FOXP FHD have also been shown to disrupt DSD formation. These kinds of mutations interfere with aromatic stabilising interactions by changing the stacking arrangement and orientation of hydrophobic amino acid residues (Byeon *et al.*, 2004; Jee *et al.*, 2008). For example, the mutation of the conserved Phe373 in the FOXP3 hinge-loop region (F373A) causes a shift in the monomer-dimer equilibrium towards the monomer by disrupting the hydrophobic core of the DSD (Figure 9) (Bandukwala *et al.*, 2011). This disruption doesn't affect DNA binding but it is possible that the mutant, which is now mainly monomeric, can no longer bring two DNA molecules together in close enough proximity to carry out FOXP3's function to suppress the proliferation of regulatory T-cells and hence results in the disease phenotype (Walters *et al.*, 2009; Bandukwala *et al.*, 2011).

1.8 Problem identification

The FOXP3 FHD is the DNA-binding domain of the FOXP3 transcription factor. The FOXP3 FHD is shown to form unique, domain-swapped dimers which are absolutely essential for the proteins suppressor function in regulating immune system homeostasis (Bandukwala *et al.*, 2011). Furthermore, the importance of the FOXP3 protein is highlighted by a severe, autoimmune condition known as the IPEX syndrome which is associated with mutations that disrupt domain swapping (Bennett *et al.*, 2001).

In addition, there is increasing evidence that suggests those domains swapped proteins are not merely artefacts of the high concentrations used in protein crystallisation but they may actually be biologically significant. However, there are only a handful of studies to date that investigated the mechanism of domain swapping. Domain swapping proteins do not have similar characteristics and are largely diverse in terms of structure and function and so a unifying mechanism remains elusive. Domain swapping, for the most part is protein-specific as it is dependent on the individual proteins unfolding/ folding pathway. This, together with the fact that very little is known about how FOXP3 functions at a molecular level, prompted an investigation into the mechanism of how the FOXP3 FHD assembles into its functional, DSD form which could shed further light as to how domain swapping occurs using the FOXP3 FHD as a model.

Main Aim

To determine the mechanism of domain swapping of the FOXP3 FHD.

1.8.1 Objectives

- To express and purify soluble wild-type FOXP3 FHD.
- To characterise the secondary, tertiary and quaternary structure of the WT FOXP3 FHD using circular dichroism, fluorescence spectroscopy and size-exclusion chromatography.
- To determine the conformational stability of the WT FOXP3 FHD by using urea-induced equilibrium unfolding, to establish if any stable intermediates are present and the cooperativity of unfolding.
- To determine the kinetic parameters for the unfolding and refolding pathways of wild-type FOXP3 FHD in order to define the mechanism of domain swapping.

CHAPTER 2: EXPERIMENTAL

2.1 Materials

The FOXP3 FHD sequence was codon-optimised for expression in a bacterial expression system. It was synthesised with an N-terminal hexaHis tag and was cloned into the pET-11a vector by Genscript ™ (Piscataway, NJ, USA). Isopropyl-β-D-thiogalactosidase (IPTG) and dithiothreitol (DTT) were purchased from Inqaba Biotec (Pretoria, South Africa). SDS-PAGE molecular weight markers were obtained from Fermentas (Ontario, Canada). 1-Anilino-8-napthalene-sulfonate and ultrapure urea were purchased from Sigma-Aldrich (St. Louis, MO, USA). Low-molecular weight gel filtration protein standards were obtained from GE Healthcare (Chicago, Illinois, US). All other chemicals used were of standard analytical grade. All solutions were filtered prior to spectroscopic studies with 0.22 µm filters (Osmonics).

2.2 Methods

2.2.1 Plasmid purification and verification

The pET-11a vector contains a BamHI cloning site as well as a T7 tag sequence. In order to confirm that the vector contained the correct DNA sequence of the FOXP3 FHD gene, the plasmid DNA was isolated and extracted using the GeneJET^{TM} Plasmid Miniprep kit (Fermentas) and sent to Inqaba Biotech (Pretoria, South Africa) for sequencing. The plasmid DNA was confirmed to contain the correct sequence encoding wild-type FOXP3 FHD with no additional mutations. The sequence was aligned with the known amino acid sequence in the database using the NCBI Blast tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The obtained sequence showed 100 % identity to the FOXP3 FHD protein sequence. Rosetta (DE3) pLysS *Escherichia coli* cells (Novagen) were transformed with the pET-11a vector encoding the FOXP3 FHD for subsequent protein expression.

2.2.2 FOXP FHD expression and purification

Since this was the first time that this protein had been expressed in our laboratory, the first step to obtaining pure protein was to determine the optimal overexpression conditions. Various induction studies were carried out using various IPTG concentrations (0.1 mM up to 1 mM), various induction temperatures (20 ^{o}C to 37 ^{o}C) and various incubation times with IPTG (4 – 6 hours).

FOXP3 FHD was purified from a 2 L batch of culture of Rosetta (DE3) pLysS *Escherichia coli* cells transformed with the pET-11a vector. Ten µl of a glycerol stock of cells (glycerol:cells = 3:1) was added to 100 ml of LB media (1 g tryptone, 0.5 g yeast and 1 g NaCl in 100 ml) containing 100 µg/ml ampicillin. These cells were left to grow for 16 hours at 30 ^{o}C in a shaking incubator at 230 rpm. A 50-fold dilution of this overnight culture was inoculated into fresh LB media containing 100 µg/ml ampicillin and was left to grow at 37 ^{o} C at 230 rpm until the late log phase ($OD_{600} \approx 0.6$). This took approximately 3 hours. Overexpression of the FOXP3 FHD protein was induced by addition of 0.3 mM IPTG. The cells were left to grow at 30 ^{o} C for 4 hours to achieve optimal expression. The cells were then harvested by centrifugation at 6000 *g* for 10 minutes at 4 ^{o} C and the pellet was resuspended in 20 mM Tris-HCl pH 7.6, 500 mM NaCl, 30 mM imidazole (Equilibration Buffer) and stored at -20 ^{o}C.

The following day, the cells were thawed and 200 µg/ml lysozyme and 5 µg/ml DNase were added. The cells were then lysed by sonication and subsequently centrifuged at 24 420 *g* for 25 minutes. The supernatant was loaded onto a HisTrap ™ purification column pre-equilibrated in Equilibration Buffer. The presence of protein was detected in each step of purification using a 15 % SDS-PAGE gel (See section 2.2.3). Once purity was established, the protein was stored at low concentrations (20 – 30 µM) in Storage Buffer (20 mM Tris-HCl pH 7.6, 150 mM NaCl, 1 mM DTT) to prevent aggregation. The protein aggregates signficantly if left at higher concentrations for more than 24 hours.

IMAC is a type of adsorption chromatography that exploits the specific formation of weak bonds between metal ions immobilised on a column and amino acid residues in order to purify proteins (Porath *et al.*, 1975). In this particular study, the HisTrap™ column (Amersham Bioscences, GE Healthcare) was used in conjunction with the ÄKTAprime^{TM} purification system (Amersham Biosciences, GE Healthcare).

The HisTrap column is prepacked with Ni Sepharose High Performance resin which consists of highly cross-linked agarose beads and charged Ni^{2+} ions (HisTrap, 1 ml and 5 ml), Amersham Biosciences, GE Healthcare, 2009) which interact with the imidazole group in histidine. The histidine-tagged fusion proteins bind tightly to the nickel ions while unwanted proteins from the cell lysate are washed off the column. This was followed by a 1.5 M salt wash to remove nonspecific cellular DNA that may be bound to the FHD, by disrupting electrostatic interactions. The purified protein was eluted using a step-gradient elution of Equilibration Buffer against Elution Buffer (20 mM Tris-HCl pH 7.6, 500 mM NaCl and 500 mM imidazole). The purified protein was stored in the Elution Buffer until further use. Small batches of protein were purified at a time and used within one week to minimise aggregation and/or degradation.

2.2.3 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

The SDS-PAGE gels were constructed using a 15 % separating gel (30 % acrylamide, 1 % bisacrylamide, 10 % sodium dodecyl sulfate (SDS), 10 % ammonium persulfate and 1.5 M Tris-HCl pH 8.8) and a 4 % stacking gel (30 % acrylamide, 1 % bisacrylamide, 10 % sodium dodecyl sulfate (SDS), 0.02 % ammonium persulfate and 0.5 M Tris-HCl pH 6.8) Samples were diluted two-fold with a 5 x reducing sample buffer (100 mM β-mercaptoethanol, 20 % (w/v) glycerol, 10 % (w/v) SDS, 0.05 % (w/v) bromophenol blue and 0.5 M Tris-HCl pH 6.8). Samples were boiled for 5 minutes at 95 ^{o}C to denature the protein prior to loading onto the gel. SDS-PAGE was used to analyse the purity of the protein, to determine whether the desired protein was expressed (according to size) and if there were other protein contaminants present. The gel was stained using Coomassie Brilliant Blue (Reisner *et al.*, 1975).

2.2.4 Protein concentration determination

The protein was subjected to a series of dilutions (2X, 4X, 6X, 8X and 10X) and the absorbance at 280 nm was measured. The absorbance corresponding to a concentration of 1X was determined by extrapolation of the dilution series to a 1X concentration.

Protein concentration was determined using the Beer-Lambert law:

$$A280 = \varepsilon cl \qquad \text{Equation 1;}$$

where ε is the molar extinction coefficient, c is the protein concentration and l is the path length ($M^{-1}.cm^{-1}$). The molar extinction coefficient of the native protein in water (Gasteiger *et al.*, 2005) was calculated using the following equation:

$$\varepsilon(\text{Protein}) = n(\text{Tyr})^* \varepsilon(\text{Tyr}) + n(\text{Trp})^* \varepsilon(\text{Trp}) + n(\text{Cystine})^* \varepsilon(\text{Cystine})$$

Equation 2;

where n refers to the number of residues and ε refers to the molar extinction coefficient (Edelhoch, 1967). The molar extinction coefficients of WT FOXP3 FHD dimer was calculated to be 49 960 $M^{-1}.cm^{-1}$.The concentration for the wild-type FOXP3 FHD has been reported in monomeric concentration throughout this thesis. The protein concentration was corrected for absorbance of aggregates at 340 nm using the equation:

$$A_{280\,(corrected)} = A_{280\,(protein)} - A_{280\,(buffer)} - (A_{340(protein)} - A_{340\,(buffer)}) \left[\frac{A_{280(buffer)}}{A_{340(buffer)}} \right]$$

Equation 3

The purity of the protein was established using the A_{280}/A_{260} ratio which indicates the amount of protein relative to the amount of DNA in the sample and was always approximately 1.6 which is acceptable.

2.2.5 Circular dichroism (CD) spectroscopy

In the far-UV region (170 nm - 250 nm), the chiral protein backbone absorbs circularly polarised light which leads to characteristic secondary structure spectra (Woody, 1994). In the far UV region, α-helices give negative troughs at 222 nm and 208 nm and a strong peak at 190 nm whereas β-sheets gives one trough near 217 nm and a peak in the 195 - 200 nm range (Woody, 1994).

All circular dichroism measurements were performed using a Jasco J-810 spectropolarimeter with a quartz cuvette with a 2 mm pathlength. Spectra recorded were an average of 7 scans at a scan speed of 100 nm/min, normal sensitivity, a bandwidth of 1 nm and a data pitch of 0.1 nm in the far-UV range (200 nm to 250 nm). A data pitch of 1 sec was used for time-course measurements. The protein concentration used varied between 4 µM and 20 µM. All CD measurements were obtained in Storage Buffer (20 mM Tris-HCl pH 7.6, 150 mM NaCl, 1 mM DTT) at 20 ^{o}C. All spectroscopy work was conducted at 20 ^{o}C for uniformity. DTT was used in all buffers to maintain the two cysteine residues in the FOXP3 DSD in a reduced state to minimse aggregation by formation of disulphide bonds.

Spectra were corrected for solvent. The mean residue ellipticity (Θ) was calculated using the following equation:

$$[\Theta] = 100\,\theta / cnl \qquad \text{Equation 4;}$$

where θ is the ellipticity (mdeg), *c* is the protein concentration in molar (M), *n* is the number of residues and *l* is the pathlength in cm.

Circular dichroism spectroscopy was used to characterise the FOXP3 FHD secondary structure and to monitor secondary structural changes during urea-induced equilibrium unfolding (See Section 2.2.10).

2.2.6 Fluorescence spectroscopy

Fluorescence is used to analyse the tertiary structure of proteins particularly changes in the local environment of innate fluorophores. Fluorescence occurs after a molecule is excited from the ground state and is exhibited on return to the ground state as emission of light at a longer wavelength. Tryptophan, tyrosine and phenylalanine, the aromatic amino acids give proteins their intrinsic fluorescent properties (Lakowicz and Masters, 2008).

Fluorescence in the near-UV range particularly focuses on and is sensitive to the environment of tryptophan residues. It depends on where in the protein the tryptophan residues are located (buried or on the surface) and how many there are. If the protein has many tryptophan residues, the absorption and emission spectra would be an average of all the environments of those residues. A blue-shifted spectrum would indicate a more buried tryptophan residue that is inaccessible in a nonpolar environment whereas a red shift indicates that the tryptophan residue is exposed in a more polar environment (Lakowicz and Masters, 2008).

Tryptophan and tyrosine generally absorb at 280 nm and have emission maximums of 350 nm and 303 nm respectively. In FOXP3, there are four tryptophan residues: Trp348 (H1), Trp366 (H2), Trp381 (H3) and Trp406 (β2). These tryptophan residues are relatively spread out which may provide information on global events in the tertiary structure of the protein. Tryptophan fluorescence was used to characterise the FOXP3 FHD tertiary structure and to monitor tertiary structural changes in urea-induced equilibrium unfolding (See Section 2.2.10). All fluorescence measurements were performed using a Perkin-Elmer Luminescence Spectrometer LS 50B at 20 ^{o} C with the FLwinlab software.

Intrinsic fluorescence

Tryptophan fluorescence was monitored by exciting the protein at 295 nm. Excitation and emission slit widths were 4 nm for all measurements. A quartz cuvette with a pathlength of 1 cm was used. The protein concentration ranged from 0.5 µM to 20 µM. All spectra obtained were an average of three accumulations with a scan speed of 500 nm/min. Data were plotted using the SigmaPlot® software v.13.0.

Extrinsic fluorescence

8-anilino-1-napthalene sulfonate (ANS) is an amphipathic dye that binds to exposed hydrophobic patches on the protein surface. It is a useful probe to study the formation of intermediates during unfolding/ refolding. ANS has been used extensively to study molten globule intermediates (Christensen and Pain, 1991; Semisotnov *et al.*, 1991; Alam *et al.*, 2015; Haque *et al.*, 2015; Rahaman *et al.*, 2015). When excited at 390 nm, ANS shows an emission maximum wavelength of 520 nm. Upon binding to an exposed hydrophobic cluster on a protein, the maximum emission wavelength shifts to a lower wavelength that is dependent on the hydrophobicity of the binding site/s on the protein. The blue-shifted emission maximum wavelength is accompanied by an increase in the quantum yield.

An ANS stock of 2 mM was prepared in 20 mM Tris-HCl pH 7.6, 150 mM NaCl and 1 mM DTT. Samples containing urea were prepared in the following way: the dilutions of protein and urea were set up according to section 2.2.10 and incubated at 20 ^{o} C for 2 hours to reach equilibrium. ANS was then added to a final concentration of 200 µM and the samples were left for another 2 hours to allow binding of ANS. The samples were excited at 390 nm at 20 ^{o} C using a 1 cm pathlength quartz cuvette. Emission spectra were an average of three accumulations at a scan speed of 500 nm/min. Excitation and emission slit widths used were kept to 5 nm. Spectra were corrected for free ANS and plotted using SigmaPlot ® v. 13.0.

2.2.7 Size-exclusion chromatography (SEC)

SEC is carried out with a stationary phase, which consists of porous gel beads packed into the column and a mobile phase, which is a solvent (Preneta, 1989). The smaller molecules are able to penetrate the pores of the gel and are retarded in migration through the column whereas the larger molecules cannot enter the gel and are therefore eluted earlier than the smaller molecules (Potschka, 1987).

A Hiload™ 16/600 75 pg size-exclusion column (GE Healthcare) together with the ÄKTA protein purification system (Amersham Bioscience, GE healthcare) was used to confirm

the quaternary structure of the WT FOXP3 FHD to be dimeric (~ 26 kDa) (Bandukwala *et al.*, 2011). This was done using protein standards: conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa). A plot of the K_{av} as a function of molecular weight (M_r) of the standards was used to determine the molecular weight of the FOXP proteins. The K_{av} value is calculated as follows: K_{av} = (elution volume – void volume)/ (column volume – void volume) (Tayyab *et al.*, 1991).

The column was equilibrated with Storage Buffer: 20 mM Tris-Cl pH 7.6, 150 mM NaCl and 1 mM DTT. The FOXP3 FHD was dialysed against the Storage Buffer and was concentrated using a P-10 membrane (Osmicon). Different concentrations ranging from 4 µM to 57 µM of FOXP3 FHD were used. 1 ml of each sample was loaded onto the column. The protein was detected by absorbance at 280 nm.

2.2.8 DNA binding studies

Electrophoretic mobility shift assays (EMSAs) are DNA binding assays that are used to detect protein-nucleic acid interactions but can also provide quantitative information such as binding stoichiometry, affinities and kinetics under particular conditions (Hellman and Fried, 2007). This technique is based upon the principle that free DNA is much smaller and thus has a higher electrophoretic mobility than protein-DNA complexes. Thus, the free DNA migrates further down into the gel than do the protein-DNA complexes (Garner, 1986; Lane *et al.*, 1992). Polyacrylamide gels were used because they show high resolution of separation of protein-DNA complexes from free DNA (Fried, 1989).

Since the FOXP3 FHD is the DNA binding domain of the transcription factor and DNA binding is required in order for this protein to function, DNA binding studies were carried out in order to confirm that the protein that was purified in this study was indeed functional. Furthermore there is a close link between structure and function of proteins and this technique provides additional information as to whether the protein was folded correctly. If the protein binds to DNA, then it serves to confirm that the protein is in its correctly folded native conformation.

The duplex cognate sequence is as follows:

5’ - TTAGG**TGTTTAC**TTTCATAG - 3’

and was used in these DNA binding studies because it is known to bind FOXP proteins (Nelson *et al.*, 2013) and has been shown to have a high affinity for these proteins (Webb, Steeb *et al.,* 2017).The duplex DNA was synthesised by Whitehead Scientific (Lethabong, South Africa).

DNA at a concentration of 0.5 µM was incubated with 1 µM FOXP3 FHD for 30 minutes. The protein-DNA complexes were resolved on a 12 % acrylamide gel (12 % (w/v) acrylamide, 0.33 (w/v) bis-acrylamide, 89 mM Tris-borate pH 8.3, 0.007 % (v/v) ammonium persulfate, 0.00075 % (v/v) TEMED) at 4 ^{o}C at 150 V for 1.5 hours in TBE buffer (89 mM Tris-borate pH 8.3). The gel was stained with SYBR® Green stain (Life Technologies).

2.2.9 Reversibility and recovery

Before equilibrium unfolding could be performed, recovery and reversibility studies had to be carried out in order to ensure that the unfolding process was indeed reversible and to ensure that the unfolding reaction had reached equilibrium. Only then could the change in free energy (ΔG) and *m*-values be calculated (Myers *et al.*, 1995; Pace and Scholtz, 1997).

The recoverability was determined by unfolding 40 µM protein in 8 M urea, followed by returning the protein to native conditions by 10 x dilution to 0.8 M urea which results in a final protein concentration of 4 µM. The circular dichroism and fluorescence spectra of the refolded protein were compared to that of a control sample: 4 µM protein incubated with 0.8 M urea. Coincidence of the control and refolded spectra would indicate that the protein had refolded and the native state was recovered. Reversibility was determined for each point along the denaturation curve. Protein at 40 µM (for circular dichroism) and 5 µM (for tryptophan fluorescence) was unfolded in 8 M urea and then diluted to different final urea concentrations ranging from 7 M to 0.8 M. The recovery was monitored using both circular dichroism and fluorescence spectroscopy.

2.2.10 Urea-induced equilibrium unfolding

Urea-induced equilibrium unfolding was used to determine the unfolding pathway and to establish whether any thermodynamically stable intermediates are present at equilibrium. Denaturant-induced unfolding in a simple, two-state model is based upon the assumption that only two states, the folded state (N) and the unfolded species (U) predominate. Three main stages are involved: the region where the folded state is most stable, the transition region, and the region where the unfolded species predominates (Pace and Scholtz, 1997). When the environment of a small, folded globular protein is gradually changed towards conditions that favour the unfolded species, the folded conformation initially changes very little and this is known as the pre-transition stage. The equilibrium constant between the folded and unfolded speciess ([N]/[U]) gradually decreases but the folded state remains predominant in this stage. As the environment changes even further, however, the ([N]/[U]) reaches approximately 1 and both folded and unfolded conformational states are present in almost equal amounts. This is the transition region. As the protein unfolds further, the unfolded species predominates and the equilibrium constant decreases even more. This stage is known as the post-transition stage (Myers *et al.*, 1995; Pace and Scholtz, 1997).

A two-state model is the simplest model because it does not address the presence of equilibrium intermediate species. If these species are present, more complex models should be used. However, in a three state model, upon addition of denaturant, the conditions that favour the native state are disrupted toward those that favour an intermediate (I), until an equilibrium is reached between the native state and the intermediate. As the protein is exposed to even further denaturing conditions, the intermediate state tends toward the unfolded species (U) until the unfolded species predominates. In this model, two transition states are present. The first one describes the structure of the native state relative to the intermediate, while the second transition describes the structure of the intermediate state relative to the unfolded species (Walters *et al.*, 2009). In this model the quaternary level adds modes of stability. Intra- and inter-chain interactions contribute significantly to the overall stability (Walters *et al.*, 2009).

Denaturation curves are used to measure the conformational stability of the native state of proteins relative to the unfolded species. It is useful to use different probes to monitor unfolding because intermediates that are not revealed by one technique may be revealed using another. Furthermore, when the data obtained from multiple techniques coincide, it serves as a method to validate the results.

Equilibrium unfolding studies were performed at pH 7.6 at 20 ^{o}C. Fresh 10 M urea stocks were used as the denaturant and were prepared using Storage Buffer (20 mM Tris-HCl pH 7.6, 150 mM NaCl, 1 mM DTT). FOXP3 FHD (0.5 µM to 20 µM) was incubated with 0 to 8 M urea in Storage Buffer and left to equilibrate at 20 ^{o}C for 2 hours prior to measurements. Far-UV CD (ellipticity at 222 nm) (Section 2.2.5) and tryptophan fluorescence spectroscopy (emission wavelength monitored at 364 nm) (Section 2.2.6) were used as secondary and tertiary structural probes, respectively. In tryptophan fluorescence, the maximum emission wavelength observed for unfolded protein was 364 nm. This particular wavelength was chosen to monitor unfolding with respect to the unfolded species. Secondary structure was monitored at 222 nm using circular dichroism which specifically monitors α-helical content. All unfolding data were obtained in triplicate.

Urea solution

Urea was prepared in Storage Buffer with fresh DTT added on the day of use after defrosting. Urea stocks were stored at – 20 ^{o} C and used within one week as urea slowly decomposes to form ammonium and cyanate ions that chemically modify amino groups. All unused urea was discarded after defrosting once. The urea concentration was confirmed using the refractive index of urea:

$$[\text{Urea}] = 117.66\Delta n + 29.753\Delta n^2 + 188.56\Delta n^3$$

Equation 5;

where Δ is the refractive index between urea and buffer (Warren and Gordon, 1966).

Two-state model

A simple, two-state dimer mechanism where the protein exists in two predominant states: folded (N) or unfolded (U) with the absence of intermediates, can be described by Equation 6.

$$N_2 \underset{}{\overset{K_{eq}}{\rightleftharpoons}} 2U \qquad \text{Equation 6;}$$

where N refers to the native dimer and U refers to the unfolded species and K_{eq} represents the equilibrium constant for the reaction.

For a two-state unfolding model, the protein will either be in the native state (N_2) or in a unfolded species (U). This implies that the sum of the fraction of native species (F_N) and fraction of the unfolded species (F_D) is equal to one (Equation 7).

$$F_N + F_D = 1$$

$$\therefore F_N = 1 - F_D \qquad \text{Equation 7;}$$

The measurements obtained from circular dichroism and fluorescence spectroscopy will be referred to as *y* for the purposes of the following discussion. Using Equation 8, it is implied that at any point along the unfolding curve, the unfolding transition will be:

$$y = Y_N F_N + Y_U F_U \qquad \text{Equation 8;}$$

where Y_N and Y_U represent the values of *y* characteristic of the folded and unfolded speciess, respectively. By combining Equation 7 and Equation 8, we obtain Equation 9:

$$F_U = \frac{(Y_N - y)}{(Y_N - Y_U)} \qquad \text{Equation 9;}$$

The change in free energy, ΔG is used as a measure of the conformational stability of a protein. Using the above equations, the ΔG and equilibrium constant, K_U can be calculated as follows:

$$K_U = \frac{F_U}{F_N}$$

$$= \frac{F_U}{(1 - F_U)}$$

$$= \frac{(Y_N - y)}{y - Y_D} \qquad \text{Equation 10;}$$

By using Equation 9:

$$\Delta G = - RT \ln K_D$$

$$= - RT \ln \left[\frac{(Y_N - y)}{y - Y_D} \right] \qquad \text{Equation 11;}$$

where *R* is the gas constant (1.987 cal. mol^{-1}. K^{-1}) and *T* is absolute temperature.

