

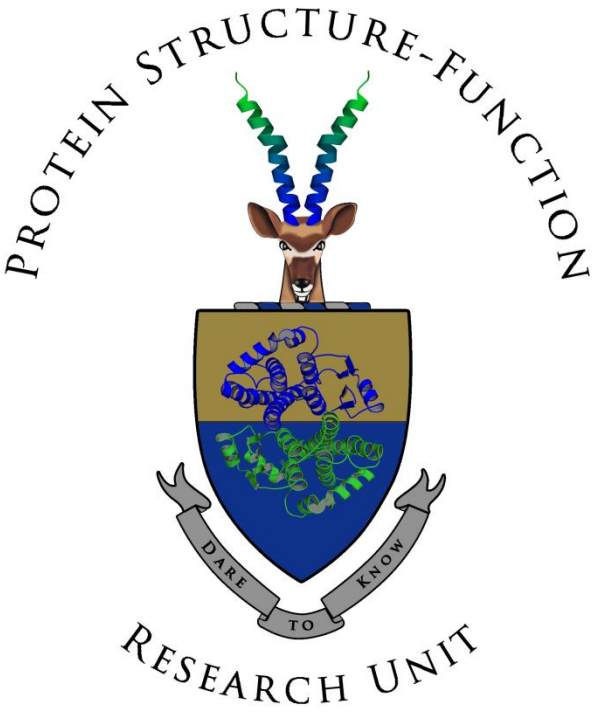
https://dspace.org/

The role of Tyr540 in dimerisation of the FOXP forkhead domain

Perumal, Kershia

2014-07-02
http://hdl.handle.net10539/14851

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



The role of Tyr540 in dimerisation of the FOXP forkhead domain

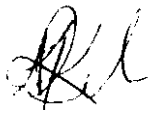
Kershia Perumal

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science.

Johannesburg, 2013

Declaration

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



Kershia Perumal

19 March 2013

Research Outputs

2012 - 2013

Conference presentations

Name	Date	Location	Title	Authors
Structural Biology in the Bioeconomy	2 - 4 Dec 2012	University of Cape Town, South Africa	Effect of an IPEX mutation on domain swapping and DNA binding in the FOXP forkhead domain	K. Perumal S. Fanucchi
Molecular Biosciences Research Thrust	5 Dec 2012	University of the Witwatersrand, South Africa	Effect of an IPEX mutation on domain swapping and DNA binding in the FOXP forkhead domain	K. Perumal S. Fanucchi
5th Cross-Faculty Postgraduate Symposium	1 - 2 Aug 2013	University of the Witwatersrand, South Africa	Effect of an IPEX mutation on domain swapping in the FOXP subfamily of transcription factors	K. Perumal S. Fanucchi
Molecular Biosciences Research Thrust	4 Dec 2013	University of the Witwatersrand, South Africa	Domain swapping in the FOXP family: A mechanism for regulating DNA binding?	K. Perumal S. Fanucchi

This work is dedicated to my parents, Shirley and Irvin, Ma and the love of my life, Rowan Gengan for their unwavering support and love.

The true sign of intelligence is not knowledge but imagination.

Albert Einstein

Abstract

The forkhead box (FOX) proteins are a family of transcription factors that interact with DNA via a winged helix motif that forms part of the forkhead domain. The FOXP (FOXP1-4) subfamily is unique in the family in that the forkhead domains of these proteins exhibit domain swapping where structural elements are exchanged via extension of the hinge-loop region. The FOXP subfamily members have high sequence homology, yet wild-type FOXP3 is a stable domain-swapped dimer in solution whereas FOXP1 and FOXP2 exist in a monomer/domain-swapped dimer equilibrium. A single amino acid difference is observed in the hinge region of the FOXP subfamily. This corresponds to Tyr540 in FOXP2 and Phe373 in FOXP3. We propose that it is the phenylalanine residue in FOXP3 that shifts the equilibrium towards dimer. Here we use FOXP2 to investigate the effect of a mutation, Y540F, on the structure and dimerisation propensity of the FOXP subfamily. Crystals for the Y540F variant in the presence of DNA have been obtained to demonstrate conclusively that domain swapping occurs. Size-exclusion chromatography indicates that the wild type FOXP2 forkhead domain is almost entirely monomeric at concentrations less than 100 μM. The Y540F variant is shown to stabilise the dimer and the ratio between monomer and dimer is concentration-dependent. DNA binding assays suggest that the Y540F variant binds less favourably to the cognate binding sequence than does the WT FOXP2 forkhead domain. Taken together, these findings suggest that domain swapping may modulate DNA binding.

Acknowledgments

Very special thanks to my supervisor, Dr. Sylvia Fanucchi for being a great mentor and for supporting and encouraging me 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 all my colleagues at the PSFRU for the stimulating discussions, assistance and friendship.

I would like to acknowledge and thank Dr. Stoyan Stoychev (Council for Scientific and Industrial Research, Pretoria, South Africa) for his assistance with the mass spectrometry work.

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 x

List of Figures..... xiii

List of Tables..... xv

Chapter 1. Introduction..... 1

1.1 Transcription factors..... 1

1.1.1 Stability of protein-protein interactions..... 2

1.1.2 Protein oligomerisation..... 3

1.1.3 Domain swapping..... 4

1.1.4 Physiological significance of domain swapping..... 6

1.1.5 Principles of DNA recognition 7

1.1.6 Regulation of DNA binding..... 8

1.1.7 Winged helix domain..... 9

1.2 Forkhead box family of transcription factors..... 11

1.3 FOXP subfamily..... 11

1.3.1 Structural similarities..... 11

1.3.2 FOXP3..... 12

1.3.3 FOXP2..... 13

1.4 The FOXP forkhead domain..... 13

1.4.1 DNA binding..... 17

1.4.2 Structure and stability of the FOXP domain-swapped dimers..... 17

1.5 Aim 22

Chapter 2. Materials and Methods 23

2.1 Materials..... 23

2.2 Methods..... 24

2.2.1 Plasmid purification..... 24

2.2.2 Oligonucleotide primer design..... 25

2.2.3 Site-directed mutagenesis..... 25

2.2.4 Plasmid identification and transformation..... 26

2.2.5 Protein production and purification..... .27

2.2.5.1 Induction studies..... 27

2.2.5.2 Protein production..... 28

2.2.5.3 Immobilised metal-ion affinity chromatography (IMAC)..... 28

2.2.6 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)..... 29

2.2.7 Protein concentration determination..... 30

2.2.8 Cleavage of tagged fusion protein..... 31

2.2.9 Tricine sodium dodecyl sulphate-polyacrylamide gel electrophoresis (Tricine SDS-PAGE)..... 33

2.2.10 Structural integrity of protein..... 34

2.2.10.1 Matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS)..... 34

2.2.10.2 Blue native-polyacrylamide gel electrophoresis (BN-PAGE)..... 35

2.2.10.3Liquid chromatography-mass spectrometry/mass spectrometry (LC- MS/MS)..... 37

2.2.10.4 Circular dichroism spectroscopy..... 38

2.2.10.5 Fluorescence spectroscopy..... 39

2.2.11 DNA binding..... 39

2.2.11.1 Electrophoretic mobility shift assay (EMSA)..... 40

2.2.11.2 Intrinsic fluorescence..... 41

2.2.12 Size-exclusion chromatography..... 41

2.2.13 Crystallisation of FOXP2 Y540F variant..... 42

Chapter 3. Results..... 44

3.1 Creation of mutant and purification..... 44

3.2 Structural characterisation of the wild-type FOXP2 FHD and Y540F variant... 53

3.2.1 Secondary structure..... 53

3.2.2Tertiary structure..... 54

3.2.3 Quaternary structure..... 56

3.2.4 Mass spectrometry..... 57

3.3 DNA binding..... 62

3.4 Propensity of the FOXP2 FHD to dimerise..... 68

3.5 Tertiary structures of the separated monomer and dimer..... 70

3.6 Crystals of the FOXP2 Y540F FHD in the presence of DNA..... 72

Chapter 4. Discussion..... 74

4.1 Phe540 does not significantly alter DNA binding but promotes dimerisation..... 74

4.2 Physiological role of the FOXP domain-swapped dimer..... 77

4.3 Conclusion..... 79

Chapter 5. References..... 80

List of Abbreviations

θ	Ellipticity
3-HPA	3-dihydroxypicolinic acid matrix
A_{280}	Absorbance at 280 nm
Å	Ångstroms
ACN	Acetonitrile
ÄKTA	Protein purification system
ASA	Accessible surface area
BN-PAGE	Blue native-polyacrylamide gel electrophoresis
BTP	Bis-tris-propane
bZIP	Basic leucine zipper domain
CD	Circular dichroism
cDNA	Complementary DNA
CNE	Clear native-polyacrylamide gel electrophoresis
Dam	DNA adenine methylase
DHAP	2, 5-dihydroxyacetophone matrix
DNA	Deoxyribonucleic acid
DSD	Domain-swapped dimer
ds-DNA	Double-stranded DNA
DTT	Dithiothreitol
ε	Molar extinction coefficient
EDTA	Ethylenediaminetetraacetic acid

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
hrCNE	High-resolution clear native-polyacrylamide gel electrophoresis
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
KE family	British family with a speech and language disorder
LB	Luria Bertani
LC-MS/MS	Liquid chromatography-mass spectrometry/mass spectrometry
lcpl	Left-circularly polarised light
LIC	Ligation-independent cloning
MALDI-TOF	Matrix-assisted laser desorption/ionisation-time of flight
MPD	2-methyl-2, 4-pentanediol
MQH_2O	Deionised and sterile Milli-Q water

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
p*I*	Isoelectric point
rcpl	Right-circularly polarised light
SA	Sinapinic acid matrix
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SgrAI	Type of restriction endonuclease
SEC	Size exclusion chromatography
SOC	Super optimal broth with catabolite repression
TFA	Trifluoroacetic acid
UV	Ultraviolet
WT	Wild-type

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

List of Figures Page number

Figure 1. Schematic of domain swapping 5

Figure 2. Schematic of the helix-turn-helix motif and the ‘winged’ helix motif 10

Figure 3. Schematic of a typical FOXP protein 12

Figure 4. Sequence alignment of the FHD in FOX family members 15

Figure 5. Structure of the wild-type FOXP2 FHD monomer bound to DNA 16

Figure 6. Structural alignment of the FOXP2 FHD monomer with the FOXP2 FHD DSD 19

Figure 7. Structural differences between the FOXP3 and FOXP2 domain-swapped dimers 19

Figure 8. The network of hydrophobic interactions that stabilises the FOXP DSD 20

Figure 9. The His-tagged FOXP2 FHD fusion protein 32

Figure 10. Representation of a protein crystallisation phase diagram 43

Figure 11. Confirmation of wild-type (WT) FOXP2 FHD and verification of the Y540F variant sequences 44

Figure 12. Induction trials for optimal protein production 45

Figure 13. Purification of the WT FOX2 FHD 48

Figure 14. Purification of the Y540F FHD variant 49

Figure 15. Estimation of the molecular weight of FOXP2 FHD 50

Figure 16. Cleavage of the Y540F variant fusion protein 52

Figure 17. Secondary structure of WT FOXP2 FHD and the Y540F variant 53

Figure 18. Tertiary structure of the WT FOXP2 FHD and the Y540F variant 55

Figure 19. BN-PAGE of the WT FOXP2 monomer-dimer mixture 57

Figure 20. LC/MS-MS data from in-gel trypsin digest fragment of BN-PAGE gel 58

Figure 21. MALDI-TOF of protein standard mix 60

Figure 22. MALDI-TOF analysis 61

Figure 23.Tryptophan fluorescence emission spectra of free FOXP2 and FOXP2/DNA complexes 64

Figure 24. Location of the Trp residues in the FOXP2 FHD 65

Figure 25. Electrophoretic mobility shift assay (EMSA) to determine optimal protein/DNA ratio 66

Figure 26. EMSA showing titration of cognate DNA with WT FOXP2 FHD and Y540F variant proteins 67

Figure 27. Molecular weight calculation using size-exclusion chromatography 68

Figure 28. Propensity of the FOXP2 FHD to dimerise 69

Figure 29. Fluorescence emission spectra of the purified monomer and dimer species of FOXP2 FHD 71

Figure 30. Crystals of the FOXP Y540F variant in the presence of DNA 73

Figure 31. Interaction of the FOXP2 FHD with DNA 75

Figure 32. Interaction of the WT FOXP2 FHD dimer and monomer with DNA 77

List of Tables **Page number**

Table 1. Non-standard materials and suppliers 23

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' **Error! Bookmark not defined.**. Transcription factors have crucial roles 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 structure usually contains specific domains that perform specific functions. These domains are typically dimerisation, protein-protein interaction, or DNA binding domains. In order to function, transcription factors must bind to nearby or distant DNA sequences (Ptashne and Gann 2002) which results in the downstream activation or repression of target genes (Latchman 1997).

Activation of target genes by transcription occurs through direct or indirect interaction of the transcription factor with the basal transcription complex whereas repression occurs by one of four mechanisms. Repressors may block the activity of the activating transcription factor by binding directly to the activator's binding site, preventing the activator from binding or they may bind directly to the activator and thereby prevent it from binding to DNA; they may interact with the activator-DNA complex, thus blocking its activity (known as quenching) or they may by bind directly to DNA in the absence of an activator (Latchman 1997).

A careful balance must be obtained between activated and repressed target genes which is achieved by the regulation of transcription factors (Latchman 1997). All cells in a multicellular organism contain the same genes but the transcription factors are often only expressed at high levels in particular tissues (tissue-specific). 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 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). For some transcription factors, dimerisation is required prior to binding DNA and 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). Taken together, protein-protein interactions can be considered to be just as important in the gene regulation networks as protein-DNA interactions.

1.1.1 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 highly conserved (Valdar and Thornton 2001; Moreira *et al.* 2007) and are usually 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 (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 1200 - 2000 Å^2 (Moreira *et al.* 2007) and a minimum of 600 Å^2 of buried surface area is necessary to form a stable protein-protein complex (Bogan and Thorn 1998; Moreira *et al.* 2007).

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) in the form of electrostatic guides (Janin 1997). 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 are found close together at the interface within 4 Å of each other (Barlow and Thornton 1983). Destabilisation of protein complexes can occur when an unfulfilled hydrogen bond donor or acceptor is buried in the interface (Xu *et al.* 1997) or by a minor change in the charge network (Chacko *et al.* 1995). Electrostatic interactions are thus essential in the binding specificity and stability of protein-protein interactions and may determine which proteins interact with each other (Fersht 1987).

Protein-protein interfaces are found to contain pockets of highly structurally conserved residues such as tryptophan, tyrosine and arginine that are termed 'hot spots' (Young *et al.* 1994; Moreira *et al.* 2007). Arginine is frequently observed near aromatic residues and typically forms hydrogen bonds at the protein surface (Glaser *et al.* 2001). Tryptophan is a relatively bulky residue that participates mainly in aromatic-pi interactions (Samanta *et al.* 2000) whereas tyrosine has a hydrophobic surface along with a hydroxyl group which allows it to participate in aromatic-pi interactions as well as in hydrogen bonding (Bogan and Thorn 1998). Mutations of residues at these interacting interfaces have been shown to cause disease (Seeliger *et al.* 1999; Wang and Moult 2001) thus highlighting the functional significance of the interaction. Understanding the stability and dynamics of protein-protein interactions will thus provide a better understanding of the link between structure and function.

1.1.2 Protein oligomerisation

Proteins can oligomerise to form dimers, trimers, or higher order multimers. These oligomers may be symmetrical (Marianayagam *et al.* 2004). Protein dimerisation, in particular, involves the association of two identical monomers in homodimerisation whereas heterodimerisation involves 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). Transcription factors frequently dimerise to bind to palindromic DNA sequences (Abel and Maniatis 1989). Dimerisation of transcription factors is facilitated by structural motifs such as the basic leucine zipper domain (bZIP) found within c-Fos and c-Jun transcription factors (Glover and Harrison 1995), the helix-loop-helix (HLH) motif (Jones 2004) and the helix-turn-helix (HTH) motif found in the mammalian telomeric protein TRF1 (Bianchi *et al.* 1997). This dimerisation event may be necessary for their DNA binding function. An example of this is found within transcription factors containing a HTH motif. Each monomer recognises half of the DNA binding sequence, termed a half-site (Pabo and Sauer 1984). Therefore formation of the symmetrical dimer is required to bind the full-length DNA sequence.

1.1.3 Domain swapping

Domain swapping is a type of oligomerisation event in which specific regions of proteins are exchanged to form an intertwined dimer (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). The structures of the monomer and the subunits of the domain-swapped dimer (DSD) are identical except for the hinge-loop region connecting the two subunits (Ogihara *et al.* 2001; Rousseau *et al.* 2003) (Figure 1). 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).

There is no clear pattern in terms of sequence or structure as to whether or not a protein will domain swap, making it difficult to predict domain swapping based on sequence alone (Liu and Eisenberg 2002). A recent study showed that domain swapping is quite frequent and occurs in many different proteins with different functions (Huang *et al.* 2012). Furthermore, proteins within a family can also domain swap to different extents, in terms of the size of the region that is swapped. The ability to domain swap relies mainly on the conformational flexibility of the hinge loop region (Liu and Eisenberg 2002).

Much focus on the mechanism behind domain swapping has been on the hinge loop region. Hinge-loop length, either by shortening or lengthening, has been shown to promote domain swapping in some cases (Ogihara *et al.* 2001).In addition, specific amino acid residues found in this region have been shown to play important roles in stabilising the domain-swapped dimer, particularly hydrophobic residues (Byeon *et al.* 2003; Byeon *et al.* 2004).

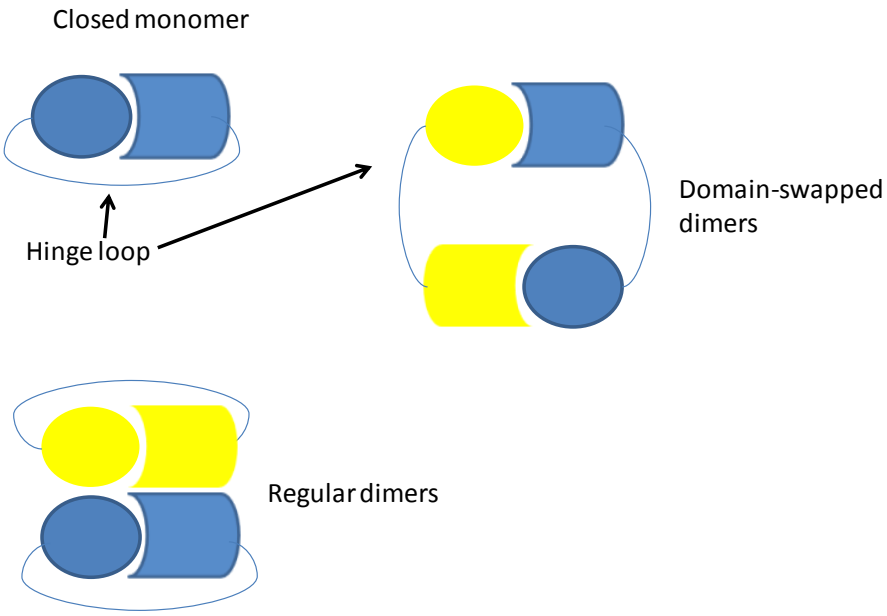


Figure 1. Schematic of domain swapping. The subunits in the domain-swapped dimer are usually identical. The hinge loop conformations are different in the closed monomer and open monomers in the domain-swapped dimer (Bennett *et al.* 1995). The hinge-loop conformation does not change between closed monomers and regular dimers.

In general, 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. Rousseau *et al.* (2001) showed that in the case of p13suc1, a cell cycle regulatory protein, 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 free energy change in the monomer-dimer transition (Heringa and Taylor 1997; Robertson 2002). The activation energy barrier for monomer-dimer inter-conversion may be large which can result in a slow re-equilibration rate following separation of monomer and dimer species by size-exclusion chromatography (Bennett *et al.* 1995).

It has been proposed that the following characteristics can affect three-dimensional domain swapping: macromolecular crowding (Lindner and Ralston 1995; Rivas *et al.* 1999), protein concentration (Yang

et al. 2004), mutations (including insertions and deletions) in the hinge region (Ercole *et al.* 2003; Picone *et al.* 2005), binding of a ligand (Schymkowitz *et al.* 2001) as well as changes in pH (Carroll *et al.* 1986) and temperature (Shameer *et al.* 2011). Although dimerisation can be shown by numerous means, that the dimer is domain-swapped can only be confirmed by solving the structure of the dimer.

1.1.4 Physiological significance of domain swapping

The physiological relevance of domain swapping is still unclear and some domain-swapped proteins may merely be artefacts of the high concentrations needed for crystallisation (Liu *et al.* 2002; Rousseau *et al.* 2003). However, domain swapping has been associated with 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 domain-swapped conformation is favoured. Thus regulation of the concentration of certain proteins can regulate when DSDs form which could be significant in regulating the activity of that particular protein. 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 (octamer 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.* 2010a). 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 domain-swapped tetramers that, upon binding to DNA, form active enzyme which cleaves at an even higher rate than the DSDs (Park *et al.* 2010a). Two members of the FOX (forkhead box) P-subfamily of transcription factors (FOXP2 and FOXP3) have been shown to form DSDs (Stroud *et al.* 2006; Bandukwala *et al.* 2011). The FOXP3 DSD has been shown to bind two distant strands of DNA and bring them together within close proximity which is believed to be essential for its suppressive function of regulatory T cells (Bandukwala *et al.* 2011). Since domain swapping is implicated in highly important 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, including DNA-binding proteins (Remenyi *et al.* 2001; Park *et al.* 2010a), it is implied that the DSD could have a unique physiological function, different to that of the monomer (Changeux and Edelstein 1998).

1.1.5 Principles of protein-DNA recognition

A prerequisite of gene regulation is the binding of transcription factors to a specific DNA sequence or to promoter regions via their DNA binding domains (Garvie *et al.* 2001). Protein-DNA interactions can be characterised according to their contribution to specificity, stability and binding affinity (Luscombe *et al.* 2001). Specificity of a protein-DNA interaction is due to the ability of the protein to distinguish its particular binding site from numerous other DNA sequences whereas the affinity arises from both the number of contacts made between the protein and DNA and the strength of these interactions. It was initially proposed that protein-DNA interactions could be governed by a one-to-one 'code' where specific amino acids would interact with particular DNA bases (Seeman *et al.* 1976), however this is not the case as indicated by extensive studies (Matthews 1988; Choo and Klug 1997; Pabo and Nekludova 2000; Luscombe *et al.* 2001; Benos *et al.* 2002). A simple recognition code is not practical because protein-DNA recognition is a highly complex process and is dependent upon the context of the interaction as specific amino acid residues may simultaneously interact with different DNA bases.

The three main interactions involved in the protein-DNA interface are hydrogen bonding, van der Waals interactions and water-mediated interactions. Van der Waals forces comprise the majority of these interactions, (Luscombe *et al.* 2001) and are mainly involved in stabilising the protein-DNA complex. Hydrogen bonds often involve specific interactions between amino acid side chains and the nitrogenous bases in DNA (Mirny 2002). The most common interactions are hydrogen bonds of amino acids lysine and arginine with the base adenine, amino acids asparagine and glutamine with the base guanine and to a lesser extent, amino acids glutamate and aspartate, with the base cytosine (Seeman *et al.* 1976; Mandel-Gutfreund *et al.* 1995; Luscombe *et al.* 2001). In addition to direct hydrogen bonds, water-mediated hydrogen bonds contribute significantly to the specificity of protein-DNA interactions (Luscombe *et al.* 2001) which indicates that water molecules are not just involved in solvating the macromolecules (Reddy *et al.* 2001; Jayaram and Jain 2004).

Protein-DNA recognition is complex in the sense that many different factors such as DNA shape and chemical complementarity (Rohs *et al.* 2009), geometry and position of the alpha carbons in the polypeptide backbone (Kono and Sarai 1999) as well as the promiscuity of side chain interactions with DNA nitrogenous bases all play a role in modulating DNA binding and dictate which interactions occur (Pabo and Nekludova 2000).

1.1.6 Modulation of DNA binding

Prokaryotic transcription factors bind to palindromic DNA sequences as homodimers whereas eukaryotic transcription factors bind to non-symmetrical DNA sequences as monomers, homodimers or heterodimers which allow recognition of a diverse range of sequences (Luscombe *et al.* 2000). DNA binding specificity or affinity can be modulated by a variety of factors including the DNA structure, DNA curving or bending (Olson *et al.* 1998), binding of co-activators, protein-protein dimerisation (Halazonetis *et al.* 1988) and post-translational modifications (Grönholm *et al.* 2012), to name a few.

DNA binding can be regulated by controlling access of the DNA binding domain to the DNA binding site. Numerous transcription factors are homooligomeric thus enabling them to bind to two or more DNA sequences simultaneously. Oligomerisation can control DNA binding affinity as the relative orientations of the domains will be different in the monomer compared to in the oligomer (Schleif 2013). In other words, binding affinity can be high if all DNA binding domains are orientated correctly whereas if they are not, the DNA binding domains cannot simultaneously contact the DNA binding sites and the affinity will be lower. Dimerisation, for example, may promote DNA binding as the affinity may be higher for a dimer than the affinity for either monomer. A study investigated the link between DNA binding affinity and homo- and hetero-dimerisation with c-Fos and c-Jun transcription factors. They found that c-Jun homodimers interacted specifically with AP 1 promoter sites with high affinity whereas the homologous c-Fos homodimers showed little to no affinity for the same site. Interestingly, the c-Fos-c-Jun heterodimer showed 25 times higher DNA binding affinity to the AP 1 binding site than the c-Jun homodimer. This indicates that different binding partners can modulate DNA binding affinity (Halazonetis *et al.* 1988).

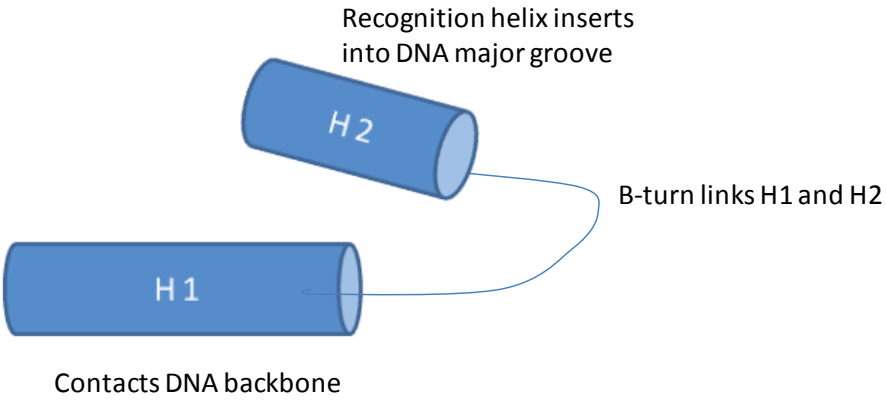
1.1.7 Winged helix motif

DNA-binding motifs are highly conserved and allow protein family members to interact with similar target sequences in a similar manner (Luscombe and Thornton 2002). The helix-turn-helix motif (HTH) was first discovered in the lambda-operator-repressor complex (Jordan and Pabo 1988). Unlike other motifs, the HTH motif always forms part of a larger, stable DNA binding domain (Pabo 1992). The HTH motif (Figure 2) is defined as a 20-amino acid region that consists of two almost perpendicular α-helices (H1 and H2) joined by a β-turn which can vary in length (Pabo and Sauer 1984; Luscombe *et al.* 2000). Upon binding to DNA the recognition helix (H2) is inserted into the major groove of the DNA double helix and supporting contacts are provided by the interaction of helix H1 with DNA. The position and orientation of the recognition helix with respect to the DNA and the

mechanism of interactions with the major groove varies between different transcription factors displaying the HTH motif.