Because Δ*G* is linearly proportional to denaturant concentration, the conformational stability of the protein in the absence of denaturant, $\Delta G_{(H2O)}$ can be extrapolated back to zero molar (0 M) denaturant concentration using Equation 11. This is known as the linear extrapolation method (Pace, 1986).

$$\Delta G = \Delta G_{(H2O)} - m\,[D] \qquad \text{Equation 12;}$$

where *m* is a measure of the dependence on denaturant concentration at the midpoint of the denaturant curve, $[D]_{1/2} = \frac{\Delta G_{(H2O)}}{m}$ and D refers to denaturant.

The *m*-value relates the free energy to the denaturant concentration, it is obtained by the slope of the plot and provides information on the surface area that is exposed to the solvent upon denaturation (Myers *et al.*, 1995).

Three-state model

For a three-state dimer model used in this study, the protein may exist in the folded (N), intermediate (I) or unfolded (U) states. Furthermore, unfolding of the dimeric protein may involve the formation of a monomeric (2I) or dimeric intermediate (I_2) species.

For a monomeric intermediate:

K_1 K_2

$N_2 \leftrightarrow 2I \leftrightarrow 2U$ Equation 13;

Where K_1 represents the equilibrium constant for the N ↔ I transition and K_2 represents the equilibrium constant for the I ↔ U transition.

K_1 K_2

$N_2 \leftrightarrow I_2 \leftrightarrow 2U$ Equation 14;

Where K_1 represents the equilibrium constant for the N ↔ I_2 transition and K_2 represents the equilibrium constant for the I_2 ↔ U transition.

In the three-state model, the native dimeric protein can first dissociate into a monomeric intermediate (2I) which further unfolds in a second reaction step into unfolded monomers. An alternative is that the native dimeric protein unfolds first into a dimeric intermediate (I_2) which proceeds to form unfolded monomers in a second reaction step (Warren and Gordon, 1966).

2.2.11 Stopped-flow kinetics

During equilibrium unfolding of a small, single-domain protein, any intermediates that form are usually thermodynamically unstable and transient species (Pace and Scholtz, 1997). However, in order to determine the mechanism and pathway of unfolding and refolding, it is necessary to identify and characterise these intermediates. In this study, the kinetic parameters were determined for both the folding and unfolding pathways of the FOXP3 FHD to determine the mechanism of domain-swapping.

In the simplest kinetic unfolding model, protein unfolding seems to be an all-or-none process with very little to no partially unfolded speciess preceding complete unfolding.

This suggests that there is a single kinetic phase, a single rate constant (Equation 15) and no lag period which implies that there is a single rate-limiting step and all molecules have equal probability of unfolding (Creighton, 1993).

$$A(t) - A(\infty) = \sum_{i=1}^{n} A_i e^{-k_i t} \qquad \text{Equation 15;}$$

where *A(t)* refers to the amplitude, *A(∞)* refers to the offset value, A_i refers to the phases, *t* refers to time and $-k_i$ refers to the rate (s^{-1}).

In order to define the folding pathway of the FOXP3 FHD dimer, kinetic experiments were used to follow events that occur during both folding and unfolding using fluorescence spectroscopy as a tertiary structural probe. Since the FOXP3 FHD contains four tryptophan residues located at different regions around the protein, fluorescence will probe global structural changes, rather than local structural changes.

All kinetics experiments were conducted on a BioSequential SX-18MV stopped-flow reaction analyser from Applied Photophysics (Leatherhead, UK). Excitation pathlength used was 10 mm and the emission pathlength used was 2 mm. The excitation bandpass was 4.63 nm. The photomultiplier voltage was 500 V for all experiments. The sample handling unit was kept at a temperature of 20 ^{o} C throughout. Analysis of the kinetics data was done using SigmaPlot software which uses the Levenberg-Marquardt algorithm of non-linear least squares method (Marquardt, 1963).

Single-jump unfolding and refolding kinetics

Unfolding kinetics can provide information about the unfolding pathway and the sequence of events that occur during unfolding on a millisecond timescale. All kinetics experiments were done by monitoring changes in intrinsic tryptophan fluorescence emission with an excitation wavelength of 295 nm and a cut-off filter of 320 nm to minimise light scatter from the excitation wavelength. Baseline values were obtained for 4 µM protein at equilibrium at various urea concentrations (from 0 to 8 M) to determine the end points of the reactions and to confirm that the reactions had run to completion.

All experiments were conducted at 20 ^{o} C and three traces were averaged for each urea concentration. The final average trace was fitted using SigmaPlot. The fit of the data to the general equation were shown to be good by the residual plots.

To produce the unfolding reactions, 12 µM native FOXP3 FHD was mixed in an asymmetric ratio of 1:5 with Storage Buffer containing final urea concentrations of 3 M to 5 M urea to reach the intermediate species or 7 M to 8 M urea to reach the unfolded species.

Refolding reactions were performed by rapidly mixing unfolded protein in a 1:5 ratio with buffer containing 0 M to 0.8 M urea to reach the native state or 3 M to 5 M final urea concentration to reach the intermediate state. Unfolded protein solution was prepared by incubating 12 µM FOXP3 FHD with 8 M urea for 1 hour at 20 ^{o} C. The final protein concentration was 2 µM.

Kinetic data analysis and fitting

Unfolding data were fitted to a single exponential or double exponential model while refolding data were fitted to a double exponential model. A plot of the residuals was obtained for each fit and shows the error between the experimental points and the theoretical curve which is used to determine which exponential model (single or double) best fit the data. The fitting process was performed for data collected at each urea concentration.

To determine the unfolding rate in the absence of denaturant, the following equation was used:

$$\ln k_u = \ln k_u^{H2O} + m_u[D] \qquad \text{Equation 16;}$$

where k_u is the apparent rate constant at different urea concentrations, k_u^{H2O} represents the apparent rate of unfolding in the absence of denaturant and m_u represents the change in accessibility of the unfolded species to solvent.

To determine the folding rate in the absence of denaturant, the following equation was used:

$$\ln k_f = \ln k_f^{H2O} + m_f[D] \qquad \text{Equation 17;}$$

where k_f is the apparent rate constant at different urea concentrations, k_f^{H2O} represents the apparent rate of folding in the absence of denaturant and m_f represents the accessibility of the folded state to solvent.

From the above data, a chevron plot is constructed (Walters *et al.*, 2009) which includes information about the folding and unfolding rates in the absence of denaturant, the *m*-values for folding and refolding as well as information on any transition or intermediate states that may be present (Walters *et al.*, 2009).

CHAPTER 3. RESULTS

The main aim of this study was to investigate the mechanism of domain swapping of the FOXP3 FHD. Since this protein had not been studied prior in our lab, the purification and overexpression of the FOXP3 FHD had to be optimised. This was doing using induction trials which were undertaken at various conditions in order to produce an optimal yield of soluble protein. The secondary, tertiary and quaternary structure of the protein was then characterised using circular dichroism, tryptophan fluorescence and size-exclusion chromatography which indicated that the structure of the FOXP3 FHD in solution is consistent with the crystal structure, however, it is not as compact in terms of the tertiary structure. Urea-induced equilibrium unfolding was performed to determine the equilibrium unfolding pathway of the FOXP3 DSD and to establish if any thermodynamically stable intermediates were present as well as the cooperativity of unfolding. Unfolding and refolding were then performed to derive the complete folding and unfolding pathways of the FOXP3 DSD.

3.1 Purification of wild-type FOXP3 FHD

3.1.1 Plasmid verification

The sequencing results confirmed that the pET-11a vector contained the correct cDNA sequence encoding wild-type FOXP3 FHD. The pET-11a vector contained the sequence encoding the FOXP3 FHD contains a hexahistidine tag at the N-terminus as the protocol established by Bandukwala *et al*. (2011) so as to minimise interference with domain swapping.

3.1.2 Expression and purification of the FOXP3 FHD

Initially, T7 Express Competent Cells were transformed with the pET 11-a vector encoding the wild-type FOXP3 forkhead domain (FHD) sequence. Induction trials were carried out to determine the optimal conditions for maximal protein production. Concentrations of IPTG (isopropyl β-D-1-thiogalactopyranoside) used were 0.1 mM, 0.2 mM, 0.3 mM, 0.5 mM and 1 mM. At 37 ^{o}C, the FOXP3 FHD was predominantly

insoluble and at 30 ^{o}C and 20 ^{o}C, the FOXP3 FHD protein was not overexpressed well. Rosetta (DE3) pLysS Competent Cells (Novagen) were subsequently used, as in a protocol previously established by Bandukwala *et al*. 2011, which dramatically improved the solubility of the FOXP3 FHD. Small-scale induction trials were carried out at 37 ^{o}C and 20 ^{o}C to determine the optimal conditions for soluble protein production. The FOXP3 FHD protein was overexpressed but solubility did not increase significantly at 37 ^{o}C. At 20 ^{o}C, the protein did not overexpress well. Further induction trials were carried out at 30 ^{o}C and the optimal conditions for induction were established to be 0.2 mM IPTG for 4 hours at 30 ^{o}C (Figure 11).

The solubility of the FOXP3 FHD was assessed by SDS-PAGE of the cytosol and pellet fractions (Figure 13). FOXP3-FHD was found to be predominantly in the supernatant while approximately 30 % appeared in the insoluble fraction or pellet (Figure 11). The lack of the overexpressed band in the flow-through lane (Figure 13) suggests that most of the fusion protein bound to the purification column. Monomeric FOXP3 FHD was found through SDS-PAGE (Figure 12), to have a molecular mass of approximately 13 kDa which corresponds to what was previously reported in the literature (Bandukwala *et al.*, 2011).

The FOXP3 FHD was purified using immobilised metal ion affinity chromatography (Figure 13). Since only a single band was detected in the purified fractions, and corresponds to a molecular mass of 13 kDa, it can be concluded that the FOXP3 FHD had been successfully purified. The experimentally determined molecular weight of 13 kDa of the FOXP3 FHD monomer is consistent with the calculated molecular weight of 12.9 kDa using the Expasy ProtParam tool (https://web.expasy.org/protparam/).

3.1.3 Protein concentration determination

It is important to note that all protein stocks were kept at a relatively low concentration (15 – 30 µM) in Elution Buffer (20 mM Tris-HCl pH 7.6, 500 mM NaCl and 500 mM imidazole) prior to dialysis against Storage Buffer (20 mM Tris-HCl pH 7.6, 150 mM NaCl, 1 mM DTT) just before use. This was done as the FOXP3 FHD is prone to aggregation if stored at higher protein concentrations (40 – 150 µM) in Elution Buffer for longer than 1 week. Concentration determinations were consistent with R^2 values between 0.95 and 1. The yield of purified protein obtained was 2 mg/L of culture was obtained in a typical purification.

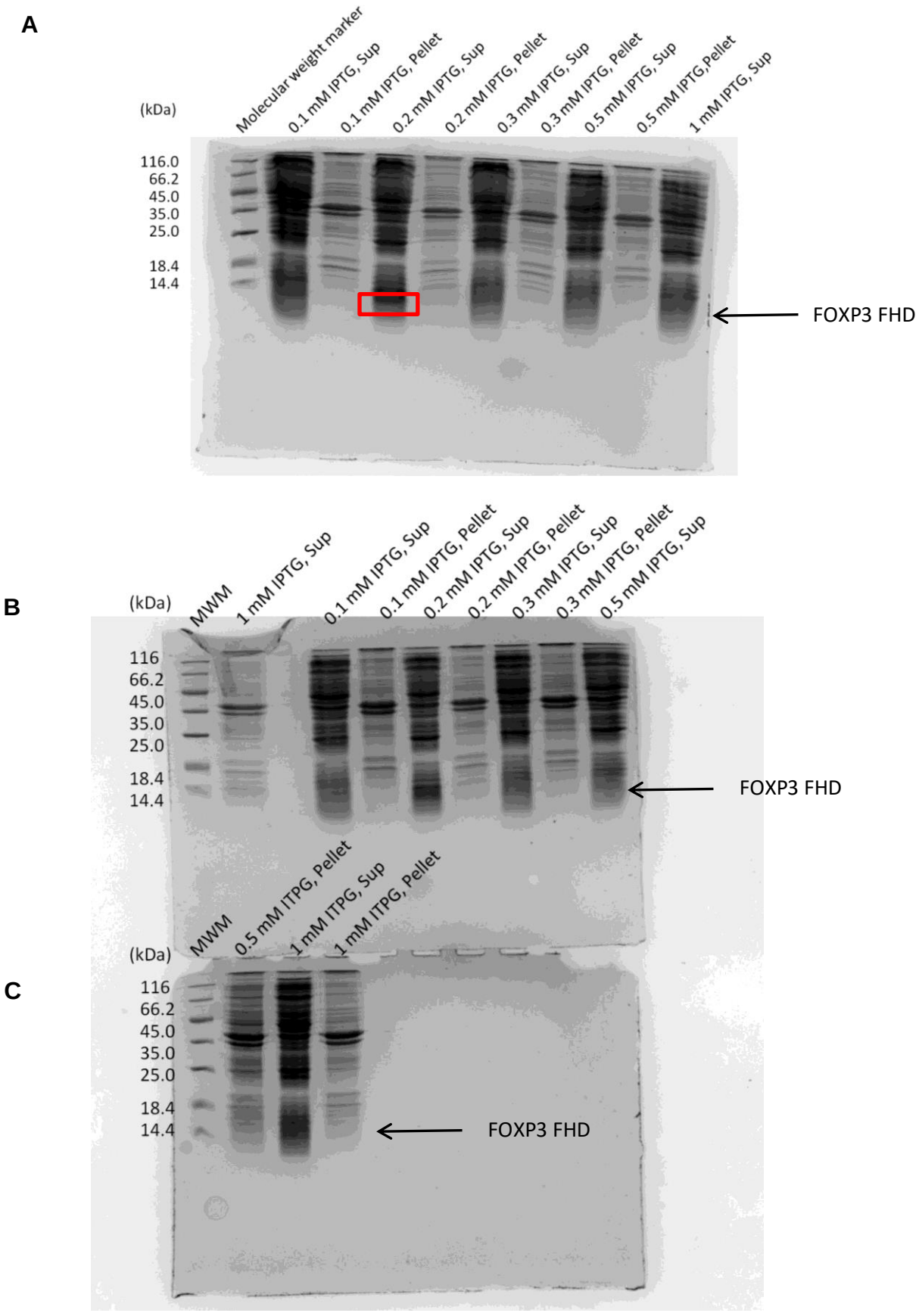


Figure 11. 15 % SDS-PAGE gel of small-scale induction trials at 30 ^{o}C for FOXP3 FHD at **A.** 4 hours and **B.** 5 hours and **C.** 6 hours. The optimal condition was determined to be 0.2 mM IPTG for 4 hours at 30 ^{o}C (red block) as the highest amount of soluble protein was produced.

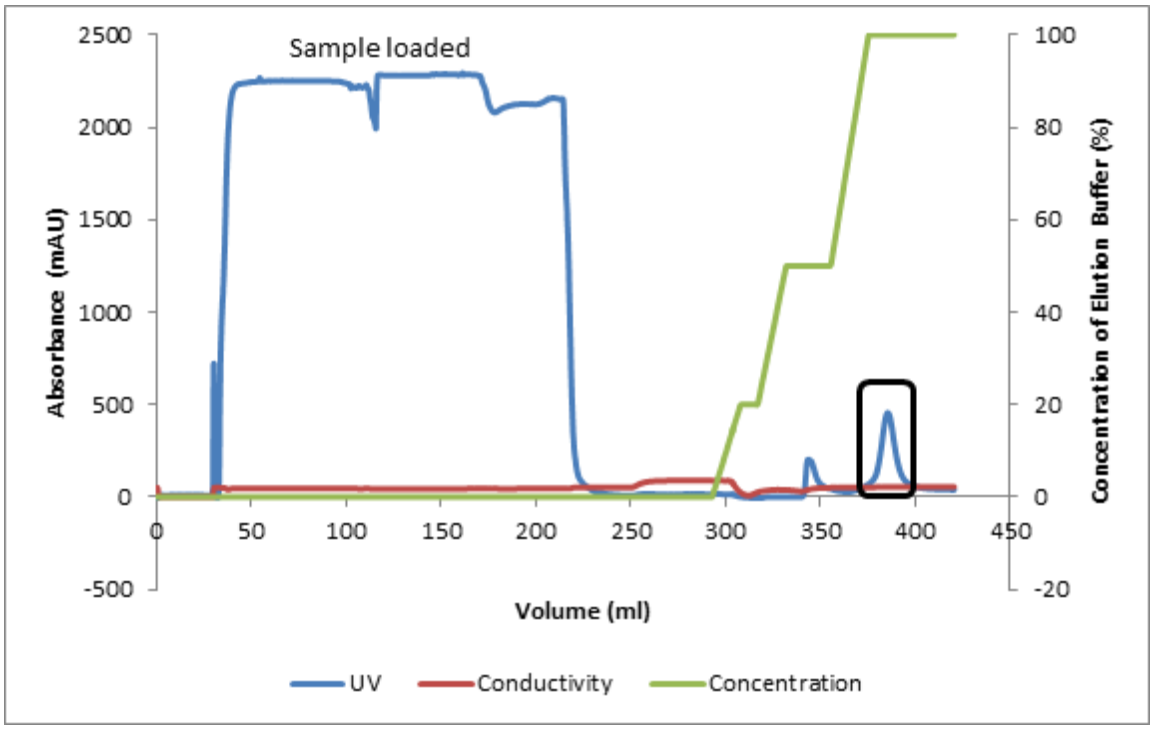


Figure 12. **Elution profile of the FOXP3 FHD using the IMAC HisTrap™ column**. The elution profile of the FOXP3FHD off the IMAC HisTrap™ purification column (GE Healthcare) recorded using absorbance at 280 nm (blue) on the Äktaprime system (GE Healthcare). A 1.5 M NaCl salt wash (red) was used to eliminate contaminating DNA fragments. Pure FOXP3 FHD was eluted in 70 % of Elution buffer (green) (350 mM imidazole) in a single symmetrical peak (black block).

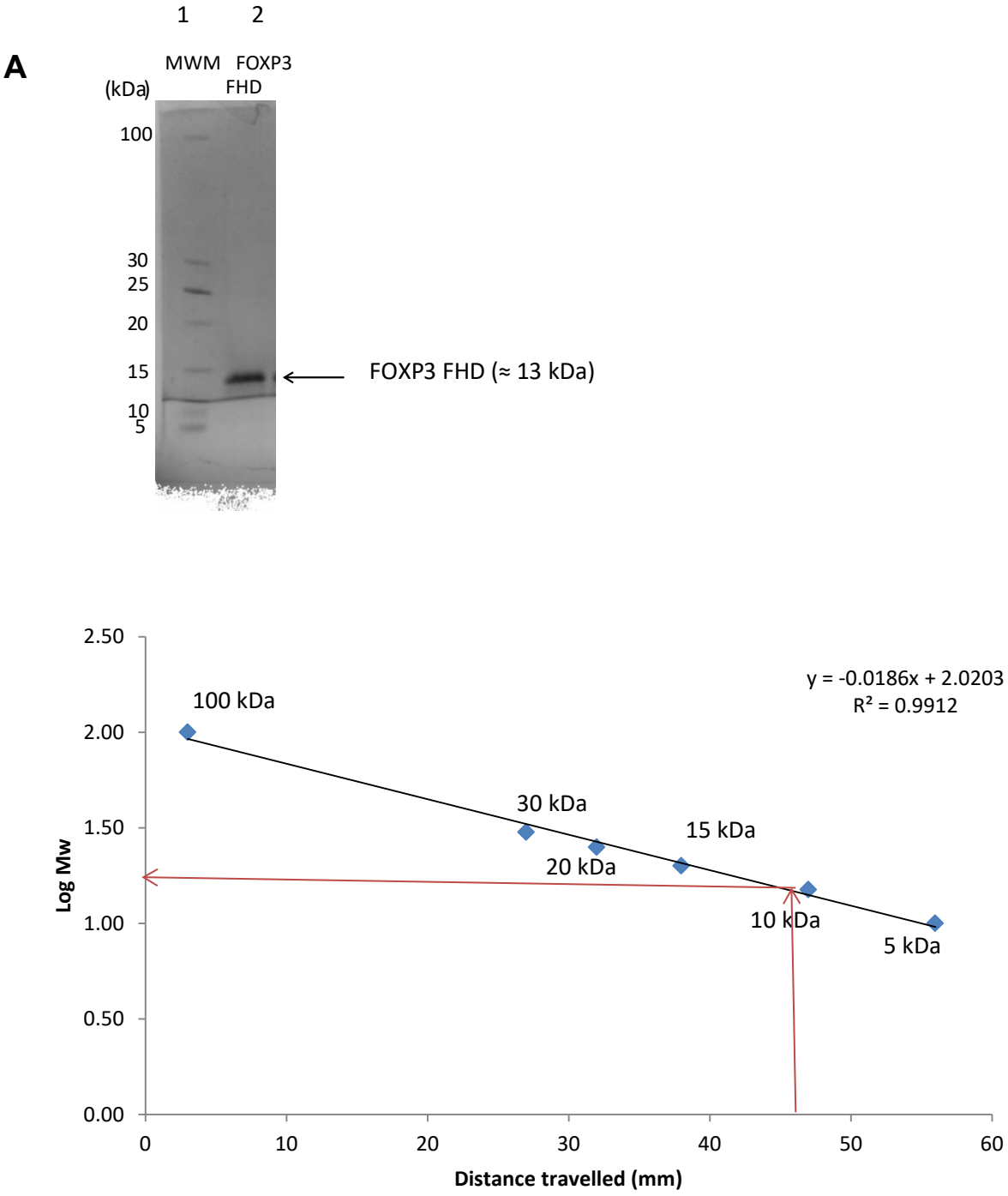


Figure 13. SDS-PAGE and calibration curve. **A.** 15 % SDS-PAGE gel stained with Coomassie Brilliant Blue illustrating a pure fraction obtained from a representative purification of the FOXP3 FHD (black arrow). Lane1 contains the molecular weight marker and lane 2 contains successfully purified FOXP3 FHD with a molecular mass of 13 kDa (black arrow). **B.** Calibration curve indicating the molecular mass of the FOXP3 FHD (red arrow) relative to that of the molecular weight markers (R^2 = 0.99).

3.2 Structural integrity of the wild-type FOXP3 FHD

Circular dichroism and tryptophan fluorescence was used to determine whether the secondary and tertiary structure, respectively was consistent with the crystal structure. Furthermore, these techniques were used to confirm that the denaturant, in this case, urea disrupted the secondary and tertiary structure significantly. This is done to confirm that the protein is fully unfolded in urea and to establish the native and unfolded baselines for urea-induced equilibrium unfolding.

3.2.1 Secondary structure

The secondary structure of the FOXP3 FHD was studied using circular dichroism (CD) spectroscopy. The far-UV CD spectra were obtained between 250 and 196 nm at 20 ^{o}C (Figure 14). The spectrum is characteristic of a protein that has high α-helical content with characteristic troughs at 208 nm and 222 nm. The spectrum of the unfolded protein shows that the α-helical secondary structure is completely abolished in the presence of 8 M urea as the characteristic minima are no longer evident, and the protein exists as a random coil under these conditions. The secondary structure percentage contribution of helical, beta sheet and random coil content was calculated using the K2d algorithm of Dichroweb (Whitmore and Wallace, 2008) with an NRMSD of 0.086 and found to consist of 27 % helix, 25 % beta structures and 48 % random coil. The crystal structure of the FOXP3 FHD is shown to have 48 % helical content and 10 % beta sheet content, the rest of the protein (42 %) is indicated to have no assigned secondary structure. There is a noticeable difference between the secondary structure of the crystal structure and that calculated using Dichroweb (Whitmore and Wallace, 2008). These discrepancies can be due to the secondary structural differences between the flexible structure in solution and the crystal structure (Hontz *et al.*, 2006).