A modified HTH motif was discovered in the DNA binding domain of the murine HNF-3 (hepatic nuclear factor-3) protein (Clark *et al.* 1993b) which shared a high degree of amino acid sequence similarity to the *Drosophila* forkhead gene and was therefore 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) (Figure 2) and is therefore known as the 'winged' helix motif. In the 'winged' helix motif, the β-turn in the classical HTH motif is replaced either by a loop/coil conformation (Tsai *et al.* 2007) or a 3_{10}-helix (Tsai *et al.* 2006). Strand S1 is situated directly after helix H1 and along with the other two strands, S2 and S3, forms antiparallel beta-sheets.

A



B

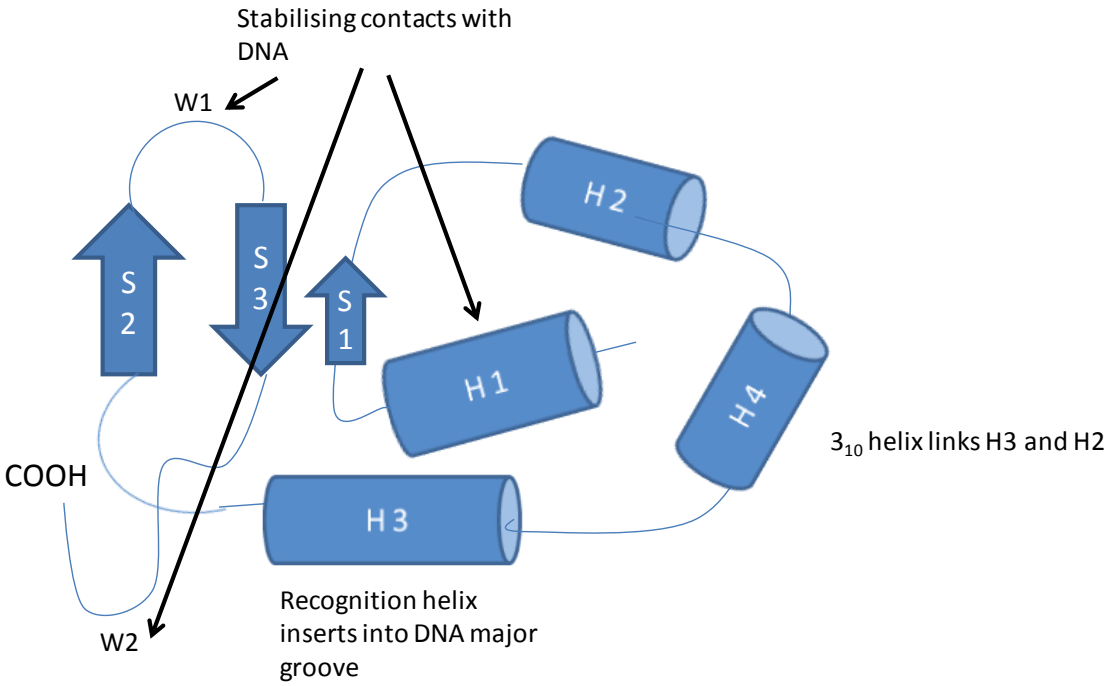


Figure 2. Schematic of the helix-turn-helix motif and the 'winged' helix motif. A. Helix-turn-helix (HTH) motif. The helix-turn-helix motif contains a recognition helix (H2) which inserts into the major groove of DNA and a helix (H1) which contacts the DNA backbone, joined by a β-turn. **B. The 'winged' helix motif** is a modified version of the HTH motif that contains an additional α-helix and three antiparallel β-strands (S1, S2 and S3). The recognition helix (H3) is connected to helix H2 by a 3_{10}-helix (H4) instead of the β-turn found in the canonical HTH motif. The wings (W1 and W2) as well as helix H1 are involved in stabilising DNA contacts.

1.2 Forkhead box 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). Even though the structure of the forkhead domain between these proteins is virtually unchanged, the FOX proteins have largely diverse yet important functions in the immune system (Su *et al.* 2003; Chae *et al.* 2006), metabolism (Friedman and Kaestner 2006), regulation of cellular proliferation and differentiation (Clark *et al.* 1993b; 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 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 are mainly monomeric (Li *et al.* 2004) and this is reflected in the solved crystal structures of FOXA3 (Clark *et al.* 1993a), FOXK1a (Tsai *et al.* 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 (Stroud *et al.* 2006; Bandukwala *et al.* 2011).

1.3 FOXP subfamily

1.3.1 Structural similarities

The FOXP subfamily consists of four members namely FOXP1, FOXP2, FOXP3 and FOXP4. All FOXP members have unique structural domains (Figure 3) including an amino-terminal polyglutamine stretch, C_2H_2 zinc finger domain and leucine zipper domain. They also contain the highly conserved forkhead domain (FHD) characteristic of the FOX superfamily (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

whereas in most other FOX proteins, the FHD is located near the N-terminus (Clark *et al.* 1993b; Carson *et al.* 2006).

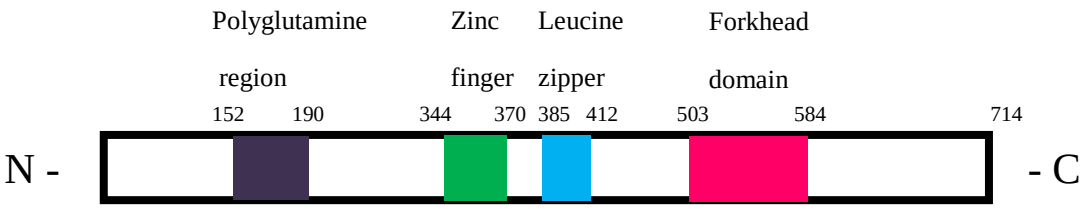


Figure 3. **Schematic of a typical FOXP protein.** Specialised motifs are highlighted. FOXP2 numbering is used. The polyglutamine region (dark purple) is thought to be responsible for transcriptional repression (Chamberlain *et al.* 1994) whereas the leucine zipper (blue) is involved in homo- or hetero-dimerisation 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).

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 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).

1.3.2 FOXP3

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). FOXP3 functions to repress genes that are involved in effector T cell proliferation and regulates homeostasis in the immune system by doing so (Koh *et al.* 2009).

Structurally, FOXP3 is slightly different to other FOXP members as it has an N-terminal proline-rich region instead of a polyglutamine stretch (Lopes *et al.* 2006; Zhou *et al.* 2008) and the full-length protein is shorter than the other FOXP members (Carson *et al.* 2006).

IPEX syndrome is a fatal autoimmune disorder that is characterised by immune dysregulation, polyendocrinopathy, enteropathy and X-linked inheritance (IPEX) (Bennett *et al.* 2001). Symptoms of IPEX include anaemia, insulin-dependent diabetes and chronic diarrhoea (Levy-Lahad and Wildin 2001). Certain missense mutations in the FOXP3 FHD, namely R347H, A384T, F371C and F373A are 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).

1.3.3 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 has been 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).

1.4 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 4). The FOXP FHD forms a 'winged helix' structure consisting of 3 α-helices, 3 β-sheets and 2 loop conformations or 'wings' (Figure 2B). 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, making fewer contacts with DNA (Figure 5) (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 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. The FHD is able to fold autonomously and bind DNA in the absence of the rest of the protein (Stroud *et al.* 2006).

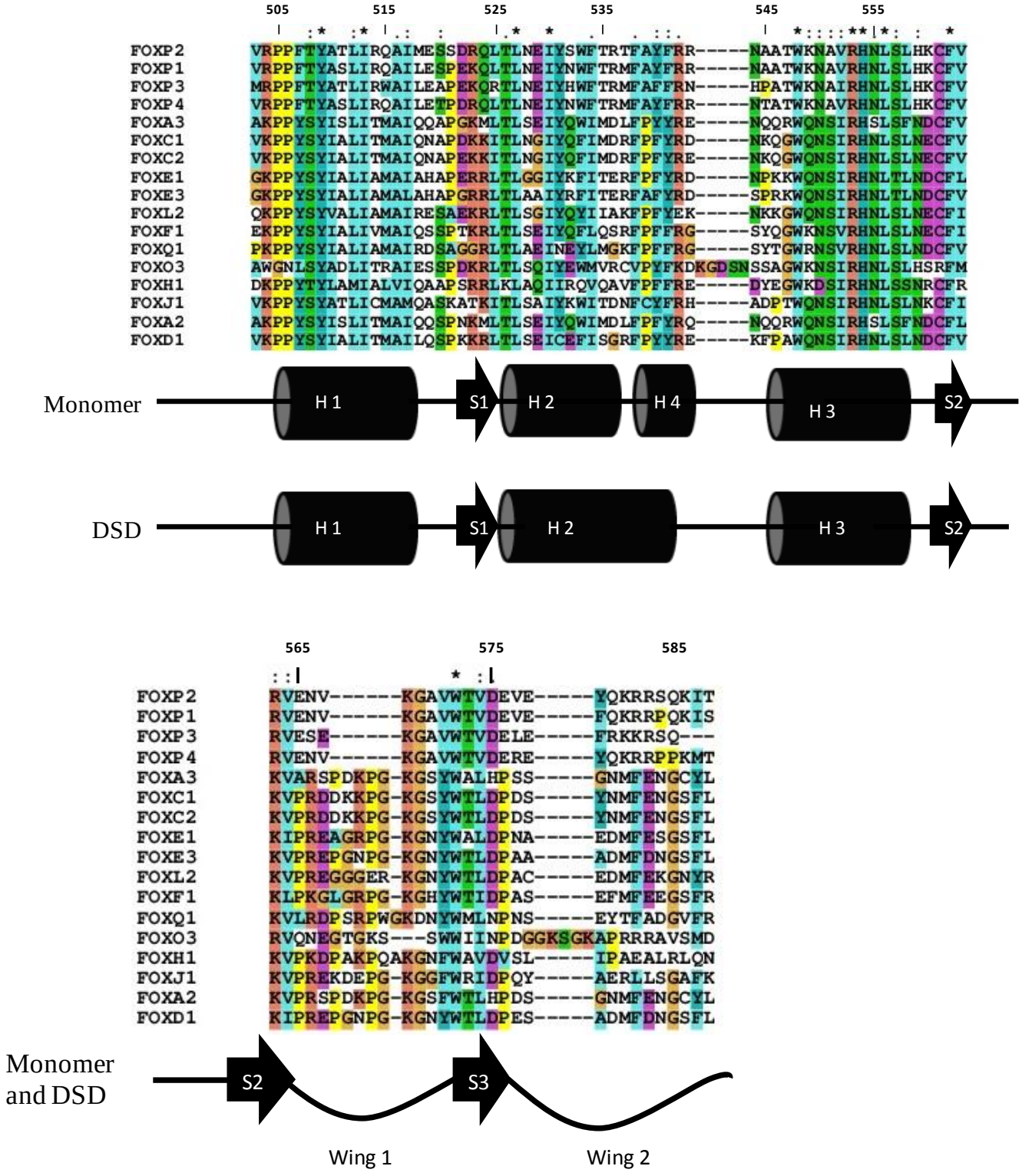


Figure 4. Sequence alignment of the FHD in FOX family members. The alignment was performed using ClustalX2 (Thompson *et al.* 2002). The '*' indicates fully conserved residues and the ':' indicates where strong groups of residues are conserved. The residues are numbered with respect to FOXP2. The secondary 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.

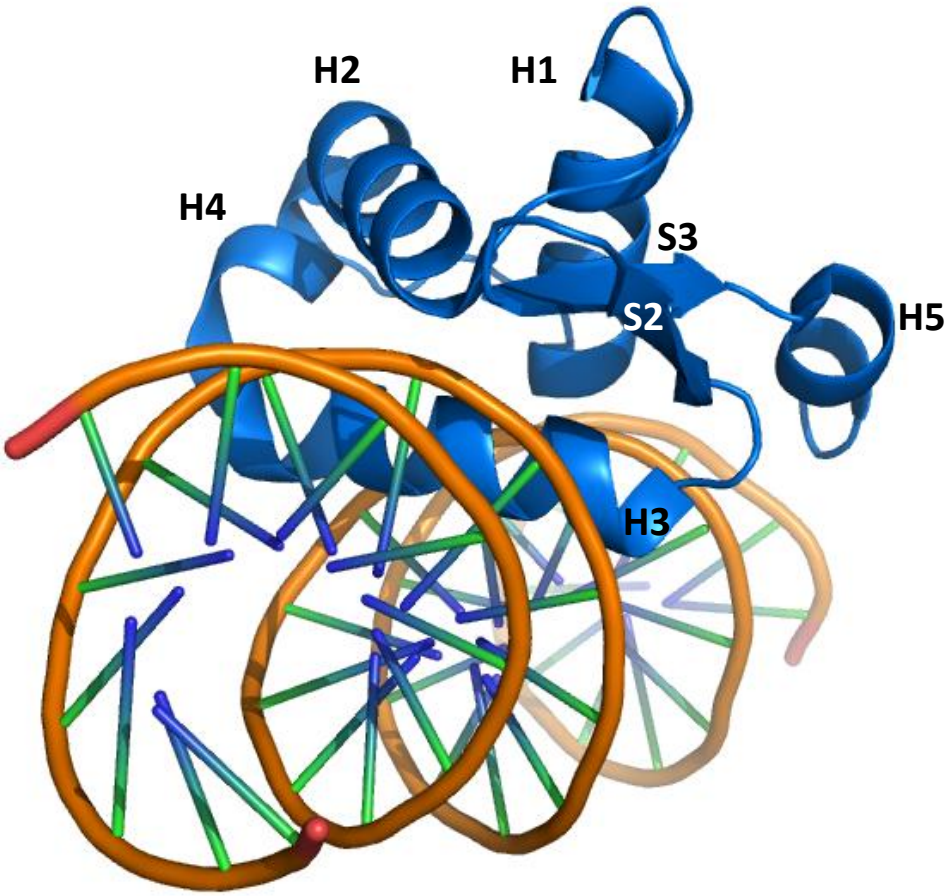


Figure 5. Structure of the wild-type FOXP2 FHD monomer bound to DNA. Cartoon representation of wild-type FOXP2 FHD in the presence of cognate DNA (PDB entry 2A07) (Stroud *et al.* 2006). The recognition helix (H3) inserts into the major groove of the cognate double-stranded DNA whereas helix H1 makes limited stabilising contacts with the DNA. A 3_{10} helix (H4) joins helices H2 and H3 and is located in close proximity to the DNA. In comparison to the helix-turn-helix motif, the FHD contains an additional helix (H1) and wing regions which are involved in stabilising the DNA interaction. This figure was generated using PyMOL (DeLano 2002).

1.4.1 DNA binding

The FOXP3 dimer has been shown to bind two strands of DNA simultaneously at two separate DNA binding sites. This supports the idea that the FOXP3 dimer functions by bringing two distant strands of DNA together within close proximity to form higher-order transcription complexes (Bandukwala *et al.* 2011). The FOXP2 dimer, on the other hand, binds to a single molecule of DNA in the crystal structure (Stroud *et al.* 2006). However, it is suggested that the FOXP2 dimer may also bind and bring together two distant strands of DNA similar to the FOXP3 dimer (Bandukwala *et al.* 2011).

The crystal structure of the FOXP2 FHD bound to DNA indicates that the monomer binds more intimately to the DNA than the dimer. Furthermore, we see that the FOXP FHD monomer has a lower affinity for DNA than other FOX family members such as FOXA3 as shown by electrophoretic gel mobility shift assays (Clark *et al.* 1993b; Li *et al.* 2004). The truncated wings and fewer hydrogen bonds account for the low affinity for DNA as they make limited contacts with the DNA (Stroud *et al.* 2006).

1.4.2 Structure and stability of the FOXP domain-swapped dimers

In the FOXP2 and FOXP3 DSDs, helix H3 and β-strands S2 and S3 are exchanged (Stroud *et al.* 2006; Bandukwala *et al.* 2011; Chu *et al.* 2011). The hinge-loop region (H4) changes conformation upon domain swapping to form an extended helix H2 in the DSD (Stroud *et al.* 2006; Bandukwala *et al.* 2011) (Figure 6).

Although FOXP1, FOXP2 and FOXP3 all dimerise, only FOXP3 exists solely as a stable dimer (Bandukwala *et al.* 2011). The dissociation constant (K_d) for dimerisation of the FOXP1 FHD was established to be 27.3 μM (Chu *et al.* 2011) which implies that at physiological concentrations the protein exists as a mixture of monomer and DSD species. Dissociation constants for the FOXP2 monomer-dimer equilibrium that would indicate the propensity of the DSD to separate into monomers have yet to be established. In contrast to FOXP1 and FOXP2, the FOXP3 FHD does not show formation of monomer at concentrations of less than 0.1 mg/ml (Bandukwala *et al.* 2011). DNA may be required to stabilise the DSD (Bandukwala *et al.* 2011). The FOXP1 FHD was shown to form a mixture of monomer and dimer species at physiological concentrations (Chu *et al.* 2011) whereas the FOXP2 FHD has been shown to form a monomer-dimer mixture under crystallisation conditions (Stroud *et al.* 2006) where the protein concentration is extremely high. Thus a comparison between FOXP2 and FOXP3 FHD domain-swapped dimers (DSDs) might give clues as to what causes the FOXP3 DSD to be more stable despite the high sequence conservation (Figure 4). There are noticeable

structural differences between the FOXP2 DSD and the FOXP3 DSD (Figure 7). The FOXP3 DSD is more arched and more compact than the FOXP2 DSD and the hinge loop region is relatively longer in FOXP3 than in FOXP2 (Stroud *et al.* 2006; Bandukwala *et al.* 2011). 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 in the FOXP3 DSD (Bandukwala *et al.* 2011). Mutation of the hydrophobic and non-conserved Trp348 and Met370 residues, located in the secondary interface, decreases the stability of the dimer and increases the proportion of monomer present (Bandukwala *et al.* 2011), indicating that the interactions these residues form at the secondary interface are crucial to stabilising the DSD.

In addition to the secondary interface, a network of hydrophobic core residues surrounding the primary interface is suggested to control the formation of the FOXP3 DSD (Bandukwala *et al.* 2011). There is a similar corresponding network in FOXP2: Tyr364, Trp366, Phe367, Phe373, Phe374 and Trp381 (Figure 8). Sequence differences in this hydrophobic network may account for the difference in stabilities of their DSDs.

Mutations in the hydrophobic core of the FOXP FHD are also shown to disrupt DSD formation which may lead to disease. These mutations interfere with aromatic stabilising interactions by changing the stacking arrangement and orientation of hydrophobic amino acid residues (Byeon *et al.* 2003; Jee *et al.* 2008). For example, a mutation of the conserved Phe373 (Tyr540 in FOXP2) in the FOXP3 hinge-loop region (F373A) is implicated in the IPEX (immune-dysregulation, polyendocrinopathy, X-linked) syndrome (Bennett *et al.* 2001). The IPEX-causing F373A mutation causes a shift in the monomer-dimer equilibrium towards the monomer by disrupting the hydrophobic core of the DSD (Figure 8) (Bandukwala *et al.* 2011). This disruption doesn't affect DNA binding but it is possible that the mutant, which is now mostly monomeric, can no longer bring two DNA molecules together in close approximation to carry out FOXP3's function to suppress the proliferation of regulatory T cells and hence results in the disease phenotype (Bandukwala *et al.* 2011).

The hinge-loop region of the FOXP FHD, as with all DSDs, plays a significant role in domain swapping. The Ala539 residue located in the hinge region of the FOXP2 FHD is substituted with a proline in all other FOX proteins (Figure 4). Interestingly, mutation of Ala539 to proline (A539P) results in an exclusively monomeric FOXP2 FHD species. 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).

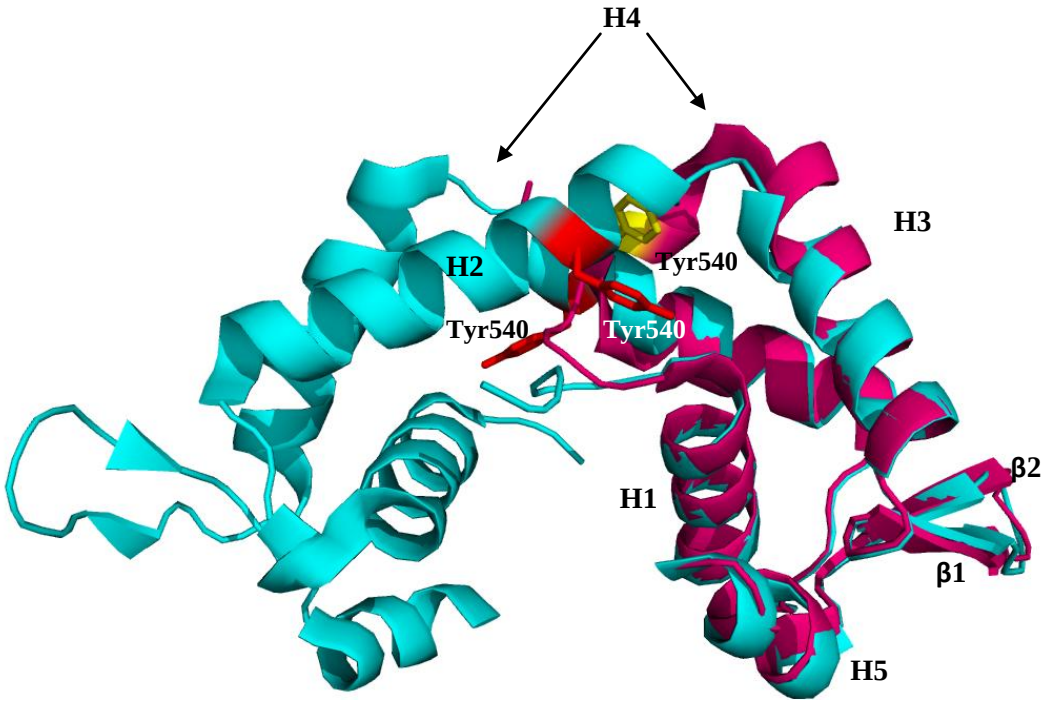


Figure 6. Structural alignment of the FOXP2 FHD monomer with the FOXP2 FHD DSD. The FOXP2 monomer (pink) superimposes almost perfectly onto the FOXP2 domain-swapped dimer (cyan) except in the helix (H4) hinge region. The Tyr540 residue (red) in the domain-swapped dimer resides in helix (H2) whereas Tyr540 (yellow) resides in helix H4 in the monomer. Image generated using PyMOL. PDB code 2A07 (DeLano 2002; Stroud *et al.* 2006).

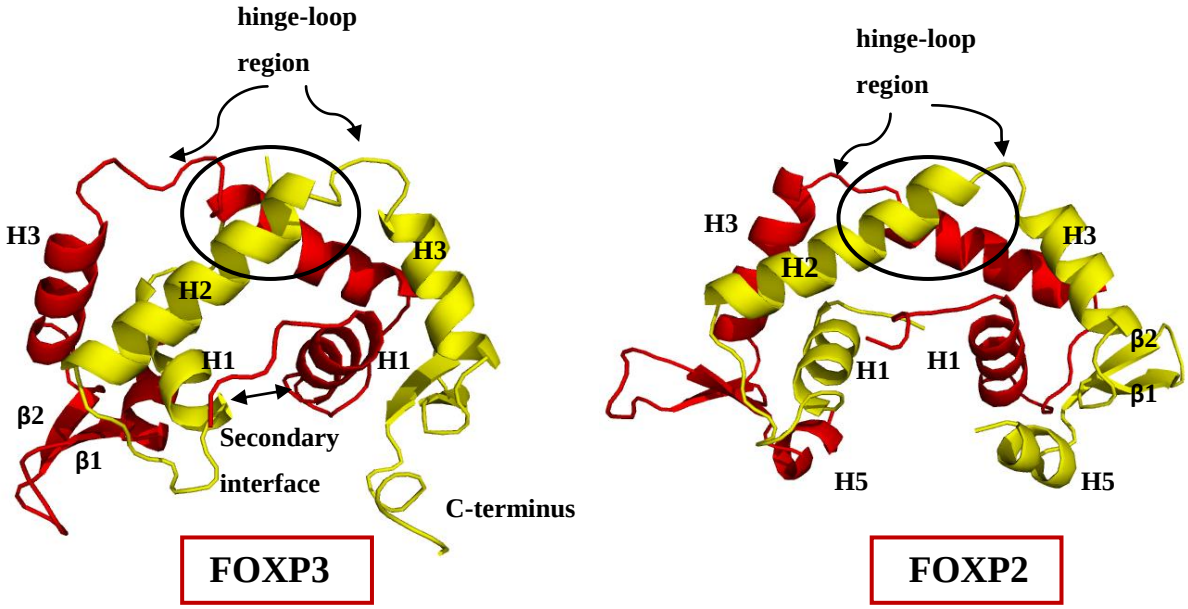
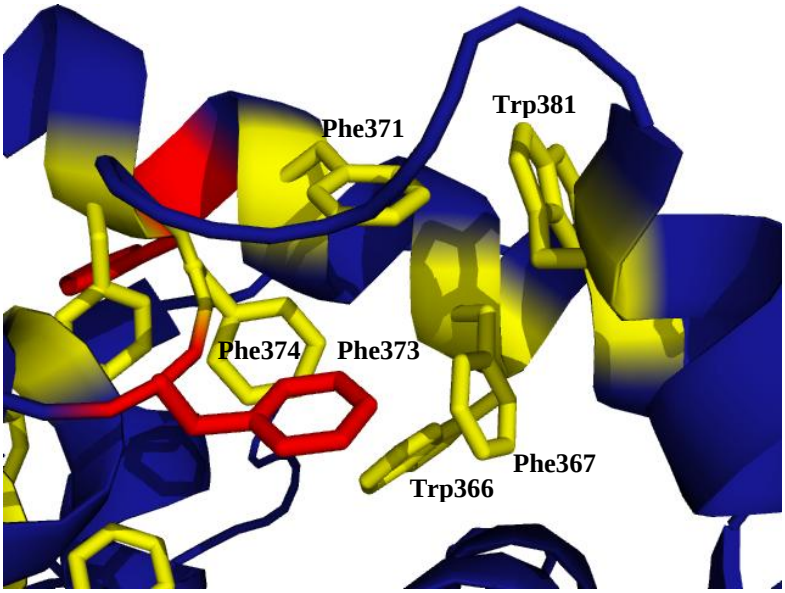


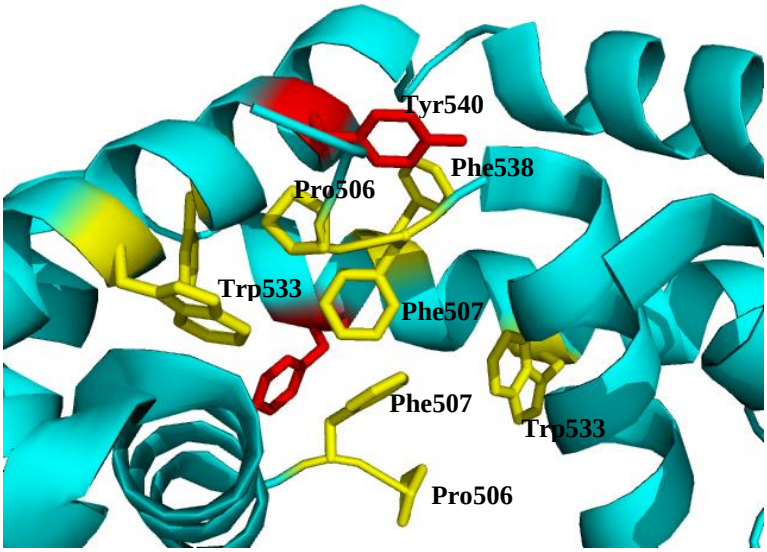
Figure 7. **Structural differences between the FOXP3 and FOXP2 domain-swapped dimers.** The FOXP3 DSD (left) (Bandukwala *et al.* 2011) is slightly more arched than the FOXP2 DSD (PDB entry 2A07) (Stroud *et al.* 2006) (right). Interestingly, the hinge (H4) region is relatively longer in FOXP3 than in FOXP2. The FOXP3 DSD (PDB entry 3QRF) 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).