3.2.2 Tertiary structure

The tertiary structure of the FOXP3 FHD was investigated using intrinsic fluorescence spectroscopy. This technique is sensitive to the immediate microenvironments of tryptophan residues and is thus a powerful tool used to probe local tertiary structural changes (Lacowicz, 1999). The FOXP3 DSD contains four tryptophan residues per subunit, namely Trp348 (H1), Trp366 (H2), Trp381 (H3) and Trp406 (S2). The Trp406 residue is located in strand S2 which is fully solvent-exposed. Trp381 is located in the primary hydrophobic interface and is partially exposed to solvent while Trp348 and Trp366 are located in the secondary stabilising interface of the DSD, are partially solvent-exposed and are in close proximity to charged residues Arg358, Arg369 and Glu352 according to the solved crystal structure (Figure 15)

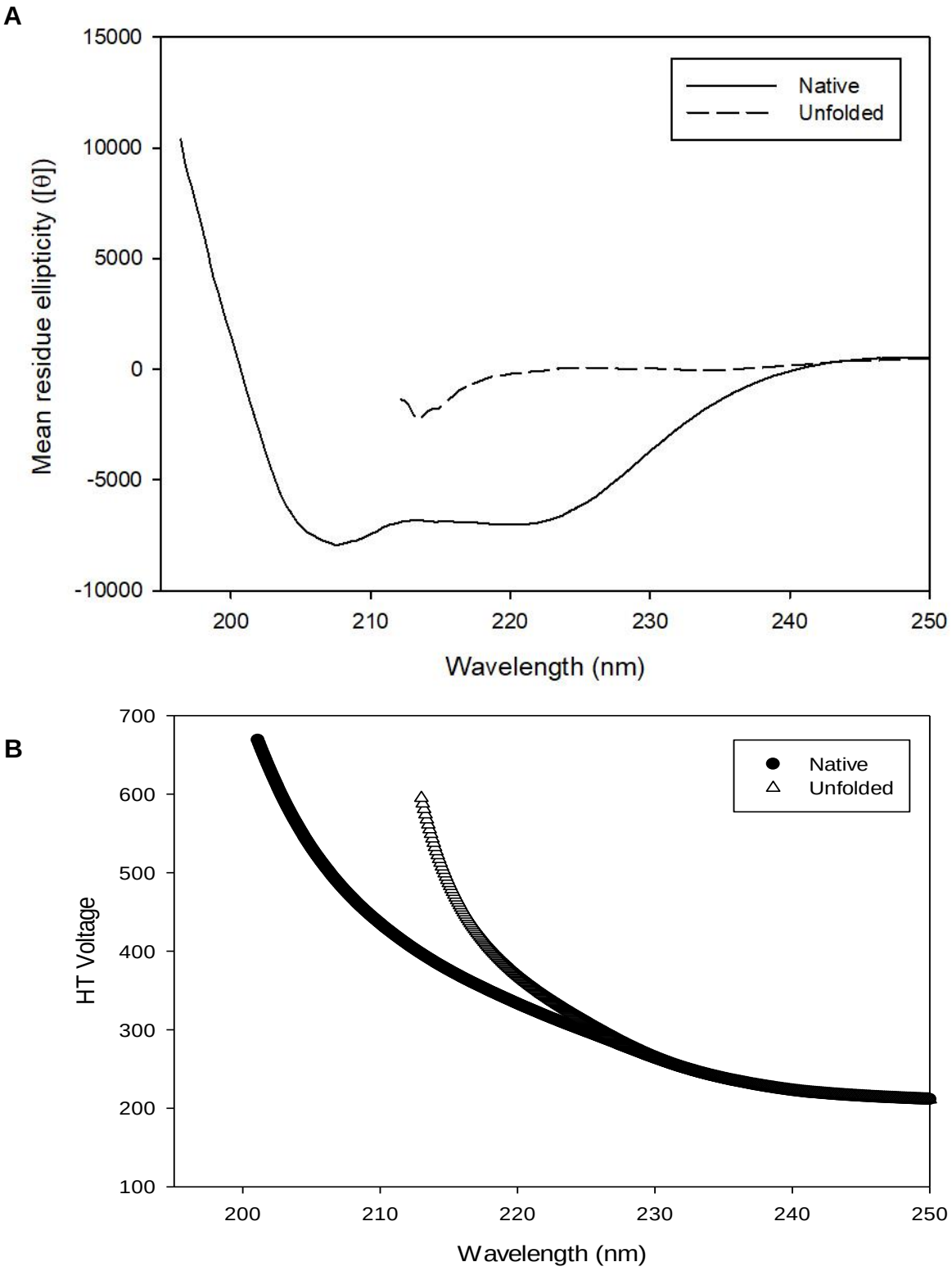


Figure 14. A. Far-UV spectra of the wild-type FOXP3 FHD. Native FOXP3 FHD is represented by a solid line while FOXP3 FHD unfolded in 8 M urea is represented by a dashed line. The spectral analyses were performed at 20 ^{o}C using 8 µM FOXP3 FHD in 20 mM Tris-HCl pH 7.6, 100 mM NaCl, 1 mM DTT. The data were smoothed using a negative exponential methodology.**B. HT voltage profile.** Voltage for the native spectrum is indicated by the solid circle and the voltage for the spectrum of unfolded FOXP3 FHD is shown by open triangles.

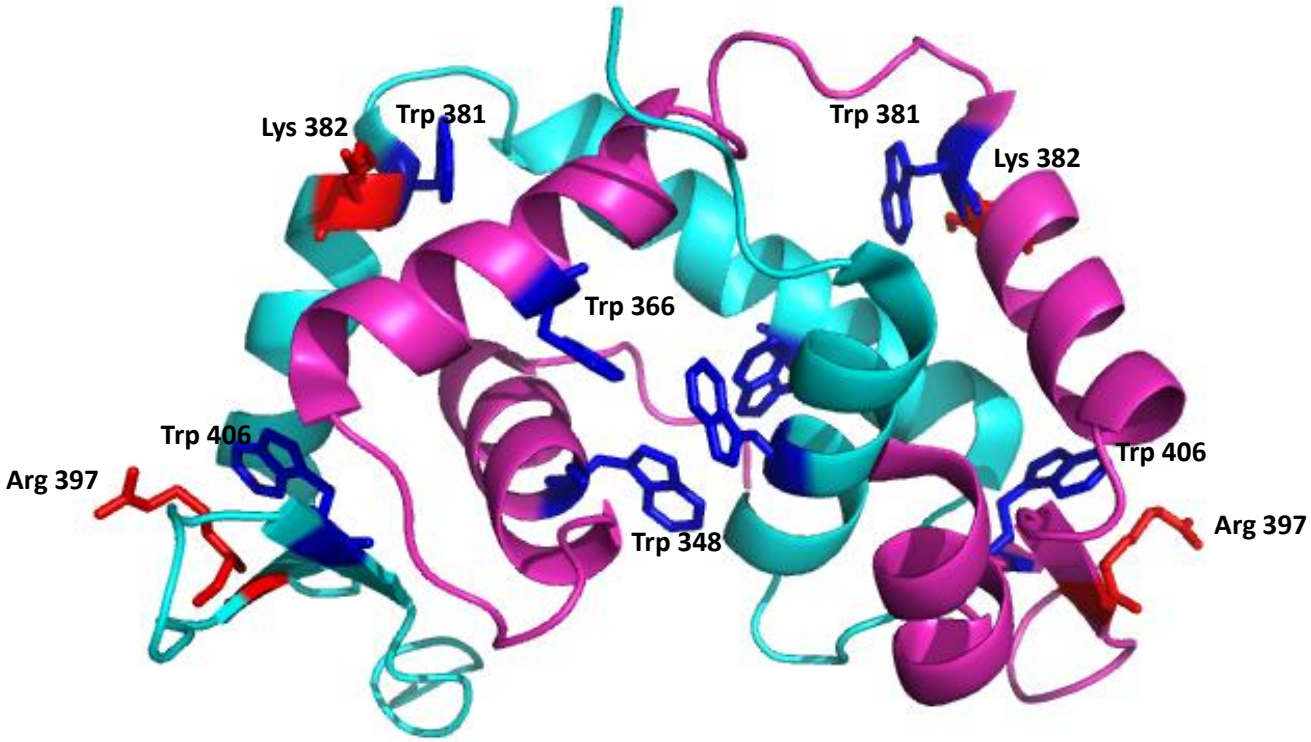


Figure 15. Crystal structure of the FOXP3 domain-swapped dimer indicating the locations of the tryptophan residues. The Trp406 residues are fully exposed to solvent while Trp381, which is involved in the primary network of hydrophobic residues is partially-buried. The Trp348 and Trp366 residues form part of the secondary interface that stabilises the FOXP3 DSD. The tryptophan residues are shown in dark blue. Positively-charged residues are indicated in red. Trp406 and Trp382 are in close proximity to Arg397 and Lys382, respectively. Figure generated using PyMOL (DeLano Scientific).

Tryptophan residues that are fully water-accessible emits at wavelengths near and above 350 nm whereas completely buried tryptophan residues in a hydrophobic microenvironment emits near 320 nm or below (Ahmad *et al.*, 2010). The indole group is particularly sensitive to changes in solvent and is thus highly dependent upon solvent polarity and/ local environment which makes it a highly-sensitive probe for tertiary structural changes upon unfolding (Lacowicz, 1999).

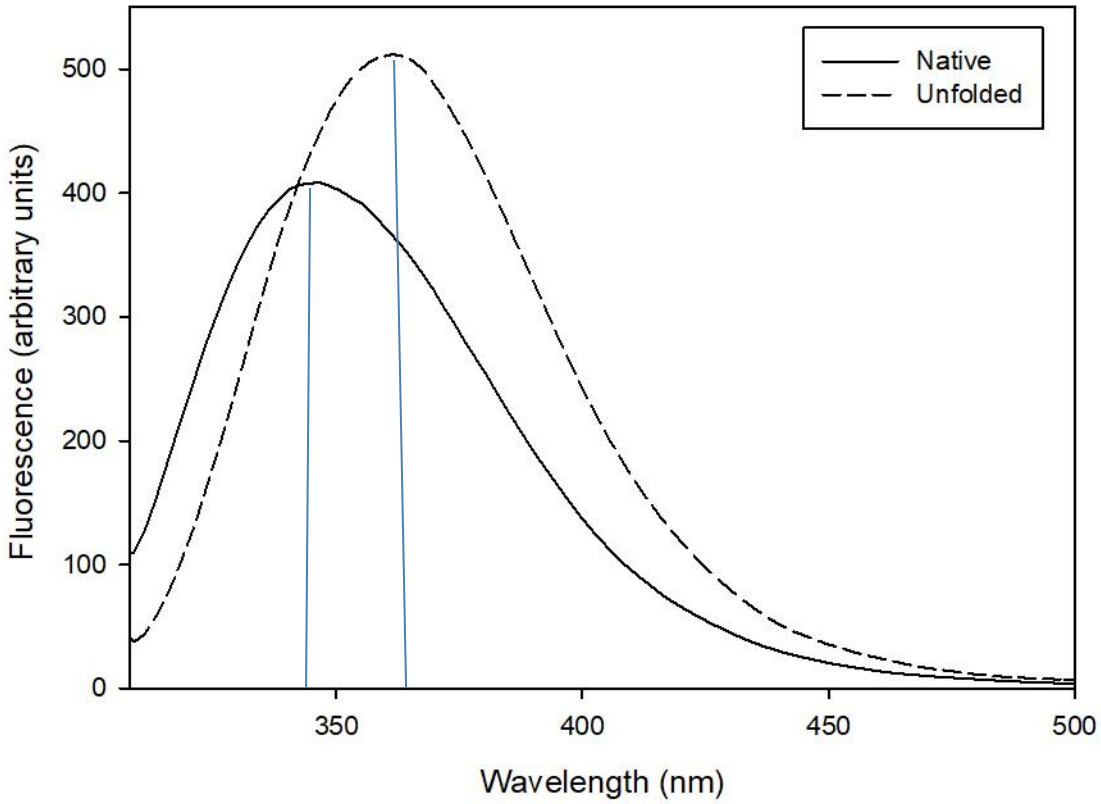


Figure 16. Fluorescence spectra of the wild-type FOXP3 FHD. The tryptophan residues were selectively excited at 295 nm. Maximum emission of the native protein (solid line) occurred at 343 nm, while that of unfolded protein (dashed line) occurred at 361 nm. Spectra were obtained for 4 µM FOXP3 FHD at 20 ^{o}C in 20 mM Tris-HCl pH 7.6, 150 mM NaCl and 1 mM DTT and were smoothed using the negative exponential methodology.

The fluorescence maximum emission wavelength of 341 nm for native protein is consistent with the average of the local environments of all the tryptophan residues in the crystal structure (Figure 16). Since the FOXP3 DSD contains 8 tryptophan residues in total, it is difficult to deconvolute this data in order to determine the relative contribution of each tryptophan residue. Further *in silico* experiments can be performed in order to do so. Upon unfolding in 8 M urea, the spectrum shows a red-shifted

emission maximum to 361 nm and an increase in the quantum yield (Figure 16). This is indicative of all tryptophan residues being exposed to solvent upon unfolding. Furthermore, the red-shifted emission maximum indicates that the tryptophan residues have shifted to a more polar microenvironment upon unfolding (Gorinstein *et al.*, 2000). Urea molecules have been shown to interact preferably with amino acid residues containing aromatic or non-polar side chains and forms a shield between the side chains and the surrounding water (Gorinstein *et al.*, 2000). This could potentially be the reason for the red-shifted emission maximum of the unfolded protein being greater than 350 nm. Scattering from protein aggregation could also cause a significant red-shifted emission maximum, however aggregation can be ruled out in this study as the Raleigh peak obtained for each spectrum did not indicate aggregation. Furthermore, emission maximum peaks around 360 nm for unfolded protein are not uncommon (Gorinstein *et al.*, 2000; Ahmad *et al.*, 2010). The quenched fluorescence in the native state compared to the unfolded species could be attributed to positively-charged residues such as Arg397 and Lys382 being closer in proximity to Trp406 and Trp381, respectively, in the native state than in the unfolded species or it could be due to solvent quenching as Trp406 is fully solvent-exposed while Trp381 is only partially solvent-exposed. In the native state, Trp366 and Trp348 are located in close proximity to positively-charged residues which are known fluorescence quenchers (Lacowicz, 1999). Upon unfolding, the Trp348 and Trp366 residues are no longer in close proximity to the positively-charged residues and thus an increase in the quantum yield is observed. Fluorescence intensity is not a good probe to monitor changes in tertiary structure upon denaturant-induced equilibrium unfolding as small changes in protein concentration can cause it to fluctuate significantly. The emission maximum wavelength, or the wavelength at which the spectra peaks, is not dependent on protein concentration and is sensitive to the polarity of the microenvironments of the tryptophan residues and is therefore a better probe to monitor tertiary structural changes and was thus used in this study.

3.2.3 Quaternary structure

The quaternary structure of the FOXP3 FHD was studied using size-exclusion chromatography (Figure 17A). A single peak was obtained on the Hiload 16/600 Superdex 75 pg column for various concentrations of protein, ranging from 4 µM to 57

µM FOXP3 FHD, which indicates that the protein is pure and exists as a single species in solution as previously reported (Bandukwala *et al.*, 2011). The FOXP3 domain-swapped dimer was shown to have a molecular mass of approximately 26 kDa (Figure 17B) which is approximately double the size of the monomer (13 kDa) as shown by SDS-PAGE.

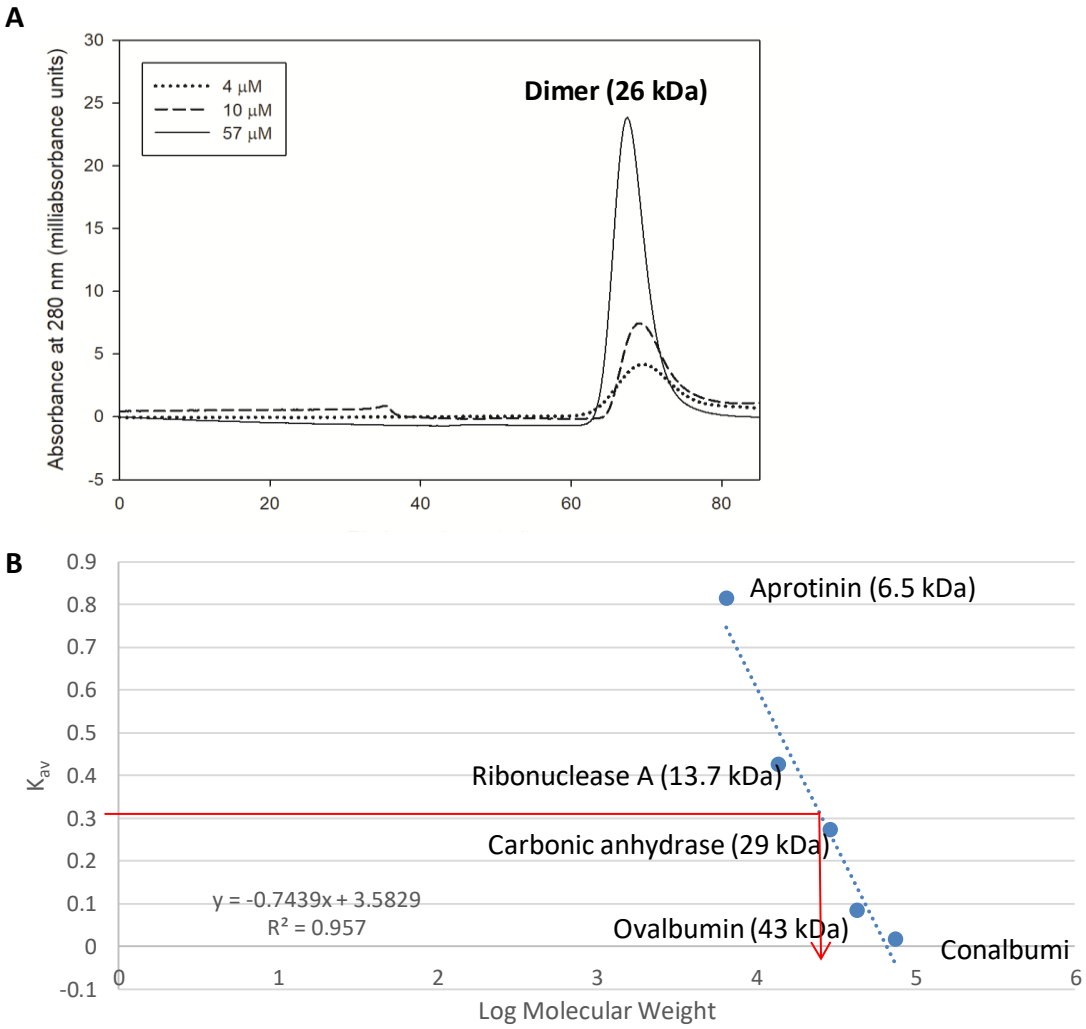


Figure 17. Size-exclusion chromatography of the FOXP3 FHD. 1 ml of FOXP3 FHD (20 mM Tris-HCl pH 7.6, 500 mM NaCl) was loaded onto the Hiload 16/60 Superdex 75 pg column (GE Healthcare). **A.** The elution profile indicates one prominant peak which eluted at a volume of 75 ml which indicates that the molecular weight is 26.2 kDa as calculated from a B. calibration curve using protein standards (red line). This indicates that the FOXP3 FHD is dimeric in solution in the micromolar concentration range.

3.3 DNA binding of the FOXP3 FHD

An electrophoretic mobility shift assay (EMSA) was used here to confirm that the FOXP3 FHD protein was indeed functional. The larger, protein-DNA complexes have lower electrophoretic mobilities and migrate more slowly through the polyacrylamide gel than does free DNA (Fried, 1989). The FOXP3 FHD binds to Nelson DNA at a protein:DNA ratio of 2:1 and shows that the protein is functional (Figure 18). Free DNA was used as a negative control (Lane 1 in Figure 18). The protein only binds to DNA when there is an excess of protein as no binding occurs when the protein:DNA ratio is 1:1 (Lane 2 Figure 18). This is consistent with previous studies performed on the FOXP2 FHD (Perumal, 2014; Morris and Fanucchi, 2016).

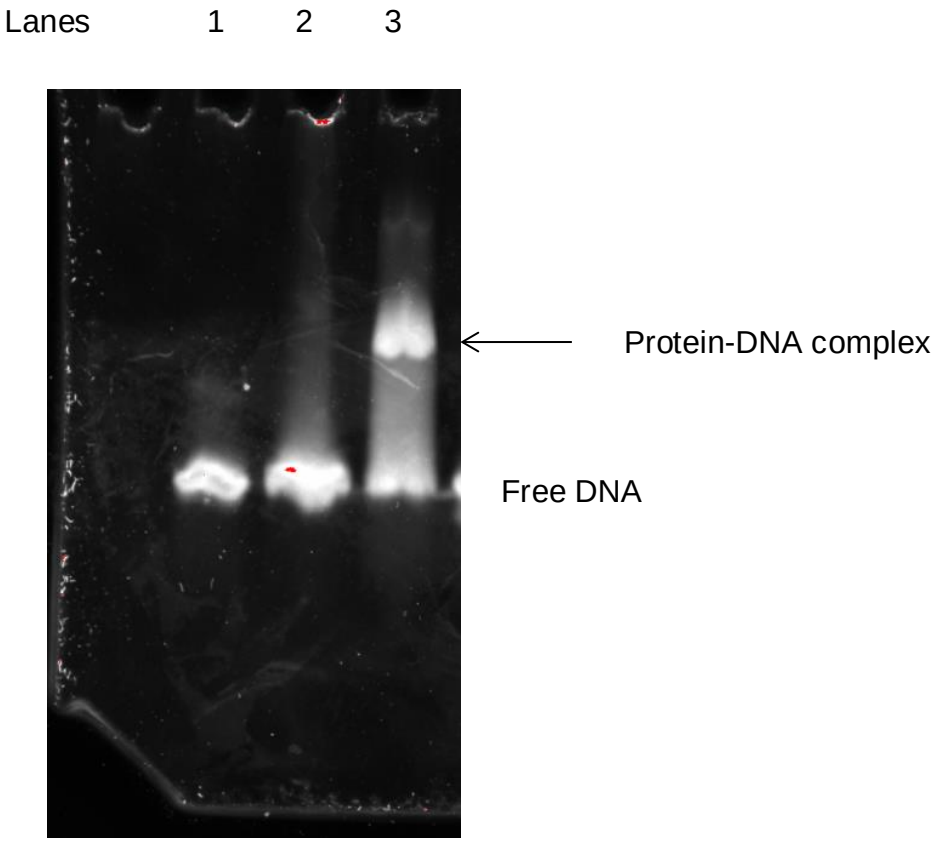


Figure 18. Electrophoretic mobility shift assay (EMSA) of the FOXP3 FHD bound to Nelson DNA. 0.5 µM DNA was incubated with 1 µM protein in binding buffer (89 mM Tris-borate pH 8.3, 20 % glycerol, 0.01 % bromophenol blue) for 30 minutes on ice prior to electrophoresis. Free DNA was used as a controlLane 1 contains 0.5 µM free DNA, lane 2 contains 0.5 µM DNA and 0.5 µM protein and lane 3 contains 0.5 µM DNA and 1 µM protein. Protein-DNA complexes are indicated (black arrow) and binding occurs at a protein:DNA ratio of 2:1.

3.4 Conformational stability of the FOXP3 DSD

Denaturation curves are used to measure the conformational stability of the native state of proteins relative to the unfolded species. Urea-induced equilibrium unfolding is usually used to determine the conformational stability of proteins (Pace, 1975). However in this study, this technique was used to determine the equilibrium unfolding pathway of the FOXP3 FHD, if any stable intermediates are present in the unfolding pathway and to determine the cooperativity of unfolding.

3.4.1 Recovery and reversibility

Before equilibrium unfolding could be performed, recovery and reversibility studies were carried out in order to ensure that the unfolding process was indeed reversible and to ensure that the unfolding reaction had reached equilibrium. This is necessary because equilibrium is needed in order to calculate the change in free energy (ΔG) and *m*-values from the equilibrium unfolding data (Myers *et al.*, 1995). When the FOXP3 FHD was unfolded in 8 M urea it was shown to exhibit 126 % recovery of tertiary structure (Figure 19) upon dilution of the urea to 0.8 M. The area under the peak for the refolded control was 126 % greater than the area of the peak under the refolded protein. The fluorescence spectrum of the refolded protein peaks at a lower intensity and has a smaller area under the peak than for the refolded control. Upon refolding, the intensity decreases and since the intensity of the refolded protein is lower than the refolded control, it suggests that the refolded protein is more folded than the refolded control.

3.4.2 Urea-induced equilibrium unfolding

Far-UV CD and intrinsic fluorescence were used to monitor the unfolding of the FOXP3 FHD. Protein concentrations of 0.5 µM to 20 µM were unfolded in 0 M to 8 M urea at 20 ^{o}C and allowed to achieve equilibrium over 2 hours. The CD unfolding curves (Figure 21A) showed distinct pre- and post-transition regions with a single transition, indicative of a highly cooperative system, from 4 M to 7 M urea. Furthermore, the CD unfolding curves

were all sigmoidal. The curves are indicative of a two-state ($N_2 \leftrightarrow 2U$) model where N_2 refers to the native dimer and U refers to the unfolded monomers, respectively.

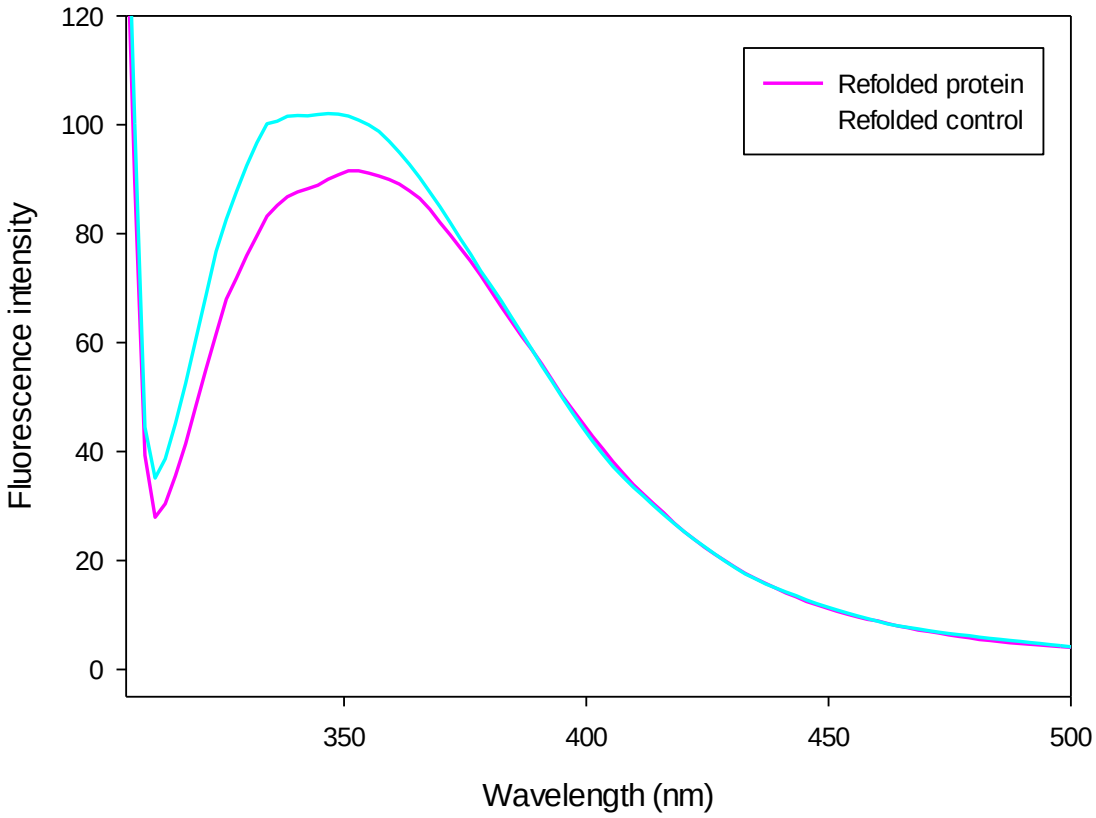


Figure 19. Recovery of the FOXP3 FHD. The protein was unfolded in 8 M urea for 2 hours prior to refolding by 10 times dilution to a final protein concentration of 0.4 µM and a urea concentration of 0.8 M. The refolded protein (pink) is shown to have a recovery of 126 % relative to the refolded control (cyan) which consists of 0.4 µM FOXP3 FHD incubated with 0.8 M urea for 1 hour.

In this study, tryptophan fluorescence was used to monitor tertiary structural changes upon unfolding and is a far more sensitive technique than circular dichroism as local, rather than global structural changes can be detected. The tryptophan unfolding curves (Figure 21B) showed no distinct pre-transition region however, a clear post-transition region was observed. The tryptophan unfolding curves, in direct contrast to the CD unfolding curves were not sigmoidal and showed a noncooperative mechanism with two

clear transition regions; the first transition is from 0.8 M to 4 M urea and the second transition is from 5.8 M to 7 M urea.

Furthermore, noncoincidence of the unfolding curves monitored by secondary and tertiary structural changes (Figure 21C) indicates that the unfolding mechanism may be more complex than a two-state unfolding mechanism. Together, these results suggest formation of a thermodynamically stable intermediate that forms between 4 M and 7 M urea.

The noncoincidence of the CD and fluorescence unfolding curves suggests that loss of secondary structure and the altered tertiary environments of the tryptophan residues do not occur simultaneously. The individual data sets separate from 1 M to 6 M urea which indicates formation of an intermediate species during unfolding. A plot of the maximum emission wavelength was plotted (Figure 22) which is independent of the fluorescence intensity and provides information on wavelength shifts associated with the formation of an intermediate species. This additional plot shows the relationship between the maximum emission peak wavelength of the protein with increasing urea concentration. This plot can assist in validating the models previously determined or may provide further information such as show the presence of another hidden intermediate. During urea-induced unfolding, the maximum emission wavelength of the FOXP3 FHD undergoes red wavelength shifts. This correlates to a protein undergoing a $N \rightarrow U$ transition, however Figure 22 shows that the maximum emission wavelength levels off between 3 M and 5 M urea. This is consistent with the three-state model observed for the fluorescence unfolding curve (Figure 21B) and suggests that the protein undergoes a transition from the native, folded (N_2) state to an intermediate state (red-shifted wavelength), and finally to a unfolded, unfolded species (greater red-shifted wavelength).

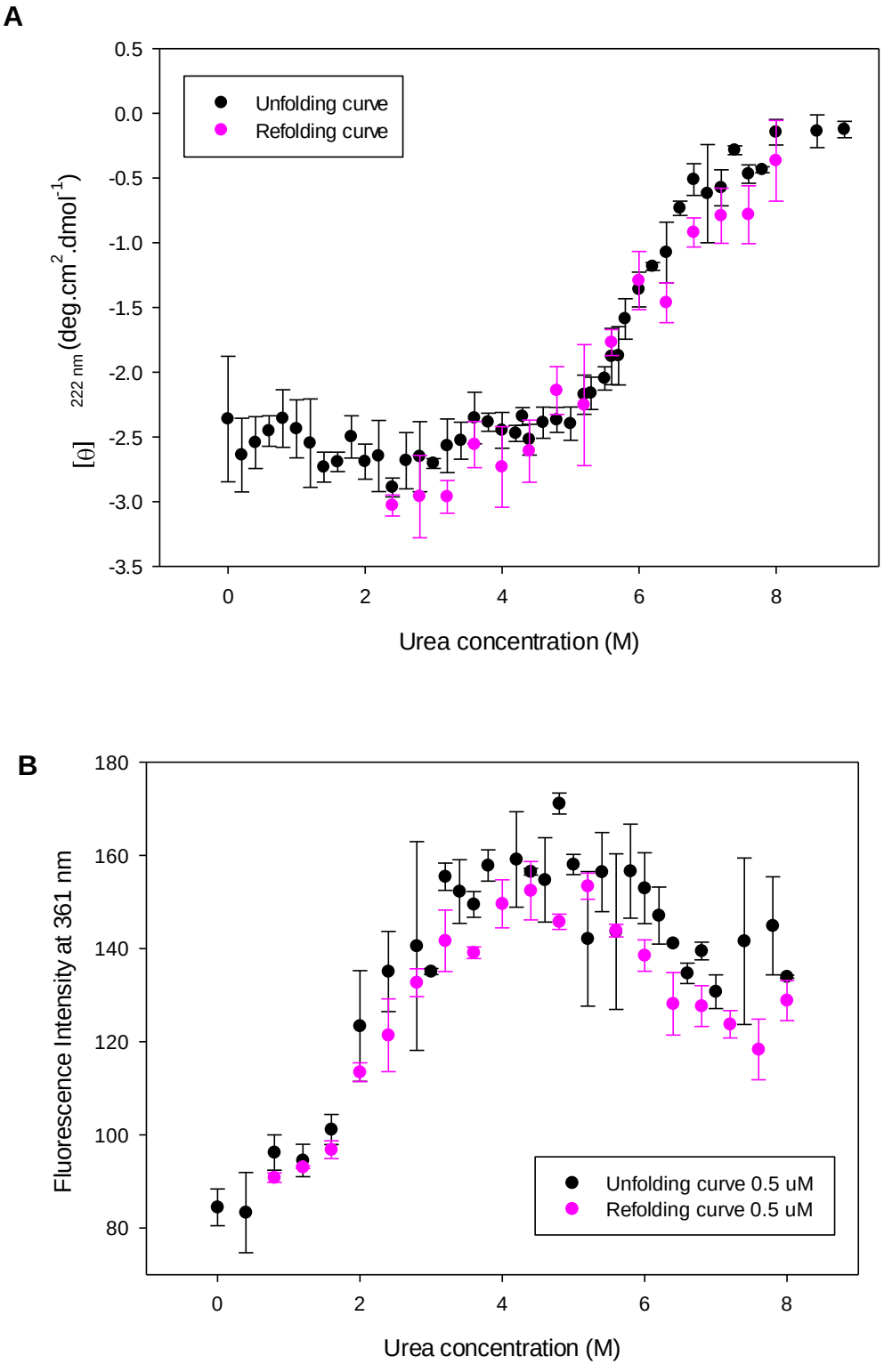


Figure 20. Unfolding and refolding curves of the FOXP3 FHD. Unfolding and refolding was monitored by **A.** circular dichroism (ellipticity at 222 nm) and **B.** fluorescence emission at 361 nm following excitation at 295 nm. A final concentration of 4 µM was used for unfolding (black dots) and refolding (pink dots) monitored by circular dichroism whereas a final protein concentration of 0.5 µM was used for unfolding (black dots) and refolding (pink dots) monitored by fluorescence spectroscopy. A ten-fold dilution was used to refold the protein to a final concentration of 4 µM. The unfolding and refolding curves are fully superimposable for CD and fluorescence.

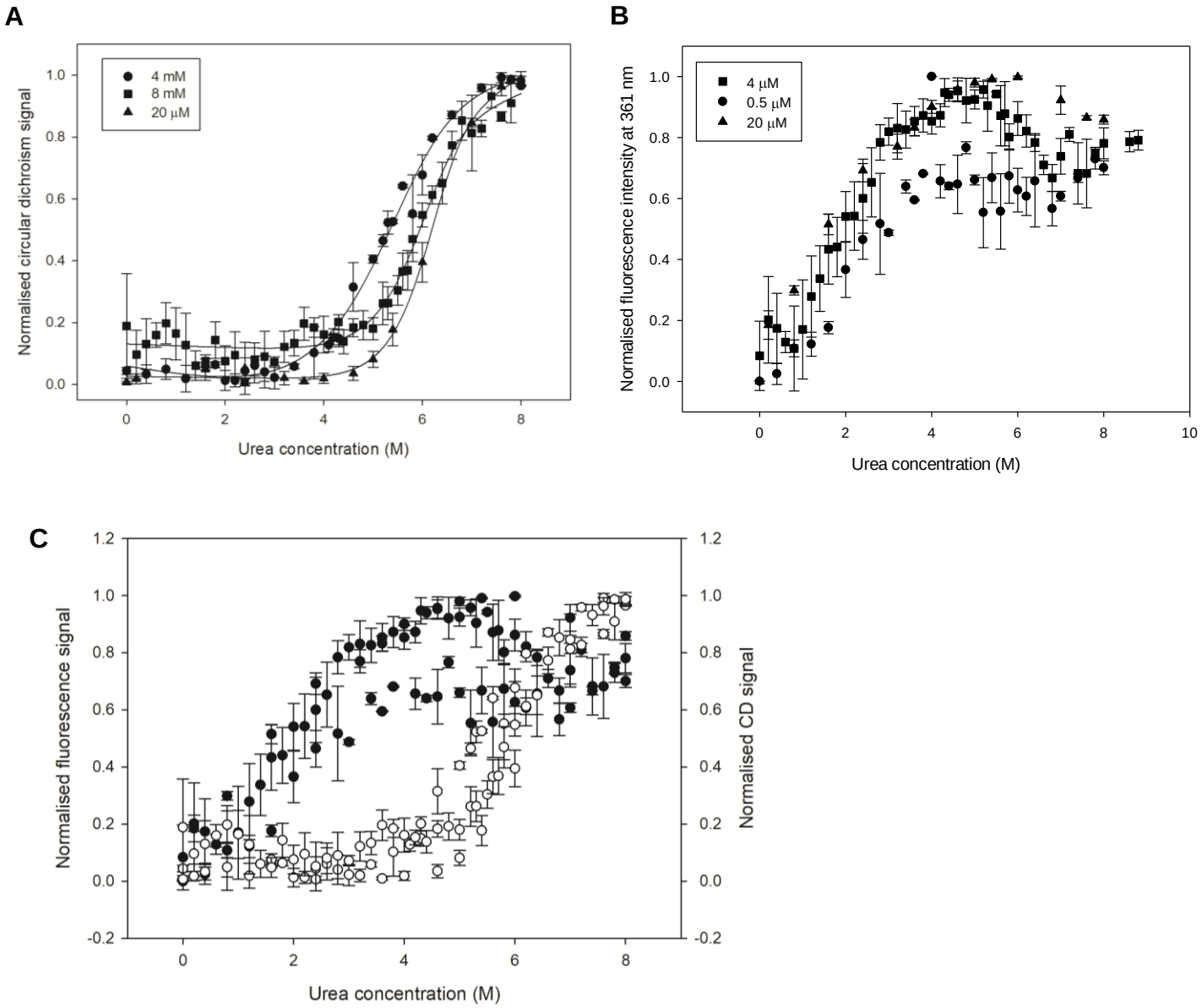


Figure 21. Urea-induced equilibrium unfolding of FOXP3 FHD monitored by **A.** far-UV CD and **B.** fluorescence spectroscopy with increasing protein concentration (●<▲). Experiments were conducted at 20 ^{o}C in 20 mM Tris-HCl pH 7.6, 150 mM NaCl and 1 mM DTT. The CD unfolding curve shows a two-state model while the fluorescence unfolding suggests that it is three-state and proceeds via formation of an intermediate. **C.** Noncoincidence of the CD (open circles) and tryptophan fluorescence (closed circles) unfolding curves indicate a more complex unfolding mechanism than a simple two-state ($N_2 \leftrightarrow 2U$) model. The data were normalised to values between 0 and 1 using the following equation: (y value – minimum)/ (maximum – minimum).

The alternative scenario is that since there is no pre-transition region in the fluorescence unfolding curve, there could be a non-linear pre-transition region whereby there is a steep dependence of fluorescence signal on urea concentration. This would

mean that the relatively flat region could correspond to a transition region and the post transition could correspond to the region from 7 to 8 M urea which appears to be non-linear as well.

Due to the law of mass action, the unfolding of oligomers, including dimers, is concentration dependent and a shift in the unfolding transition is observed upon increased protein concentration (Perrett *et al.*, 1999; Walters *et al.*, 2009). There is a concentration dependence observed for the CD unfolding curves which validates the two-state dimer model. Different spectroscopic techniques are used to monitor unfolding as formation of an intermediate that is not revealed by one technique may be revealed by another (Walters *et al.*, 2009). This is shown to be the case here as there is also a concentration dependence on the tryptophan unfolding curve which shows an apparent shift to higher transition midpoints from 4 M to 7 M urea.

In addition, the tryptophan unfolding curves show a concentration dependence in the second transition. Physiologically relevant FOXP3 concentrations are in the nanomolar to low micromolar The noncoincidence of the circular dichroism and tryptophan fluorescence unfolding curves made data fitting more complex. The CD and fluorescence unfolding curves could not be globally fitted to a three-state model and so the CD unfolding curves were fitted to a two-state model. This was done to provide thermodynamic parameters of the the native dimer relative to the unfolded species, even though the fluorescence data shows formation of the intermediate. Thermodynamic parameters could not be obtained for the $N_2 \leftrightarrow I_2$ and $I_2 \leftrightarrow 2U$ transitions.

The curves were fit to a two-state ($N_2 \leftrightarrow 2U$) model where N_2 refers to the native dimer and U refers to the unfolded monomers, respectively. The total free-energy change between the native dimer and unfolded monomer is 16 ± 2.5 kcal.mol-1. The *m*-value was calculated to be 1.55 ±0.61 kcal.mol-1.M-1. The fits to a two-state model were good with R^2 values greater than 0.98. Midpoint (C_m) values of 5.29 M, 6.16 M and 6.21 M urea were obtained for concentrations of 4 µM, 8 µM and 20 µM, respectively.

3.5 Kinetic studies of the FOXP3 DSD

During equilibrium unfolding of a small, single-domain protein, any intermediates that form are usually thermodynamically unstable and transient species (Pace, 1975). However, in order to determine the mechanism and pathway of unfolding and refolding, it is necessary to identify and characterise these intermediates. In this study, the kinetic parameters will be determined for both the folding and unfolding pathways of the FOXP3 FHD to determine the mechanism of domain-swapping. It is necessary to determine the unfolding and refolding pathways as the protein may not refold *via* the exact same route as it took to unfold. Furthermore, there may be additional transient species that form upon refolding that are not present in the unfolding pathway.

Intrinsic tryptophan fluorescence was used as a probe for tertiary structural changes occurring during (un)folding. It was chosen because it is a highly-sensitive technique which can provide information on local changes to the tryptophan environments.

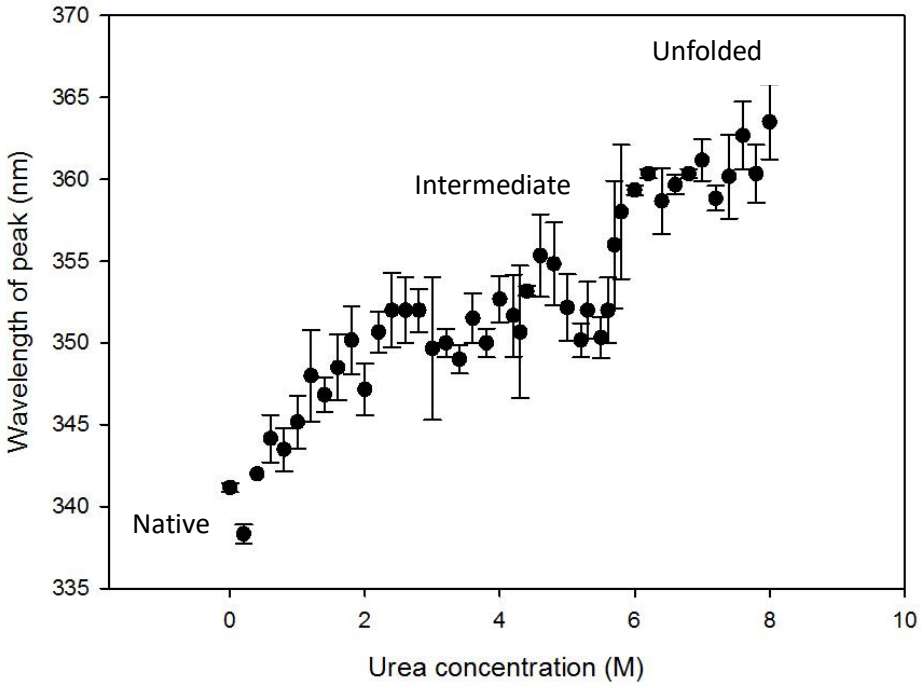


Figure 22. **Maximum emission wavelength of 4 µM FOXP3 FHD upon unfolding.** The study is independent of the fluorescence intensity but is used to identify wavelength shifts with increasing urea concentration. This model

validates the three-state model observed in the fluorescence unfolding curve and suggests formation of an intermediate species along the unfolding pathway of the FOXP3 FHD.

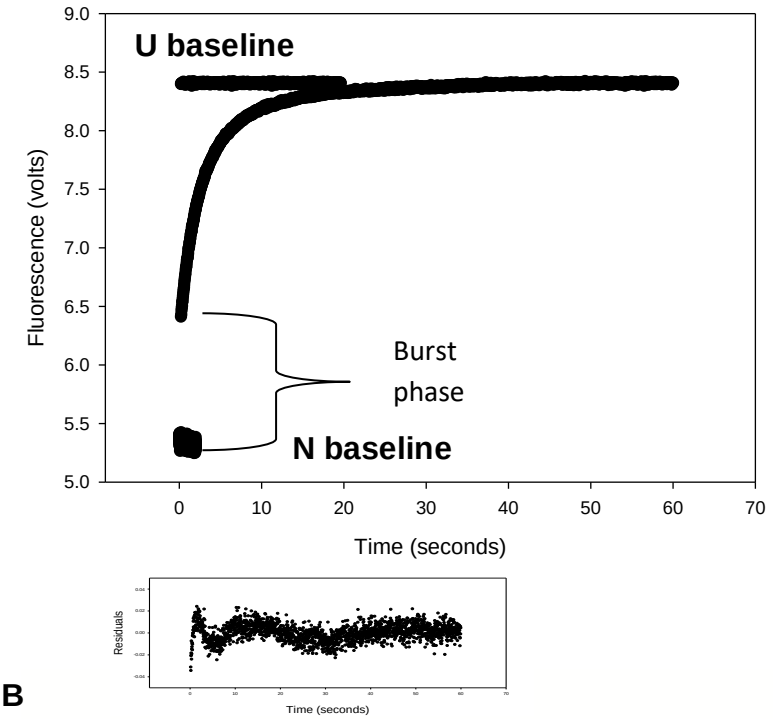
3.5.1 Unfolding kinetics

From the equilibrium unfolding data it can be seen that the FOXP3 FHD is in its native conformation only at very low concentrations of urea. It is predominantly in an intermediate conformation at urea concentrations from approximately 4 M – 6 M urea and at urea concentrations between 6 M and 8 M, the predominant species is that of unfolded FOXP3 FHD (Figure 21).

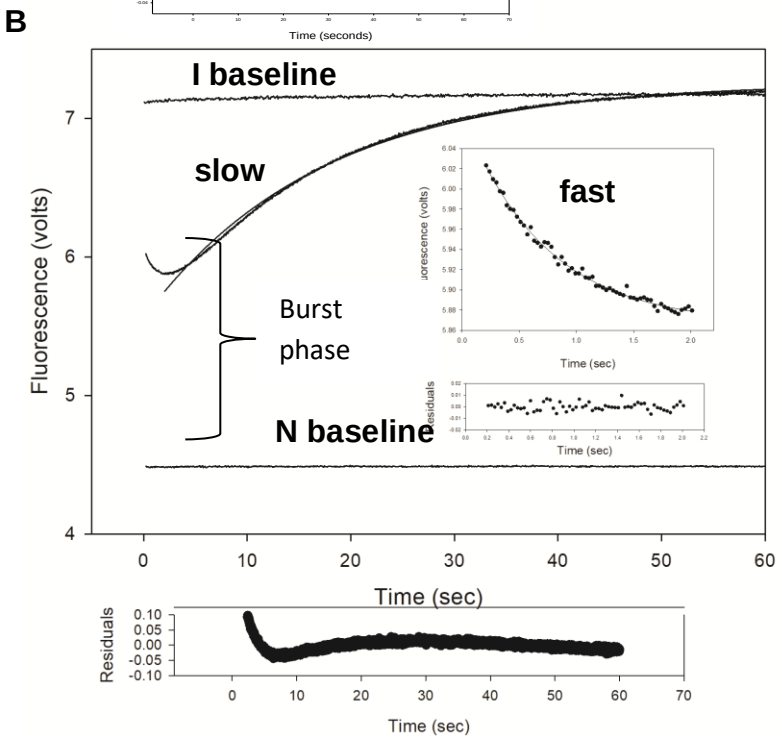
To monitor the unfolding kinetics of the FOXP3 FHD, native protein was unfolded at concentrations of urea representative of the unfolded transition (6.5 – 8 M urea). The kinetics of unfolding of the native state to the unfolded species are biphasic. Unfolding from the native state to urea concentrations representative of the intermediate (5 M – 6 M urea) are also biphasic. There is a minor fast phase and a major slow phase in this unfolding event. Interestingly, kinetics of unfolding from the intermediate to the unfolded species is monophasic (Figure 21B). The minor, fast phase refers to a small population of species that unfold via that route and occurs within the first few seconds, whereas the major slow phase refers to the significant increase in signal whereby the majority of species unfold. Minor and major refer specifically to the relative amount of species that unfold *via* that particular pathway.

For all unfolding kinetics traces, there is a notable burst phase between the native baseline and the start of kinetic trace (in the dead time of the instrument) which suggests that an event is occurring within 2 ms but is unable to be detected by the stopped-flow instrument. The kinetic trace therefore does not extend from the native protein baseline and thus the phases observed do not account for the total amplitude change.