A



FOXP3

B



FOXP2

Figure 8. **The network of hydrophobic interactions that stabilises the FOXP DSD. A.** Phe373 (red) and corresponding FOXP2 residue Tyr540 (red) are surrounded by hydrophobic residues that are thought to be involved in stabilising the DSD (Stroud *et al.* 2006; Bandukwala *et al.* 2011). Images were generated using PyMOL (DeLano 2002). PDB codes are 2A07 (FOXP2) (Stroud *et al.* 2006) and 3QRF (FOXP3) (Bandukwala *et al.* 2011).

Comparison of the sequences of the hinge loop region of the FOXP-subfamily with the other FOXPs shows only two amino acid differences in FOXP3 which are located at positions Phe373 and Asn376 (Figure 4) and are conserved as tyrosine and arginine, respectively in other members of the FOXP subfamily. Since Phe373 is buried deep in a stabilising hydrophobic pocket at the dimer interface and an F373A mutation has been linked to IPEX, we decided to study the significance of having a phenylalanine at this position for dimerisation. In FOXP2 there is a tyrosine residue (Tyr540) located at the equivalent position to Phe373 in FOXP3. Therefore, we created the Y540F FOXP2 FHD variant which resembles FOXP3 at position 373. We propose that this mutation will stabilise the FOXP2 dimer. Here we plan to study the effect of a mutant (Y540F) on dimerisation propensity of the FOXP2 FHD by monitoring the shift in the monomer-dimer equilibrium in comparison to the WT FOXP2 FHD in order to draw conclusions about the importance of the hydrophobic interface in dimerisation. This will provide information as to why the members of the FOXP subfamily are the only FOX proteins observed to form dimers.

1.5 Aim

The FOXP3 FHD is highly stable and exists solely in dimeric form at concentrations under 0.1 mg/ml, (Bandukwala *et al.* 2011), the FOXP2 FHD dimer is shown to exist in equilibrium with a proportion of monomer under similar conditions (Stroud *et al.* 2006). The difference in the structure of the two FHD dimers has been shown, in part, to be in the hydrophobic pocket at the domain interface. In order to investigate the role of a key hydrophobic residue, Phe373, in stabilising the FOXP3 DSD, we mutated the corresponding residue in FOXP2, Tyr540, to a phenylalanine. This work plans to study the effect of the Y540F variant on dimerisation propensity by monitoring the monomer-dimer equilibrium at various protein concentrations.

The objectives were:

- To obtain pure soluble wild-type FOXP2 FHD.
- To create and purify soluble Y540F FOXP2 FHD variant.
- To use circular dichroism and fluorescence spectroscopy to characterise the secondary and tertiary structure of the Y540F variant and compare with the wild-type to establish whether the mutation induced any major structural changes to the protein.
- To investigate the dimerisation propensity of the Y540F variant compared to wild-type using size-exclusion chromatography at increasing protein concentrations.
- To use electrophoretic mobility shift assays to determine whether the mutation affected DNA binding.
- To obtain crystals of the Y540F variant so that the structure can ultimately be solved so as to provide evidence at the atomic level of any structural changes in the Y540F variant.

CHAPTER 2: EXPERIMENTAL

2.1 Materials

The WT FOXP2 FHD, cloned into a pET-30 LIC vector was a generous gift from Lin Chen (University of Southern California, Los Angeles, CA). The pET-30 LIC vector contains a hexahistidine tag and an S-tag which are used for detection and purification. The histidine tag is used in this study for purification and is linked to the N-terminus of the WT FOXP2 FHD by an enterokinase cleavage site. All non-standard chemicals are listed in Table 1. Plasmids were sequenced by Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa).

Table 1. Non-standard materials and suppliers

Material	Source	Location
Primers for mutagenesis	Inqaba Biotec	Pretoria, South Africa
Quikchange® Lightening Mutagenesis Kit	Stratagene	La Jolla, CA, USA
GeneJET™ Plasmid Miniprep Kit	Fermentas	Ontario, Canada
SDS-PAGE molecular weight markers	Fermentas	Ontario, Canada
DNA ladders	Fermentas	Ontario, Canada
E.cloni® 10G Competent Cells	Lucigen	Middleton, WI, USA
Isopropyl β-D-thiogalactopyranoside (IPTG)	Sigma-Aldrich	St. Louis, MO, USA
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma-Aldrich	St. Louis, MO, USA
CM Sepharose® Fast Flow resin	Sigma-Aldrich	St. Louis, MO, USA
Coomassie Brilliant Blue G-250	Sigma-Aldrich	St. Louis, MO, USA
L-Histidine	Sigma-Aldrich	St. Louis, MO, USA
Dithiothreitol (DTT)	Melford Laboratories Ltd.	Suffolk, UK
5 ml Histrap purification column	GE Healthcare	Buckinghamshire, UK
Low Molecular Weight Gel Filtration Marker kit	GE Healthcare	Buckinghamshire, UK
Hiload 16/600 Superdex 75 pg size-exclusion column	GE Healthcare	Buckinghamshire, UK
FOXP2 Cognate DNA sequence: 5'-AACTATGAAACAAATTTTCCT-3' 5'-TTAGGAAAATTTGTTTCATAG-3'	IDT, Whitehead Scientific	Stikland, South Africa
Hampton Index Crystallisation kit	Hampton Research	Aliso Viejo, CA, USA
Amicon Ultra-15 10K centrifugal concentration columns	Merck Millipore	Darmstadt, Germany
Light chain enterokinase	New England Biolabs	Ontario, Canada
T7 Express Competent Cells	New England Biolabs	Ontario, Canada

2.2 Methods

2.2.1 Plasmid purification

The pET-30 enterokinase/ligation-independent cloning (Ek/LIC) plasmid is 5.4 kb in size and encodes a gene for kanamycin resistance (Braman *et al.* 1996) which allows selection only for cells containing the plasmid. The fusion protein contains the protein of interest fused to an N-terminal His-Tag® and S-tag® sequence which allows for efficient detection and purification. The pET-30 plasmid vector uses the T7 expression system that is strong so that when expression of the target gene is induced, almost all of the cell's resources are devoted to producing the fusion protein (Novagen 2011).

A glycerol stock was made up of T7 Express competetent *E.coli* cells transformed with the pET-30 LIC plasmid encoding the WT FOXP2 forkhead domain (FHD). Subsequently the glycerol stock was innoculated into lysogeny broth (LB) (1% (w/w) tryptone, 0.5% (w/w) yeast extract and 1% (w/w) NaCl) containing 30 μg/ml kanamycin. The culture was grown at 37 ^{o}C overnight (16 hours) with shaking at 230 rpm.

The plasmid was then isolated and purified using the GeneJet™ Plasmid Miniprep Kit which utilises the alkaline-lysis method (Birnboim and Doly 1979). Two rounds of 1.5 ml of cell culture were centrifuged at 27000 x g to harvest the cells after which 100 μl of Resuspension Solution was added to the cells. The Resuspension Solution is a buffer that maintains the optimal pH and contains EDTA which chelates divalent cations so as to inhibit enzymes from cleaving the plasmid DNA in preparation for cell lysis. The cells were resuspended by gentle aspiration using a pipette tip. The cells were then lysed under alkaline conditions by the addition of 100 μl of Lysis Solution containing SDS (sodium-dodecyl-sulfate) and NaOH whereby SDS disrupts the cell membranes by denaturing the proteins embedded in the membrane and NaOH denatures the genomic and plasmid DNA. Thereafter, 125 μl of Neutralization Solution containing sodium acetate was added to neutralise the reaction and to precipitate out the proteins and larger genomic DNA as there is random association between the longer strands as opposed to the plasmid DNA which reanneals correctly. Centrifugation at 27000 x g allowed the cellular debris to collect in the pellet while the plasmid DNA remained soluble in the supernatant. The supernatant was then decanted into a spin column which contains a silica-based membrane that binds the plasmid DNA. The supernatant was then centrifuged at 27000 x g once more and the flow-through was discarded. The plasmid DNA bound to the column was washed twice with Washing Solution that contained ethanol to further remove any contaminants. The plasmid DNA was then eluted with sterilised and deionised water and stored at - 20 ^{O}C. The purified plasmid was subsequently sent for sequencing (see section 2.2.3).

2.2.2 Oligonucleotide primer design

The primers used to generate the mutant were designed using Primer X (http://www.bioinformatics.org/primerx/) in order to incorporate the desired mutation during site-directed mutagenesis. The codon for tyrosine at position 540 was mutated to that of phenylalanine. The primers were analysed by Gene Runner (V3.01, Hastings Software In., NY, USA) for the propensity to form hairpins and loop structures prior to synthesis. The mutant primers were designed for use with the Stratagene Quikchange® mutagenesis system and had the following sequences:

Y540F mutation:

Forward: 5' CA CGG ACA TTT GCG TTC TTC AGG CGT AAT GCA GCA AC 3'

Reverse: 5' GT TGC TGC ATT ACG CCT GAA GAA CGC AAA TGT CCG TG 5'

The underlined regions represent the substitution mutations. The codon encoding tyrosine (TAC) was mutated to TTC which encodes phenylalanine. The mutant primers were synthesised by Inqaba Biotech (Inqaba Biotech, Pretoria). The mutant primers contain flanking sequences that are complementary to the target sequence (the WT FOXP2 cDNA) and contain the desired mutation in the centre. The lyophilised oligonucleotide primers were resuspended in sterilised and deionised water (MQH_2O) to a final concentration of 100 µM stock solution. The stock solution was subsequently diluted to a working concentration of 125 ng/µl for site-directed mutagenesis.

2.2.3 Site-directed mutagenesis

In this study, site-directed mutagenesis was used to mutate the FOXP2 gene to encode for the Y540F FOXP2 FHD variant protein using the Quikchange® Lightening Kit (Stratagene, La Jolla, CA) as follows:

The reaction mixtures consisted of 5 µl (10X) Quikchange® Lightening Buffer, 2 µl (100 ng) ds-DNA template, 1 µl (125 ng) forward primer, 1 µl (125 ng) reverse primer, 1 µl dNTP mix, 1.5 µl of Quiksolution reagent, 38.5 µl of sterile MilliQ water (MQH_2O) and 1 µl of *Pfu Turbo* DNA Polymerase (Quikchange® Lightening Enzyme).

The cycling parameters used were: 1 amplification cycle of 2 minutes at 95 ^oC to denature the template DNA (pET encoding WT FOXP2 ds-DNA) in order to allow the mutant primers to anneal, followed

by 18 cycles of 20 seconds at 95 ^{o}C, 10 seconds at 60 ^{o}C and 162 seconds at 68 ^{o}C. This was followed by extension using a high-fidelity DNA polymerase (*PfuUltra* HF DNA polymerase) via thermal cycling to ensure both strands contain the desired mutations. The resulting mutant plasmid has staggered nicks. 2 µl of *Dpn I* was added to each reaction mixture and incubated at 37 ^{o}C for 5 minutes to cleave methylated DNA (Braman *et al.* 1996). *E.coli* strains are usually methylated by dam (DNA adenine methylase) which was cleaved to ensure the parental DNA is digested. The reaction products were then used to transform *E.cloni*® 10G Chemically Competent Cells (Lucigen, Middleton, WI, USA).

The *E.cloni*® 10G cells were transformed as follows:

After 100 µl of cells were thawed on ice, 250 ng of the Y540F FOXP2 FHD ds-DNA was added to the reaction mixture which was then left on ice for 30 minutes to stabilise the lipid membranes of the cells. The cells were heat-shocked for 45 seconds at 42 ^{o}C on a heating block, which alters the state of the fluid membrane by increasing its permeability and allowing the DNA to enter the cell, after which they were immediately put back on ice to cool down. Thereafter, 500 µl of SOC media (1 % (w/v) tryptone, 0.5 % (w/v) yeast, 10 % (w/v) NaCl and 0.02 M glucose) was added to the transformed cells to provide nutrients and allow the cells to recover from the stress of being heat-shocked. The cells were then incubated at 37 ^{o}C for 1 hour with shaking at 230 rpm. The cells were then plated onto LB-agar (1 % tryptone, 0.5 % yeast extract and 0.5 % NaCl) plates containing 30 µg/ml kanamycin sulfate and were subsequently incubated overnight at 37 ^{o}C. Transformants were selected at random and the plasmid DNA was extracted as previously mentioned. Plasmid encoding the Y540F FOXP2 FHD variant protein was extracted using the GeneJET™ Plasmid Miniprep Kit (see section 2.2.1).

2.2.4 Plasmid identification and transformation

The plasmids encoding the WT FOXP2 FHD and the FOXP2 Y540F mutant cDNA were sequenced to confirm that the sequence for the WT was correct, that the mutation was incorporated into the pET-30 plasmid, and also to ensure that no other mutations were introduced during the polymerase chain reaction (PCR) step in the mutagenesis reaction (see section 2.2.2). DNA sequencing was performed by Inqaba Biotechnical Industries Pty (Ltd) (Pretoria, South Africa).

Once the plasmid containing the correct insert sequence for the FOXP2 Y540F FHD was confirmed by sequencing, the plasmid was used to transform the T7 Express cells for protein overexpression and

subsequent purification of each of the FOXP2 FHD proteins. The T7 Express cells were transformed with WT FOXP2 or the Y540F variant according to the protocol in Section 2.2.3.

For WT FOXP2 FHD and the Y540F variant, a colony from the agar plates was picked and added to 20 ml of fresh LB broth with 30 μg/ml kanamycin and left to grow for 16 hours. Only cells containing and expressing the pET-30 plasmids grew. Glycerol stocks were prepared by the addition of 500 μl of each overnight culture to 200 μl of 80 % glycerol after which they were stored at -80 ^{o}C.

2.2.5 Protein production and purification

2.2.5.1 Induction studies

Induction studies were carried out in order to determine the induction conditions for optimal expression and protein production of the WT FOXP2 FHD and the FOXP2 Y540F variant proteins. 5 μl of a glycerol stock of transformed T7 Express cells were added to separate flasks containing 20 ml of LB broth. The cells were left to grow for 16 hours at 37 ^{o}C, thereafter 20 ml of fresh LB was innoculated with a 50-fold dilution of overnight culture containing 30 μg/ml kanamycin sulfate. These cultures were then left to grow at 37 ^{o}C with shaking at 230 rpm to mid-log phase (OD_{600} = 0.6) which took about 2.5 hours. Production of the Y540F variant and the WT FOXP2 FHD proteins was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to different final concentrations (0.3 mM to 1 mM) and at different temperatures (23 ^{o}C and 37 ^{o}C). Aliquots of 500 μl of cell culture were removed every hour after the addition of IPTG for 5 hours and an overnight sample was also collected. The cells were pelleted by centrifugation at 4221 x g at 10 ^{o}C. The harvested cells were then resuspended in 20 μl of HisTrap Equilibration Buffer (20 mM HEPES, pH 7.6, 500 mM NaCl and 30 mM imidazole) and were sonicated to lyse the cells. The sample was then centrifuged at 27000 x g at 10 ^{o}C to separate the soluble fraction. An equal volume of SDS-PAGE reducing sample buffer was added to each 20 μl sample which was then subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) according to the Laemmli method (1970) so as to determine the conditions that were most ideal for both wild-type and Y540F variant expression.

2.2.5.2 Protein production

For WT FOXP2 FHD and the Y540F variant proteins:

Overnight cultures were used to innoculate fresh LB broth (1:50 dilution) with 30 μg/ml kanamycin sulfate. This was grown at 37 ^{o}C with shaking at 230 rpm until mid-log growth phase (OD_{600} = 0.6) at which point expression of the WT FOXP2 FHD or the Y540F variant proteins was induced with 1 mM IPTG for 3 hours at 37 ^{O}C. The cells were harvested by centrifugation for 20 minutes at 4221 x g at 10 ^{o}C and resuspended in HisTrap Equilibration Buffer (20 mM HEPES, 500 mM NaCl and 30 mM imidazole, pH 7.6) with 10 μg/ml lysozyme. The cells were then stored overnight at -20 ^{o}C. The following morning, the cells were thawed at 20 ^{o}C on a rotator, and then lysed on ice by sonication for 6 rounds of 30 second pulses, using a power output of 9 with a Sonicator Ultrasonic Processor (Misonix Incorporated). The lysed cells were then centrifuged at 27000 x g for 20 minutes at 10 ^{o}C to pellet the insoluble fraction. A sample of the supernatant (soluble fraction) was subjected to SDS-PAGE to confirm that the protein was indeed expressed and soluble.

2.2.5.3 Immobilised metal-ion affinity chromatography (IMAC)

The supernatant was subsequently loaded onto a 5 ml HisTrap purification column (GE Healthcare), pre-equilibrated with HisTrap Equilibration Buffer (20 mM HEPES, pH 7.6, 30 mM imidazole and 500 mM NaCl) and purified with the ÄKTA FPLC (fast protein liquid chromatography) purification system (GE Healthcare). The tagged fusion protein binds non-covalently to the nickel ions on the HisTrap column. The column was washed with a minimum of 10 column volumes of HisTrap Equilibration Buffer to allow protein contaminants to pass through the column, followed by a salt wash (20 mM HEPES, 1.5 M NaCl, and 30 mM imidazole, pH 7.6) to allow cellular DNA fragments that may be bound to the protein after sonication to be released from the protein and pass through the column. The FOXP2 FHD proteins were then eluted with a one-step elution using the HisTrap Elution Buffer (20 mM HEPES, 500 mM NaCl and 500 mM imidazole, pH 7.6). The relatively high concentration of imidazole in the HisTrap Elution Buffer outcompetes the tagged protein for binding to the nickel ions in the HisTrap column, thus eluting the protein.

The FOXP2 FHD proteins were subsequently dialysed into Storage buffer (20 mM Tris-Cl, 150 mM NaCl, 0.5 mM EDTA and 1 mM DTT). For all experiments, the FOXP2 FHD proteins were dialysed weekly into Storage buffer with fresh DTT to ensure that the cysteines remained reduced and did not form disulphide bonds. All experiments hereafter were performed in Storage buffer, unless otherwise stated.

2.2.6 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS (sodium dodecyl sulphate) is an anionic detergent that binds to proteins giving them a net uniform negative charge (Pitt-Rivers and Impiombato 1968) and denatures them into their individual polypeptide units. Electrophoresis is based upon the movement of charged particles in an electric field and since the proteins will all have a net negative charge, they will be separated solely on the basis of size due to the sieving effect of the pores created by the acrylamide and bis-acrylamide (Summers *et al.* 1965). Discontinuous gel electrophoresis was performed using the Laemmli method (Laemmli 1970) where the stacking gel composition is different to the resolving gel. The proteins 'stack up' according to their densities prior to separation which results in enhanced separation (Holmes *et al.* 2003). SDS-PAGE was used to analyse the purity of the protein and to determine whether the desired protein was expressed according to size. β-mercaptoethanol was used to maintain the denatured and reduced state.

The fractions collected from purification were subjected to discontinuous SDS-PAGE according to the Laemmli method (1970) with 4 % acrylamide stacking gels (4 % (w/v) acrylamide, 0.36 % (w/v) bis-acrylamide, 0.05 M Tris-Cl pH 6.8, 0.01 % (w/v) SDS, 0.005 % (v/v) ammonium persulphate and 0.2 % (v/v) TEMED) and 15 % separating gels (15 % (w/v) acrylamide, 1.35 % (w/v) bisacrylamide, 0.25 M Tris-Cl pH 8.8, 0.1 % (w/v) SDS, 0.05 % (v/v) ammonium persulphate and 0.1 % (v/v) TEMED). The samples were diluted with an equal volume of reducing sample buffer (125 mM Tris-HCl, pH 6.5, 4 % (w/v) SDS, 20 % (v/v) glycerol, 10 % (v/v) β-mercaptoethanol and 3.5 μg/ml bromophenol blue) and was boiled at 95 ^{o} C for 5 minutes prior to loading the gels. The gels were electrophoresed using the electrode buffer (anode and cathode): 250 mM Tris-Cl, 192 mM glycine, 0.1 % (w/v) SDS (pH 8.3) at 160 V for approximately 1.5 hours to achieve separation of proteins. The following proteins were used as molecular weight markers: β-galactosidase (116 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), lactate dehydrogenase (35 kDa), restriction endonuclease Bsp981 (25 kDa), β-lactoglobulin (18.4 kDa) and lysozyme (14.4 kDa). The gels were then stained with Coomassie Brilliant Blue R-250 for 1 hour, followed by destaining with a 1:5:4 (acetic acid : methanol : water) solution overnight.

2.2.7 Protein concentration determination

The protein was subjected to a series of dilutions and the absorbance at 280 nm, which is the wavelength at which the side chains of the aromatic residues absorb light maximally, was measured for each dilution. The absorbance corresponding to the undiluted stock was determined by extrapolation of the standard curve to a 1X concentration. Absorbance was measured using a Jasco V-630 UV-VIS spectrophotometer.

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 contaminating DNA which is an important measurement when working with transcription factors. A ratio of 1.8 or higher shows that the protein is relatively pure and has a relatively low amount of DNA contamination.

The absorbance values were corrected for the light scattering contributed by the buffer and aggregates in solution using Equation 2.

$$A_{280protein} = A_{280protein} - A_{280buffer} - [(A_{340protein}-A_{340buffer})\ (A_{280buffer} / A_{340buffer})] \quad \text{(Equation 2);}$$

where absorbance at 280 nm is indicative of protein and absorbance at 340 nm is indicative of aggregation.

Protein concentration was determined by using the Beer-Lambert law:

$$A280 = \varepsilon cl \quad \text{(Equation 3);}$$

where ε is the molar extinction coefficient ($M^{-1}.cm^{-1}$), c is the protein concentration (M) and l is the path length (cm).

The molar extinction coefficient was predicted using the ExPASy ProtParam tool (http://www.expasy.org/tools/protparam.html) (Gasteiger *et al.* 2003). ProtParam calculates physico-chemical properties of a protein using an amino acid sequence as input (Gasteiger E. 2005). The molar extinction coefficient was calculated from the amino acid composition from the protein sequence using Equation 4.

ε (Protein) = n (Tyr)* ε (Tyr) + n (Trp)* ε (Trp) + n(Cystine)* ε (Cystine) (Equation 4);

where n refers to the number of residues, ε refers to the molar extinction coefficients of the amino acids and ε (Protein) refers to the molar extinction coefficient of the native protein in water (Gasteiger E. 2005).

Monomeric concentration was used in all instances. The molar extinction coefficient of the tagged protein was calculated to be ε_{280} = 22 460 M^{-1} cm^{-1} (Gasteiger *et al.* 2003; Graslund *et al.* 2008).

2.2.8 Cleavage of tagged fusion protein

Tagged fusion proteins are commonly used to efficiently detect and purify proteins. The pET-30 plasmid encodes a His-tag and an S-tag for protein purification. Active ribonuclease A (RNase A) is made up of the S-tag (residues 1 - 20) and the S-protein (residues 21 - 124) (Kim and Raines 1993). The S-tag binds with high affinity (K*d* = 0.1 μM) to the S-protein which can be attached to a matrix and forms the basis of S-tag purification (Terpe 2003). The histidine-tagged method is one of the most popular as it enables one-step purification of the protein of interest. Purification tags are also cleavable; meaning that once pure protein has been obtained the tags can be removed using different proteases such as enterokinase or thrombin, depending on the cleavage sites cloned into the vector. The pET-30 LIC vector contains cleavage sites for both enterokinase and thrombin (Figure 9). The thrombin cleavage site cannot be used as it would only cleave the His-tag and would not cleave off the S-tag sequence therefore enterokinase was used as the cleavage site is located just N-terminal to the FOXP2 FHD amino acid sequence and hence the entire tag would be cleaved off.

The source of the enterokinase used in this study is purified from *Kluyveromyces lactis*, which is a yeast vector containing the clone of the light chain bovine enterokinase gene (New England Biolabs, Ontario, Canada). This particular type of enterokinase was used as it was cost-effective and efficient. The recombinant light chain enterokinase has been shown to cleave at the Asp-Asp-Asp-Asp-Lys site of artificial fusion proteins (Collins-Racie *et al.* 1995).

The cleavage reaction was prepared as follows: The FOXP2 Y540F variant was dialysed against the Cleavage buffer (20 mM Tris-Cl, pH 8.0, 120 mM NaCl, 2 mM $CaCl_2$). The 120 mM NaCl was used because the accessible surface area of the FOXP2 FHD has a relatively high amount of solvent-exposed charged amino acids. When their electric charge is not satisfied, the proteins form electrostatic interactions with the surfaces of other proteins resulting in aggregation. The slightly higher salt concentration helps to maintain solubility by interacting with the charges on the surface of the protein thereby minimising protein-protein interactions. An amount of 0.003 μg of enterokinase was

incubated with 1500 µg of the Y540F variant protein in Cleavage buffer for 16 hours at room temperature.

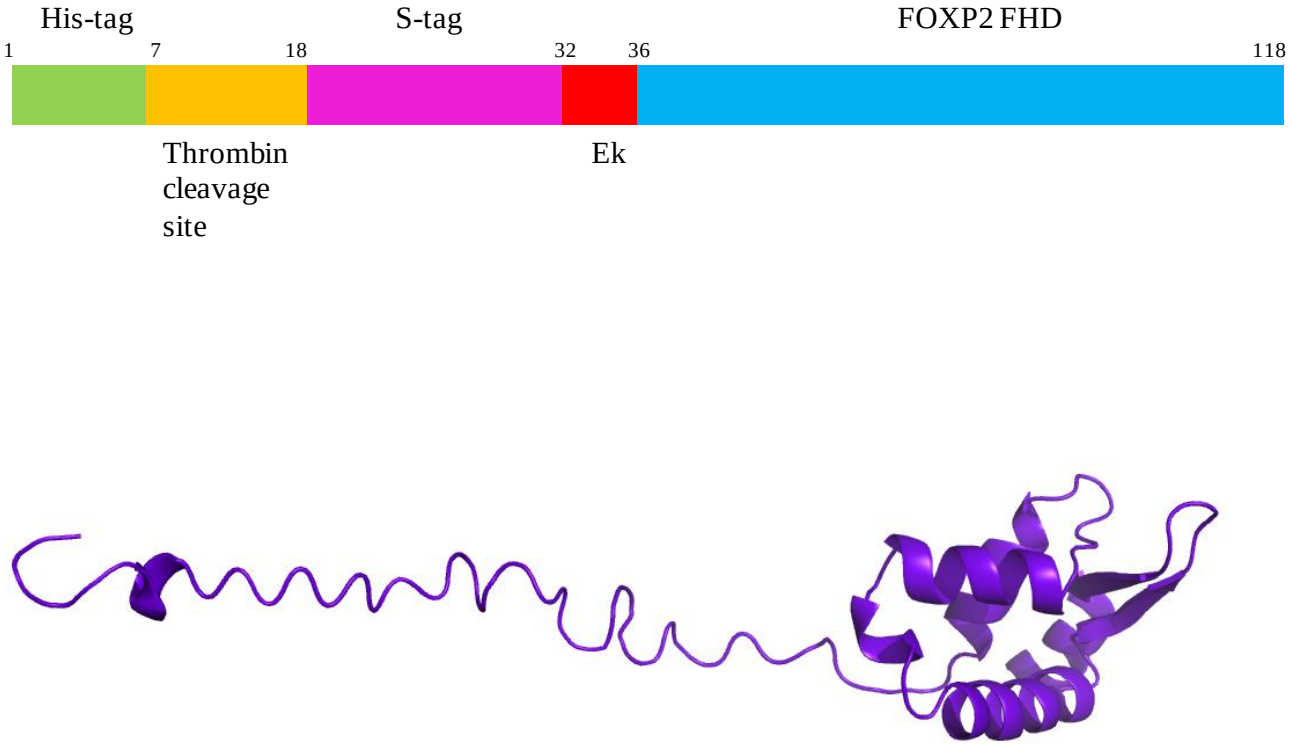


Figure 9. The His-tagged FOXP2 FHD fusion protein. A. Schematic showing the pET-30 LIC vector contains a hexahistdine (His-tag) tag (green) and an S-tag (pink) for detection and purification of the protein of interest in addition to thrombin (orange) and enterokinase (red) cleavage sites. **B. Model of the tagged FOXP2 FHD.** The sequence of the WT FOXP2 FHD, including the tag, was used as input in order to predict the structure of the tagged region. MUFOLD (Zhang *et al.* 2010) was used to predict the structure and to construct the model of the tagged FOXP2 FHD. The image was generated using PyMOL.