A


U baseline
Burst
phase
N baseline
Fluorescence (volts)
Time (seconds)
Residuals
Time (seconds)

B


I baseline
slow
fast
Burst
phase
N baseline
Fluorescence (volts)
Time (sec)
Residuals
Time (sec)
Time (sec)

C

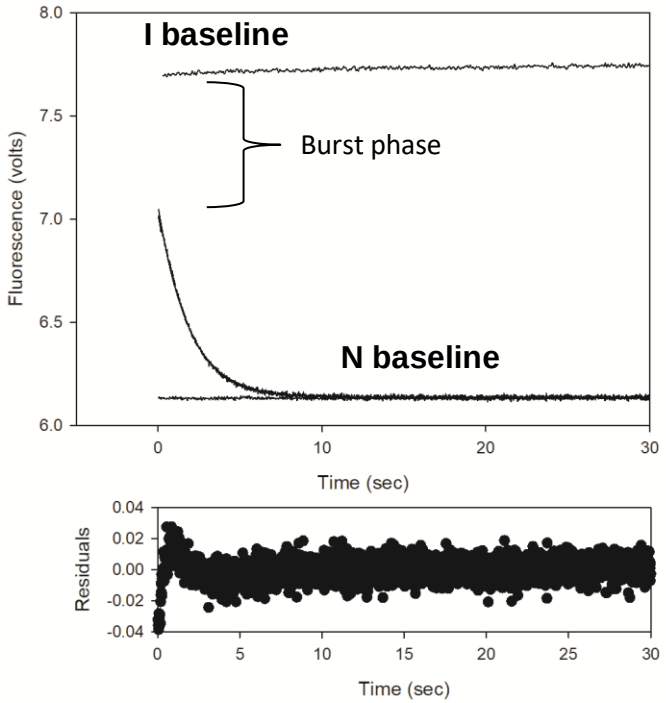


Figure 23. Typical traces observed for unfolding of the FOXP3 DSD monitored by stopped-flow fluorescence. An excitation wavelength of 295 nm was used, and fluorescence was collected above 320 nm using a cut-off filter. Native and unfolded (8 M urea) fluorescence signals are indicated by N and U, respectively. **A.** Unfolding of the native protein in 8 M urea. The trace fits best to a double exponential model which gives amplitudes of -1.54 and -0.56 and rates of 0.45 sec^{-1} and 0.10 sec^{-1} **B.** Unfolding of the protein from the native state to the intermediate state (5 M urea) occurs in two distinct phases. The fast phase (inset) fits to a single exponential model which provides an amplitude of 0.20 and a rate of 1.39 sec^{-1}.The slow phase also fits to a single exponential model and gives an amplitude of -1.73 and a rate of 0.06 sec^{-1}. **C.** Unfolding of the intermediate species (preequilibrated in 5 M urea) to the unfolded species (8 M urea) fits to a single exponential model with an amplitude of 0.92 and a rate of 0.53 sec^{-1}. . A burst phase is detected for each of the traces. Residuals of the fits are shown below each of the plots.

$N_2 \leftrightarrow$ 2U transition

The kinetics of unfolding from native to unfolded protein was fitted to a double exponential and has a time constant (inverse of rate constant) of 2.2 sec for the fast phase and 10 seconds for the slow phase. The shape of the kinetic trace for the single jump from native to unfolded protein (Figure 23A) shows an increase in the fluorescence signal until the unfolded baseline is reached. Upon denaturation in 8 M urea, two phases are exhibited. The rate constant of phase 1 increases with increasing urea concentration while that of phase 2 decreases with increasing urea concentration.

$N_2 \leftrightarrow I_2$ transition

The jump from native to the intermediate (in 5 M urea) exhibits a fast phase with a time constant of 0.7 sec and the slow phase has a time constant of 16.6 sec. The kinetic trace for the native protein to the intermediate (5 M urea) (Figure 23B) shows a decrease in the fluorescence signal within 4 seconds during the fast phase, and then a slow increase in fluorescence signal towards the 5 M baseline in the slow phase. The fast phase of the $N_2 \leftrightarrow I_2$ transition was fitted to a single exponential while the slow phase was also fitted to a single exponential. Three phases were observed in formation of the intermediate from the native state. One major intermediate phase and a minor fast and minor slow phase were exhibited. The rate constants and m_u values obtained for each phase is shown in Table 1. The rate constants of the unfolding phases increase with increasing urea concentration. In contrast, the rate constant of the slowest phase (phase 3) does not increase with urea concentration but remains fairly constant with increasing urea concentration (Figure 24A).

$I_2 \leftrightarrow$ 2D transition

For the unfolding kinetics from the intermediate to unfolded protein, the time constant is similar to the fast phase of native to unfolded protein and was shown to be 1.8 sec. The jump from the intermediate to unfolded protein is monophasic which suggests that a single event is taking place as the fluorescence signal decreases from the 5 M baseline toward the 8 M baseline (Figure 23C).

The rate constant of the major phase of unfolding decreased with increasing urea concentration in a linear manner (Figure 24). The data fits well to a linear function of the type:

$ln k_u = ln k_u^{H2O} + m_u[D]$ Equation 14;

where k_u is the rate constant for unfolding at a denaturant concentration, D, and $ln k_u^{H2O}$ is the rate constant of unfolding in water, m_u is the m-value that reflects the change in solvent exposure between the native state and the rate-determining transition state for unfolding. The values of k_u^{H2O} andm_u, determined using Equation 14 are shown in Table

1 and give an indication of the rate of unfolding in the absence of denaturant and the degree of solvent exposure upon unfolding, respectively.

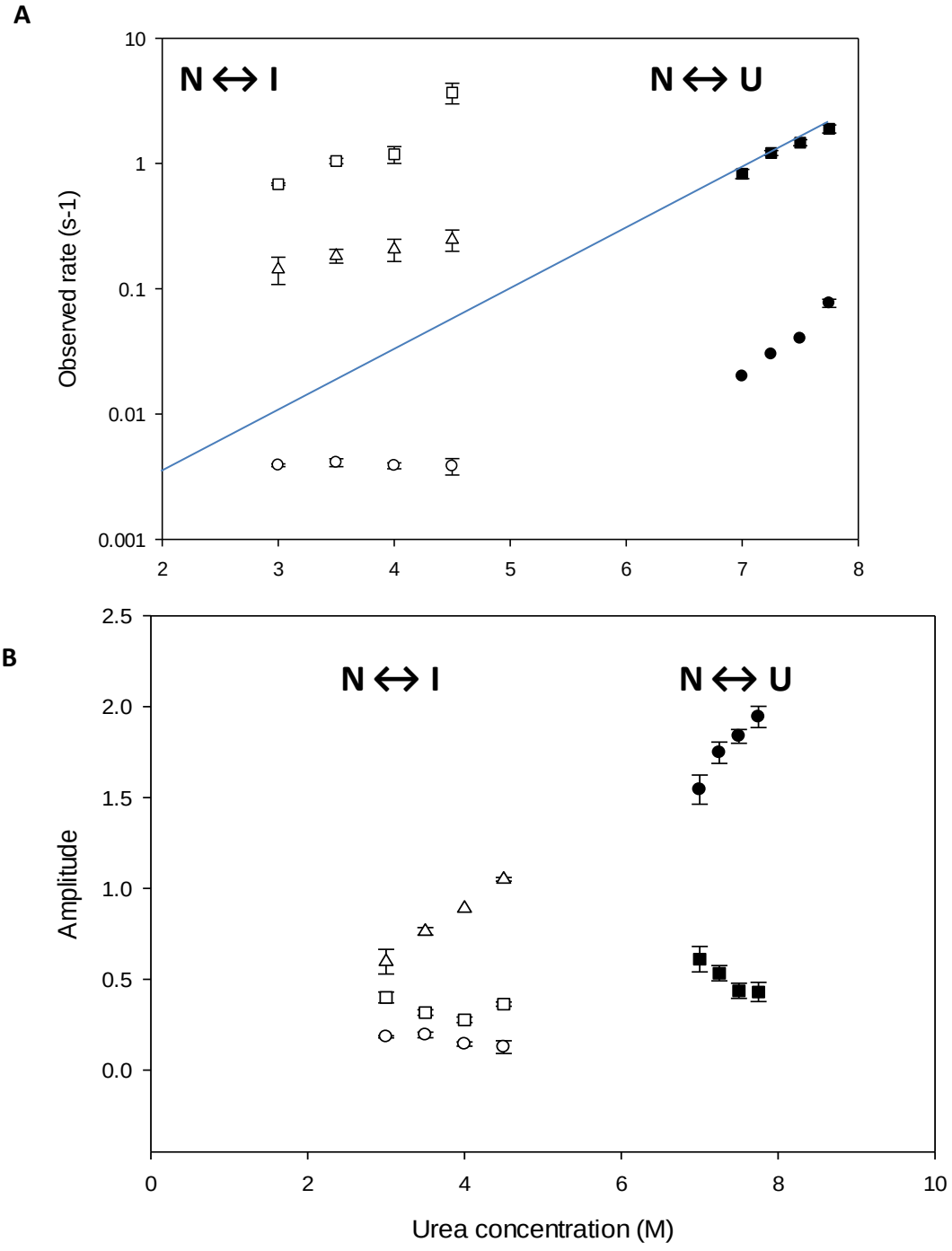


Figure 24. Unfolding rate and amptlitude plots of the FOXP3 DSD. A. Urea dependence on the unfolding rate constant for 3 µM FOXP3 FHD. The data was fit to a linear regression fit using Equation 14. An example of a linear regression fit is depicted here (blue line). R^2 = 0.99. **B.** Absolute amplitudes for unfolding. Three phases are observed in the N to I transition while two phases are observed in the N to U transition. Open symbols represent the $N_2 \leftrightarrow I_2$ transition while closed symbols represent the $N_2 \leftrightarrow 2U$ transition. Phase 1 is represented by squares, phase 2 by triangles and phase 3 by circles.

Table 1. Kinetic parameters of unfolding obtained for the FOXP3 DSD

	Phase	$k_u^{(H2O)}$ (s^{-1})	m_u	**Time constant**
N ↔ I	1	2.8×10^{-3}	0.44	35.5 s
N ↔ I	2	4.3×10^{-3}	-0.11	4 min
N ↔ I	3	1.4×10^{-3}	0.37	12 min
N ↔ D	1	1×10^{-3}	0.3	16 min
N ↔ D	2	5×10^{-4}	0.46	33 min

Upon unfolding of the native protein to the intermediate, the majority of the species shows a significant exposure of surface area (m_u value of 0.37) relative to the native state. The minor, fast phase also shows a significant exposure of surface area (m_u value of 0.44) to the solvent while the minor, slow phase shows a decrease in exposure to solvent (m_u value of -0.011) or rather shows a burial of surface area relative to the native state as indicated by the negative slope.

Upon unfolding of the protein (N ↔U transition), the major population of species unfolds via the slow phase (closed circles Figure 24B) while the minor, small population unfolds via a fast phase (closed squares Figure 24B). The rate constant for both phases show an increase with increasing urea concentration. The major, slow phase has a positive slope (m_u value of 0.46) and shows a significant increase in exposure to solvent upon complete unfolding while the minor, fast phase exhibits solvent exposure to a lesser extent (m_u value of 0.3) (Table 1).

3.5.2 Refolding kinetics

The equilibrium unfolding data indicate that there is very little to no significant pre-transition baseline. Thus, refolding kinetic experiments were conducted from 8 M to 4 M (to monitor the U ↔ I transition) and from 4 M to 0.4 M (to monitor the I ↔ N transition). At a final concentration of 0.4 M, the equilibrium unfolding curve indicates a significant change in fluorescence signal compared to the native protein which suggests that the

tertiary structure has been perturbed. However, it was the lowest urea concentration that could be obtained upon refolding to the 'native' state.

Refolding was monitored by tryptophan fluorescence and three kinetic refolding traces were averaged per urea concentration. The average trace was fitted to a double exponential and the fit was good according to the random distribution of points around zero on the residuals plot (Figure 25).

The kinetic trace for the I to N transition (Figure 25a) showed an initial decrease in the fluorescence signal from the intermediate baseline to below that of the native protein baseline, in the burst phase, and subsequent increase toward the native protein baseline. The data was fitted to a double exponential with amplitudes of -0.19 and -0.04 and rate constants of 0.39 and 0.02 sec^{-1}.

Similarly, the kinetic trace for the U to I transition (Figure 25b) showed a decrease in the fluorescence signal from 8 M to below that of the 5 M baseline, in the burst phase, and then an increase of the fluorescence signal towards the 5 M baseline until it reached equilibrium. The data was fitted to a double exponential with amplitudes of -1.42 and -0.33 and rate constants of 8.37 and 1.69 sec^{-1}.

The urea dependence of the amplitude for the fast and slow phases of refolding are shown in Figure 26. The relative percentage of the two species decreased from 4 to 6 M urea.

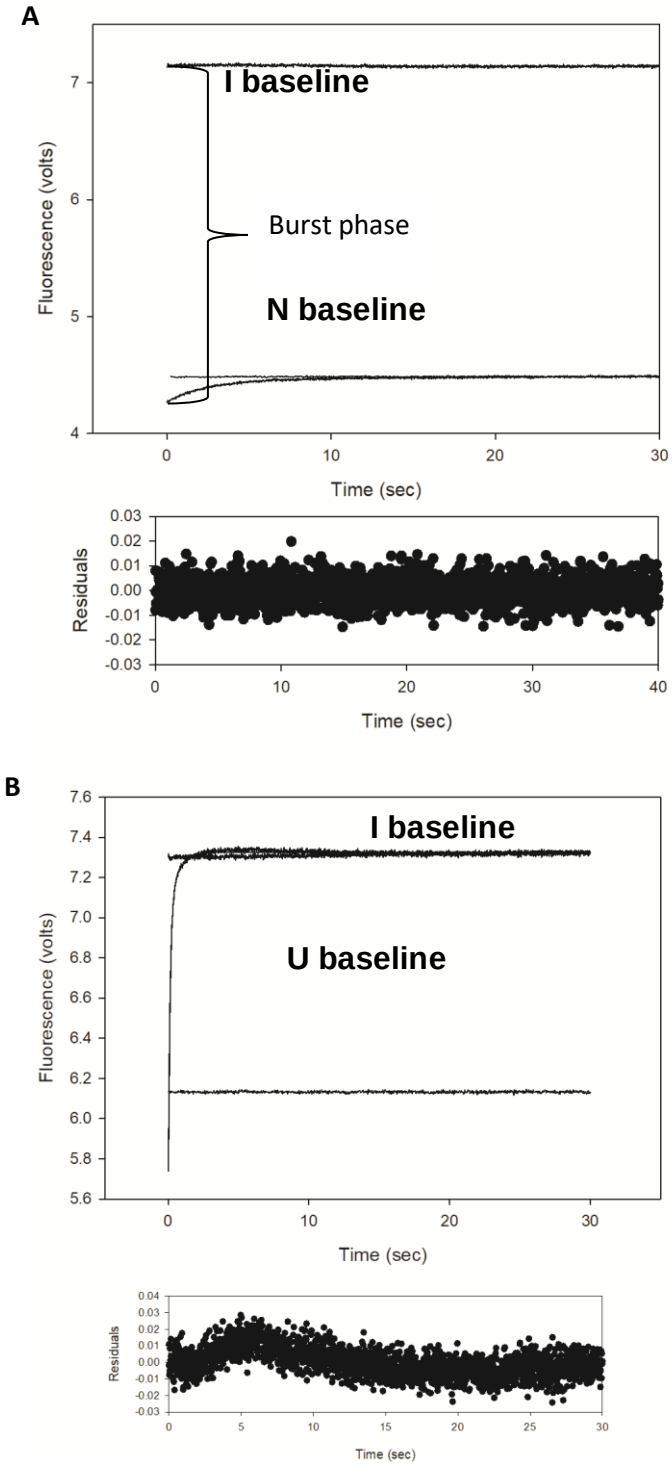


Figure 25. Kinetic traces observed for refolding of the FOXP3 DSD monitored by stopped-flow fluorescence. An excitation wavelength of 295 nm was used, and fluorescence was collected above 320 nm using a cut-off filter. **A.** Refolding of the FOXP3 FHD preequilibrated in 5 M urea. The final urea concentration was 0.83 M. The data was fitted to a double exponential process providing amplitudes of -0.19 and -0.04 and rates of 0.39 sec^{-1} and 0.02 sec^{-1}, respectively. Residuals are shown below the plot. **B.** Refolding of the FOXP3 FHD preequilibrated in 8 M urea. The final urea concentration is 5 M. Residuals for the fit to a double exponential with amplitudes of -1.42 and -0.33 and rates of 8.37 and 1.69 sec^{-1}. Residuals are shown below the plot.

The rate constant of the fast, major phase of refolding decreased with increasing urea concentration in a linear manner (Figure 26A). The data fits well to a linear function of the type:

$$\ln k_f = \ln k_f^{H2O} + m_f[\mathrm{D}] \qquad \text{Equation 15}$$

where k_f is the rate constant for refolding at a denaturant concentration, D, and k_f^{H2O} is the rate constant of refolding in water, m_f is the *m*-value that reflects the change in solvent exposure or compactness of the unfolded species relative to the rate-determining transition state for refolding.

The refolding kinetics from the U to the I transition shows three phases, a major, slow refolding phase and a minor, fast refolding phase and a minor, intermediate unfolding phase and is indicative of three events occurring during refolding. This is in agreement with three phases observed in refolding from the intermediate state to the native state (Figure 26). The k_f^{H2O} and m_f values for these phases are indicated in Table 2. The major, slow refolding phase, of the D to I transition, shows that there is a large burial of exposed surface area of the unfolded protein relative to the intermediate state (m_f value of -0.57). The minor, intermediate phase also shows a burial of exposed surface area according to the negative slope but could not be fitted accurately to Equation 15. The minor, fast phase showed a positive slope (m_f value of 0.45) which indicates exposure of surface area upon refolding from the unfolded species to the intermediate which could indicate aggregation or association of open-ended domain-swapped structures (Janowski *et al.*, 2001).

Table 2. Kinetic parameters of refolding of the FOXP3 DSD

	Phase	**k_f^{H2O} (s^{-1})**	***m_u***	**Time constant**
D → I	1	4×10^{-3}	0.45	24 s
D → I	2	4×10^{-2}	-0.57	4 min
I → N	1	6×10^{-1}	-0.27	1.6 s
I → N	2	7×10^{-3}	-0.16	2 min
I → N	3	1×10^{-4}	-0.06	2 hrs 20 min

In the I to N transition, the major, slow phase showed a less steep negative slope upon refolding from the intermediate to the native state (m_f value of -0.06) which indicates that the major population of species does not bury a significant amount of surface area upon formation of the native state. The minority of species follow a fast phase route and show a more significant burial of exposed surface area of the intermediate relative to the native state (m_f value of -0.27) and the minor, intermediate phase shows burial of exposed surface area to a lesser extent (m_f value of -0.16).

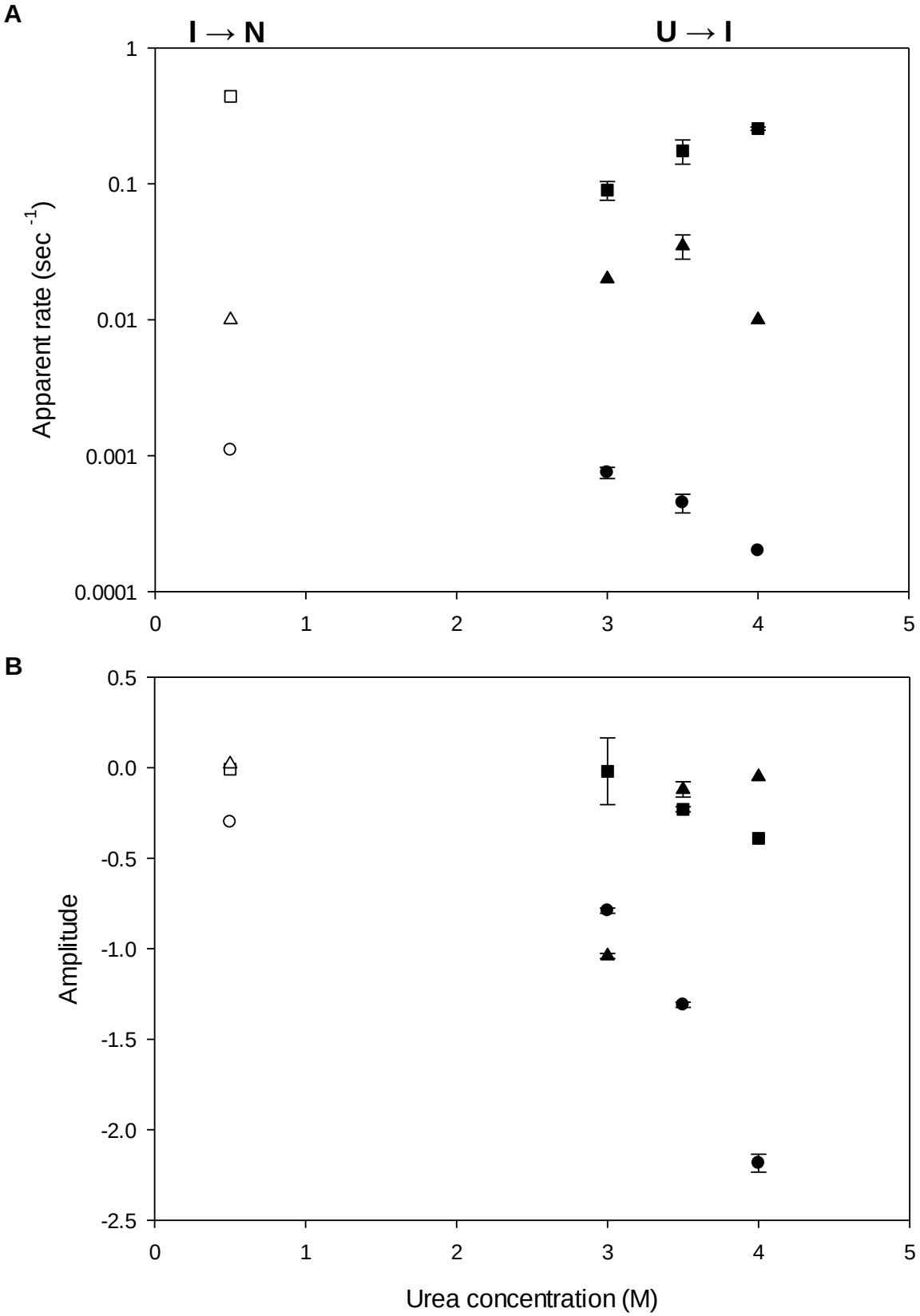


Figure 26. Effect of varying urea concentrations on refolding of the FOXP3 DSD. **A.** Urea dependence of the rate constants of refolding and the **B.** refolding amplitude plot. Refolding experiments were performed at 20 ^{o}C in 20 mM Tris-HCl pH 7.6, 150 mM NaCl and 1 mM DTT. The final protein concentration was 3 µM. Refolding of the FOXP3 FHD shows three phases. The U → I transition is shown in closed symbols while the N → I transition is shown by the open symbols. Phase 1 is indicated by squares, phase 2 is represented by triangles and phase 3 is indicated by circles.

3.6 Properties of the equilibrium unfolding intermediate of the FOXP3 DSD

The FOXP3 FHD shows formation of a stable intermediate species under equilibrium conditions. From the tryptophan unfolding curves (Figure 21B), there appears to be a concentration dependence of the second transition which suggests that the intermediate is likely be dimeric. It is important to characterise the intermediate and to establish whether it resembles the native or unfolded species, structurally to gain further insight into how the FOXP3 FHD assembles into its functional, three-dimensional domain-swapped dimer form which is the crux of this study.

3.6.1 Secondary structure

The far-UV circular dichroism spectra of the FOXP3 FHD in the native (0 M urea), unfolded (8 M urea) and intermediate (5 M urea) states were compared so as to see whether the secondary structure changes in the three states. The native spectrum shows two distinct troughs at 222 nm and 208 nm, respectively which indicate the presence of α-helices. As would be expected, this is completely abolished in the unfolded protein spectrum (Figure 14). The spectrum obtained of the intermediate protein is almost identical to that of the native protein and also exhibits a trough at 222 nm. This suggests that the secondary structure of the intermediate is almost identical to that of the native protein (Figure 27A).

3.6.2 Tertiary structure

As with CD, the fluorescence spectra of the FOXP3 FHD in the native (0 M urea), intermediate (5 M urea) and unfolded (8 M urea) states were compared. The unfolded protein spectrum shows enhanced fluorescence intensity compared to the native protein as well as a red-shifted emission maximum from 343 nm in the native state to 361 nm in the unfolded species (Figure 27B). The spectrum obtained for the intermediate state shows higher fluorescence intensity than either the native state or the fully unfolded protein and there is also a red-shifted emission maximum from 343 nm in the native state to 351 nm in the intermediate (Figure 27). Furthermore, the red shifted emission maximum indicates that the intermediate species is perhaps not as compact as the

native protein and as a result of the partial unfolding; there is enhancement of the fluorescence of the average of all the tryptophan residues.

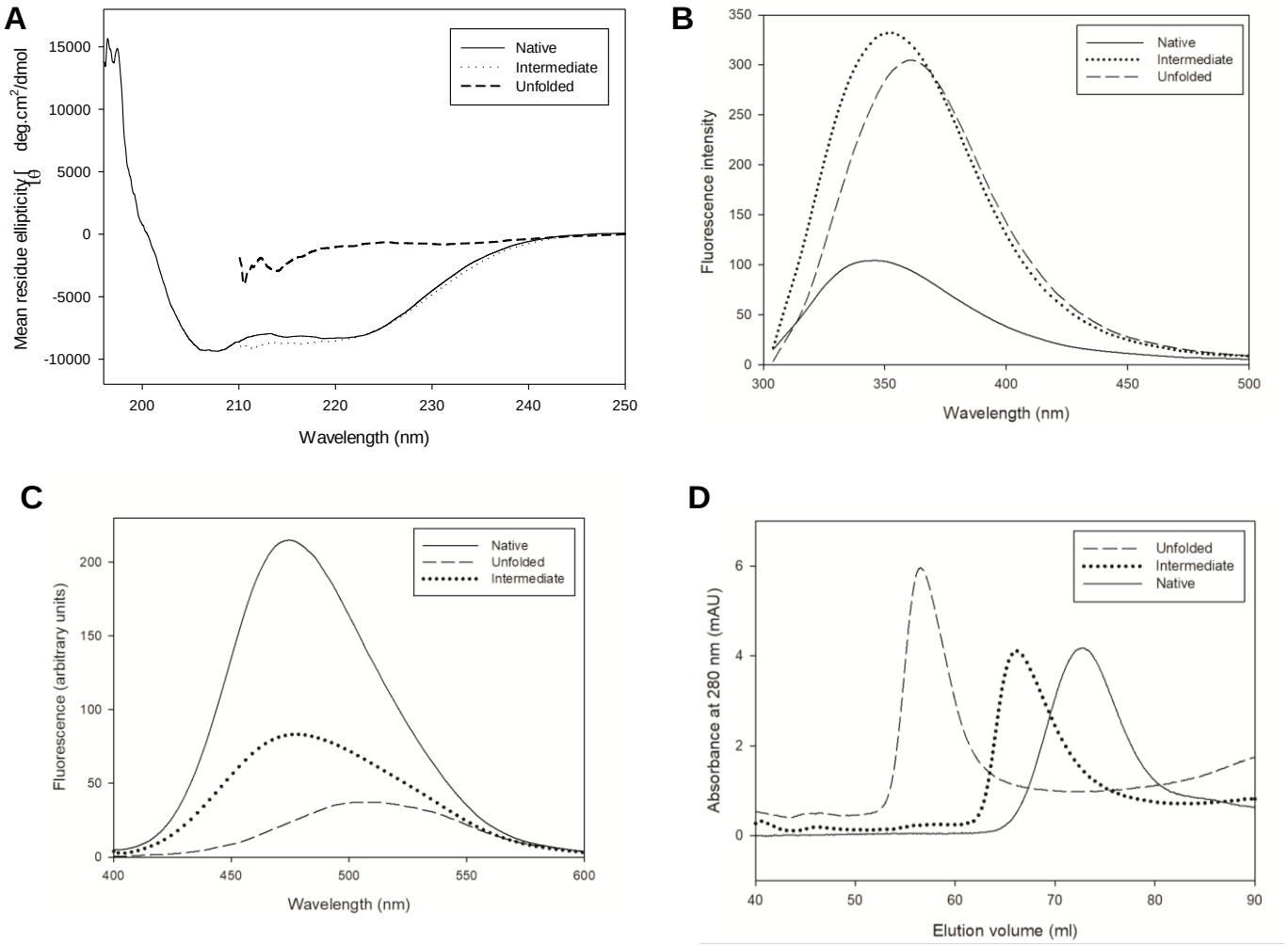


Figure 27. Structural characterisation of the intermediate observed in equilibrium unfolding. **A.** circular dichroism, **B.** intrinsic fluorescence spectroscopy, **C.** extrinsic fluorescence spectroscopy using 200 µM ANS and **D.** size-exclusion chromatography performed using a Hiload 16/600 Superdex 75 pg column (GE Healthcare). The intermediate species was characterised according to secondary, tertiary and quaternary structure and compared to the native and unfolded protein. All spectroscopy work was performed in 20 mM Tris-HCl pH7.6, 150 mM NaCl, 1 mM DTT at 20 ^{o}C.

3.6.3 Quaternary structure

The hydrodynamic volume and compactness of the intermediate species was studied using size-exclusion chromatography. Larger proteins with a higher hydrodynamic volume are eluted before smaller proteins. Unfolded FOXP3 FHD eluted at 56 ml which is close to the void volume of the column (45 ml). The unfolded protein was expected to elute either within the void volume of the column or closes to it as it is in a random coil conformation. The unfolded protein has a larger hydrodynamic volume than the native protein as it is far less compact than the native protein. The intermediate species eluted at a larger volume of 66 ml compared to native protein which is eluted at 71 ml. The intermediate shows a larger hydrodynamic or 'tumbling' volume compared to the native protein and this suggests that it is slightly less compact than the native protein but not as expanded as unfolded protein (Figure 27D). The native protein corresponds to a molecular weight of 25 kDa, while that of the intermediate species corresponds to a molecular weight of 27 kDa and the unfolded protein corresponds to a molecular weight of 41 kDa. Since the molecular weight of the unfolded protein is not near double that of the dimer, which would be approximately 50 kDa, higher-order oligomers such as a tetramer can be ruled out.

3.6.4 ANS binding to the intermediate

8-anilino-1-napthalene sulfonate (ANS) is an amphipathic dye that binds to exposed hydrophobic patches on the protein surface. It is a useful probe to study the formation of intermediates during unfolding/ refolding. ANS has been used extensively to study molten globule intermediates (Christensen and Pain, 1991; Semisotnov *et al.*, 1991; Alam *et al.*, 2015; Haque *et al.*, 2015; Rahaman *et al.*, 2015). When excited at 390 nm, ANS shows an emission maximum wavelength of 520 nm. Upon binding to an exposed hydrophobic cluster on a protein, the maximum emission wavelength shifts to a lower wavelength that is dependent on the hydrophobicity of the binding site/s on the protein. The blue-shifted emission maximum wavelength is accompanied by an increase in the quantum yield. This technique can be used to assess the relative 'openness' of the intermediate relative to the native and unfolded speciess. If there are exposed hydrophobic patches in the intermediate, it may indicate that the structure of the intermediate is more exposed to the solvent than in the native state for example.

The native protein bind to ANS which is shown by the blue-shifted emission maximum from 510 nm in free ANS (Stevens and Augusteyn, 1997; Ali *et al.*, 1999) to 474 nm as well as an increase in fluorescence intensity which indicates exposure of hydrophobic patches on the protein surface (Figure 27C). The intermediate species also binds ANS and shows a blue-shifted emission maximum to 478 nm with a decreased fluorescence intensity compared to that of the native protein. Unfolded protein is not expected to bind ANS because the hydrophobic patches that were close together in the intact tertiary structure is disrupted in the unfolded protein and no longer form patches. Since the intermediate species binds to ANS, it indicates that there are exposed hydrophobic patches in the intermediate state. The intermediate spectrum is not as blue-shifted as in the native state which suggests that it some of the exposed hydrophobic patches in the native state have been disrupted. Together, these results indicate that the intermediate is native-like, but has properties that are different to either of the native and unfolded species.

.

CHAPTER 4. DISCUSSION

With the ever-increasing number of domain-swapped structures being elucidated, there is substantial reason to believe that this phenomenon is biologically significant. However, to date there is no general mechanism of domain swapping. This could be attributed to domain-swapped proteins being extremely diverse in structure as well as function and the mechanism depends on the particular protein and its folding/ unfolding pathway. The FOXP3 FHD has been shown to form a domain-swapped dimer that is crucial for its suppressor function to regulate homeostasis of the immune system (Bandukwala *et al.*, 2011). The importance of FOXP3 is highlighted by a severe autoimmune disease (IPEX syndrome) caused by mutations that disrupt domain swapping (Bennett *et al.*, 2001; Bandukwala *et al.*, 2011). In addition, the FOXP3 FHD is the only FOXP member to show formation of DSD only, rather than a monomer-dimer equilibrium which is present in the FOXP1 and FOXP2 FHDs (Stroud *et al.*, 2006; Chu *et al.*, 2011). This together with the fact that very little is actually known about how FOXP3 functions at a molecular level prompted this investigation into the mechanism of domain swapping of the FOXP3 FHD. Furthermore, there are only a handful of studies which have investigated the mechanism of domain swapping and this particular study using the FOXP3 FHD as a model can shed more light into how these proteins elegantly fold into these exquisite domain-swapped structures.

4.1 Equilibrium unfolding of the FOXP3 FHD proceeds via a three-state model

Most domain-swapped proteins follow a two-state equilibrium folding model (N↔D) (Bennett *et al.*, 1995; Rousseau *et al.*, 2003; Gronenborn, 2009). In contrast, the equilibrium unfolding curves (Figure 21) indicate that the FOXP3 FHD follows a three-state model ($N_2 \leftrightarrow I_2 \leftrightarrow 2D$). Furthermore, a concentration dependence is observed in the second transition which indicates that the protein unfolds *via* a dimeric intermediate.

The noncoincidence of the CD and fluorescence unfolding curves suggest that the tertiary structure is more susceptible to disruption by low urea concentrations than is the secondary structure (Figure 21B) and the tryptophan residues are likely to become exposed to solvent as the protein starts to unfold prior to any secondary structure

dissolving (Lee and Richards, 1971). Indeed, the secondary structure remains intact until approximately 4.4 M urea which is in contrast to the weak pre-transition region in the fluorescence-monitored curve and indicates that the secondary structure of the FOXP3 DSD is more resistant to the denaturant than the tertiary structure.

Interestingly most domain swapping proteins have been shown to unfold via a two-state model with no formation of thermodynamically stable intermediates, except for the FOXP1 DSD (Rousseau *et al.*, 1998; Staniforth *et al.*, 2001; Topping *et al.*, 2004). Furthermore, all domain-swapped proteins studied to date require either partial or complete unfolding of the monomeric state in order to attain the domain-swapped structure (Rousseau *et al.*, 1998; Staniforth *et al.*, 2001; Rousseau *et al.*, 2004; Chen *et al.*, 2005; Liu *et al.*, 2012). Indeed, the only other case reported of a three state unfolding transition for a domain swapping protein is that of the FOXP1 FHD which was shown to unfold via a monomeric intermediate (Medina *et al.*, 2016). In this particular study, the FOXP1 DSD dissociates prior to unfolding and involves monomer stability (Medina *et al.*, 2016). Additionally, most domain-swapped proteins, there is a large kinetic barrier that separates the monomer and domain-swapped dimer and this barrier can only be overcome by favouring the unfolded species using salt, pH, temperature or denaturants.

The equilibrium unfolding results for the FOXP3 DSD are in direct contrast with all other domain-swapped proteins. Here, it is shown that the FOXP3 DSD exists as a stable dimer in solution and there is no need for additional perturbations to favour the unfolded species in order for the protein to domain swap. The FOXP3 DSD is shown to unfold *via* a three-state mechanism with formation of a dimeric equilibrium intermediate compared to the FOXP1 DSD which unfolds *via* formation of a monomeric intermediate. Thermodynamic parameters and thus the conformational stability could not be determined as the data could not be fit to a three-state model due to the lack of a defined pre-transition region in the fluorescence unfolding curve (Figure 21).

4.2 The FOXP3 DSD intermediate

From the equilibrium unfolding studies, it is clear that the FOXP3 FHD forms a thermodynamically stable intermediate species. In order to assess the role of this

intermediate in the FOXP3 folding pathway, it is necessary to characterise this intermediate in terms of its secondary, tertiary and quaternary structure and compare it to the native state

The CD spectrum of the intermediate (FOXP3 FHD in the presence of 5 M urea) superimposes perfectly onto the CD spectrum obtained for the native protein (Figure 27). This suggests that the secondary structure of the intermediate remains intact and very much native-like.

Tryptophan fluorescence of the intermediate species shows a red-shifted emission maximum to 350 nm while the native and unfolded protein fluorescence spectra have emission maximum wavelengths of 341 nm and 361 nm, respectively (Figure 27B). The emission maximum of 341 nm is reflective of the average local environments of all the tryptophan residues in the FOXP3 DSD. The tryptophan Trp406 residue is not as solvent-exposed as indicated in the crystal structure in Figure 15 and as shown by the molecular surface in Figure 28. As shown in Figure 29, the FOXP3 DSD is largely flexible except for helices H2 and H3 as well as the wing regions (S2 and S3). The red-shifted emission maximum of the intermediate state indicates that the environments of the tryptophan residues have changed and are now more exposed to solvent in the intermediate state compared to the native state. The tryptophan residues in the intermediate state, however are not as exposed as is in the fully-unfolded protein (Lacowicz, 1999).

ANS is a well-known hydrophobic dye that binds to the exposed hydrophobic surfaces of proteins through its aniline naphthalene group (Semisotnov *et al.*, 1991). It was used here as a probe to detect the intermediate species that was populated during urea-induced equilibrium unfolding (Figure 21B) because due to their more ‘open’ nature, they are known to exhibit exposed hydrophobic patches which can interact with and bind to ANS. Free ANS, in the absence of protein, was found to have an emission maximum of 520 nm. A blue-shifted emission maximum to 474 nm was observed upon binding to native protein, in the absence of urea. This suggests that the native protein contains hydrophobic patches on the surface of the protein that is solvent-exposed (Figure 28).

Size-exclusion chromatography was used here to assess the compactness of the intermediate state compared to that of the native and unfolded proteins. The

intermediate state is eluted at a similar volume to that of the native protein. This suggests that the intermediate is very similar to the native protein in terms of compactness but is very different to that of unfolded protein. The unfolded protein is eluted close to the void volume which indicates that the hydrodynamic radius or 'tumbling' volume has increased approximately 5-fold upon unfolding.

The structural characteristics of the FOXP3 FHD intermediate are similar to those found in molten globule intermediates. According to Christensen and Pain, 1991, molten globule intermediates are shown to possess the following traits:

- Secondary structure is always present. The molten globule intermediate is found to contain as much as 95 % of the secondary structural content as the native protein.
- Tertiary interactions are significantly reduced as the intermediate has a greater access to solvent than does the native state.
- The intermediate is largely globular as the intermediate state retains most of its secondary structure which implies that it is more compact than the unfolded species.
- Stable intermediates tend to aggregate as thermodynamically stable intermediates may not be kinetically stable. This particular characteristic is reflected in the binding of the intermediate state to ANS.
- The states behave thermodynamically as if "molten" as the transition between the native and intermediate states are mostly two-state mechanisms.

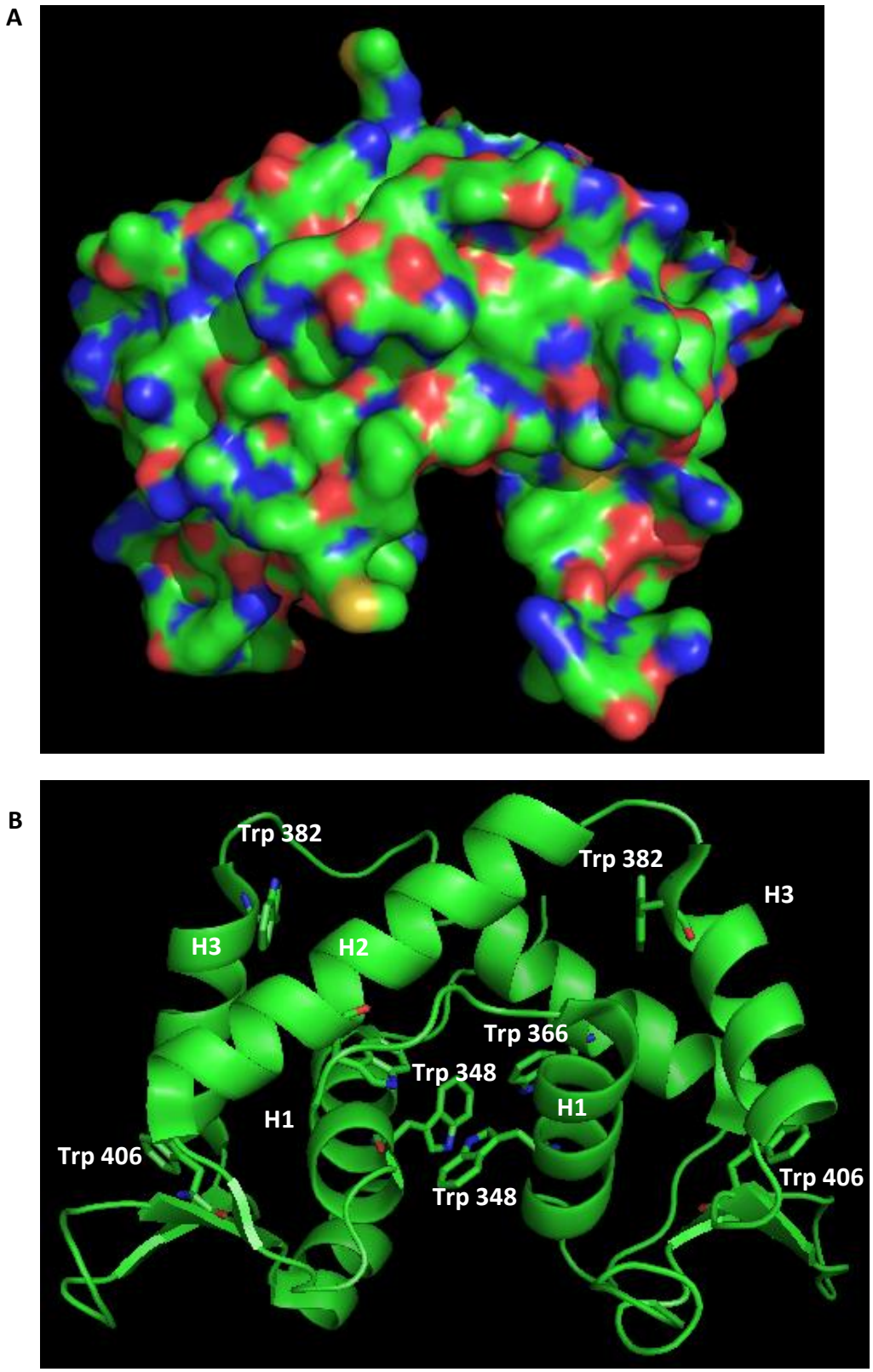
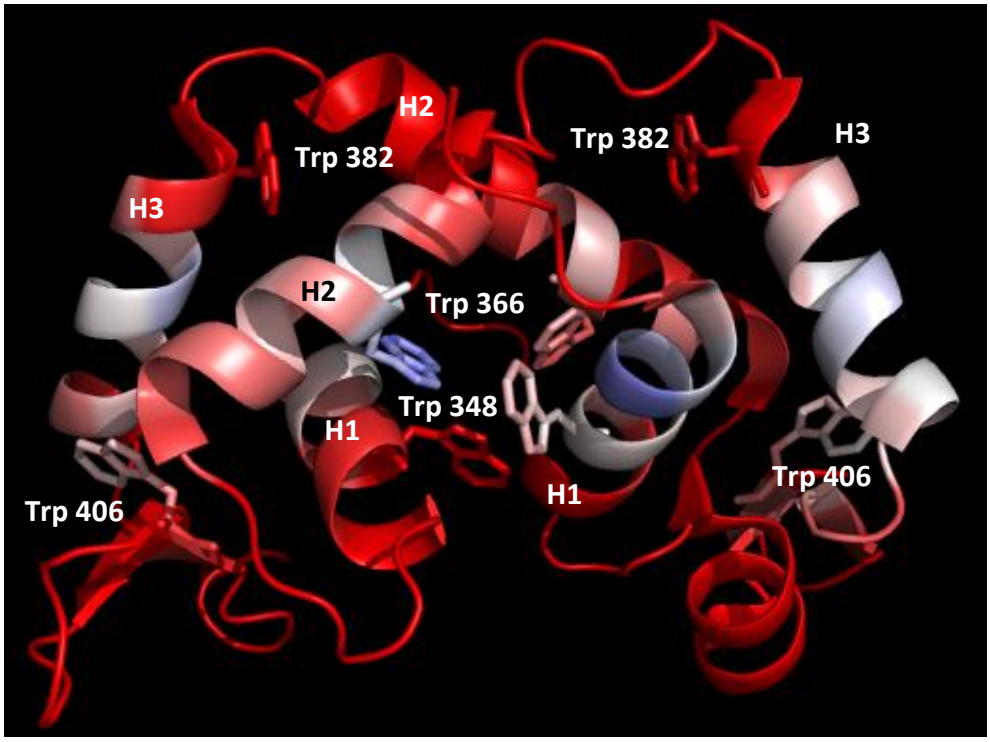


Figure 28. Exposed hydrophobic surface of the FOXP3 DSD. A. The exposed hydrophobic surface is shown in red while polar regions are shown in blue. **B.** Positions of the tryptophan residues in relation to the exposed surface hydrophobic patches. Figure generated using PyMOL (Delano Scientific). PDB code 3QRF.

The FOXP3 FHD intermediate state is native-like in some aspects. The intermediate has secondary structure that is almost identical to the native state with significantly disrupted tertiary structure that shows exposed tryptophan residues. The intermediate state is shown to bind to ANS although to a lesser extent than the native protein which indicates disruption of exposed hydrophobic surface patches. The quaternary structure of the intermediate is shown to be largely globular and similar to the native state. It is still dimeric in nature as no protein was eluted at the expected size for monomer (13 kDa). Together, these results suggest that the FOXP3 FHD intermediate forms a molten globule intermediate.

A



B

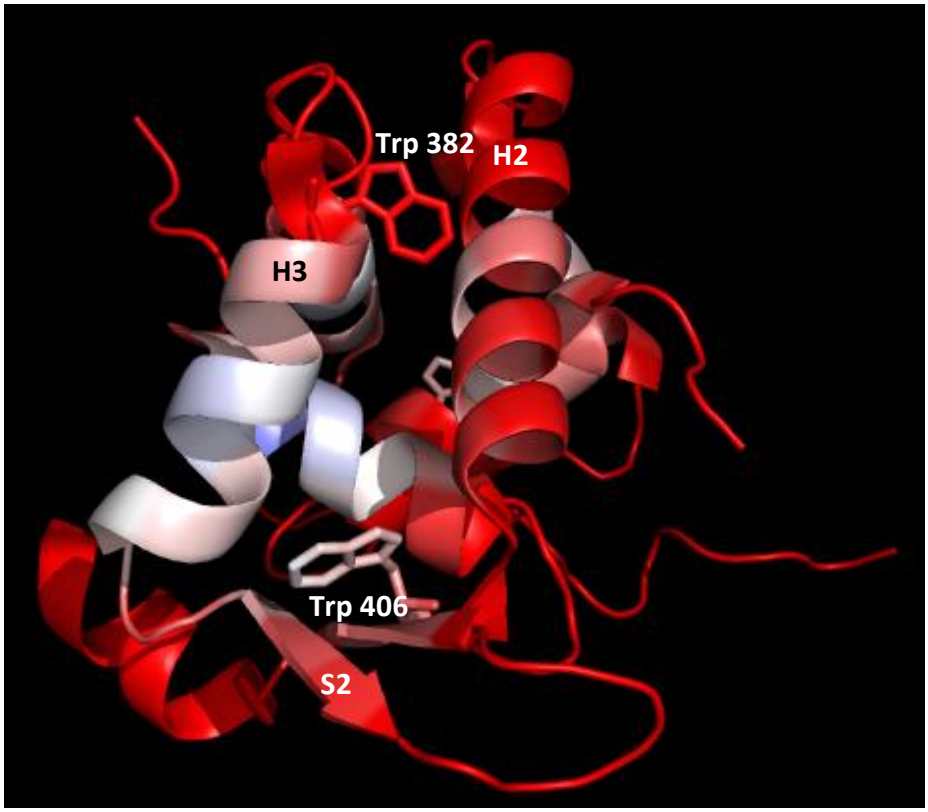


Figure 29. The crystal structure of the FOXP3 FHD coloured according to B-factor. A. front view and **B.** side view coloured according to B-factor. The regions in red are highly flexible while the regions in blue are most rigid. The regions in white are somewhat rigid. Figure generated using PyMol (Delano Scientific). PDB code 3QRF.

4.3 Mechanism of (Un)folding of the FOXP3 DSD

The conclusion drawn from equilibrium unfolding is that unfolding of the FOXP3 DSD is a noncooperative, three-state mechanism via the formation of a stable intermediate species. Based on this, one would expect to observe biphasic kinetics for unfolding with at least formation of one intermediate. The kinetic unfolding of the FOXP3 DSD in 8 M urea, however is composed of two unfolding phases as shown by the parameters obtained by fitting the data to a double exponential curve. The residuals indicate that the fit was good. Furthermore, the unfolding kinetics showed a burst phase within the dead-time of the instrument (< 2 ms) for all the single-jump unfolding and refolding kinetics. The burst phase could indicate the formation of an intermediate species that is forming too quickly to be detected.