The enterokinase was then separated from the protein solution by ion-exchange chromatography. Ion-exchange chromatography is used to separate or purify proteins by exploiting their ability to bind differently to charged resins due to differences in their p*I*. Proteins are charged molecules due to positive and negative charges on the amino acid side chains. The net charge therefore depends on the relative numbers of positively- and negatively-charged residues. When a protein experiences a pH higher than that of its unique p*I*, it will have a net negative charge due to hydrogen ions being donated in a more alkaline environment.

The cleavage reaction mixture containing the FOXP2 Y540F variant and enterokinase was loaded onto the negatively-charged CM-Sepharose Fast Flow® resin (GE Healthcare, Buckinghamshire, UK) pre-equilibrated with 20 mM Tris-HCl, pH 8.0. The Y540F variant has a theoretical p*I* of about 10.9 whereas bovine enterokinase (light chain) has a theoretical p*I* of 5.35, calculated using ProtParam (Gasteiger *et al.* 2003). Since the binding buffer had a pH of 8.0, the Y540F variant will have a net positive charge and hence bind to the CM-Sepharose™ resin while the enterokinase will have a net negative charge and thus pass through the column. A minimum of 10 column volumes of 20 mM Tris-Cl, pH 8.0 were used to wash the column which allowed the enterokinase to pass through the column. The protein was eluted with a linear gradient from 0 mM to 700 mM NaCl. The peak fractions were collected and subsequently analysed by Tricine SDS-PAGE.

2.2.9 Tricine sodium dodecyl sulphate-polyacrylamide gel electrophoresis

This technique was used to confirm that the cleavage of the fusion protein was successful. The principle of this technique is similar to SDS-PAGE except that tricine is used as the trailing ion instead of glycine to allow optimal separation of small proteins and peptides (Schagger 2006). A 10 % separating gel (10 % (w/v) acrylamide, 0.6 % (w/v) bis-acrylamide, 1 M Tris-Cl pH 8.45, 0.1 % (w/v) SDS, 10 % (v/v) glycerol, 0.05 % (v/v) ammonium persulphate and 0.005 % (v/v) TEMED) and a 4 % stacker gel (4 % (w/v) acrylamide, 0.25 % (w/v) bis-acrylamide, 0.75 M Tris-Cl pH 8.45, 0.075 % (w/v) SDS, 0.075 % (v/v) ammonium persulphate and 0.00075 % (v/v) TEMED) were used. The samples were prepared as described in Section 2.2.6. The anode buffer consisted of 100 mM Tris, 22.5 mM HCl, pH 8.9) and the cathode buffer contained 100 mM Tris, 100 mM Tricine, 0.1 % (w/v) SDS, pH 8.25,. The samples were subjected to electrophoresis at an initial voltage of 30 V until the sample entered the stacking gel and thereafter at 190 V for 1 hour to achieve good separation.

2.2.10 Structural integrity of protein

It is necessary to confirm the integrity of the protein at the secondary, tertiary and quaternary structural levels to provide different types of information. Circular dichroism spectroscopy is used to analyse the secondary structure of the protein and to establish whether the protein is folded. Fluorescence spectroscopy provides information on the tertiary structure by looking at changes to the local environment of tryptophan residues. Blue native-polyacrylamide gel electrophoresis is used to determine the quaternary structure of the protein and mass spectrometry can be used to confirm that the protein is the correct size and does not have any DNA bound from the purification step.

2.2.10.1 Matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS)

MALDI-TOF is frequently used to analyse various biomolecules including proteins and oligonucleotides. This technique can be used to accurately determine the molecular weight of proteins, detect post-translational modifications as well as protein-protein and protein-DNA interactions. The critical initial step involves co-crystallisation of the analyte with a matrix, usually a weak organic acid. The analyte-matrix mixture is then irradiated with a nitrogen laser beam which leads to ionisation of the analyte (Kaufmann 1995).

The choice of matrix is crucial as matrices facilitate ionisation of the analyte and some matrices may be more sensitive than others which may lead to better-quality spectra with improved resolution (Gusev *et al.* 1995). The matrix donates protons in the positive mode and accepts protons in the negative mode. In positive mode, $[M+H]^+$ species form and in the negative mode $[M-H]^-$ species form during deprotonation (Zaluzec *et al.* 1995). Upon irradiation with a laser beam at a specific wavelength, the ions are given the same amount of kinetic energy and are accelerated in a long vertical tube to a detector which allows separation of ions with different masses as the ions reach the detector at different times. The smaller ions reach the detector first as they migrate faster than the heavier ions which take longer. The time taken to reach the detector is dependent on mass, charge and kinetic energy (Kaufmann 1995; Zaluzec *et al.* 1995). Doubly protonated species $[M+2H]^+$, with a mass-to-charge ratio of 1:2 can occur but usually at a lower relative abundance (Zaluzec *et al.* 1995) and will travel two times slower than the $[M+H]^+$ ions with a mass-to-charge ratio of 1:1.

We chose to use the sinapinic acid (SA) matrix, commonly used in protein and peptide analysis, in positive mode to determine if the protein was of the correct size and 3-hydroxypicolinic acid (3-HPA) in negative mode to detect whether all contaminating DNA had been removed. Oligonucleotides have

a negatively-charged backbone and thus prefer deprotonated states during ionisation ($[M-H]^-$) in negative mode (Kaufmann 1995) where the cations present in the positive amino acids ($[M+H]^+$) or the buffer ($[M+Na]^+$ and $[M+K]^+$) are detected in positive mode (Beavis *et al.* 1989; Bonk and Humeny 2001).

MALDI-TOF MS was performed at the CSIR (Council for Scientific and Industrial Research) (Pretoria, South Africa) on WT FOXP2 FHD to establish i) that the protein was the correct size, had not incorporated any post translational modifications and was indeed FOXP2 FHD and ii) that the protein had been successfully purified in the absence of DNA and no contaminating cellular DNA was bound to the protein. The samples and matrices were prepared using the dried-droplet method. SA was prepared by mixing acetonitrile (ACN) and 0.1% (v/v) trifluoroacetic acid (TFA) (1:2). Equal volumes of protein sample and SA were mixed vigorously and 1 µl of sample/matrix mixture was spotted onto the target metal plate (Bruker Daltonics, Germany). The drop was allowed to dry at room temperature. The 3-HPA matrix was saturated in an ACN/water (1:1) mixture containing 10 mg/ml diammonium hydrogen citrate. 0.5 µl of the saturated matrix was then spotted onto the AnchorChip™ metal plate and allowed to dry completely at room temperature. 0.5 µl of protein sample was then spotted onto the dried droplet and allowed to dry completely.

The MALDI-TOF mass spectrometer used to collect data was the Autoflex™ (Bruker Daltonics). Each spotted sample was subjected to approximately 100 - 140 laser shots with an N_2 laser at 337 nm which were then averaged to form the final mass-to-charge spectra.

2.2.10.2 Blue native-polyacrylamide gel electrophoresis (BN-PAGE)

Native polyacrylamide gel electrophoresis (native-PAGE) separates proteins according to their charge, mass, isoelectric point (p*I*) and shape under nondenaturing conditions (Schagger *et al.* 1994), unlike SDS-PAGE which is a denaturing technique and separates on the basis of mass only. Native-PAGE allows determination of the proteins' quaternary structure (Braz and Howard 2009) and was used here prior to size-exclusion chromatography to confirm that the FOXP2 FHD was indeed a mixture of monomer and dimer species as reported by Stroud *et al.,* (2006).

There are a variety of native-PAGE techniques that have their own advantages and limitations. These include clear native-PAGE (CNE) (Wittig and Schägger 2005), high-resolution clear native-PAGE (hrCNE) (Wittig *et al.* 2007) and blue native-PAGE (BN-PAGE) (Wittig *et al.* 2006). The CNE technique is identical to SDS-PAGE but does not contain SDS in the gels or tank buffer. CNE is solely dependent upon the p*I* of the protein, which is required to be less than 5.4; otherwise the protein

migrates toward the cathode and is subsequently lost. This technique is advantageous in that physiological protein complexes are maintained and the gel can be used for further studies which require non-stained proteins such as in-gel enzymatic assays (Wittig and Schagger 2005). However, this low-resolution technique is not useful for basic proteins (p*I* >7.5) which includes the FOXP2 FHD as it has a theoretical p*I* of 10.9 (Gasteiger *et al.* 2003). The hrCNE technique, an improvement of CNE, was developed to improve the solubility and hence the resolution of membrane proteins by using mixed micelles of sodium deoxycholate, an anionic detergent, to impose a charge shift (Wittig *et al.* 2007). This does not apply to the FOXP2 FHD as it is soluble. BN-PAGE is exactly the same as CNE, except the anionic dye, Coomassie Brilliant Blue G-250, is used in the sample buffer and in the cathode buffer to induce a net negative charge on the protein by binding to exposed hydrophobic surface patches on the protein (Georgiou *et al.* 2008). In doing so, BN-PAGE allows basic proteins such as the FOXP2 FHD to migrate toward the anode in their native state (Wittig *et al.* 2006).

A modified discontinuous BN-PAGE technique has been shown to improve the resolution and separation of protein species (Niepmann and Zheng 2006) compared to the original continuous method (Wittig *et al.* 2006). The 'discontinuous' system refers to the difference between the buffers in the tank and the gel. In the Tris-glycine buffer system, the predominant form of glycine upon encountering the stacking gel (pH 6.8) is a slow-migrating neutral zwitterion species, whereas the negatively-charged chloride ions from the Tris-HCl buffer migrate much faster and are always ahead of the glycine. A narrow zone occurs between the glycine and chloride ions which allow the protein to 'stack' together before entering the separating gel simultaneously. When glycine enters the separating gel (pH 8.8), it becomes slightly more negatively-charged and migrates slightly faster than in the stacking gel. The protein has an electrophoretic mobility that lies between the fast-migrating Cl^- ions and the slow-moving glycine which allows for concentration of the protein sample into distinct bands and separation at high resolution. In the modified BN-PAGE method, histidine is used as the slow dipolar ion instead of glycine which has been shown to improve the separation of basic proteins (Niepmann and Zheng 2006) and is used in this study.

The sample was prepared as follows:

Wild-type FOXP2 FHD was added to BN-PAGE sample buffer (100 mM Tris-Cl pH 8.0, 40 % glycerol, 0.5 % Coomassie Blue G-250) and incubated for 30 minutes on ice under reducing conditions. The sample was subjected to discontinuous native-PAGE on a 15 % polyacrylamide separating gel and a 4 % stacking gel as used in SDS-PAGE but without SDS. The cathode buffer contained 100 mM histidine-Tris pH 8.0 and 0.002 % Coomassie Brilliant Blue G-250 which was mixed with the buffer after the sample had been loaded. The anode buffer consisted of 100 mM Tris-Cl

pH 8.8 (Niepmann and Zheng 2006). The gel was run on ice at 100 V for the first hour and 150 V thereafter for 2 hours. The ice was used to prevent protein aggregation during electrophoresis and to achieve good separation as the solubility of the FOXP2 FHD is temperature sensitive and decreases with increasing temperature. Both bands were then excised and subjected to digestion by trypsin followed by liquid-chromatography mass spectrometry/mass spectrometry (LC-MS/MS) to confirm both bands were indeed from the FOXP2 FHD and not due to the laddering effect (Braz and Howard 2009). Furthermore it was used to confirm that these bands did not represent a mixture of DNA-bound and unbound protein or protein contaminants.

2.2.10.3 Liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS)

The 15 % BN-PAGE gel was sent to the CSIR (Council for Scientific and Industrial Research, Pretoria, South Africa) to perform an in-gel trypsin digest and LC-MS/MS in order to determine the protein content of the bands. The bands were excised out of the gel and were subsequently destained using 50 mM ammonium bicarbonate (NH_4HCO_3) / 50 % methanol. This was followed by alkylation using 55 mM iodoacetamide in 25 mM NH_4HCO_3. The bands of the gel were subjected to in-gel trypsin digest overnight at 37 ^oC using 5 - 50 µl, 10 ng/µl trypsin depending on the gel piece size (Shevchenko *et al.* 2006).

The digests were then resuspended in 2 % acetonitrile/0.2 % formic acid and analysed using a Dionex Ultimate 3000 RSLC (rapid separation liquid chromatography) system coupled to a QSTAR ELITE mass spectrometer. The peptides were then desalted on an Acclaim PepMap C18 trap column using 2 % acetonitrile/0.2 % formic acid and then separated using the Acclain PepMap C18 RSLC column. The peptides were then eluted using a flow rate of 500 nl/min with a 4 - 60 % linear gradient of 80 % acetonitrile.

The peptides were subsequently ionised by electrospray ionisation and the mass of the ions were measured using a QSTAR ELITE mass spectrometer. The MS scans were acquired from *m/z* 400 - 1500 and the three most intense ions were fragmented automatically. Protein Pilot (version 4.0.8085) using the Paragon search engine (AB Sciex) (Shilov *et al.* 2007) was used to compare the obtained MS/MS spectra with proteins in a UniSwiss database supplemented with the WT FOXP2 FHD amino acid sequence:

IVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNAATWKNAVRHNLSLHKCFVR
VENVKGAVWTVDEVEYQKRRSQKIT

Proteins with a confidence above 99.9 % confidence were reported.

2.2.10.4 Circular dichroism

Circular dichroism (CD) spectroscopy measures the difference in the absorption of left- and right-circularly polarised light. If there is a difference in the left-circularly polarised light and right-circularly polarised light, it results in the formation of an ellipse, hence the term ellipticity (θ).

In the far-UV region (170 nm - 250 nm), the protein backbone absorbs strongly and leads to characteristic secondary structure spectra. 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 give one trough near 217 nm and a peak in the 195 - 200 nm range (Woody 1995).

The FOXP2 FHD is an α/β protein (Stroud *et al.* 2006) and the wild-type protein has type I β-turns, α-helices, antiparallel β-sheets and a 3_{10} - helix as well. The secondary structure of the wild-type FOXP2 FHD monomer was calculated to be 48.4 % alpha-helix, 12.9 % beta-sheet and 38.7 % random coil by a programme called 2Struc (Klose *et al.* 2010) whereas the dimer was calculated to contain 51.6 % alpha-helix, 7.5 % beta-sheet and 40.9 % random coil. The CD spectra of the wild-type and the Y540F variant were measured in the far-UV region and compared in order to determine if there were any major structural changes introduced by the mutation with respect to the protein backbone or secondary structure.

WT FOXP2 FHD and the Y540F variant at a concentration of 8 μM, where both proteins are almost entirely monomeric were dialysed against 5 mM HEPES pH 7.6, 150 mM NaF, 1 mM DTT) and were subsequently filtered using a 0.2 micron filter to remove aggregates or dust particles which can affect the polarisation of light. NaCl is opaque and thus will interfere with the spectra so NaF was used instead as it is transparent. The spectra were collected with a Jasco J-810 spectropolarimeter at 20 ^{o}C with a data pitch of 0.2 nm, scanning speed of 200 nm/min and a band width of 0.5 nm. 10 spectra were collected in the far-UV region (205 nm to 250 nm) for both wild-type FOXP2 FHD and the Y540F variant which were then corrected by subtraction of the buffer spectrum from the sample spectra, averaged and smoothed using the negative exponential method (SigmaPlot version 11).

2.2.10.5 Fluorescence spectroscopy

When a molecule is excited from the ground state, fluorescence is exhibited on return to the ground state as emission of light at a longer wavelength (Lakowicz 1999). Fluorescence is used to analyse the tertiary structure of proteins particularly changes in the local environment of innate fluorophores (Lakowicz 1999). Tryptophan and tyrosine, the aromatic amino acids in proteins, have intrinsic fluorescent properties (Schmid 1997).

Fluorescence, in the near-UV range particularly focuses on and is sensitive to the environment of tryptophan residues. It depends on the region of the protein where the tryptophan residues are located (buried or on the surface) and on how many there are (Schmid 1997). A blue-shifted spectrum indicates a more buried tryptophan residue that is in a nonpolar environment and inaccessible to solvent whereas a red shift indicates that the tryptophan residue is exposed in a more polar environment (Lakowicz 1999). If the protein has many tryptophan residues, the emission spectra would represent an average of all the environments of those residues.

In the FOXP2 FHD, there are three tryptophan residues at Trp533 (H2), Trp548 (H3) and Trp572 (β2). These tryptophan residues are relatively spread out in the structure so that the fluorescence emission spectra may provide information on global changes in the tertiary protein structure of the Y540F variant compared to the WT FOXP2 FHD. To analyse the tertiary structure, 4 μM of WT FOXP2 FHD was used to selectively excite the three tryptophan residues (Trp533, Trp548 and Trp572) at 295 nm on a Perkin-Elmer LS50B luminescence spectrofluorimeter. The emission spectra were recorded from 300 nm to 500 nm using a data pitch of 0.5 nm, an excitation band width of 5 nm and an emission band width of 2.5 nm. Three spectra were collected for each protein, averaged and smoothed using the negative exponential method (SigmaPlot version 11).

2.2.11 DNA binding

Since the FOXP2 FHD functions by binding to DNA, DNA binding studies were carried out to confirm that the protein was functional, to determine the optimal protein/DNA binding ratio and to determine whether the Y540F mutation affected DNA binding. Cognate DNA can be defined as a DNA sequence that is recognised by closely related proteins such as those within a protein family or superfamily (Harris *et al.* 1993).

The FOX cognate sequence, 5’ -AACTATGAAACAAATTTTCCT- 3’

3’-GATACTTTGTTTAAAAGGATT- 5’ was used in these DNA binding studies because it is known to bind to the FOXP2 FHD (Stroud *et al.* 2006) as well as to other FOX proteins (Wang *et al.* 2003). The DNA binding propensity was compared between the WT

FOXP2 FHD and the Y540F variant using electrophoretic mobility shift assays and intrinsic fluorescence spectroscopy.

2.2.11.1 Electrophoretic mobility shift assay (EMSA)

This technique is used to detect protein-nucleic acid interactions but can provide quantitative information such as binding stoichiometry, affinities and kinetics under particular conditions (Hellman and Fried 2007). Polyacrylamide gels are used because they show a higher resolution of separation of protein-DNA complexes from free DNA than agarose gels (Fried 1989). The large protein-DNA complexes have a lower electrophoretic mobility and migrate more slowly than their small free nucleic acid counterparts through the pores in the polyacrylamide gel (Garner and Revzin 1986; Lane *et al.* 1992).

WT FOXP2 FHD and the Y540F variant were added to DNA at protein/DNA ratios to final concentrations of 10 μM: 10 μM, 10 μM: 20 μM, 10 μM :30 μM , 20 μM :10 μM and 30 μM :10 μM, respectively, in 100 mM Tris-Cl, pH 7.6, 100 mM NaCl, 1 mM DTT, 1 mM $MgCl_2$. This was used to determine the optimal protein/DNA binding ratio of the wild-type compared to the Y540F variant. The protein and DNA were incubated in binding buffer (89 mM Tris-borate pH 8.3, 5 % glycerol) on ice for 1 hour to allow for binding. Prior to electrophoresis, 0.01 % bromophenol blue was added to each sample. The protein-DNA complexes were resolved on a 10 % acrylamide gel (10 % (w/v) acrylamide, 0.33 (w/v) bis-acrylamide, 89 mM Tris-borate pH 8.3, 0.007 % (v/v) ammonium persulphate, 0.00075 % (v/v) TEMED) at 4 ^{o}C at 50 V for 2.5 hours in TBE buffer (89 mM Tris-borate pH 8.3). The gel was visualised under ultraviolet light after being stained with 0.5 μg/ml ethidium bromide. The gels were analysed by densitometry using the LabWorks Image Acquisition and Analysis Software version 4.6 (UVP Bio-Imaging) to quantify the proportion of DNA bound to the WT FOXP2 FHD and to the Y540F variant. The proportion of bound DNA was calculated as follows:

The sum of the intensity of all the bands in each lane is equal to 100 %. In other words:

[Protein/DNA] + [Unbound DNA] = $[DNA]_{TOTAL}$ = 100 %. The proportion of bound DNA was calculated using Equation 5.

% Bound DNA = (100 %) - (% Unbound DNA) (Equation 5)

2.2.11.2 Fluorescence to detect DNA binding

If a protein is mixed with DNA and exposed to an excitation wavelength of 295 nm, a change in the fluorescence spectrum may be observed upon the protein binding to DNA. The nitrogenous bases in DNA are not able to fluoresce under these conditions and thus the change in the signal between bound and unbound protein arises solely from that of the tryptophan residues in the protein that are involved in DNA binding (Kneale 1994).

The fluorescence can be quenched through collision or formation of the protein-DNA complex. This change in fluorescence can be used to determine the binding stoichiometry and equilibrium binding constants (Kneale 1994; Bogdarina *et al.* 1998; Chu *et al.* 2011). In this instance, fluorescence spectra of DNA-bound and unbound protein were compared so as to determine whether binding affected the tryptophan environment. Furthermore, DNA-bound spectra were compared for the WT FOXP2 FHD and the Y540F variant to determine whether the mutation caused the FHD to interact differently with DNA.WT FOXP2 FHD and the Y540F variant were filtered using a 0.22 micron filter. DNA was added to the protein to a final concentration of 4 μM protein: 4 μM DNA. The protein was incubated with the DNA in 5 mM HEPES pH 7.6, 105 mM NaF, 1 mM DTT for 1 hour at room temperature (20 ^{o}C) prior to excitation at 295 nm and the emission spectra were collected from 300 nm to 500 nm. 4 μM DNA was used as a blank. Three accumulations were collected using a Perkin-Elmer LS50B luminescence spectrofluorimeter for each sample; the spectra were corrected for buffer and DNA, averaged and smoothed using the negative exponential method (SigmaPlot version 11). The ratio of the bound to unbound fluorescence spectra for the WT FOXP2 FHD and the Y540F variant were plotted to determine whether the mutation affected DNA binding.

2.2.12 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 1993). 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) was used to determine the relative amounts of monomer and dimer in both the wild-type FOXP2 FHD and the Y540F variant and to compare the wild-type to

the variant. The matrix consists of highly cross-linked agarose with an exclusion range of 3 kDa to 75 kDa which is able to effectively separate monomer (20.9 kDa) from dimer (37.2 kDa).

The column was equilibrated with S75 buffer: 20 mM Tris-Cl pH 7.6, 150 mM NaCl and 1 mM DTT. Both wild-type FOXP2 FHD and the Y540F variant were dialysed against the S75 buffer and were concentrated using 3K ultrafiltration spin columns (Amicon). Different concentrations ranging from 15 μM to 100 μM of both wild-type FOXP2 FHD and the Y540F variant were left to equilibrate for 24 hours at 4 ^{o}C. 1 ml of each sample was loaded onto the column. The protein was detected by absorbance at 280 nm. The elluent was collected for the WT FOXP2 FHD and the Y540F variant at final concentrations of 93 μM and 70 μM, respectively. The tertiary structures of the separate monomer and dimer fractions were monitored using fluorescence spectroscopy and the methodology in Section 2.2.10.5. The data was smoothed and normalised between 0 and 1 as the concentration of the monomer and dimer species when collected off the column was different and also if the concentrations were made the same by dilution, the equilibrium between the dimer and monomer would be affected.

2.2.13 Crystallisation of FOXP2 Y540F variant

One of the most widely used techniques to study protein structures at the atomic level is x-ray crystallography. There are various methods of growing protein crystals which include the hanging drop- and the sitting drop-vapour diffusion methods. In the hanging drop vapour diffusion method, a siliconised cover slide with supersaturated protein solution is mixed with mother liquor and positioned over a reservoir of mother liquor of a higher concentration. The well is sealed by wax or grease. The vapour from the mother liquor in the reservoir and the vapour from the drop above equilibrate by diffusion toward each other. As water leaves the drop it causes the protein concentration to increase and the solubility of the protein solution to drop gradually until the protein comes out of solution in the form of an ordered crystal (Drenth 1999; McPherson 2004).The crystallisation process can be explained by using a phase diagram (Figure 10). If a highly supersaturated protein solution is initially used, the protein will precipitate because the solubility becomes lowered even further. If an undersaturated protein solution is used, then no crystals will form at all. If a protein that is moderately supersaturated is used, then as the solubility is lowered, many nuclei may form in the nucleation zone. Nuclei are ordered protein ‘aggregates’ that facilitate crystal growth by acting as ‘seeds’. However, if too many nuclei form, this may lower the protein concentration too much to allow growth of a single, large crystal. The best crystals form in a region of the phase diagram called the metastable zone. In the metastable zone, few nuclei form and can grow into large protein crystals (Drenth 1999; Chayen 2004; McPherson 2004; Chayen 2005).

Crystal trials were performed using the hanging drop method with the Hampton Research Index HT crystallisation kit. The FOXP2 Y540F variant was mixed with cognate ds-DNA in a protein/DNA ratio of 3:1 and concentrated using 3K ultrafiltration spin columns to a final concentration of 8 mg/ml of protein. Different reservoir buffer volumes (0.5 ml and 1 ml) were pipetted into the wells, thereafter 3 µl of the protein/DNA complex was added to 3 µl of reservoir buffer. The reservoir buffer (0.1 M Bis-tris pH 6.5, 0.2 M calcium chloride dihydrate, 45 % (v/v) 2-methyl-2, 4-pentanediol) was then mixed with the protein sample by aspiration.