Unfolding of the FOXP3 DSD exhibited triphasic kinetics except for the urea concentration range of 7 M – 7.75 M urea which exhibited biphasic kinetics (Figure 24). The formation of the intermediate showed one major, intermediate phase, a minor, slow phase and a minor, fast phase. The major, intermediate phase indicated a significant exposure of surface area to solvent upon unfolding to the intermediate state. However, the minor, fast phase showed a higher exposure to solvent than that of the intermediate phase and suggests that an additional intermediate forms. It is interesting to note that the jump from native to 8 M urea does not show formation of this phase which indicates that this intermediate with a higher exposure of surface area to solvent could be transient. The minor, slow phase of unfolding to the interemediate shows a negative slope and indicates burial of surface area relative to the native state. This suggests that a minor population of species could be forming protein aggregates; however the majority of the species unfolds to form the intermediate state.

The major, slow phase in unfolding indicates that a large amount of surface area is exposed to solvent upon unfolding completely (Tanford, 1968; Myers *et al.*, 1995) as indicated by the m_u value of 0.46. The minor, fast phase shows that there is exposure of surface area to the solvent but to a lesser extent than in the major, slow phase. This suggests formation of a species that is more compact than fully unfolded protein but less compact than the native state, however, it is a minor species along the unfolding pathway.

The refolding kinetics consists of three phases. The kinetics data were fitted to a double exponential and the residuals indicated that the fit was good. A burst phase was also observed within the instrument dead-time for all kinetic refolding jumps which suggests that an event is occurring but the instrument is unable to detect it. The amplitude of the slow phase and rates of all three phases are dependent on urea concentration. Only the slow phase shows a substantial decrease in absolute amplitude with increasing urea concentration as the native state population diminishes in the equilibrium transition region (Park *et al.*, 1999). Furthermore, at lower urea concentrations (< 4 M urea) the refolding kinetics is dominated by the slow, refolding phase. The rate of the slow phase of refolding shows an increase with increasing urea concentration at the expense of the fast refolding phase while the rate constants of the fast and intermediate phases show a decrease with increasing urea concentration. This suggests that the majority of the population forms the intermediate that is more compact than fully unfolded protein while there are two other minor populations that are becoming less compact.

A three-state mechanism is proposed for the (un)folding pathway of the FOXP3 DSD (Figure 31). Unfolding from the native species to the intermediate, at low urea concentrations (3 M to 5 M urea) where the equilibrium intermediate is present, shows three phases, a fast, intermediate and a slow phase. This suggests formation of three structures in the transition state, however it is unclear whether these intermediates form in sequence or in parallel. Unfolding of the intermediate to the unfolded species occurs in two phases with the major, fast phase dominating the kinetics. This suggests that two events are occurring with the majority of species completely unfolding while the minor species forming a kinetic intermediate that is less compact than the native state but more compact than the unfolded species. This minor population could be due to dissociated dimer that subsequently unfolds completely.

Refolding from the unfolded species to the intermediate state involves three phases and a burst phase (Figure 25). These results suggest formation of at least four kinetic intermediates during refolding. The refolding kinetics from the unfolded species to the intermediate state shows a major, slow refolding phase which results in the formation of a kinetic intermediate that is significantly more compact than the unfolded protein. This is attributed to formation of the equilibrium intermediate. Two minor populations both show a positive slope in the refolding rate plot (Figure 26) which suggests that these kinetic intermediates that form along this refolding pathway to form the intermediate (2U

↔ I_2) contain more exposed surface area to solvent relative to the unfolded species. This suggests that a small population could be forming aggregates upon refolding (Figure 30)

Refolding from the intermediate state to the native state shows three phases. The major, slow refolding phase is attributed to formation of the compact native state while the fast, minor phase is attributed to formation of a kinetic intermediate that is less compact than the native state and the equilibrium intermediate. This could be attributed to formation of a transient kinetic intermediate

Structural characterisation of the equilibrium unfolding intermediate suggests that it has substantial secondary structure, with a disrupted or more 'open' tertiary structure but is still quite compact in nature. Equilibrium unfolding shows formation of one stable intermediate species, however the kinetic unfolding pathway suggests that at least two transient intermediate species are present in the formation of the intermediate species from the native state.

According to the positions and environments of the tryptophan residues (Figure 28B), it is implied that the primary hydrophobic interface is disrupted upon formation of the intermediate. This results in a highly-flexible, yet less compact molten-globule intermediate which then dissociates into monomers (burst phase in I ↔ U transition) and subsequently unfolds completely. The I ↔ U transition shows a decrease in the fluorescence signal which implies that the tryptophan fluorescence is quenched upon exposure to the solvent upon complete unfolding. Since the crystal structure of the FOXP3 FHD only exhibits domain-swapped dimers which interact in a non-covalent manner, the protein should unfold into unfolded monomers. There is no evidence to suggest that any other type of dimer, such as non-domain-swapped dimers are able to form.

4.4 Conclusion and future work

All other domain-swapped proteins to date show a two-state mechanism with the absence of thermodynamically stable intermediates (Rousseau *et al.*, 1998; Staniforth *et al.*, 2001; Rousseau *et al.*, 2004; Chen *et al.*, 2005; Liu *et al.*, 2012) while that of the FOXP proteins exhibit a three-state equilibrium unfolding mechanism. It is interesting to

note that the FOXP1 DSD unfolds *via* a monomeric intermediate (Medina *et al.*, 2016) while the FOXP3 DSD unfolds *via* a dimeric intermediate. Furthermore, the unfolding and refolding kinetics of the FOXP3 DSD, reveal that there are transient intermediates along the folding and unfolding pathway that are not detected during equilibrium unfolding. These hidden intermediates may be fleeting and may be on- or off-pathway which means that they may or may not lead to the correctly folded conformation. Here, we propose a complete folding and unfolding pathway of the FOXP3 DSD (Figure 30) which shows that domain swapping occurs from the unfolded species to form a native-like, dimeric intermediate which then undergoes minor structural rearrangements to form the native FOXP3 DSD. This work is in direct contrast with other domain swapping proteins which have to undergo significant partial or complete unfolding in order to attain its domain-swapped form. Furthermore, other domain-swapped proteins exist in a monomer-dimer equilibrium whereas the FOXP3 DSD exists only as a dimer in solution. No monomer is detected at concentrations as low as 4 µM. This could possibly explain why theFOXP3 DSD folds and unfolds *via* a dimeric intermediate as opposed to a monomeric intermediate. Future work needs to be directed towards investigating whether these intermediates are on- or off-pathway and to establish whether they are obligatory, as well as if they occur sequentially or simultaneously using double-jump kinetics. Furthermore, since the secondary interface and the hinge-loop region have been shown to be crucial for domain swapping, it is important to establish the roles of these regions in the stability and mechanism of domain swapping.

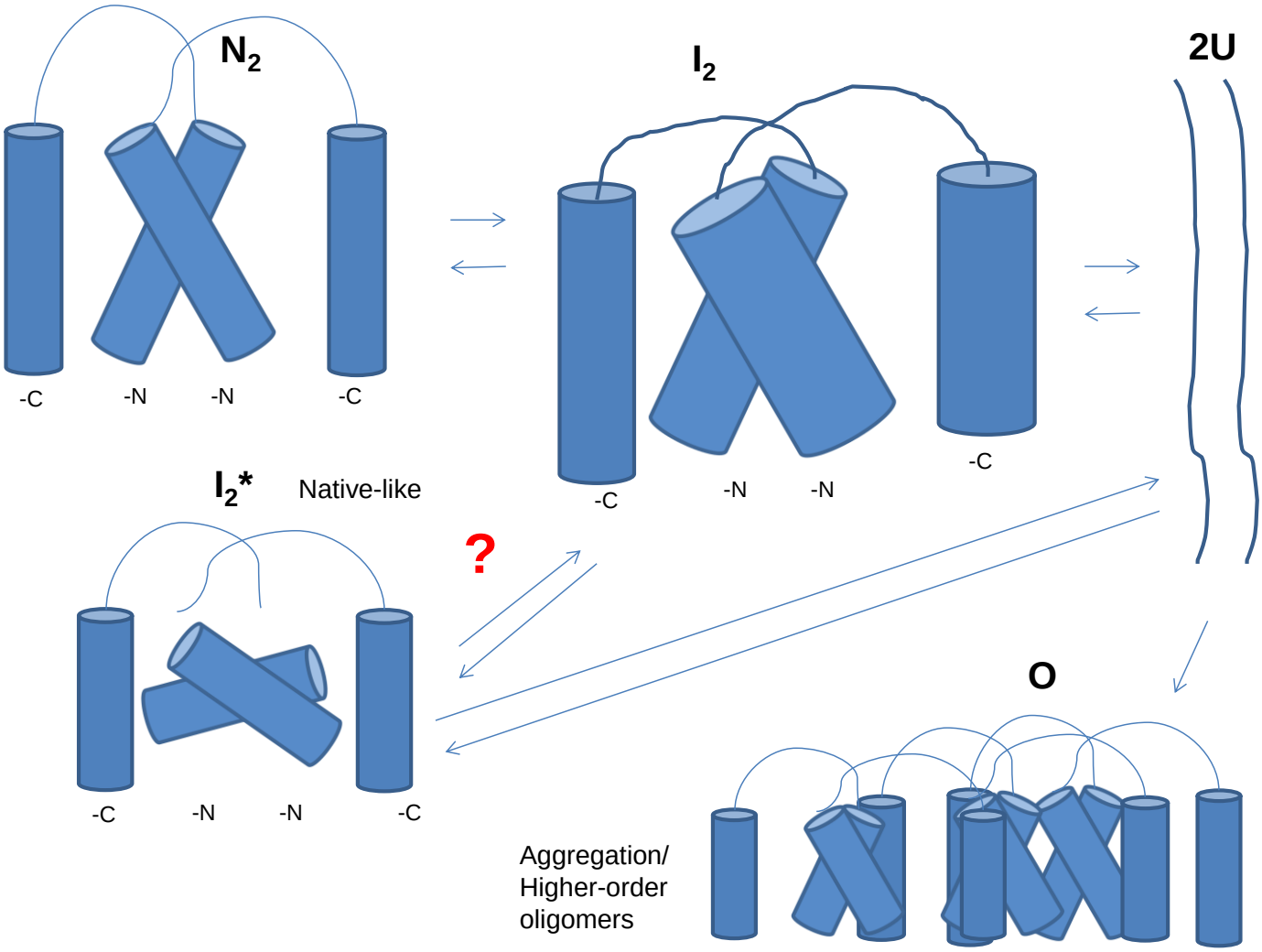


Figure 30. Proposed scheme of the (un)folding pathway of the FOXP3 DSD. The protein unfolds from the native state to the intermediate via two phases, a fast and slow phase which could represent two structures in the transition state. The FOXP3 DSD unfolds from the native state to the unfolded species in a fast phase but also shows a burst phase. This suggests that a highly transient, kinetically unstable intermediate may be forming too quickly to be detected. The burst phase is observed in both refolding and unfolding kinetics. The refolding of the FOXP3 DSD from the unfolded species to the intermediate is monophasic which implies a single folding event. Similarly the refolding from the intermediate to the native state is monophasic. Figure was generated using Microsoft PowerPoint 2010.

References

Abel, T. and T. Maniatis (1989). "Gene regulation. Action of leucine zippers." Nature **341**(6237): 24-25.

Ahmad, E., S. K. Rahman, et al. (2010). "Phytolacca americana lectin (Pa-2; pokeweed mitogen): an intrinsically unordered protein and its conversion into partial order at low pH." Biosci Rep **30**(2): 125-134.

Alam, P., F. Naseem, et al. (2015). "Effect of galactose on acid induced molten globule state of soybean agglutinin: biophysical approach." Journal of Molecular Structure **1099**: 149-153.

Ali, V., K. Prakash, et al. (1999). "8-anilino-1-naphthalene sulfonic acid (ANS) induces folding of acid unfolded cytochrome c to molten globule state as a result of electrostatic interactions." Biochemistry **38**(41): 13635-13642.

Amoutzias, G., A. Veron, et al. (2007). "One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity." Mol Biol Evol **24**(3): 827-835.

Amoutzias, G. D., D. L. Robertson, et al. (2004). "Convergent evolution of gene networks by single-gene duplications in higher eukaryotes." EMBO Rep **5**(3): 274-279.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(4096): 223-230.

Anthony-Cahill, S. J., P. A. Benfield, et al. (1992). "Molecular characterization of helix-loop-helix peptides." Science **255**(5047): 979-983.

Antonin, W., D. Fasshauer, et al. (2002). "Crystal structure of the endosomal SNARE complex reveals common structural principles of all SNAREs." Nature Structural & Molecular Biology **9**(2): 107-111.

Baba, M., S. Nakajo, et al. (1998). "Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies." The American journal of pathology **152**(4): 879.

Baldwin, E. P. and B. W. Matthews (1994). "Core-packing constraints, hydrophobicity and protein design." Curr Opin Biotechnol **5**(4): 396-402.

Bandukwala, H. S., Y. Wu, et al. (2011). "Structure of a domain-swapped FOXP3 dimer on DNA and its function in regulatory T cells." Immunity **34**(4): 479-491.

Bandukwala, H. S., Y. Wu, et al. (2011). "Structure of a domain-swapped FOXP3 dimer on DNA and its function in regulatory T cells." Immunity **34**(4): 479-491.

Banham, A. H., N. Beasley, et al. (2001). "The FOXP1 winged helix transcription factor is a novel candidate tumor suppressor gene on chromosome 3p." Cancer Res **61**(24): 8820-8829.

Barlow, D. J. and J. Thornton (1983). "Ion-pairs in proteins." J Mol Biol **168**(4): 867-885.

Bence, N. F., R. M. Sampat, et al. (2001). "Impairment of the ubiquitin-proteasome system by protein aggregation." Science **292**(5521): 1552-1555.

Bennett, C. L., J. Christie, et al. (2001). "The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3." Nat Genet **27**(1).

Bennett, C. L., J. Christie, et al. (2001). "The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3." Nat Genet **27**(1): 20-21.

Bennett, M. and D. Eisenberg (1994). "Refined structure of monomelic diphtheria toxin at 2.3 Å resolution." Protein science **3**(9): 1464-1475.

Bennett, M. J., M. P. Schlunegger, et al. (1995). "3D domain swapping: a mechanism for oligomer assembly." Protein Sci **4**(12): 2455-2468.

Bennett, M. J., M. P. Schlunegger, et al. (1995). "3D domain swapping: a mechanism for oligomer assembly." Protein science **4**(12): 2455-2468.

Bergdoll, M., M.-H. Remy, et al. (1997). "Proline-dependent oligomerization with arm exchange." Structure **5**(3): 391-401.

Berman, H. M., J. Westbrook, et al. (2000). "The protein data bank." Nucleic Acids Res **28**(1): 235-242.

Bettelli, E., M. Dastrange, et al. (2005). "Foxp3 interacts with nuclear factor of activated T cells and NF-kappa B to repress cytokine gene expression and effector functions of T helper cells." Proc Natl Acad Sci U S A **102**(14): 5138-5143.

Bianchi, A., S. Smith, et al. (1997). "TRF1 is a dimer and bends telomeric DNA." EMBO J **16**(7): 1785-1794.

Bogan, A. A. and K. S. Thorn (1998). "Anatomy of hot spots in protein interfaces." J Mol Biol **280**(1): 1-9.
Boura, E., L. Rezabkova, et al. (2010). "Structure of the human FOXO4-DBD-DNA complex at 1.9 A resolution reveals new details of FOXO binding to the DNA." Acta Crystallogr D Biol Crystallogr **66**(Pt 12): 1351-1357.
Bourne, Y., A. S. Arvai, et al. (1995). "Crystal structure of the cell cycle-regulatory protein suc1 reveals a beta-hinge conformational switch." Proceedings of the National Academy of Sciences **92**(22): 10232-10236.
Brent, M. M., R. Anand, et al. (2008). "Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification." Structure **16**(9): 1407-1416.
Bullock, A. N., J. Henckel, et al. (1997). "Thermodynamic stability of wild-type and mutant p53 core domain." Proceedings of the National Academy of Sciences **94**(26): 14338-14342.
Byeon, I. J., J. M. Louis, et al. (2003). "A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping." J Mol Biol **333**(1): 141-152.
Byeon, I. J., J. M. Louis, et al. (2004). "A captured folding intermediate involved in dimerization and domain-swapping of GB1." J Mol Biol **340**(3): 615-625.
Campbell, A. J., L. Lyne, et al. (2010). "Aberrant expression of the neuronal transcription factor FOXP2 in neoplastic plasma cells." Br J Haematol **149**(2): 221-230.
Carson, B. D., J. E. Lopes, et al. (2006). "Insights into transcriptional regulation by FOXP3." Front Biosci **11**: 1607-1619.
Castanet, M., S. M. Park, et al. (2002). "A novel loss-of-function mutation in TTF-2 is associated with congenital hypothyroidism, thyroid agenesis and cleft palate." Hum Mol Genet **11**(17): 2051-2059.
Chae, W. J., O. Henegariu, et al. (2006). "The mutant leucine-zipper domain impairs both dimerization and suppressive function of Foxp3 in T cells." Proc Natl Acad Sci U S A **103**(25): 9631-9636.
Chamberlain, N. L., E. D. Driver, et al. (1994). "The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function." Nucleic Acids Res **22**(15): 3181-3186.
Changeux, J. P. and S. J. Edelstein (1998). "Allosteric receptors after 30 years." Neuron **21**(5): 959-980.
Chen, C. P., S. Posy, et al. (2005). "Specificity of cell-cell adhesion by classical cadherins: Critical role for low-affinity dimerization through beta-strand swapping." Proc Natl Acad Sci U S A **102**(24): 8531-8536.
Chen, J., N. Sawyer, et al. (2013). "Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area." Protein Sci **22**(4): 510-515.
Chothia, C. (1974). "Hydrophobic bonding and accessible surface area in proteins." Nature **248**(446): 338-339.
Chothia, C. and J. Janin (1975). "Principles of protein-protein recognition." Nature **256**(5520): 705-708.
Christensen, H. and R. H. Pain (1991). "Molten globule intermediates and protein folding." Eur Biophys J **19**(5): 221-229.
Chu, Y. P., C. H. Chang, et al. (2011). "Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: insight into its domain swapping and DNA binding." Protein Sci **20**(5): 908-924.
Chu, Y. P., C. H. Chang, et al. (2011). "Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: Insight into its domain swapping and DNA binding." Protein science **20**(5): 908-924.
Clark, K. L., E. D. Halay, et al. (1993). "Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5." Nature **364**(6436): 412-420.
Clark, K. L., E. D. Halay, et al. (1993). "Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5."
Coffer, P. J. and B. M. Burgering (2004). "Forkhead-box transcription factors and their role in the immune system." Nature Reviews Immunology **4**(11): 889-899.
Coffer, P. J. and B. M. Burgering (2004). "Forkhead-box transcription factors and their role in the immune system." Nat Rev Immunol **4**(11): 889-899.
Cole, N. and G. B. Ralston (1994). "Enhancement of self-association of human spectrin by polyethylene glycol." Int J Biochem **26**(6): 799-804.
Cowan-Jacob, S. W., G. Fendrich, et al. (2005). "The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation." Structure **13**(6): 861-871.
Creighton, T. E. (1993). Proteins: structures and molecular properties, Macmillan.
D'Alessio, G. (1999). "The evolutionary transition from monomeric to oligomeric proteins: tools, the environment, hypotheses." Prog Biophys Mol Biol **72**(3): 271-298.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. San Carlos, CA, USA, DeLano Scientific.
Desvergne, B., L. Michalik, et al. (2006). "Transcriptional regulation of metabolism." Physiological reviews **86**(2): 465-514.
Dill, K. A. (1985). "Theory for the folding and stability of globular proteins." Biochemistry **24**(6): 1501-1509.
Dill, K. A., S. Bromberg, et al. (1995). "Principles of protein folding—a perspective from simple exact models." Protein science **4**(4): 561-602.
Dobson, C. M. (2000). Mechanisms of Protein Folding. New York, United States, Oxford University Press.
Drozdetskiy, A., C. Cole, et al. (2015). "JPred4: a protein secondary structure prediction server." Nucleic Acids Res **43**(W1): W389-394.
Edelhoch, H. (1967). "Spectroscopic determination of tryptophan and tyrosine in proteins." Biochemistry **6**(7): 1948-1954.
Endicott, J. A. and P. Nurse (1995). "The cell cycle and suc1: from structure to function?" Structure **3**(4): 321-325.
Englander, S. W. and L. Mayne (2014). "The nature of protein folding pathways." Proc Natl Acad Sci U S A **111**(45): 15873-15880.
Fagagnini, A., A. Pica, et al. (2017). "Onconase Dimerization Through 3D Domain Swapping: Structural Investigations and Increase Of The Apoptotic Effect In Cancer Cells." Biochemical Journal: BCJ20170541.
Fersht, A. R. (1987). "The hydrogen bond in molecular recognition." Trends Biochem Sci **12**: 301-304.
Fersht, A. R. (1997). "Nucleation mechanisms in protein folding." Curr Opin Struct Biol **7**(1): 3-9.
Fisher, S. E. and C. Scharff (2009). "FOXP2 as a molecular window into speech and language." Trends Genet **25**(4): 166-177.
Foguel, D., J. L. Silva, et al. (1998). "Characterization of a partially folded monomer of the DNA-binding domain of human papillomavirus E2 protein obtained at high pressure." J Biol Chem **273**(15): 9050-9057.
Frank, M. K., F. Dyda, et al. (2002). "Core mutations switch monomeric protein GB1 into an intertwined tetramer." Nature Structural & Molecular Biology **9**(11): 877-885.
Frankel, A. D. and P. S. Kim (1991). "Modular structure of transcription factors: implications for gene regulation." Cell **65**(5): 717-719.
Fried, M. G. (1989). "Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay." Electrophoresis **10**(5-6): 366-376.
Friedman, J. and K. Kaestner (2006). "The Foxa family of transcription factors in development and metabolism." Cellular and Molecular Life Sciences CMLS **63**(19-20): 2317-2328.
Garner, M. M., Revzin, A. (1986). "The use of gel electrophoresis to detect and study nucleic acid-protein interactions." Trends Biochem Sci **11**: 395-396.
Gasteiger, E., C. Hoogland, et al. (2005). Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook, Springer**:** 571-607.
Gething, M.-J. and J. Sambrook (1992). "Protein folding in the cell."
Glaser, F., D. M. Steinberg, et al. (2001). "Residue frequencies and pairing preferences at protein-protein interfaces." Proteins **43**(2): 89-102.
Gong, X., M. Jia, et al. (2004). "Association between the FOXP2 gene and autistic disorder in Chinese population." Am J Med Genet B Neuropsychiatr Genet **127B**(1): 113-116.
Gorinstein, S., I. Goshev, et al. (2000). "Intrinsic tryptophan fluorescence of human serum proteins and related conformational changes." J Protein Chem **19**(8): 637-642.
Gotte, G., M. Bertoldi, et al. (1999). "Structural versatility of bovine ribonuclease A. Distinct conformers of trimeric and tetrameric aggregates of the enzyme." Eur J Biochem **265**(2): 680-687.
Gotte, G., F. Vottariello, et al. (2003). "Thermal Aggregation of Ribonuclease AA CONTRIBUTION TO THE UNDERSTANDING OF THE ROLE OF 3D DOMAIN SWAPPING IN PROTEIN AGGREGATION." Journal of Biological Chemistry **278**(12): 10763-10769.
Grandori, C., S. M. Cowley, et al. (2000). "The Myc/Max/Mad network and the transcriptional control of cell behavior." Annual review of cell and developmental biology **16**(1): 653-699.
Green, S. M., A. G. Gittis, et al. (1995). "One-step evolution of a dimer from a monomeric protein." Nature Structural & Molecular Biology **2**(9): 746-751.

Gronenborn, A. M. (2009). "Protein acrobatics in pairs--dimerization via domain swapping." Curr Opin Struct Biol **19**(1): 39-49.
Haque, M. A., S. Ubaid-Ullah, et al. (2015). "Characterization of pre-molten globule state of yeast iso-1-cytochrome c and its deletants at pH 6.0 and 25 C." International journal of biological macromolecules **72**: 1406-1418.
Harrison, S. (1991). "A structural taxonomy of DNA-binding domains." Nature **353**: 715-719.
Harrison, S. C. and R. Durbin (1985). "Is there a single pathway for the folding of a polypeptide chain?" Proceedings of the National Academy of Sciences **82**(12): 4028-4030.
Hartl, F. U. and M. Hayer-Hartl (2009). "Converging concepts of protein folding in vitro and in vivo." Nature Structural & Molecular Biology **16**(6): 574-581.
Hayes, M. V., R. B. Sessions, et al. (1999). "Engineered assembly of intertwined oligomers of an immunoglobulin chain." Journal of molecular biology **285**(4): 1857-1867.
Hellman, L. M. and M. G. Fried (2007). "Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions." Nat Protoc **2**(8): 1849-1861.
Heringa, J. and W. R. Taylor (1997). "Three-dimensional domain duplication, swapping and stealing." Curr Opin Struct Biol **7**(3): 416-421.
Hontz, J. S., M. T. Villar-Lecumberri, et al. (2006). "Differences in crystal and solution structures of the cytolethal distending toxin B subunit: Relevance to nuclear translocation and functional activation." J Biol Chem **281**(35): 25365-25372.
Hu, J. C. and R. T. Sauer (1992). The Basic-Region Leucine-Zipper Family of DNA Binding Proteins. Nucleic Acids and Molecular Biology. F. Eckstein and D. M. J. Lilley. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 82-101.
Huang, Y., H. Cao, et al. (2012). "Three-dimensional domain swapping in the protein structure space." Proteins **80**(6): 1610-1619.
Huang, Y., M. Gao, et al. (2018). "Exploring the Roles of Proline in Three-Dimensional Domain Swapping from Structure Analysis and Molecular Dynamics Simulations." Protein J **37**(1): 13-20.
Janowski, R., M. Kozak, et al. (2001). "Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping." Nat Struct Biol **8**(4): 316-320.
Janowski, R., M. Kozak, et al. (2001). "Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping." Nature Structural and Molecular Biology **8**(4): 316.
Janowski, R., M. Kozak, et al. (2001). "Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping." Nature Structural & Molecular Biology **8**(4): 316-320.
Jee, J., I. J. Byeon, et al. (2008). "The point mutation A34F causes dimerization of GB1." Proteins **71**(3): 1420-1431.
John-Baptiste, A., A. Vitsky, et al. (2012). "Comparison of 3 kidney injury multiplex panels in rats." Int J Toxicol **31**(6): 529-536.
Jones, S. (2004). "An overview of the basic helix-loop-helix proteins." GENOME BIOLOGY. **5**(6): 226-226.
Jones, S. and J. M. Thornton (1996). "Principles of protein-protein interactions." Proc Natl Acad Sci U S A **93**(1): 13-20.
Karplus, M. and D. L. Weaver (1994). "Protein folding dynamics: The diffusion-collision model and experimental data." Protein science **3**(4): 650-668.
Kaufmann, E. and W. Knochel (1996). "Five years on the wings of fork head." Mech Dev **57**(1): 3-20.
Khazanovich, N., K. Bateman, et al. (1996). "Crystal structure of the yeast cell-cycle control protein, p13 suc1, in a strand-exchanged dimer." Structure **4**(3): 299-309.
Khazanovich, N., K. Bateman, et al. (1996). "Crystal structure of the yeast cell-cycle control protein, p13suc1, in a strand-exchanged dimer." Structure **4**(3): 299-309.
Knaus, K. J., M. Morillas, et al. (2001). "Crystal structure of the human prion protein reveals a mechanism for oligomerization." Nat Struct Biol **8**(9): 770-774.
Koh, K. P., M. S. Sundrud, et al. (2009). "Domain requirements and sequence specificity of DNA binding for the forkhead transcription factor FOXP3." PLoS One **4**(12): e8109.
Kortt, A. A., R. L. Malby, et al. (1994). "Recombinant anti-sialidase single-chain variable fragment antibody." European journal of biochemistry **221**(1): 151-157.
Kuhlman, B., J. W. O'Neill, et al. (2001). "Conversion of monomeric protein L to an obligate dimer by computational protein design." Proceedings of the National Academy of Sciences **98**(19): 10687-10691.

Kurt, S., S. E. Fisher, et al. (2012). "Foxp2 mutations impair auditory-motor association learning." PLoS One **7**(3): e33130.
Lacowicz, J. R. (1999). Principles of fluorescence spectroscopy. New York, Kluwer Academic/Plenum.
Lai, C. S., S. E. Fisher, et al. (2001). "A forkhead-domain gene is mutated in a severe speech and language disorder." Nature **413**(6855): 519-523.
Lakowicz, J. R. and B. R. Masters (2008). "Principles of fluorescence spectroscopy." Journal of Biomedical Optics **13**(2): 9901.
Lane, D., P. Prentki, et al. (1992). "Use of gel retardation to analyze protein-nucleic acid interactions." Microbiol Rev **56**(4): 509-528.
Latchman, D. S. (1997). "Transcription factors: an overview." The international journal of biochemistry & cell biology **29**(12): 1305-1312.
Latchman, D. S. (1997). "Transcription factors: an overview." Int J Biochem Cell Biol **29**(12): 1305-1312.
Lee, B. and F. M. Richards (1971). "The interpretation of protein structures: estimation of static accessibility." J Mol Biol **55**(3): 379-400.
Levinthal, C. (1969). "How to fold graciously." Mossbauer spectroscopy in biological systems: 22-24.
Levy-Lahad, E. and R. S. Wildin (2001). "Neonatal diabetes mellitus, enteropathy, thrombocytopenia, and endocrinopathy: Further evidence for an X-linked lethal syndrome." J Pediatr **138**(4): 577-580.
Levy, D. E. and J. Darnell (2002). "Stats: transcriptional control and biological impact." Nature Reviews Molecular Cell Biology **3**(9): 651-662.
Li, S., J. Weidenfeld, et al. (2004). "Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions." Mol Cell Biol **24**(2): 809-822.
Libonati, M., M. Bertoldi, et al. (1996). "The activity on double-stranded RNA of aggregates of ribonuclease A higher than dimers increases as a function of the size of the aggregates." Biochem. J **318**: 287-290.
Lindner, R. and G. Ralston (1995). "Effects of dextran on the self-association of human spectrin." Biophys Chem **57**(1): 15-25.
Littler, D. R., M. Alvarez-Fernandez, et al. (2010). "Structure of the FoxM1 DNA-recognition domain bound to a promoter sequence." Nucleic Acids Res **38**(13): 4527-4538.
Liu, L., I.-J. L. Byeon, et al. (2012). "Domain Swapping Proceeds via Complete Unfolding: A 19F- and 1H-NMR Study of the Cyanovirin-N Protein." J Am Chem Soc **134**(9): 4229-4235.
Liu, L., I.-J. L. Byeon, et al. (2012). "Domain swapping proceeds via complete unfolding: a 19F-and 1H-NMR study of the Cyanovirin-N protein." Journal of the American Chemical Society **134**(9): 4229-4235.
Liu, Y. and D. Eisenberg (2002). "3D domain swapping: as domains continue to swap." Protein Sci **11**(6): 1285-1299.
Liu, Y. and D. Eisenberg (2002). "3D domain swapping: as domains continue to swap." Protein science **11**(6): 1285-1299.
Liu, Y., G. Gotte, et al. (2001). "A domain-swapped RNase A dimer with implications for amyloid formation." Nat Struct Biol **8**(3): 211-214.
Liu, Y., G. Gotte, et al. (2002). "Structures of the two 3D domain-swapped RNase A trimers." Protein Sci **11**(2): 371-380.
Liu, Y., P. J. Hart, et al. (1998). "The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-A resolution." Proc Natl Acad Sci U S A **95**(7): 3437-3442.
Lopes, J. E., T. R. Torgerson, et al. (2006). "Analysis of FOXP3 reveals multiple domains required for its function as a transcriptional repressor." The Journal of Immunology **177**(5): 3133-3142.
Lopes, J. E., T. R. Torgerson, et al. (2006). "Analysis of FOXP3 reveals multiple domains required for its function as a transcriptional repressor." J Immunol **177**(5): 3133-3142.
Lovejoy, B., S. Choe, et al. (1993). "Crystal structure of a synthetic triple-stranded alpha-helical bundle." Science **259**(5099): 1288-1293.
Lu, M. M., S. Li, et al. (2002). "Foxp4: a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues." Mech Dev **119 Suppl 1**: S197-202.
Luisi, B., W. Xu, et al. (1991). "Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA."
Luisi, B. F., W. X. Xu, et al. (1991). "Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA." Nature **352**(6335): 497-505.

Marianayagam, N. J., M. Sunde, et al. (2004). "The power of two: protein dimerization in biology." Trends Biochem Sci **29**(11): 618-625.
Marquardt, D. W. (1963). "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." Journal of the Society for Industrial and Applied Mathematics **11**(2): 431-441.
Massari, M. E. and C. Murre (2000). "Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms." Mol Cell Biol **20**(2): 429-440.
Mazet, F., J. K. Yu, et al. (2003). "Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria." Gene **316**: 79-89.
Mecklenburg, L., M. Nakamura, et al. (2001). "The Nude Mouse Skin Phenotype: The Role of< i> Foxn1</i> in Hair Follicle Development and Cycling." Experimental and molecular pathology **71**(2): 171-178.
Medina, E., C. Cordova, et al. (2016). "Three-Dimensional Domain Swapping Changes the Folding Mechanism of the Forkhead Domain of FoxP1." Biophys J **110**(11): 2349-2360.
Medina, E., J. Reyes, et al. (2017). "Biophysical and Evolutionary Aspects of Domain Swapping in the Forkhead Domain of Human FoxP Proteins." Biophys J **112**(3): 168a.
Miller, K. H., J. R. Karr, et al. (2010). "A Hinge Region cis-Proline in Ribonuclease A Acts as a Conformational Gatekeeper for C-Terminal Domain Swapping." Journal of molecular biology **400**(3): 567-578.
Moreira, I. S., P. A. Fernandes, et al. (2007). "Hot spots--a review of the protein-protein interface determinant amino-acid residues." Proteins **68**(4): 803-812.
Morris, G. and S. Fanucchi (2016). "A Key Evolutionary Mutation Enhances DNA Binding of the FOXP2 Forkhead Domain." Biochemistry **55**(13): 1959-1967.
Murawska, M., A. Szymańska, et al. (2017). "Overall conformation of covalently stabilized domain-swapped dimer of human cystatin C in solution." Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms **411**: 136-140.
Murray, A. J., S. J. Lewis, et al. (1995). "One sequence, two folds: a metastable structure of CD2." Proceedings of the National Academy of Sciences **92**(16): 7337-7341.
Murrell, J., M. Farlow, et al. (1991). "A mutation in the amyloid precursor protein associated with hereditary Alzheimer's disease." Science **254**(5028): 97-99.
Myers, J. K., C. Nick Pace, et al. (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." Protein science **4**(10): 2138-2148.
Myers, J. K., C. N. Pace, et al. (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." Protein Sci **4**(10): 2138-2148.
Nagarkar, R. P., R. A. Hule, et al. (2010). "Domain swapping in materials design." Peptide Science: Original Research on Biomolecules **94**(1): 141-155.
Nandwani, N., P. Surana, et al. (2017). "Amino-acid composition after loop deletion drives domain swapping." Protein Science **26**(10): 1994-2002.
Neet, K. E. and D. E. Timm (1994). "Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation." Protein Sci **3**(12): 2167-2174.
Nelson, C. S., C. K. Fuller, et al. (2013). "Microfluidic affinity and ChIP-seq analyses converge on a conserved FOXP2-binding motif in chimp and human, which enables the detection of evolutionarily novel targets." Nucleic Acids Res **41**(12): 5991-6004.
Newcomer, M. E. (2002). "Protein folding and three-dimensional domain swapping: a strained relationship?" Curr Opin Struct Biol **12**(1): 48-53.
Nilsson, M., X. Wang, et al. (2004). "Prevention of Domain Swapping Inhibits Dimerization and Amyloid Fibril Formation of Cystatin C USE OF ENGINEERED DISULFIDE BRIDGES, ANTIBODIES, AND CARBOXYMETHYLPAPAIN TO STABILIZE THE MONOMERIC FORM OF CYSTATIN C." Journal of Biological Chemistry **279**(23): 24236-24245.
O'Neill, J. W., D. E. Kim, et al. (2001). "Single-Site Mutations Induce 3D Domain Swapping in the B1 Domain of Protein L from< i> Peptostreptococcus magnus</i>." Structure **9**(11): 1017-1027.
Ogihara, N. L., G. Ghirlanda, et al. (2001). "Design of three-dimensional domain-swapped dimers and fibrous oligomers." Proc Natl Acad Sci U S A **98**(4): 1404-1409.

Onuchic, J. N., H. Nymeyer, et al. (1999). "The energy landscape theory of protein folding: insights into folding mechanisms and scenarios." Adv Protein Chem **53**: 87-152.
Pabo, C. a. S., R. (1992). "Transcription factors: structural families and principles of DNA recognition." Annual Review of Biochemistry **61**: 1053-1095.
Pabo, C. O. and R. T. Sauer (1984). "Protein-DNA recognition." Annu Rev Biochem **53**: 293-321.
Pace, C. (1986). "[14] Determination and analysis of urea and guanidine hydrochloride denaturation curves." Methods in enzymology **131**: 266-280.
Pace, C. N. (1975). "The stability of globular proteins." CRC Crit Rev Biochem **3**(1): 1-43.
Pace, C. N. and J. M. Scholtz (1997). "Measuring the conformational stability of a protein." Protein structure: A practical approach **2**: 299-321.
Pace, C. N., B. A. Shirley, et al. (1996). "Forces contributing to the conformational stability of proteins." FASEB J **10**(1): 75-83.
Park, C. and R. T. Raines (2000). "Dimer formation by a "monomeric" protein." Protein Sci **9**(10): 2026-2033.
Park, C. K., H. K. Joshi, et al. (2010). "Domain swapping in allosteric modulation of DNA specificity." PLoS Biol **8**(12): e1000554.
Park, S. H., M. C. Shastry, et al. (1999). "Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing." Nat Struct Biol **6**(10): 943-947.
Partridge, L. and J. C. Bruning (2008). "Forkhead transcription factors and ageing." Oncogene **27**(16): 2351-2363.
Perisic, O., P. A. Webb, et al. (1994). "Crystal structure of a diabody, a bivalent antibody fragment." Structure **2**(12): 1217-1226.
Perrett, S., S. J. Freeman, et al. (1999). "Equilibrium folding properties of the yeast prion protein determinant Ure2." J Mol Biol **290**(1): 331-345.
Perumal, K. (2014). "The role of Tyr540 in dimerisation of the FOXP forkhead domain." University of The Witwatersrand.
Perumal, K., H. W. Dirr, et al. (2015). "A Single Amino Acid in the Hinge Loop Region of the FOXP Forkhead Domain is Significant for Dimerisation." Protein J **34**(2): 111-121.
Pingoud, A. and A. Jeltsch (2001). "Structure and function of type II restriction endonucleases." Nucleic Acids Res **29**(18): 3705-3727.
Porath, J., J. Carlsson, et al. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation." Nature **258**: 598-599.
Potschka, M. (1987). "Universal calibration of gel permeation chromatography and determination of molecular shape in solution." Analytical biochemistry **162**(1): 47-64.
Preneta, A. (1989). "Separation on the basis of size: gel permeation chromatography." Protein Purification Methods, a practical approach: 293-306.
Rahaman, H., M. K. A. Khan, et al. (2015). "Heterogeneity of equilibrium molten globule state of cytochrome c induced by weak salt denaturants under physiological condition." PLoS One **10**(4): e0120465.
Rausa, F. M., Y. Tan, et al. (2003). "Association between hepatocyte nuclear factor 6 (HNF-6) and FoxA2 DNA binding domains stimulates FoxA2 transcriptional activity but inhibits HNF-6 DNA binding." Mol Cell Biol **23**(2): 437-449.
Reisner, A. H., P. Nemes, et al. (1975). "The use of Coomassie Brilliant Blue G250 perchloric acid solution for staining in electrophoresis and isoelectric focusing on polyacrylamide gels." Analytical biochemistry **64**(2): 509-516.
Remenyi, A., A. Tomilin, et al. (2001). "Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping." Mol Cell **8**(3): 569-580.
Reményi, A., A. Tomilin, et al. (2001). "Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping." Molecular cell **8**(3): 569-580.
Rivas, G., J. A. Fernandez, et al. (1999). "Direct observation of the self-association of dilute proteins in the presence of inert macromolecules at high concentration via tracer sedimentation equilibrium: theory, experiment, and biological significance." Biochemistry **38**(29): 9379-9388.
Robertson, A. D. (2002). "Intramolecular interactions at protein surfaces and their impact on protein function." Trends Biochem Sci **27**(10): 521-526.

Roder, H. and W. Colón (1997). "Kinetic role of early intermediates in protein folding." Curr Opin Struct Biol **7**(1): 15-28.
Rousseau, F., J. Schymkowitz, et al. (1998). "Stability and folding of the cell cycle regulatory protein, p13 suc1." Journal of molecular biology **284**(2): 503-519.
Rousseau, F., J. Schymkowitz, et al. (2001). "Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues." Proceedings of the National Academy of Sciences **98**(10): 5596-5601.
Rousseau, F., J. W. Schymkowitz, et al. (2003). "The unfolding story of three-dimensional domain swapping." Structure **11**(3): 243-251.
Rousseau, F., J. W. Schymkowitz, et al. (1998). "Stability and folding of the cell cycle regulatory protein, p13(suc1)." J Mol Biol **284**(2): 503-519.
Rousseau, F., J. W. Schymkowitz, et al. (2001). "Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues." Proc Natl Acad Sci U S A **98**(10): 5596-5601.
Rousseau, F., J. W. Schymkowitz, et al. (2004). "Intermediates control domain swapping during folding of p13suc1." J Biol Chem **279**(9): 8368-8377.
Sakaguchi, S. (2005). "Naturally arising Foxp3-expressing CD25+ CD4+ regulatory T cells in immunological tolerance to self and non-self." Nature immunology **6**(4): 345-352.
Saleem, R. A., S. Banerjee-Basu, et al. (2003). "Structural and functional analyses of disease-causing missense mutations in the forkhead domain of FOXC1." Hum Mol Genet **12**(22): 2993-3005.
Saleem, R. A., S. Banerjee-Basu, et al. (2004). "Essential structural and functional determinants within the forkhead domain of FOXC1." Nucleic Acids Res **32**(14): 4182-4193.
Scherzer, C. R., J. A. Grass, et al. (2008). "GATA transcription factors directly regulate the Parkinson's disease-linked gene alpha-synuclein." Proc Natl Acad Sci U S A **105**(31): 10907-10912.
Schiene, C. and G. Fischer (2000). "Enzymes that catalyse the restructuring of proteins." Curr Opin Struct Biol **10**(1): 40-45.
Schleif, R. (1988). "DNA binding by proteins." Science **241**(4870): 1182-1187.
Schleif, R. F. (2013). "Modulation of DNA Binding by Gene-Specific Transcription Factors." Biochemistry.
Schlunegger, M. P., M. J. Bennett, et al. (1997). "Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly." Adv Protein Chem **50**: 61-122.
Schymkowitz, J. W., F. Rousseau, et al. (2001). "Observation of signal transduction in three-dimensional domain swapping." Nature Structural & Molecular Biology **8**(10): 888-892.
Semisotnov, G. V., N. A. Rodionova, et al. (1991). "Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe." Biopolymers **31**(1): 119-128.
Sheinerman, F. B., R. Norel, et al. (2000). "Electrostatic aspects of protein-protein interactions." Curr Opin Struct Biol **10**(2): 153-159.
Sicheri, F., I. Moarefi, et al. (1997). "Crystal structure of the Src family tyrosine kinase Hck." Nature **385**(6617): 602-609.
Sirota, F. L., S. Héry-Huynh, et al. (2008). "Role of the amino acid sequence in domain swapping of the B1 domain of protein G." Proteins: Structure, Function, and Bioinformatics **72**(1): 88-104.
Sonawane, A. R., J. Platig, et al. (2017). "Understanding Tissue-Specific Gene Regulation." Cell Rep **21**(4): 1077-1088.
Song, X., B. Li, et al. (2012). "Structural and biological features of FOXP3 dimerization relevant to regulatory T cell function." Cell reports **1**(6): 665-675.
Spaniel, F., J. Horacek, et al. (2011). "Genetic variation in FOXP2 alters grey matter concentrations in schizophrenia patients." Neurosci Lett **493**(3): 131-135.
Spit, A., R. H. Hyland, et al. (1998). "A role for heterodimerization in nuclear localization of a homeodomain protein." Proc Natl Acad Sci U S A **95**(11): 6228-6233.
Staniforth, R. A., S. Giannini, et al. (2001). "Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily." EMBO J **20**(17): 4774-4781.

Stevens, A. and R. C. Augusteyn (1997). "Binding of 1-anilinonaphthalene-8-sulfonic acid to alpha-crystallin." Eur J Biochem **243**(3): 792-797.
Strickler, S. S., A. V. Gribenko, et al. (2006). "Protein stability and surface electrostatics: a charged relationship." Biochemistry **45**(9): 2761-2766.
Stroud, J. C., Y. Wu, et al. (2006). "Structure of the forkhead domain of FOXP2 bound to DNA." Structure **14**(1): 159-166.
Su, D. M., Navarre, S., Oh, W.J., Condie, B.G. and Manley, N.R. (2003). "A domain of Foxn1 required for corsstalk-dependent thymic epithelial cell differentiation." Nat Immunol **4**: 1128-1135.
Sykiotis, G. P. and D. Bohmann (2010). "Stress-activated cap'n'collar transcription factors in aging and human disease." Sci Signal **3**(112): re3.
Tan, S. and M. Pepys (1994). "Histopathology." Amyloidosis **25**: 403-414.
Tanford, C. (1968). "Protein denaturation." Adv Protein Chem **23**: 121-282.
Tayyab, S., S. Qamar, et al. (1991). "Size exclusion chromatography and size exclusion HPLC of proteins." Biochemistry and Molecular Biology Education **19**(3): 149-152.
Thompson, J. D., T. J. Gibson, et al. (2002). "Multiple sequence alignment using ClustalW and ClustalX." Curr Protoc Bioinformatics **Chapter 2**: Unit 2 3.
Tisi, L. C. and P. A. Evans (1995). "Conserved structural features on protein surfaces: small exterior hydrophobic clusters." J Mol Biol **249**(2): 251-258.
Tolosa, A., J. Sanjuan, et al. (2010). "FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies." BMC Med Genet **11**: 114.
Topping, T. B., D. A. Hoch, et al. (2004). "Folding mechanism of FIS, the intertwined, dimeric factor for inversion stimulation." J Mol Biol **335**(4): 1065-1081.
Tsai, K. L., Huang, C.Y., Chang, C.H., Sun, Y.J., Chuang, W.J. and Hsiao,. C.D. (2006). "Crystal structure of the human FOXK1a-DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins." J Biol Chem **281**(25): 17400-17409.
Tsai, K. L., Y. J. Sun, et al. (2007). "Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification." Nucleic Acids Res **35**(20): 6984-6994.
Valdar, W. S. and J. M. Thornton (2001). "Protein–protein interfaces: analysis of amino acid conservation in homodimers." Proteins: Structure, Function, and Bioinformatics **42**(1): 108-124.
Van den Burg, B., B. W. Dijkstra, et al. (1994). "Protein stabilization by hydrophobic interactions at the surface." Eur J Biochem **220**(3): 981-985.
von Hippel, P. H. (1994). "Protein-DNA recognition: new perspectives and underlying themes." Science **263**(5148): 769-770.
Walters, J., S. L. Milam, et al. (2009). "Practical approaches to protein folding and assembly: spectroscopic strategies in thermodynamics and kinetics." Methods in enzymology **455**: 1-39.
Walters, J., S. L. Milam, et al. (2009). "Practical approaches to protein folding and assembly: spectroscopic strategies in thermodynamics and kinetics." Methods Enzymol **455**: 1-39.
Wang, B., D. Lin, et al. (2003). "Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors." J Biol Chem **278**(27): 24259-24268.
Warren, J. R. and J. A. Gordon (1966). "On the refractive indices of aqueous solutions of urea." The Journal of Physical Chemistry **70**(1): 297-300.
Webb, B. and A. Sali (2016). "Comparative Protein Structure Modeling Using MODELLER." Curr Protoc Protein Sci **86**: 2 9 1-2 9 37.
Webb, H., O. Steeb, et al. (2017). "The FOXP2 forkhead domain binds to a variety of DNA sequences with different rates and affinities." J Biochem **162**(1): 45-54.
Whitmore, L. and B. A. Wallace (2008). "Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases." Biopolymers: Original Research on Biomolecules **89**(5): 392-400.
Woody, R. W. (1994). "Circular dichroism." Encyclopedia of Molecular Biology.
Wu, C., T. Thalhamer, et al. (2014). "Galectin-9-CD44 interaction enhances stability and function of adaptive regulatory T cells." Immunity **41**(2): 270-282.

Yang, S., S. S. Cho, et al. (2004). "Domain swapping is a consequence of minimal frustration." Proc Natl Acad Sci U S A **101**(38): 13786-13791.
Young, L., R. L. Jernigan, et al. (1994). "A role for surface hydrophobicity in protein-protein recognition." Protein Sci **3**(5): 717-729.
Zegers, I., J. Deswarte, et al. (1999). "Trimeric domain-swapped barnase." Proceedings of the National Academy of Sciences **96**(3): 818-822.
Zhou, H., A. Cheruvanky, et al. (2008). "Urinary exosomal transcription factors, a new class of biomarkers for renal disease." Kidney Int **74**(5): 613-621.
Ziegler, S. F. (2006). "FOXP3: of mice and men." Annu. Rev. Immunol. **24**: 209-226.

)