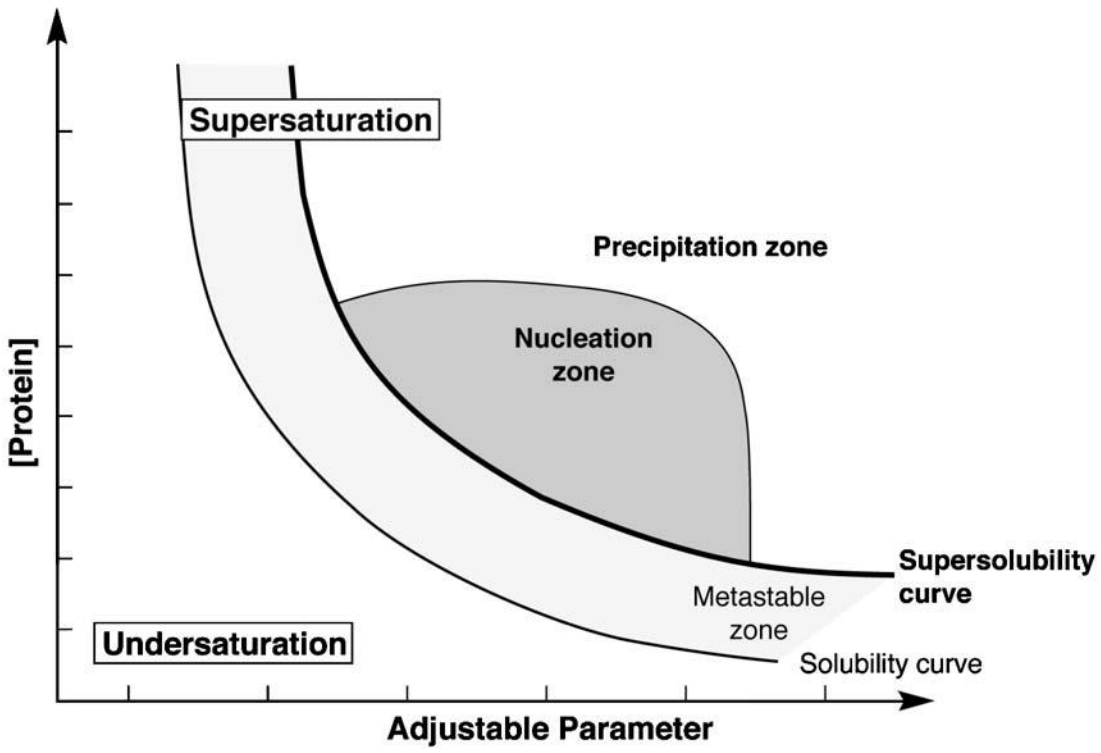


Figure 10. Representation of a protein crystallisation phase diagram. The schematic diagram indicates the regions where the protein is undersaturated at very low concentrations where the protein will not crystallise. In the supersaturation zone, the protein concentration is too high which results in precipitation or aggregation. However, when the protein is moderately supersaturated, it may enter the nucleation zone which produces many small nuclei or protein microcrystals that can inhibit growth of few, large crystals. The best crystals grow in the metastable zone, which is located beneath the nucleation zone. In the metastable zone, few nuclei grow into large crystals. This figure was adapted from Chayen, 2005.

CHAPTER 3: RESULTS

3.1 Creation of mutant and purification

Sanger sequencing performed by Inqaba Biotec (Pretoria, South Africa) confirmed that the plasmids have the correct sequences that encode for wild-type FOXP2 FHD and the Y540F variant. The codon for tyrosine at position 540 in the wild-type had been successfully mutated to phenylalanine in the Y540F plasmid (Figure 11).

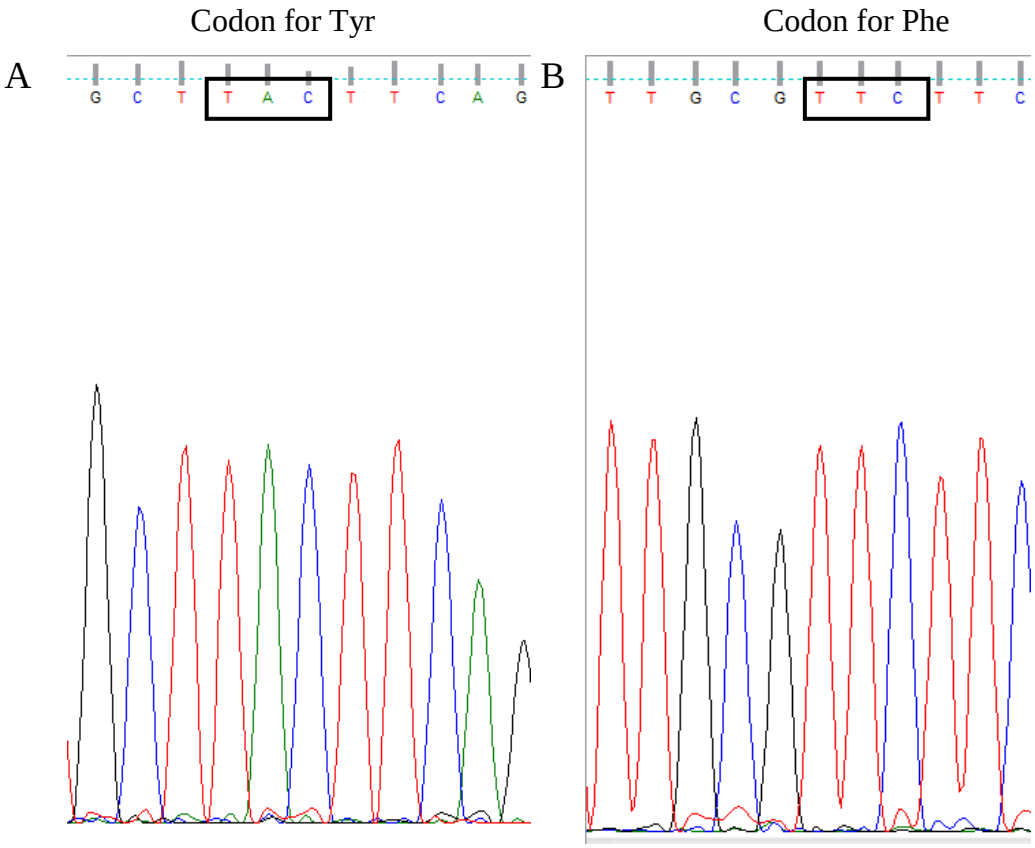


Figure 11. Confirmation of wild-type (WT) FOXP2 FHD and verification of the Y540F variant sequences. The sequencing data obtained from Inqaba Biotec confirmed that the pET-30 plasmid encoded the correct sequence for the WT FOXP2 FHD (A) and that the codon encoding Tyr540 (TAC) was successfully mutated to TTC which encodes phenlylalanine in the Y540F variant (B). The sequencing data was viewed using Finch TV v.1.4.0 (www.geospiza.com/finchtv).

The optimal conditions for producing the maximum amount of soluble protein for both WT FOXP2 and the Y540F variant (Figure 12) were found to be induction at 37 ^{o}C with a final concentration of 1 mM IPTG and growing the cells for 3-4 hours post-induction.

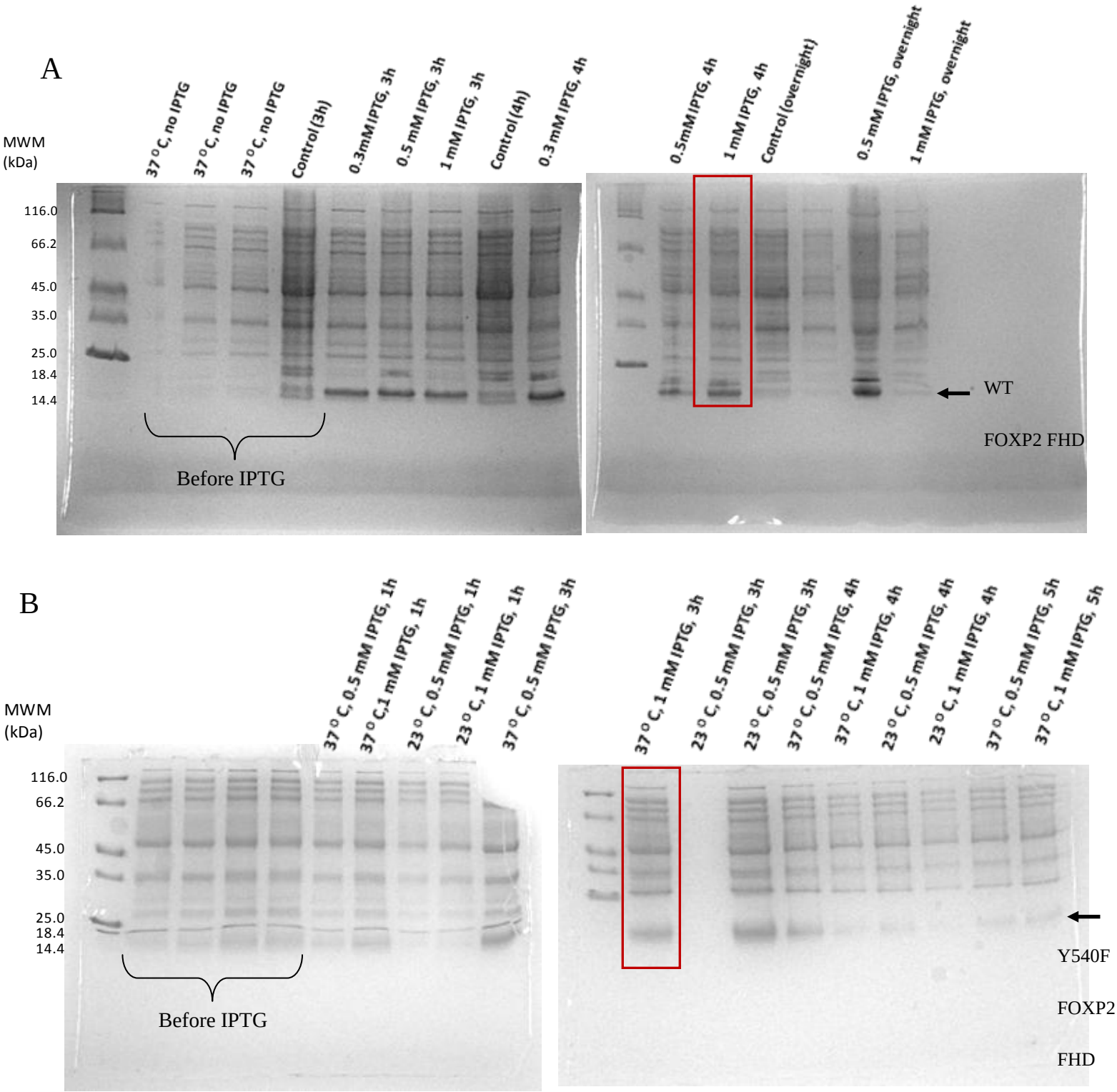


Figure 12. Induction trials for optimal protein production. A. 15 % SDS-PAGE gels showing production of the WT FOXP2 FHD (black arrow) when exposed to a range of final IPTG concentrations (0.1 mM to 1 mM) for varying time periods (1 hour - 16 hours) at 37 ^{o}C. **B.** 15 % SDS-PAGE gels showing production of the Y540F variant (black arrow) when exposed to 0.1 mM to 1 mM IPTG for 1 hour - 16 hours at 20 ^{o}C and 37 ^{o}C.The control samples contained no IPTG (0 mM). The optimal conditions for both WT FOXP2 FHD and the Y540F variant were found to be induction at 37 ^{o}C with a final IPTG concentration of 1 mM for 3-4 hours (red block).

Both the wild-type FOXP2 FHD (Figure 13A) and the Y540F variant (Figure 14A) were purified using immobilised metal-ion chromatography and the pure fractions as shown by SDS-PAGE (Figures 13B and 14B) were then pooled. On the SDS-PAGE gel, the protein migrated according to the correct size predicted for the tagged-FOXP2 FHD which is 14.9 kDa. There was a relatively small amount of both WT and the Y540F variant present in the insoluble fraction (pellet). This could be due to the incorrect folding of the protein which resulted in the formation of inclusion bodies. Sufficient pure soluble protein was obtained at a yield of approximately 3 mg/l of culture.

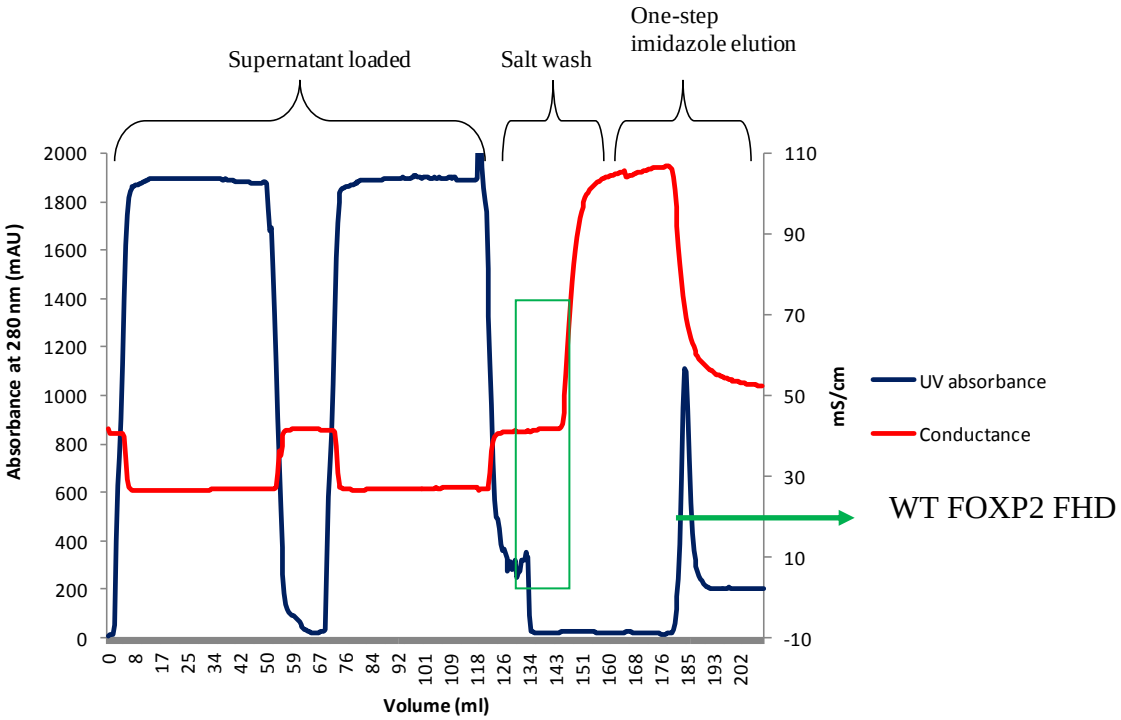
Immobilised metal-ion chromatography (IMAC) purification for both WT FOXP2 FHD and the Y540F variant was performed using a HisTrap column (GE Healthcare) which involves the binding of the hexahistidine-tagged fusion protein to the nickel ions covalently attached to the matrix. The absorbance was measured at 280 nm to detect proteins and conductance was measured to indicate the amount of salt. A low concentration of imidazole was used in the Histrap Equilibration Buffer (20 mM HEPES, pH 7.6, 30 mM imidazole, 500 mM NaCl) because the imidazole has a high affinity for the nickel ions and thus out-competes any nonspecific proteins for binding to the column. The hexahistidine-fusion protein has a stronger affinity for the nickel ions than does the low concentration of imidazole and protein contaminants. Therefore only the fusion protein should bind to the column and the other proteins should pass through as their binding affinity is relatively weak.

Upon loading the supernatant, the absorbance increased as numerous proteins that do not bind to the column passed through the detector (Figures 13A and 14A). The absorbance subsequently decreased and returned to the baseline to indicate that all the contaminants came off the column. The conductance increased slightly between loadings of the two samples as the salt concentration in the Histrap Equilibration Buffer is different to that in the Sample Buffer.

During sonication, genomic DNA is sheared into fragments which can interact non-specifically with the positively-charged surface of the FOXP2 FHD. If random DNA fragments were bound to the FOXP2 FHD, they would interfere with binding of the cognate DNA sequence therefore any DNA contaminants had to be removed in order to perform further DNA binding studies with the cognate DNA. The purpose of using a salt wash (Histrap Equilibration Buffer with 1.5 M NaCl) was therefore, to remove any cellular DNA fragments that may have bound to the protein subsequent to sonication. The high concentration of salt prevents DNA from interacting electrostatically with the protein as the sodium and chloride ions interact with the exposed charged side chains on the protein surface. The sodium and chloride ions shield and prevent the protein from interacting with the negatively-charged DNA backbone. The DNA fragments should therefore pass through the column during the salt wash. As a result of the high salt concentration, the conductance dramatically increased. The FOXP2 FHD

was then eluted with a high concentration of imidazole (500 mM) to out-compete the FOXP2 FHD for binding to the column. Pure FOXP2 FHD (as shown by lane 6 in Figure 13B for wild-type and in Figure 14C for the Y540F variant) elutes in a sharp peak. In order to determine whether the purified band was indeed the tagged FOXP2 FHD, a calibration curve was constructed from the SDS-PAGE gel to determine the molecular weight of the purified protein. The calibration curve shows the estimated molecular weight of the WT tagged FOXP2 FHD and the Y540F variant to be approximately 14.9 kDa (Figure 15).

A



B

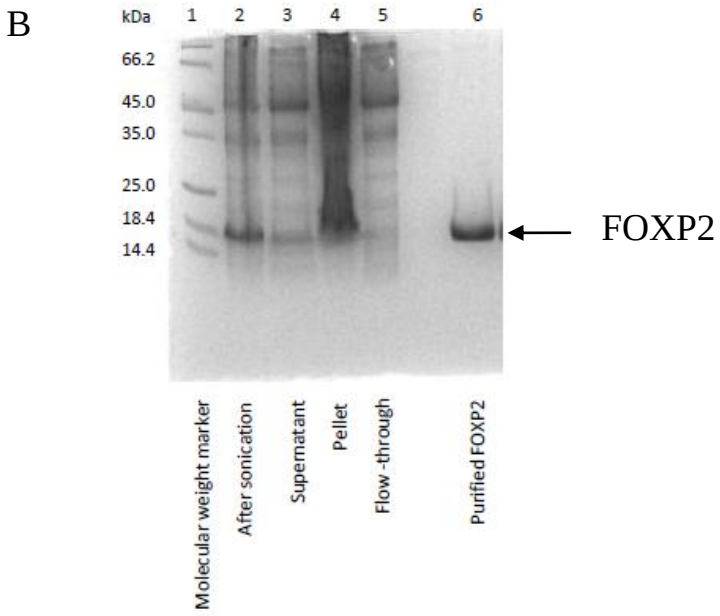


Figure 13. Purification of the WT FOX2 FHD. A. Immobilised metal-ion chromatography (IMAC) elution profile of WT FOXP2. The soluble fraction subsequent to cell lysis was loaded onto a HisTrap IMAC column (GE Healthcare) pre-equilibrated with 20 mM HEPES pH 7.6, 30 mM imidazole and 500 mM NaCl. A minimum of 10 column volumes of equilibration buffer including 1.5 M NaCl (red) was used to remove additional DNA fragments bound to FOXP2 FHD after sonication. The desired protein was eluted in 20 mM HEPES pH 7.6, 500 mM imidazole and 500 mM NaCl. Elution of WT FOXP2 FHD is indicated by green arrow. **B.** 15 % SDS-PAGE gel depicting the WT FOXP2 purification. Lane 1 contains the molecular weight marker; lane 2 indicates the protein content of the *Escherichia coli* after induction with 0.5 mM IPTG for 4 hours, lanes 3 and 4 indicate the protein content in the soluble and insoluble fractions, respectively. Lane 5 shows the protein contaminants that did not bind to the HisTrap column and lane 6 shows the fraction of protein eluted off the HisTrap column. It corresponds to pure FOXP2 FHD of the correct size (black arrow).

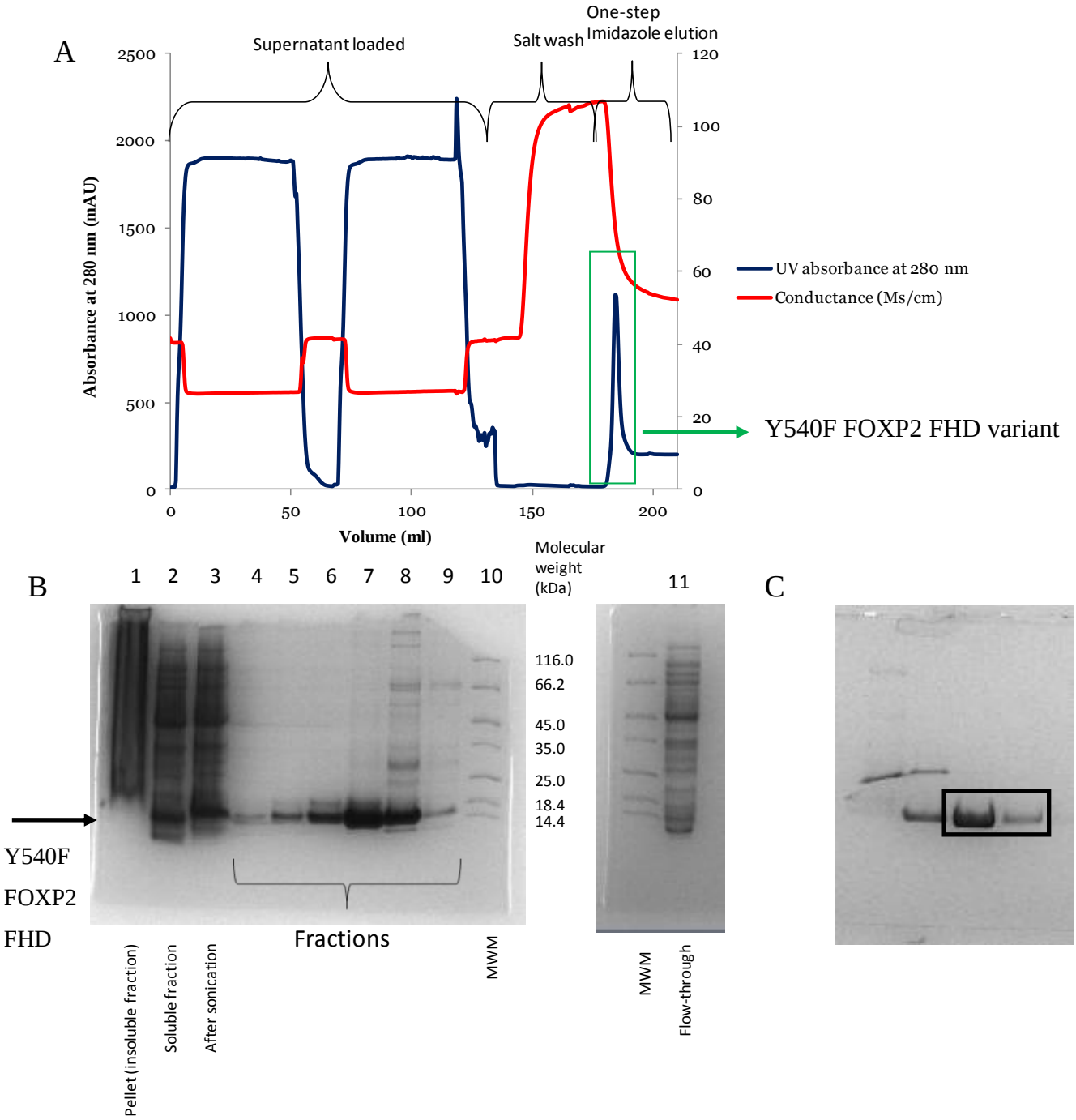


Figure 14. Purification of the Y540F FHD variant. A. IMAC elution profile of the Y540F variant. There is no difference between the purification procedure for the WT FOXP2 (Figure 12) and the variant. The soluble fraction subsequent to cell lysis was loaded onto a HisTrap column (GE Healthcare) pre-equilibrated with 20 mM HEPES pH 7.6, 30 mM imidazole and 500 mM NaCl. A minimum of 10 column volumes of salt wash (Equilibration buffer including 1.5 M NaCl) (red) was used to remove cellular DNA fragments potentially bound to the protein. The Y540F variant was eluted with 20 mM HEPES pH 7.6, 500 mM imidazole and 500 mM NaCl. **B.** 15 % SDS-PAGE gel of the Y540F variant purification. Y540F variant is indicated by the black arrow. The cells were sonicated (lane 3) and the sample was then clarified using centrifugation. The insoluble fraction (lane 1) was separated from the soluble fraction (lane 2). The soluble fraction was loaded onto the HisTrap column. The proteins that did not bind (flow-through) are shown in lane 11. The fractions containing FOXP2 FHD (lanes 4-9) were collected, pooled, dialysed into Histrap Equilibration Buffer and loaded back onto the HisTrap column for further purification. **C. 15% SDS-PAGE gel of 2nd HisTrap purification.** Fractions with pure protein (black block) were pooled. The Y540F variant is of the correct size (black arrow).

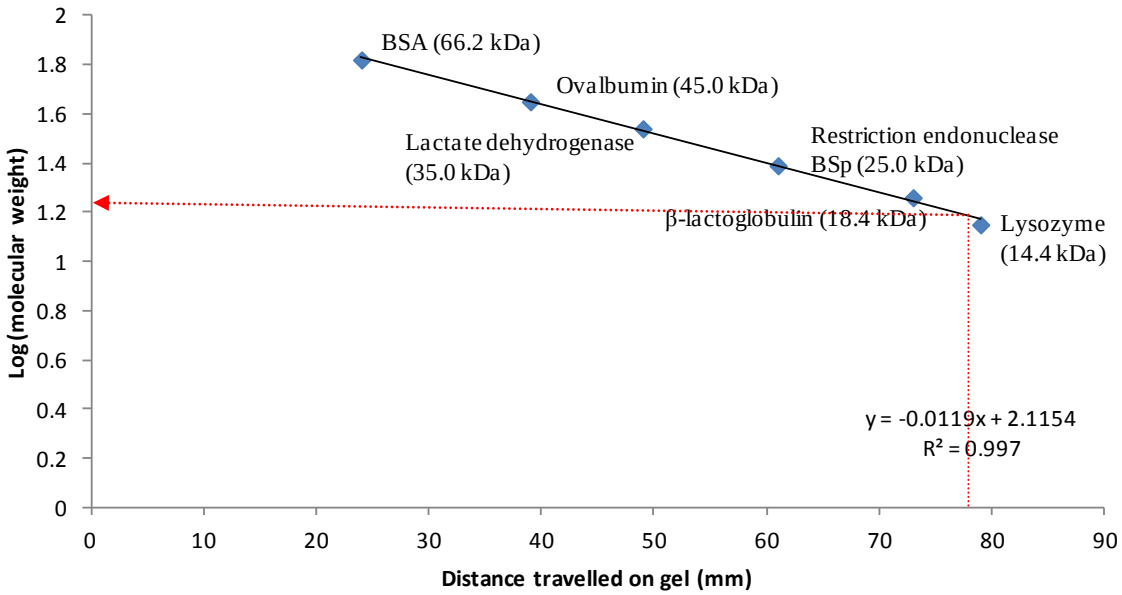


Figure 15. Estimation of the molecular weight of FOXP2 FHD. Calibration curve from an SDS-PAGE gel of the FOXP2 FHD and molecular weight markers. The molecular weight of FOXP2 FHD was calculated to be 14.9 kDa for both wild-type FOXP2 FHD and the Y540F variant. The molecular weight marker used contained the following proteins: BSA (66.2 kDa), ovalbumin (45.0 kDa), lactate dehydrogenase (35.0 kDa), Restriction endonuclease BSp (25.0 kDa), β-lactoglobulin (18.4 kDa) and lysozyme (14.4 kDa). The molecular weight of the FOXP2 FHD is approximately 14.9 kDa (red arrow).

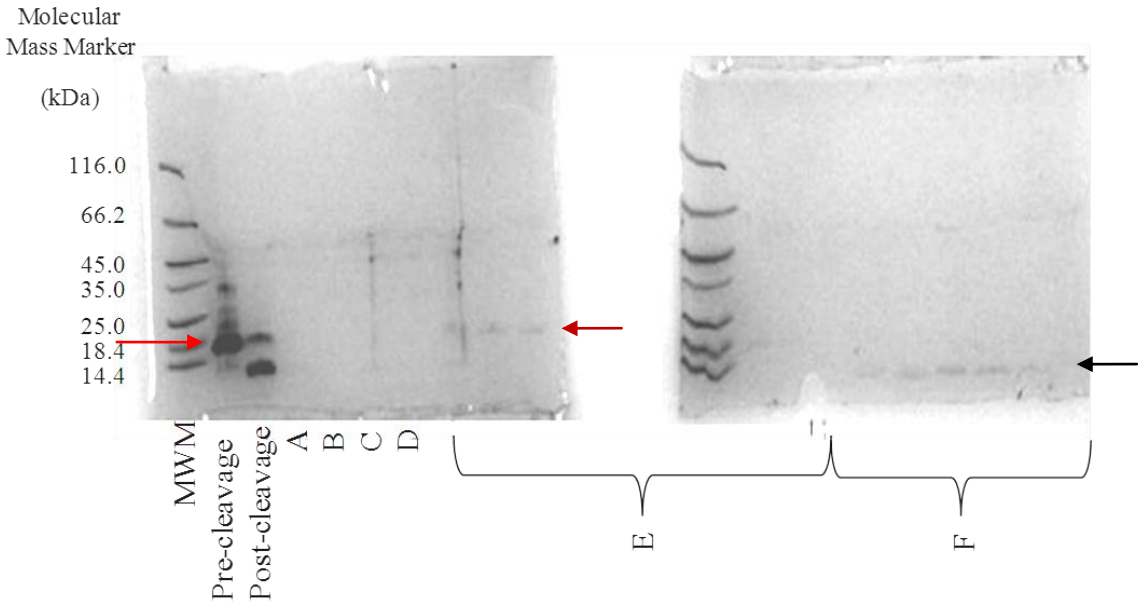
Fusion tags can be removed so as to prevent them from affecting the structure, function or behaviour of the protein in solution (Arnau *et al.* 2006). The pET-30 LIC vector contains a His-tag and an S-tag N-terminal to the FOXP2 FHD (Figure 9). In order to eliminate the possibility of the His-tag and S-tag interfering in further DNA binding studies, crystallisation trials or the folded structure of the FOXP2 FHD, we performed cleavage trials on the Y540F FOXP2 FHD variant using enterokinase to remove both the His-tag and the S-tag. Enterokinase cleavage trials were performed at 4 ^{o}C and 20 ^{o}C. The Y540F variant was incubated with different concentrations of enterokinase for various amounts of time to establish the optimal conditions for efficient cleavage of the His-tag and the S-tag (Morris 2013).The fusion protein was cleaved optimally using 0.0002 % (w/w) enterokinase for 16 hours at room temperature (20 ^{o}C). However, although cleavage occurred it was not 100 % efficient as a mixture of both cleaved and uncleaved protein was present (Figure 16A). The cleavage reaction mixture containing enterokinase, and uncleaved tagged FOXP2 FHD as well as cleaved FOXP2 FHD was loaded onto a column packed with CM-Sepharose® fast flow resin. The enterokinase did not bind to the column in 20 mM Tris-HCl, pH 8.0 (peaks A-D in Figure 16B) whereas the Y540F FOXP2 FHD variant (both cleaved and uncleaved) bound to the column. The Y540F FOXP2 FHD variant eluted with a linear gradient of 0 mM to 700 mM NaCl. The elution profile shows two peaks which

respectively correspond to uncleaved fusion protein (Figure 16B peak E) which eluted first and the cleaved fusion protein (Figure 16B peak F) which eluted afterwards. The enterokinase may have been diluted as a result of passing through the column and thus was not visible on the gel (Figure 16A).

The cleaved protein fractions were pooled and concentrated. The amount of cleaved Y540F FOXP2 FHD variant obtained was too low for further experimental work. This could be due to nonspecific cleavage which resulted in protein fragments. Furthermore, the protein may have aggregated due to the increased exposure of surface charges upon cleavage of the tag. Because of this problem, different conditions were then used to try and minimise aggregation such as incubating the fusion protein with enterokinase (0.0002% (w/w)) at 4 ^{o}C overnight. However in this case, the cleavage was inefficient and the majority of the fusion protein was uncleaved. A higher concentration of salt was used in the cleavage buffer with 0.0002% (w/w) enterokinase and incubated at 20 ^{o}C in an attempt to shield surface charges and prevent aggregation but the protein still aggregated.

In some instances, fusion tags have been shown to enhance the solubility of the fusion protein (Waugh 2005) which may be the case here. At physiological pH and temperatures above 25 ^{o}C, the S-tag has an essentially unfolded conformation (Bierzynski *et al.* 1982). This is further supported by prediction of the His-tag/S-tag region to form a flexible random coil structure by MUFOLD (Zhang *et al.* 2010) (Figure 9). The full-length FOXP2 protein has several domains (Figure 3) that are proposed to take on the 'beads on a string' conformation where the inter-domain regions are flexible. There is a flexible inter-domain region located N-terminal to the forkhead domain and is situated between the leucine zipper and FHD. The His-tag (Carson *et al.* 2007), S-tag (Kim and Raines 1993) and cleavage sites located N-terminal to the forkhead domain are unstructured and are flexible in solution, not unlike the inter-domain region between the leucine zipper and the forkhead domain in the full-length FOXP2 protein. Since the FHD naturally exists with a disordered region at its N-terminus, and it has been shown that both His-tags and S-tags have little effect on folding, structure or function of proteins (Raines *et al.* 2000; Carson *et al.* 2007), all subsequent work was performed on the tagged FOXP2 FHD so that soluble protein could be obtained in useable quantities.

A



B

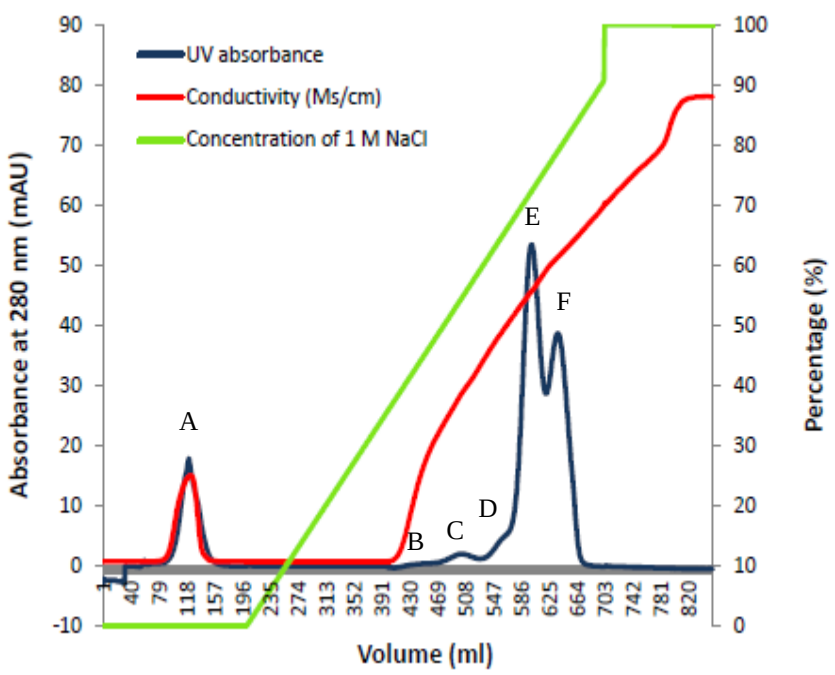


Figure 16. Cleavage of the Y540F variant fusion protein. A. 10% tricine SDS-PAGE gel. Pre-cleavage sample contains uncleaved FOXP2 FHD fusion protein (red arrow) and post-cleavage sample contains a mixture of cleaved and uncleaved fusion protein. Peaks A - D contained contaminants in low amounts that were not visible on the gel, whereas peak E contained uncleaved fusion protein (red arrows) and peak F contained cleaved fusion protein (black arrow) in small quantities. **B.** Elution profile off the CM-Sepharose® cation exchange column. 1.5 mg of fusion protein was incubated with 0.0002 % (w/w) light chain recombinant enterokinase for 16 hours at 20 °C. A sample was taken pre-cleavage and post-cleavage as shown in the 10 % tricine SDS-PAGE gel. The fusion protein-enterokinase mixture was subsequently loaded onto a CM-Sepharose column pre-equilibrated with 20 mM Tris-Cl, pH 8.0 to separate the enterokinase from the fusion protein. The column was eluted with 700 mM NaCl (red and green lines). The fractions were collected and analysed on the 15 % tricine SDS-PAGE gel.

3.2 Structures of wild-type FOXP2 FHD and the Y540F variant

3.2.1 Secondary structure

Information on the secondary structure of the WT FOXP2 FHD and Y540F variant was obtained using circular dichroism spectroscopy (Figure 17). The monomer and dimer fractions were not measured separately but at the concentration used, the proteins were predominantly, if not entirely, monomeric. The spectra were recorded over a far-UV wavelength range of 190 nm to 250 nm. The characteristic troughs of the spectra at 208 nm and 222 nm indicate that the protein is mainly α-helical and that the protein contains anti-parallel β-sheets as there is a defining peak at 195 nm (Woody 1995). The crystal structure indicates that the WT FOXP2 FHD monomer consists of four α-helices, a 3_{10} helix and three anti-parallel β-sheets (Stroud *et al.* 2006) which is consistent with the circular dichroism spectra obtained. The secondary structure of the Y540F variant differs slightly to that of WT FOXP2 FHD in that there is a decrease in the signal of the peak at 195 nm. This may indicate that the Y540F variant may contain a slightly lower amount of β-sheet structure than the WT FOXP2 FHD but is still mainly α-helical.

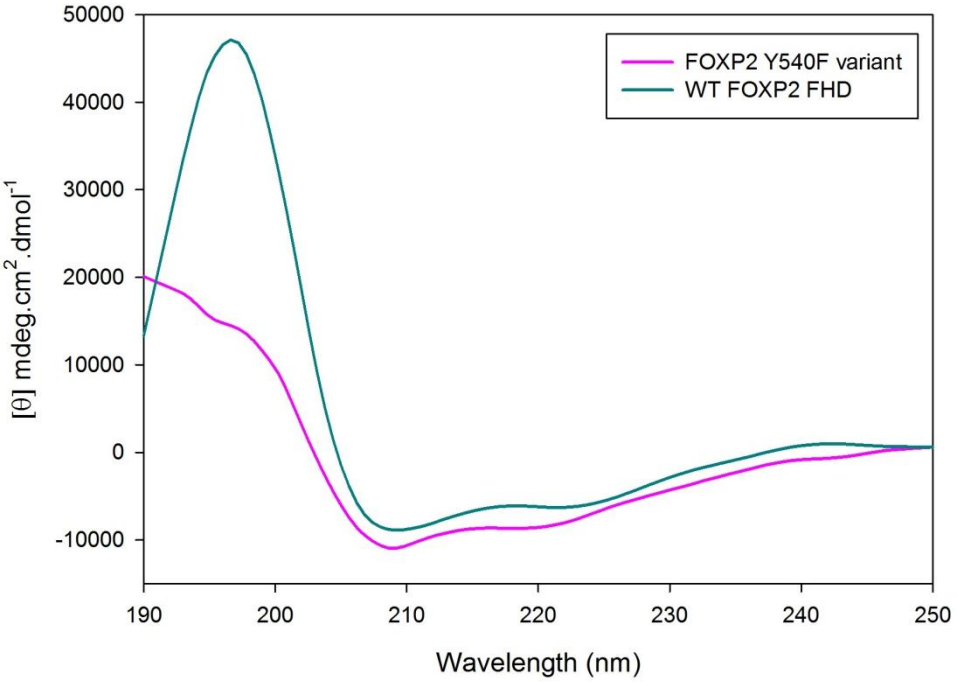


Figure 17. Secondary structure of WT FOXP2 FHD and the Y540F variant. Far-UV circular dichroism spectra of 8 μM WT FOXP2 FHD and the Y540F variant in 5 mM HEPES, 150 mM NaF, 1 mM DTT, pH 7.6. The spectra show troughs at 208 nm and 222 nm indicative of mainly α-helical protein whereas the peak at 195 nm indicates the presence of β-sheets. The spectra were converted from millidegrees to mean residue ellipticity ([θ]), corrected for buffer and smoothed.

3.2.2 Tertiary structure

The tertiary structures of the WT FOXP2 FHD and the Y540F variant were investigated using intrinsic fluorescence spectroscopy. This technique provides information on the local tertiary structure which is sensitive to the environment of the tryptophan residues when excited at 295 nm (Lakowicz 1999). There are three tryptophan residues which are located at different regions in the folded FOXP2 FHD.. In proteins with multiple tryptophan residues such as this, the fluorescence spectrum is an average of the local environments of all the tryptophans. It is difficult to separate the contribution from each tryptophan residue.

As with circular dichroism, the fluorescence studies were performed at a concentration at which the protein was predominantly monomeric. When the protein was excited at 295 nm, there was a peak at roughly 335 nm (Figure 18) for both WT FOXP2 FHD and the Y540F variant indicating no major changes to the local environment of the tryptophan residues upon mutation. The peak at 335 nm indicates that the WT FOXP2 FHD and the Y540F variant proteins are folded and that the tryptophan residues are deeply buried. A slightly lower quantum yield was observed for the WT FOXP2 FHD fluorescence spectrum compared to the spectrum of the Y540F variant. This suggests that the local environments of the tryptophan residues, and thus the tertiary structure, differ between the Y540F variant and the WT FOXP2 FHD. The Rayleigh scatter was low and indicated that there was no aggregation that would affect protein concentration. Normally fluorescence intensity is not used to show structural differences between proteins as it can vary between different days and slight changes in temperature. A more reliable measure of tertiary structural differences is the shift in wavelength of the emission and in this case there is no shift in the emission maximum wavelength between the WT FOXP2 FHD and the Y540F variant which is 335 nm. This is slightly different to that of the A39P monomeric variant of FOXP1 which is reported to show an emission maximum wavelength of 338.5 nm (Chu *et al.* 2011).

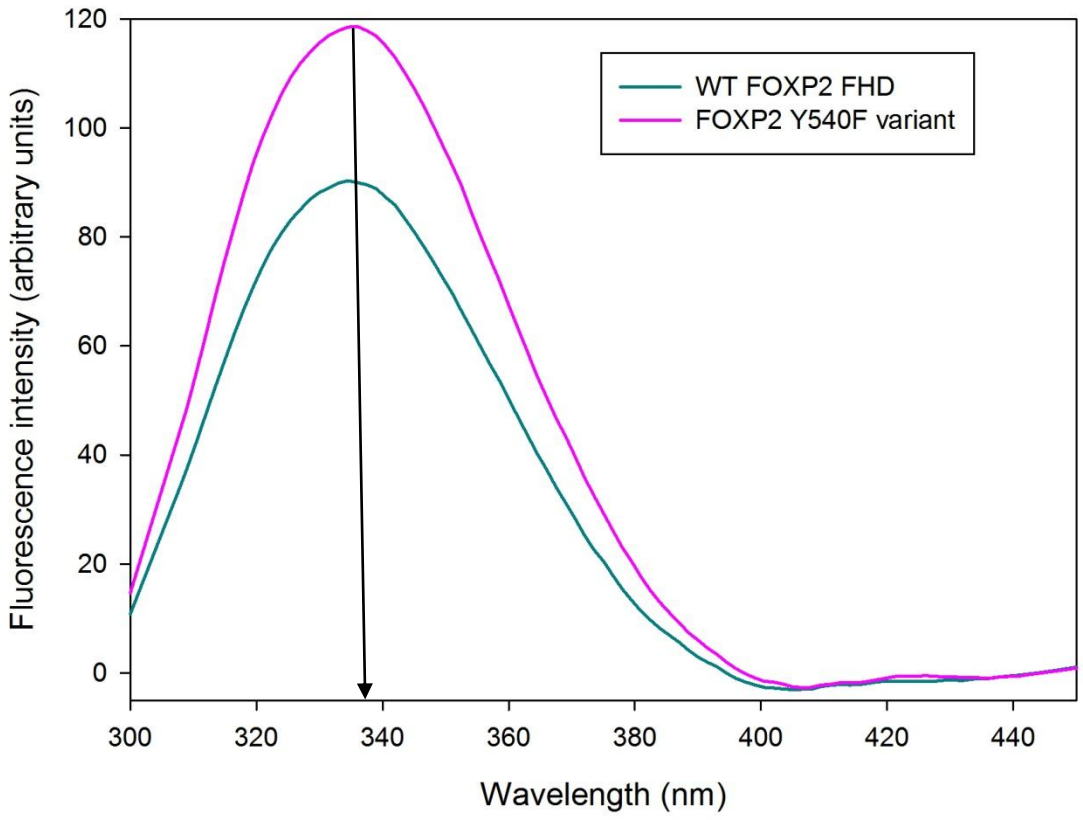


Figure 18. Tertiary structure of the WT FOXP2 FHD and the Y540F variant. Fluorescence emission spectra of 4.5 μM of both WT FOXP2 FHD and the Y540F variant (5 mM HEPES pH 7.6, 150 mM NaF and 1 mM DTT) when excited at 295 nm. The emission maximum wavelength for both the WT FOXP2 FHD and the Y540F variant is approximately 335 nm. The data was an average of 10 accumulations and was corrected for buffer and smoothed.

3.2.3 Quaternary structure

Prior to size-exclusion analysis, the quaternary structure of WT FOXP2 FHD was investigated using blue native-polyacrylamide gel electrophoresis (BN-PAGE). The FOXP2 FHD is highly positively-charged and will migrate towards the cathode in a native gel. The technique of BN-PAGE uses the anionic dye, Coomassie Brilliant Blue G-250, to induce a charge shift by binding to the protein surface, thus allowing it to migrate from the cathode to the anode. As with SDS-PAGE, BN-PAGE separates proteins according to their mass but unlike SDS-PAGE, this technique also separates proteins according to charge and furthermore, does not denature the protein. BN-PAGE therefore allows the separation of proteins in their native state and even allows proteins to be separated according to their shape or hydrodynamic volume. For example rod-shaped proteins will migrate differently to heart-shaped proteins (Fiala *et al.* 2011).

According to Figure 19A, the FOXP2 FHD migrates as two bands on the BN-PAGE gel but one band on the SDS-PAGE gel (Figure 19B) indicating that the purified protein exists in two different quaternary structures. There is also a considerable amount of laddering above these bands which is not uncommon in BN-PAGE gels (Braz and Howard 2009). The 'laddering effect' could be due to the interaction of the protein with the polyacrylamide (Braz and Howard 2009) or could be due to the formation of higher-order complexes or protein aggregation. Protein aggregation is unlikely in this case as the samples were electrophoresed under conditions to minimise aggregation, such as at low temperature and in a buffer containing 150 mM NaCl. It was also not clear whether the two BN-PAGE bands corresponded to monomer and dimer species or whether the bands represented unbound protein and protein bound to DNA. Size-exclusion chromatography and mass spectrometry were subsequently used to confirm the oligomeric state and to confirm the protein/DNA content of the bands, respectively.

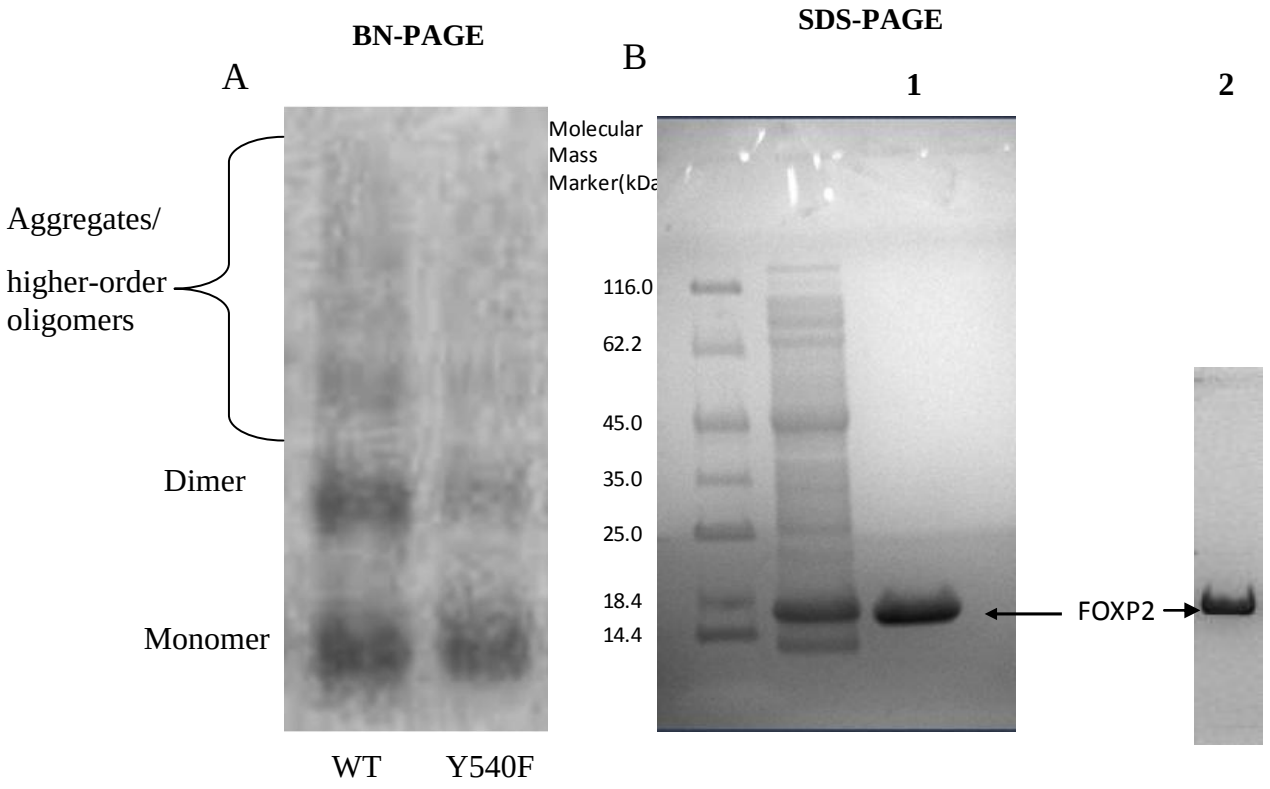
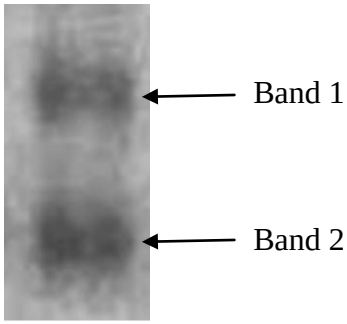
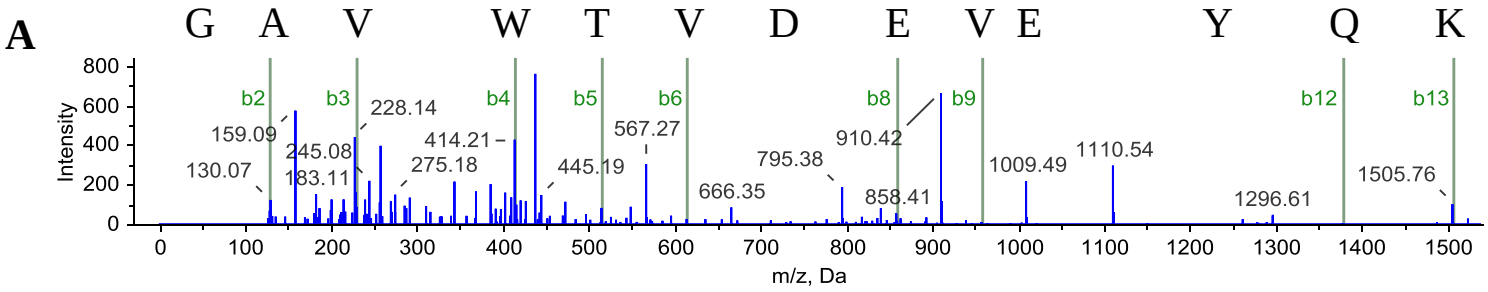


Figure 19. BN-PAGE of the WT FOXP2 monomer-dimer mixture. A. 15 % native polyacrylamide gel indicating the monomer and dimer species of 20 μM WT FOXP2 FHD and 20 μM Y540F variant. Coomassie Brilliant Blue G-250 was added directly to the sample prior to electrophoresis and there was no need to stain the gel afterwards as the bands are visible during electrophoresis. .**B.** 15 % SDS-PAGE gel of the same sample indicating one band of pure WT FOXP2 FHD (lane 1) and one band of Y540F variant (lane 2).

3.2.4 Mass spectrometry

It was still unclear at this stage whether the other bands in the BN-PAGE caused by the laddering effect were due to protein aggregation or interaction of the stain with the polyacrylamide gel. It was also unclear whether the two bands represented monomer-dimer or DNA bound/unbound protein. Two different types of mass spectrometry were used here to provide important information on this. Liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) was used to identify the two bands in the BN-PAGE gel (Figure 20) and to confirm that the protein that was purified was indeed the WT FOXP2 FHD. Matrix-assisted laser desorption/ionisation-time of flight (MALDI-TOF) mass spectrometry was used to confirm that the salt wash from the purification step had indeed eliminated all traces of DNA and to determine the exact molecular mass of the purified WT FOXP2 FHD.





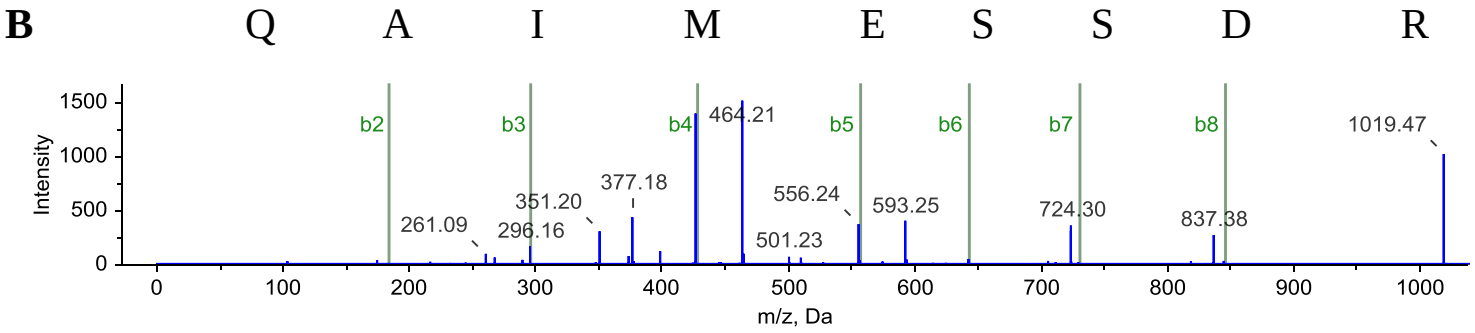


Figure 20. LC/MS-MS data from in-gel trypsin digest fragment of BN-PAGE gel (inset). Each band from the BN-PAGE gel was digested with trypsin to establish whether both bands were indeed the WT FOXP2 FHD. The sequence of the fragment from WT FOXP2 FHD is shown. **A. WT FOXP2 Band 1.** This band corresponds to a fragment from the WT FOXP2 FHD with a reported confidence of 99 % whereas the fragment from the **B. WT FOXP2 Band 2** has a reported confidence of 99 % The confidence is a probability that the sequence consists of those particular residues in that order.

In order to establish whether the two bands from the BN-PAGE gel were indeed representative of the monomer and dimer species of the WT FOXP2 FHD, an in-gel trypsin digest was done followed by LC-MS/MS (Figure 20). The LC-MS/MS data indicated that the protein in band 1 and band 2 both belonged to the WT FOXP2 FHD and that the sample was not contaminated. The confidence reported for fragments from both bands were 99 %. The confidence indicates the probability that the fragment consists of specific sequence of amino acid residues. The mass of peptide fragments generated from the digestion is compared to the mass of the known sequence fragments of the WT FOXP2 FHD.

MALDI-TOF (matrix-assisted laser/desorption-time of flight) analysis was used (i) to determine the correct molecular weight of the FOXP2 FHD as the accuracy is in the range 0.01 % (Bonk and Humeny 2001) and (ii) to ensure that no DNA fragments were still bound to the protein following purification. The 2,5 dihydroxyacetophone (DHAP) and sinapinic acid (SA) matrices were used to detect proteins as the samples ionise best using these matrices whereas the 3-HPA (3-hydroxypicolinic acid) matrix was used to detect nucleotides. Protein standards were subjected to MALDI-TOF using both the DHAP and SA matrices in order to show the difference in sensitivity between these matrices and to internally calibrate the machine so that the accurate molecular weight of FOXP2 FHD could be determined and compared to that estimated in SDS-PAGE (Figure 15). Sinapinic acid (SA) was shown to be more sensitive than the DHAP matrix to protein detection (Figure 21).

Two samples containing the (i) cognate ds-DNA and (ii) purified WT FOXP2 FHD were used. The cognate ds-DNA (12.8 kDa) shows two peaks on the 3-HPA matrix which correspond to each of the individual DNA strands (Figure 22A). The two peaks had different masses because the two strands consist of complementary sequences and are thus not identical. The DHAP and SA matrices did not detect any significant peaks in the DNA sample which firstly confirms that the DNA was pure and there was no protein contamination and secondly, provides a negative control to show that these matrices do not detect nucleotides (Figure 22A). The purified WT FOXP2 FHD sample was ionised on both the SA and the 3-HPA matrices to confirm the correct molecular weight that all cellular DNA contaminants had indeed been removed. Mass spectrometry of the WT FOXP2 FHD using the 3-HPA matrix (Figure 22B) indicates that no DNA was still bound to the protein following purification and that the 1.5 M salt wash during the HisTrap purification was sufficient to remove DNA.

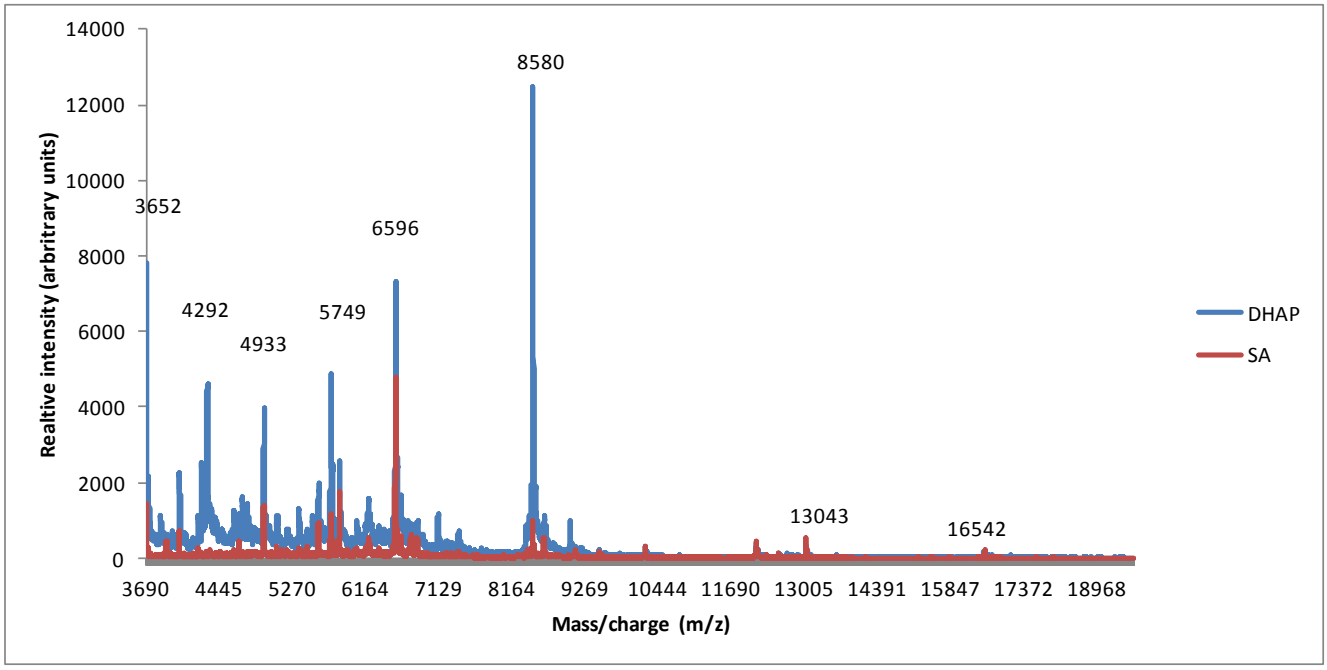


Figure 21. MALDI-TOF of protein standard mix. The protein standards were used to internally calibrate the machine for accurate molecular weight determination of the FOXP2 FHD. DHAP and SA matrices were used to ionise the sample. It is indicated that the SA matrix is more sensitive than the DHAP matrix in protein detection.

The spectrum shown for the WT FOXP2 FHD in SA indicates that the analyte has a mass/charge ratio of 14 667 (Figure 22) when it has a charge of $[M+1]^+$ and a mass-to-charge ratio of 7442 when the analyte has a charge of $[M+2H]^+$. The $[M+2H]^+$ species corresponds to the sample being doubly-protonated (mass-to-charge ratio of 1:2) and its formation is due to the way in which the sample ionises with the particular matrix (Zaluzec *et al.* 1995). The doubly-protonated species usually forms at a lower abundance than the singly-protonated species ($[M+H]^+$) (Zaluzec *et al.* 1995). The $[M+H]^+$ corresponding to 9993 is lower and is likely to be a contaminant. It may not have been visible on the SDS-PAGE gel as it was probably present in a very small amount (less than 5%) but was detected by the SA matrix as it is highly sensitive for proteins. The molecular weight determined by MALDI-TOF is 14.667 kDa is highly accurate and does not significantly differ to the predicted molecular weight of 14.683 kDa (Gasteiger *et al.* 2003) of the tagged protein, therefore confirming that we have purified the FOXP2 FHD.

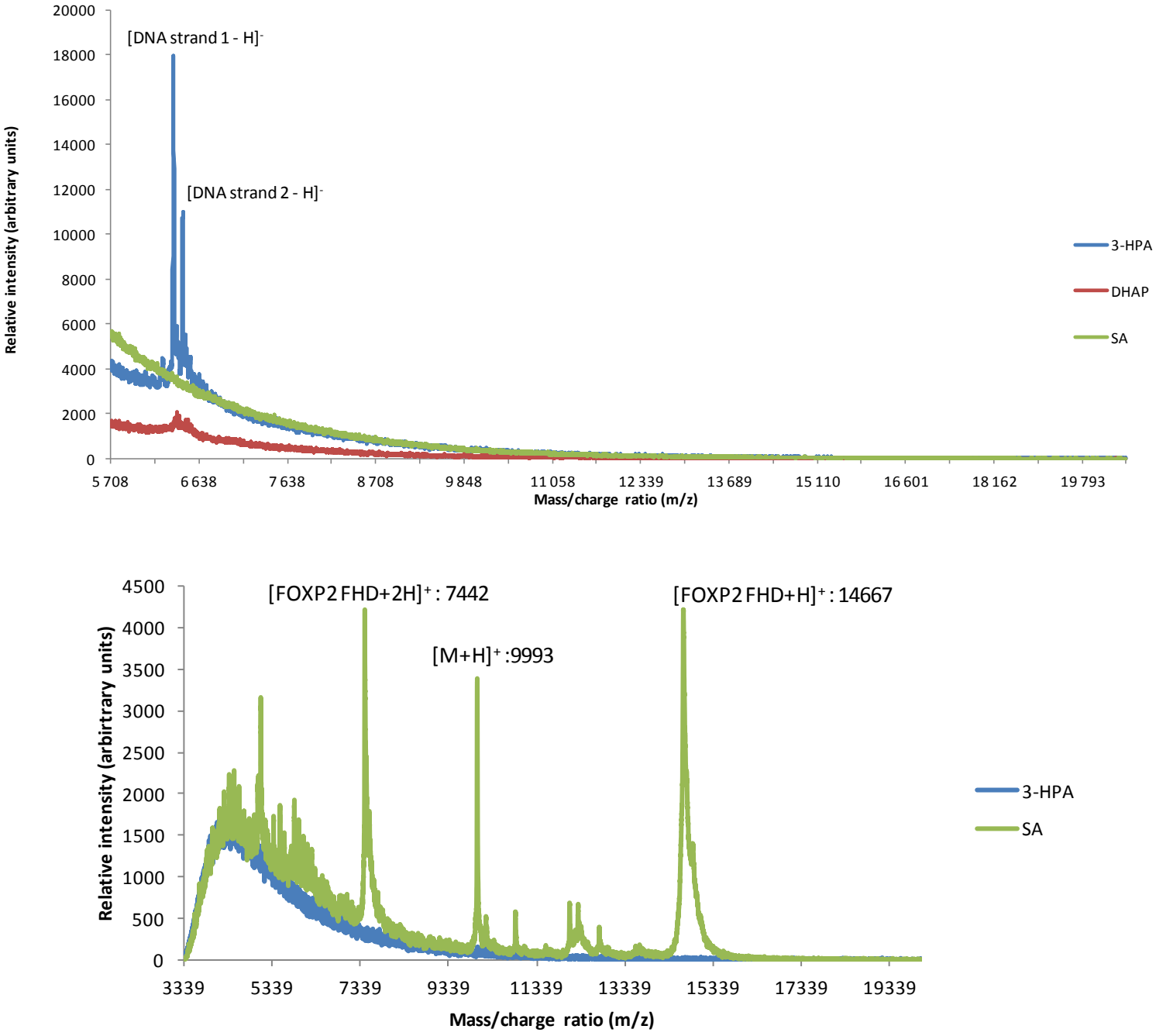


Figure 22. MALDI-TOF analysis. A. 21 bp cognate DNA. The 3-HPA matrix is used to detect nucleic acids and it is shown by the double-stranded cognate DNA (12.8 kDa) which is dissociated into two individual strands $[M-H]^-$ of 6.3 kDa and 6.4 kDa. The other matrices, DHAP and SA are used to detect proteins and to not pick up the cognate DNA. **B**. WT FOXP2 FHD. The analysis was carried out using the SA matrix to determine and confirm the size of the FOXP2 FHD which is shown to correspond to the monoisotopic mass of 14667 Da which is not significantly different to the estimation by SDS-PAGE of 14.9 kDa. The mass/charge ratio when the analyte has a charge of 2^+ is 7442. The 3-HPA matrix is used to show if any contaminating DNA fragments were still bound to the protein after purification. There are no significant peaks and thus no contaminating DNA present. The $[M+H]^+$ peak that corresponds to 9.9 kDa may be a protein contaminant that was too dilute to be visible on an SDS-PAGE gel.

3.3 DNA binding

The function of the FOXP2 FHD is to bind to DNA in order to regulate gene expression. Thus we were interested to see if the mutation affected DNA binding. Intrinsic fluorescence was used to study protein conformational changes upon binding to DNA for WT FOXP2 FHD and the Y540F variant. The advantage of this technique is that lower concentrations can be used than in non-radioactive electrophoretic mobility shift assays. Furthermore, it is highly sensitive to changes in the environment of the tryptophan residues especially if they are involved in the binding reaction itself (Kneale 1994). Although fluorescence is commonly quenched upon DNA binding (Kneale 1994), a study by Tipper and Wagner (1995) has shown that the fluorescence intensity can also increase upon binding. In this study, the increase in fluorescence indicated specific binding to a sequence because the fluorescence intensity decreased upon binding to nonspecific poly (dI/dC). They propose that when the positively-charged protein residues make specific contacts with the negatively-charged DNA, the microenvironment of the tryptophan residue changes (Tippner and Wagner 1995).

In this study, the fluorescence intensity is shown to be substantially enhanced for both the WT FOXP2 FHD (Figure 23) and the Y540F variant when bound to DNA. Furthermore, the fluorescence intensity of the Y540F variant is enhanced approximately four-fold upon binding to DNA compared to the WT FOXP2 FHD (Figure 23C). The change in fluorescence between bound and unbound protein indicates that the microenvironment of the tryptophan residues is altered in the presence of DNA. In the DNA-bound FOXP2 FHD, the Trp533 is located in close proximity to the DNA backbone and is surrounded by positively-charged lysine and arginine residues (Figure 24) whereas Trp548 and Trp572 are far away from the DNA binding site, thus the change in signal upon binding is likely to be mainly representative of the Trp533. Arginine and lysine residues are known to quench tryptophan fluorescence (Lakowicz 1999). The positively-charged Arg564 may interact with Trp533 through cation-pi contacts and may quench the fluorescence of Trp533 in the unbound state. Upon DNA binding, the Arg564 residue forms a direct bond with the DNA (Stroud *et al.* 2006), allowing the Trp533 to fluoresce. This accounts for the increase in fluorescence upon binding DNA. There is no shift in the emission maximum wavelength between the bound and unbound WT FOXP2 FHD and the Y540F variant which suggests that no major structural changes occurred to either protein upon binding to DNA.

In addition to intrinsic fluorescence, electrophoretic mobility shift assays (Figures 25 and 26) were used to assess the functionality of WT FOXP2 FHD and the Y540F variant. In this technique, the electrophoretic mobility of free DNA is greater than that of the protein-DNA complex, simply because it has a lower molecular weight. As a result, the free DNA will migrate further in the gel than the

protein-DNA complex (Hellman and Fried 2007). Different ratios of protein were mixed with DNA to determine the optimal protein/DNA ratio for binding. A protein/DNA ratio of 3:1 was found to be optimal (Figure 25) and this corresponds with what had been previously reported for WT FOXP2 (Stroud *et al.* 2006). According to the electrophoretic mobility shift assay (EMSA) in Figure 25, a protein/DNA ratio of 1:1 indicated that the WT FOXP2 FHD did indeed bind to DNA. Thus, for the fluorescence of bound and unbound WT FOXP2 FHD and the Y540F variant, a protein/DNA ratio of 1:1 was used. A key difference between the fluorescence and the EMSAs is the concentration of protein. Much higher concentrations were used for the EMSAs and the protein was far more likely to exist as a mixture of monomer and dimer species in the EMSAs than in the fluorescence work.

The electrophoretic mobility shift assay shows one band corresponding to unbound DNA at the bottom and two retarded bands. These two bands could represent monomer and dimer bound to DNA. Figure 25 shows that the WT FOXP2 FHD interacts slightly more favourably with DNA than the Y540F variant. Densitometry analysis in Figure 26 indicates that although there is no significant change to the total amount of DNA bound by both the WT FOXP2 FHD and the Y540F variant, the amount of DNA bound by the monomer and dimer fractions differs in the WT FOXP2 FHD and the Y540F variant. This cannot be quantified due smearing of the bands. Visual inspection of the gel seems to suggest that the mutation has altered the affinities of the monomer and dimer fractions for DNA with the dimer fraction binding to more DNA in the Y540F variant than the wild-type. This could also imply that there is a greater proportion of dimer present in the Y540F variant compared to the wild-type. This was investigated using SEC.

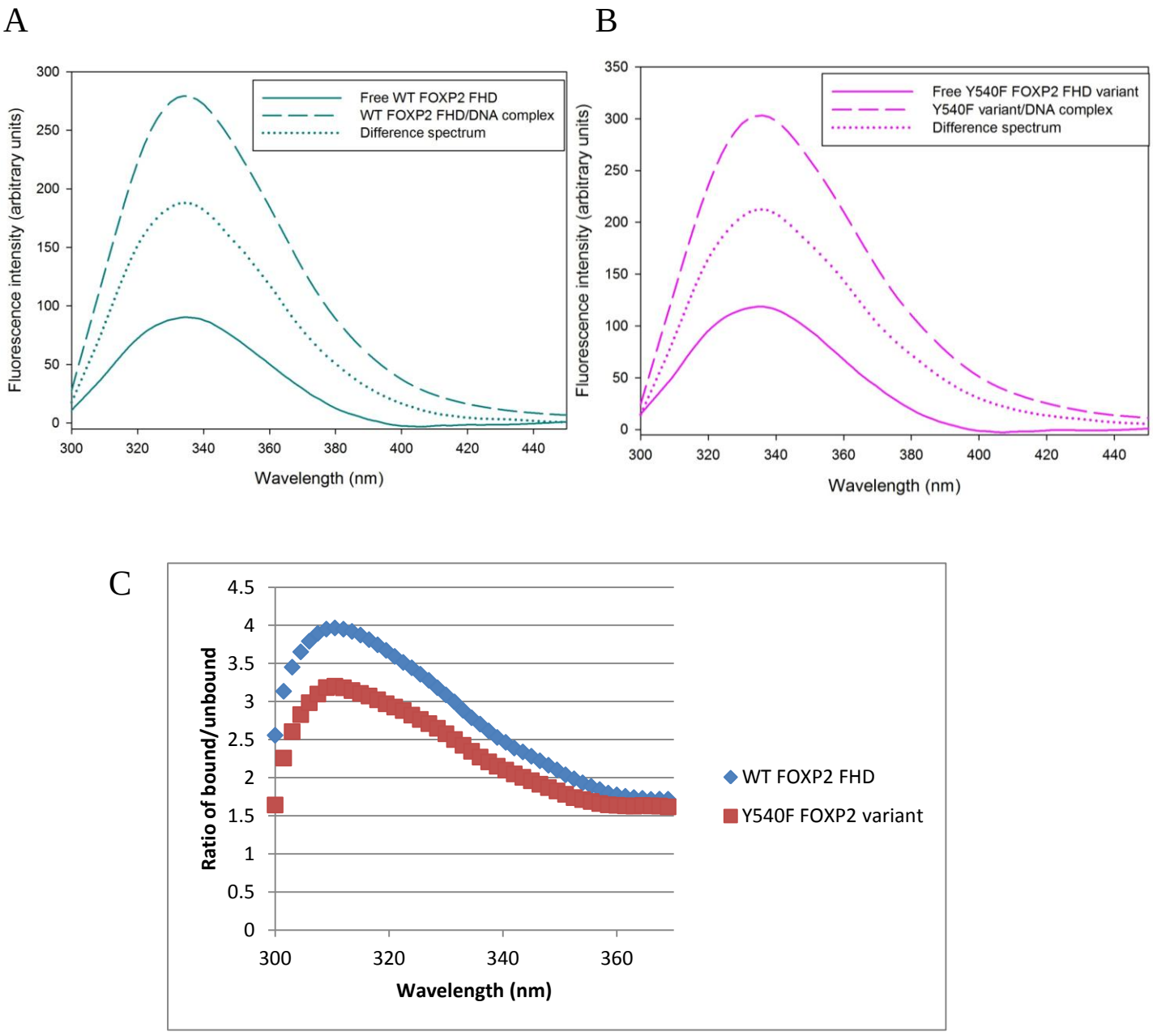


Figure 23.Tryptophan fluorescence emission spectra of free FOXP2 and FOXP2/DNA complexes. The spectra were recorded for WT FOXP2 FHD (A) and the Y540F variant (B) in 20 mM Tris-Cl pH 7.6, 150 mM NaCl and 1 mM DTT after excitation at 295 nm at 20 ^{o}C. Protein and DNA was mixed to give a final concentration of 4 µM protein: 4µM DNA. The protein is mainly monomeric at this concentration.The data were corrected for buffer and smoothed. The fluorescence intensity of the ratio of DNA-bound/unbound were plotted (C) for both the WT FOXP2 FHD and the Y540F variant to establish whether the mutation had affected DNA binding. The fluorescence of the WT FOXP2 FHD is enhanced 3.3-fold compared to the Y540F variant which is enhanced approximately 4-fold.

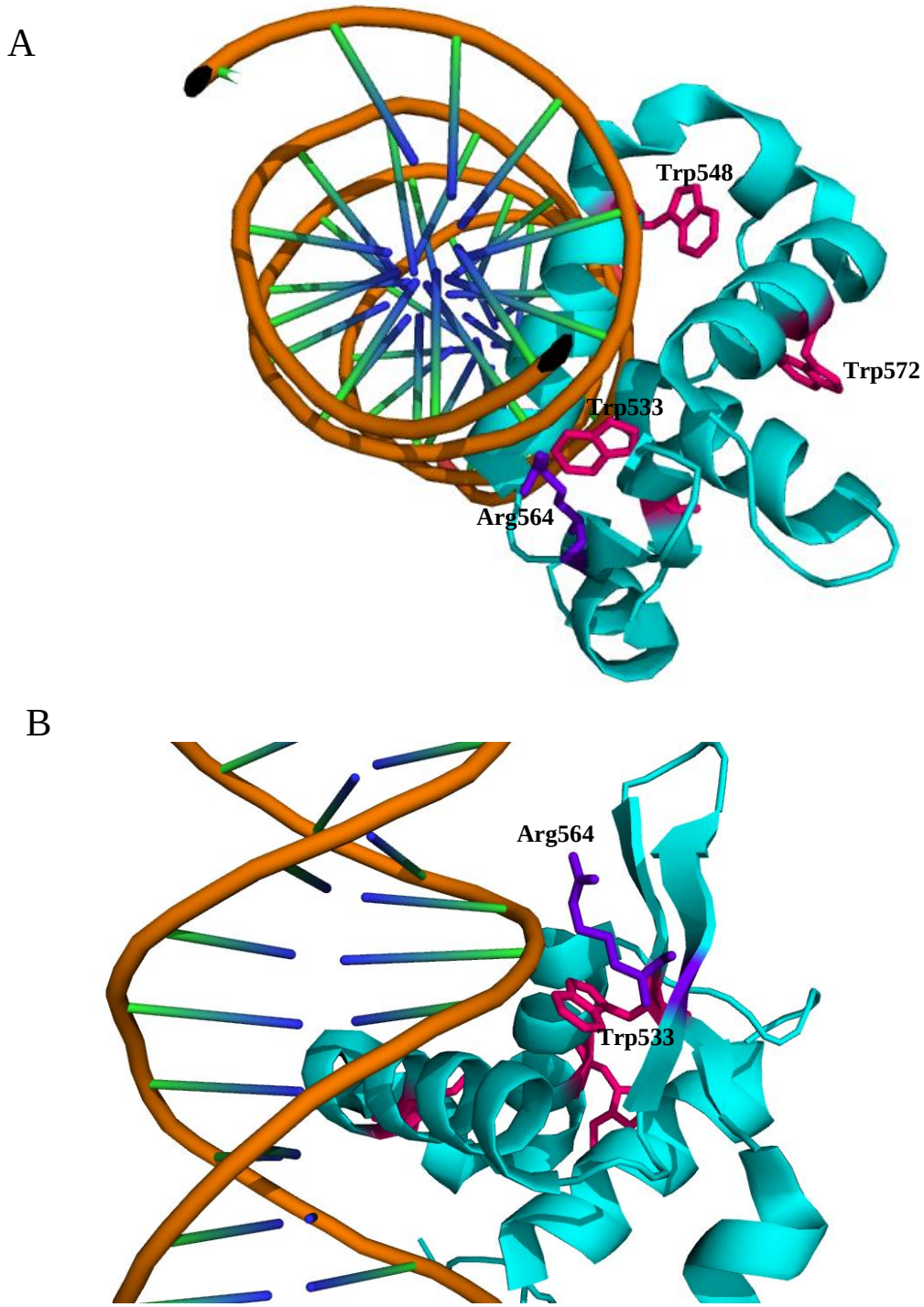


Figure 24. Location of the Trp residues in the FOXP2 FHD. A. The FOXP2 FHD (cyan) contains three tryptophan residues (pink). **B.** In the FOXP2 FHD monomer (cyan), the Trp533 residue (pink) is in close proximity to the DNA backbone (brown) and is in close proximity to Arg564 (purple). In the unbound state, Trp533 is likely to participate in cation-pi interactions with Arg564 which quenches fluorescence. Upon binding to DNA, however, Trp533 is able to fluoresce as Arg564 interacts directly with a phosphate group in the DNA backbone (Stroud *et al.* 2006).

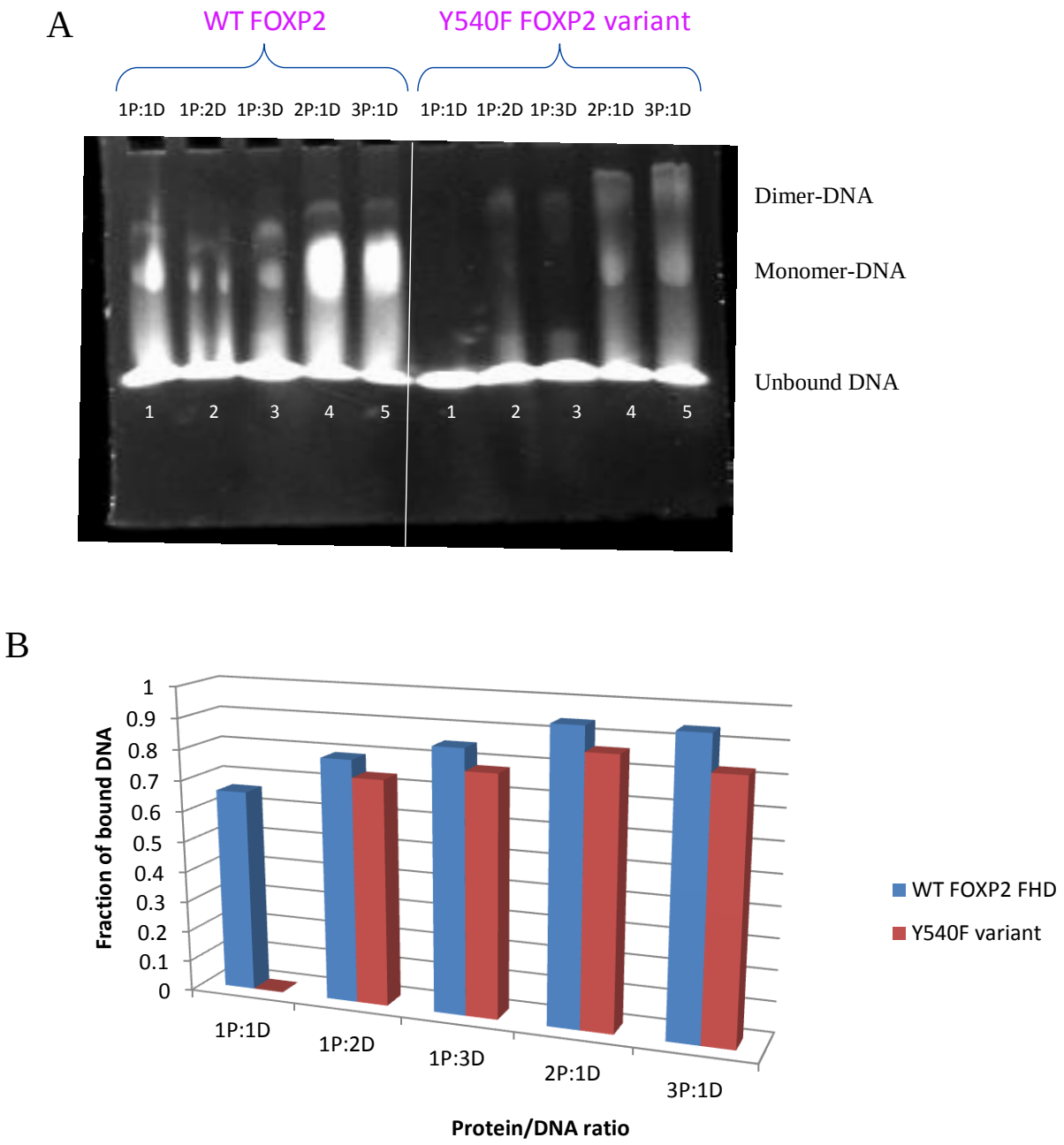
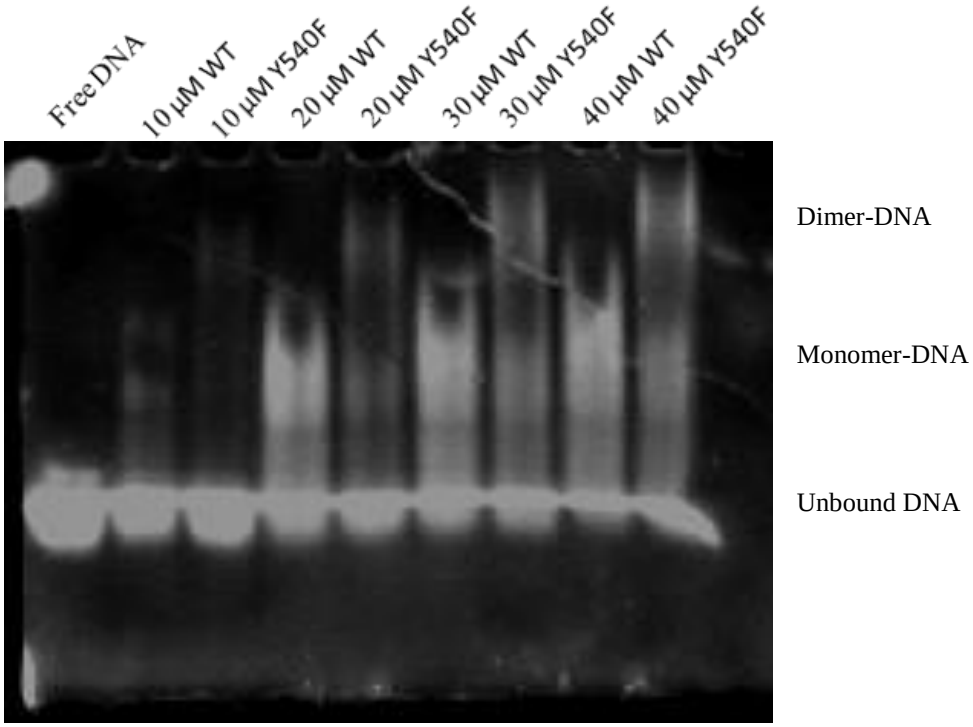


Figure 25. Electrophoretic mobility shift assay (EMSA) to determine optimal protein/DNA ratio. A. Native 15 % polyacrylamide gel depicting the protein-DNA complexes formed at different protein/DNA ratios for WT FOXP2 FHD and the Y540F variant in the presence of DNA. 1P:1D is equivalent to 10 µM protein: 10 µM DNA. Protein was mixed with the cognate DNA in 100 mM Tris-Cl pH 7.6, 100 mM NaCl, 1 mM DTT and 1 mM $MgCl_2$ for 1 hour on ice prior to electrophoresis on ice at 50 V for 2.5 hours. The gel was subsequently stained for DNA with ethidium bromide. **B. Densitometry analysis.** LabWorks (version 4.6) was used to calculate the percentage of bound DNA off the EMSA gel. The proportion of bound DNA was calculated using: fraction of bound DNA = 1- fraction of unbound DNA where the densitometry on the bottom band gives the fraction of unbound DNA. DNA binding is most favourable under conditions where protein is in excess but specifically in a 3:1 protein/DNA ratio. WT FOXP2 FHD binds slightly more favourably to DNA than does the Y540F variant at all ratios tested.

A



B

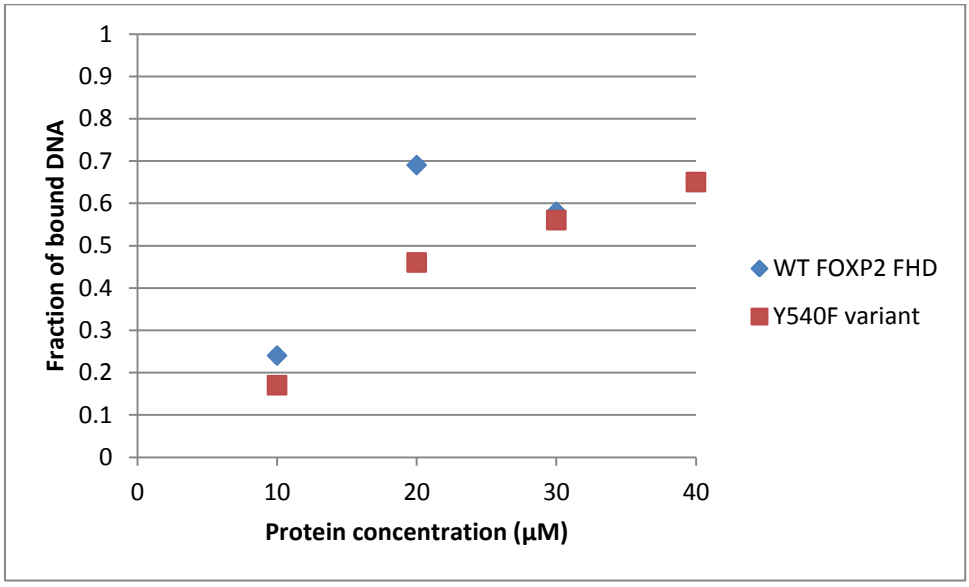


Figure 26. EMSA showing titration of cognate DNA with WT FOXP2 FHD and Y540F variant proteins. A. 15 % native polyacrylamide gel. All lanes contain cognate DNA at 10 µM. Lane 1 contains free cognate DNA; lanes 2, 4, 6 and 8 contain 10 µM, 20 µM, 30 µM and 40 µM of WT FOXP2 FHD, respectively. Lanes 3, 5, 7 and 9 contain 10 µM, 20 µM, 30 µM and 40 µM of the Y540F variant, respectively. The protein was mixed with DNA and incubated for 1 hour on ice prior to electrophoresis on ice at 100 V for 1.5 hours. The gel was subsequently stained for DNA with ethidium bromide. **B. Densitometry analysis**. The fraction of bound DNA was calculated using: fraction of bound DNA = 1 - fraction of unbound DNA where fraction unbound DNA is given by the densitometry of the bottom band in the EMSA gel. Total amount of DNA bound appears to be unaffected by the mutation but the ratio of bound monomer:dimer may be affected.

3.4 Propensity of the FOXP2 FHD to dimerise

Size-exclusion chromatography was used to study the concentration-dependence of the monomer-dimer equilibrium of WT FOXP2 FHD and the Y540F variant. This technique can also be used to estimate the sizes of proteins by using protein standards and constructing a calibration curve (Figure 27). Increasing concentrations of WT FOXP2 FHD and Y540F variant proteins were loaded onto a 16/600 Superdex 75 pg column (GE Healthcare) and the relative amounts of monomer and dimer were detected with absorbance at 280 nm (Figure 28A). The molecular weight was calculated to be 22 kDa for the monomer and 35 kDa for the dimer. Mass spectrometry showed the size of the monomer to be 14.67 kDa (Figure 22) which is slightly lower than the value obtained from SEC. In SEC, macromolecules are separated according to their molecular weight in addition to their hydrodynamic volume. The hydrodynamic volume is dependent upon the shape of the protein as it migrates through the beads and can interfere with the calculation of the correct molecular weight of proteins. The shift in equilibrium from monomer to dimer with increasing protein concentrations was monitored. The WT FOXP2 FHD remained exclusively monomeric al all concentrations tested up to 100 μM. The Y540F variant, however, displayed an increase in the amount of dimer relative to monomer as the concentration increased (Figure 28B).

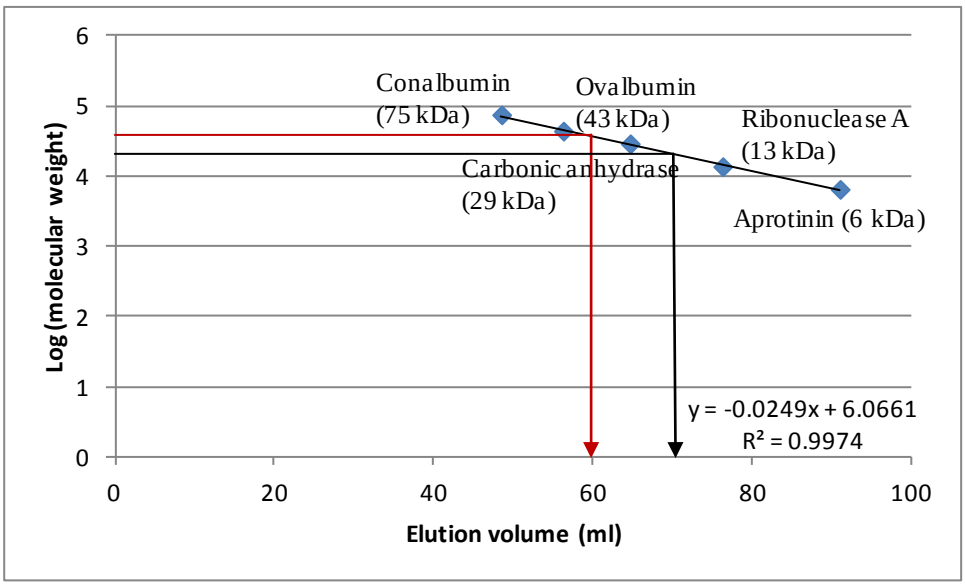


Figure 27. Molecular weight calculation using size-exclusion chromatography. The following protein standards were subjected to size-exclusion chromatography on a Hiload 16/600 Superdex 75 pg column (GE Healthcare): conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa) (GE Healthcare) and were used to calculate the molecular weight of the FOXP2 FHD monomer (black arrow) (22 kDa) and dimer (red arrow) (35 kDa) from their elution volume.

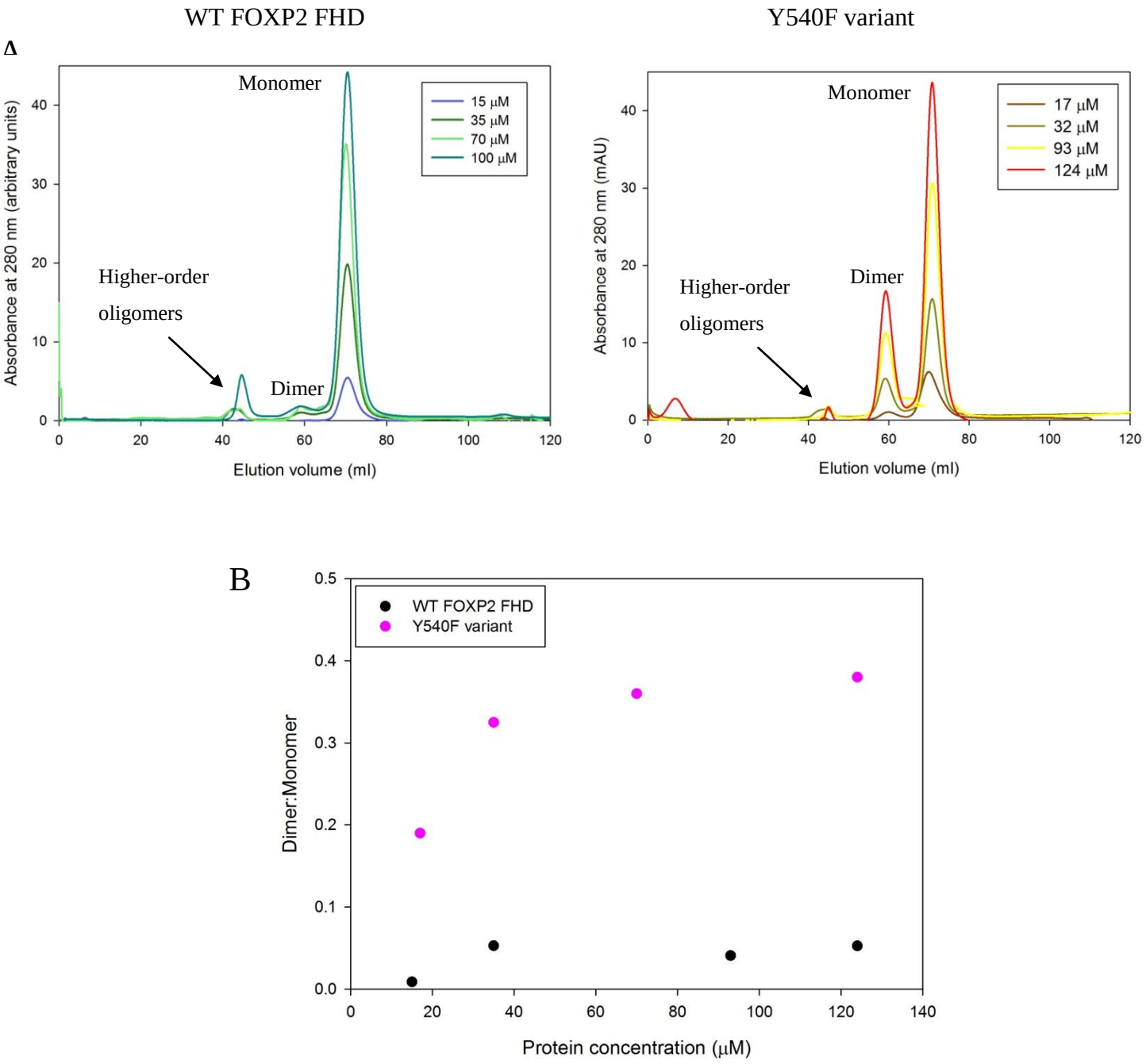


Figure 28. Propensity of the FOXP2 FHD to dimerise. A. Size-exclusion elution profiles of the WT FOXP2 FHD (left) and Y540F variant (right). Increasing concentrations of protein were subjected to size-exclusion chromatography using a 16/600 Hiload Superdex 75 prep grade column (GE Healthcare) equilibrated with 20 mM Tris-Cl pH 7.6, 150 mM NaCl, and 1 mM DTT at 20 ^{o}C. The concentrations loaded onto the column for WT FOXP2 FHD are 15 μM, 35 μM, 70 μM and 100 μM and for the Y540F variant: 17 μM, 32 μM, 93 μM and 124 μM. **B. Scatter plot of the dimer:monomer ratio** shows the amount of dimer relative to monomer. The ratio of dimer:monomer shows no significant increase for the WT FOXP2 FHD while the relative amount of Y540F dimer increases with increasing protein concentration.

3.5 Tertiary structures of the monomer and dimer

In order to determine whether the tertiary structure is altered in the monomer compared to the dimer, fluorescence spectra (excitation at 295 nm) of WT FOXP2 FHD and the Y540F variant were collected for the separated monomer and dimer fractions as they were eluted off the SEC column (Figure 28). Since the concentrations of the monomer and dimer obtained off the size-exclusion column were not the same and the concentrations could not be changed without influencing the monomer-dimer equilibrium, the intensity was normalised between 0 and 1 and only shifts in the emission maximum wavelength were monitored.

The spectrum for the WT FOXP2 FHD dimer could not be collected as the protein did not form a significant proportion of dimer even at 100 μM (Figure 28B). The spectrum for the WT FOXP2 FHD monomer peaks roughly at 335 nm compared to the Y540F variant monomer which peaks at 336 nm and the Y540F variant dimer which peaks at roughly 338 nm (Figure 29). The emission maximum for the Y540F variant dimer is thus red-shifted compared to the monomer. A red-shifted emission maximum indicates that the tryptophan residues in the Y540F variant dimer are more exposed to the solvent in the tertiary structure (Lakowicz 1999). This data suggests that the local environment of the tryptophan residues, and thus the tertiary structure, of the variant FOXP2 FHD monomer is different to that of the variant FOXP2 FHD dimer. It is not clear how they differ as detailed information on the structure can be provided only by solving the 3-D structure of the Y540F variant using x-ray crystallography.

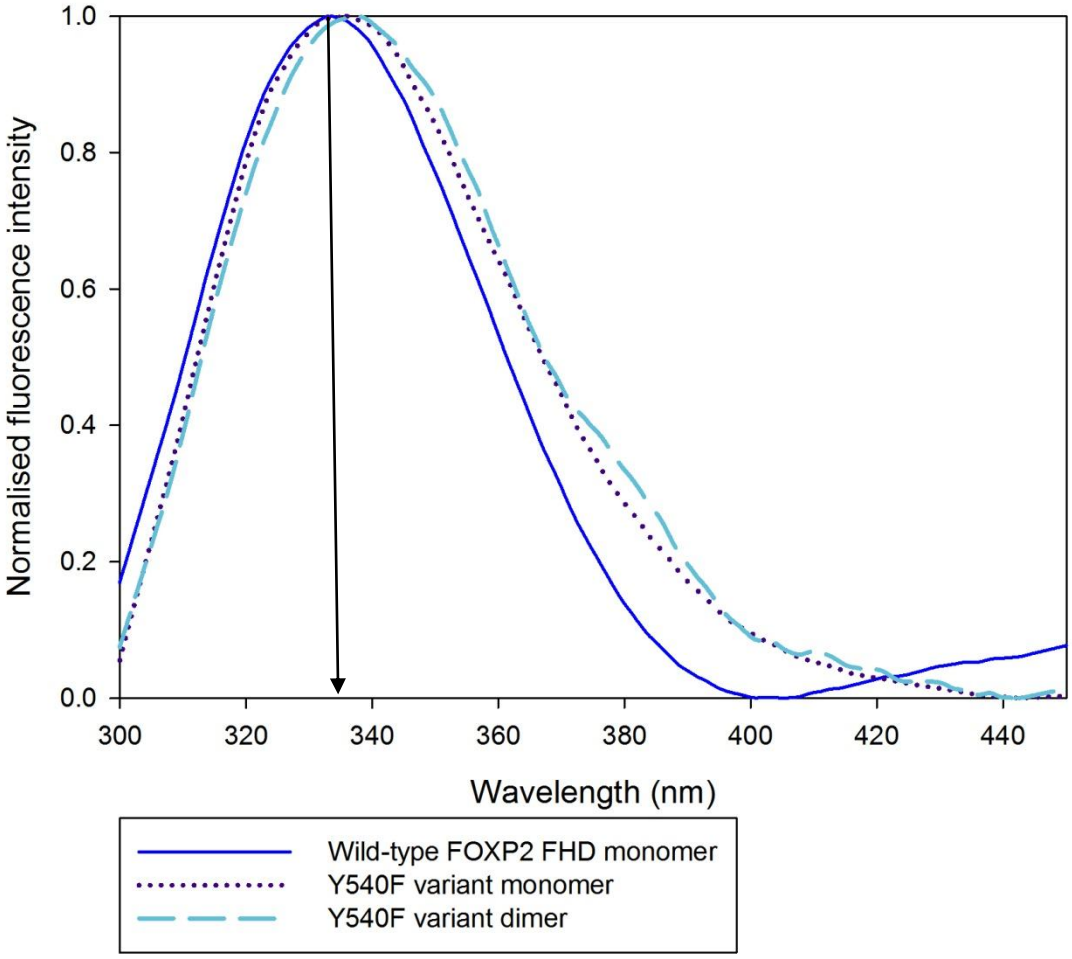


Figure 29. Fluorescence emission spectra of the purified monomer and dimer species of FOXP2 FHD. Monomer and dimer fractions were collected subsequent to size-exclusion chromatography. Fluorescence spectra were collected for monomer and dimer separately for WT FOXP2 FHD monomer (solid blue line), the Y540F variant monomer (dotted line) and the Y540F variant dimer (dashed line) at 20 ^{o}C 20 mM Tris-Cl pH 7.6, 150 mM NaCl, 1 mM DTT. The excitation wavelength used was 295 nm. Data were corrected for buffer, the intensity was normalised between 0 and 1, and then smoothed.

3.6 Crystals of FOXP2 Y540F FHD in the presence of DNA

The basic principle of protein crystallisation is to have a pure, homogenous protein sample at a concentration just below its solubility limit. The protein is gradually brought to supersaturation with the aid of various precipitating agents such as polyethylene glycol (PEG), organic solvents such as 2-methyl-2, 4-pentanediol (MPD) and salts such as ammonium sulphate. Salt does so by forming electrostatic interactions with the solvent which is no longer available to interact with the protein and thus forces the protein to interact with itself due to exposed surface charges which causes salting out. PEG molecules have chains of varying lengths and relatively large radii and serve to exclude water or the solvent from interacting with the protein. Organic solvents lower the dielectric constant of the solvent and in doing so lower protein solubility thus promoting protein-protein interactions (Drenth 1999). In the supersaturated state, small stable nuclei form which facilitate the growth of larger protein crystals by the addition of new molecules to the crystal faster than they are dissolved in the solvent. Protein crystals can be grown in a variety of different buffers, precipitants, salts, organic solvents or additives so screening is important to determine the optimal conditions for crystal growth.

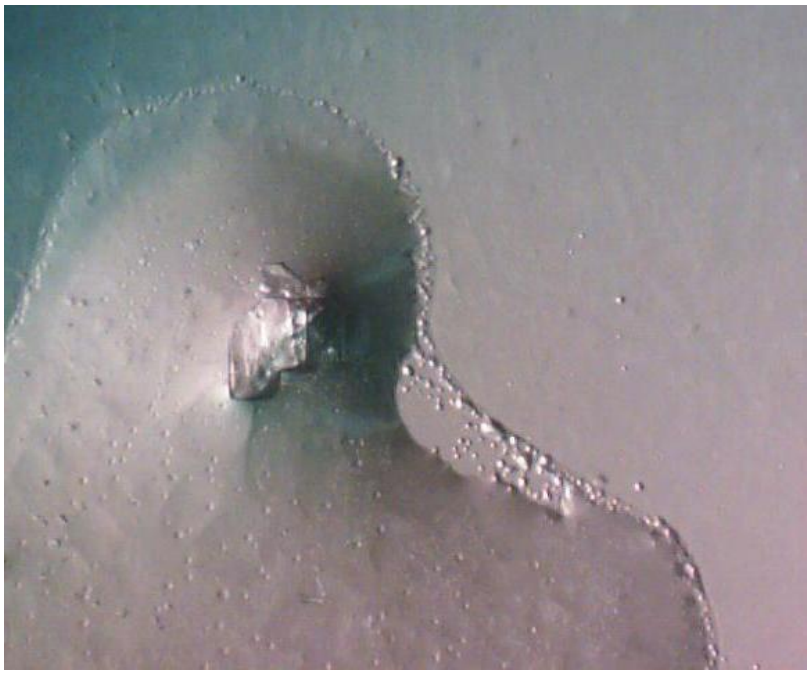
Crystallisation of proteins in the absence of additional elements presents a challenge since variables such as purity, homogeneity and other factors play a crucial role in determining whether crystals will form. Therefore, the presence of DNA in the crystallisation process adds another element of difficulty (Hollis 2007). Additional variables such as the protein/DNA ratio as well as length and composition of the DNA sequence must be taken into consideration because they can determine whether crystals will form or not. However, various additives such as carbohydrates, polyamines, alcohols and other compounds are sometimes required for nucleation and crystal growth and may improve the quality of x-ray diffraction of protein-DNA crystals (Sauter *et al.* 1999).

Crystal trials are a largely trial-and-error process but if similar proteins, such as those within the same family, have been crystallised then those conditions can be used as initial conditions. Once a ‘hit’ has been obtained, the conditions can be further optimised if necessary by altering the pH, temperature or protein or precipitant concentration to obtain better crystals.

The FOXP2 Y540F variant formed crystals in approximately six months in the presence of its cognate DNA binding sequences (5’-AACTATGAAACAAATTTTCCT-3’ and 5’-TTAGGAAAATTTGTTTCATAG-3’) using the hanging drop method at a final protein concentration of about 8 mg/ml. The crystals were grown with a protein/DNA ratio of 3:1 similar to Stroud et al., (2006) at 20 ^{o}C using the reservoir buffer: 0.1 M Bis-tris pH 6.5, 0.2 M calcium chloride dihydrate, 45 % (v/v) 2-methyl-2,4-pentanediol (MPD) (Figure 30). 45 % MPD is convenient because it also acts as a cryoprotectant and addition of further cryoprotectant was not necessary prior to snap-freezing the

crystals in liquid nitrogen. Further optimisation is needed however, as these crystals may be too small for diffraction.

A



B

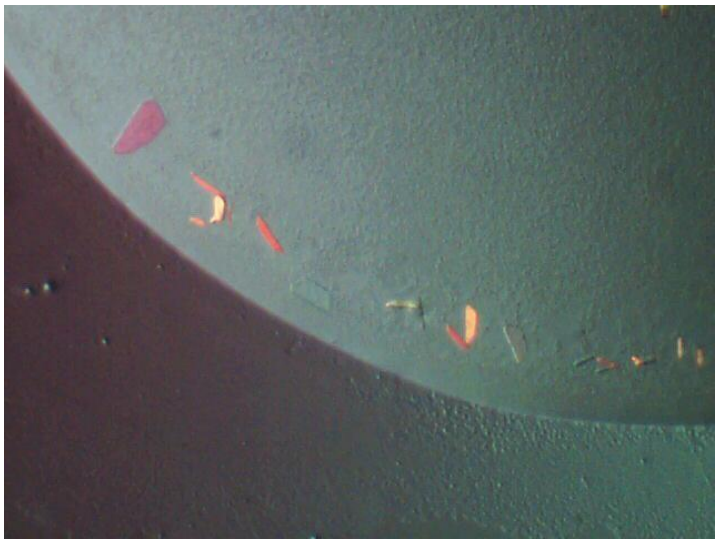


Figure 30. Crystals of the FOXP Y540F variant in the presence of DNA. A. A hexagonal-shaped crystal and **B.** crystal plates were grown in 0.1 M Bis-tris pH 6.5, 0.2 M calcium chloride dihydrate, 45 % (v/v) 2-methyl-2,4-pentanediol using a protein concentration of approximately 8mg/ml and a protein/DNA ratio of 3:1. Crystals took approximately 6 months to grow.

CHAPTER 4: DISCUSSION

4.1 Phe540 does not significantly alter DNA binding but promotes dimerisation

The hinge-loop region is one of the most diverse parts of the FHD among the FOX proteins, along with wing 1 and the C-terminus. The hinge region exists in different conformations among the FOX proteins. In FOXM1 (Littler *et al.* 2010), FoxD3 (Marsden *et al.* 1998), FoxC2 (van Dongen *et al.* 2000), FoxO4 (Weigelt *et al.* 2001), FoxQ1 (Sheng *et al.* 2002) and FOXP2 (Stroud *et al.* 2006), the hinge region forms a 3_{10}-helix, whereas in FOXK1a it consists of a type-I turn (Tsai *et al.* 2006) and in FOXO3a (Tsai *et al.* 2007) it forms a solvent-exposed loop structure with an insertion of five additional residues. It is speculated that the conformational differences in the hinge-loop region between the FOX proteins may regulate the DNA binding specificity by specifying the orientation and presentation of the recognition helix (H3) to DNA (Marsden *et al.* 1998). From the FOXO4-DNA structure, it is proposed that the hinge region affects DNA binding specificity through hydrophobic interactions of residues in the hinge region Tyr133 and Phe134 (Tyr540 and Phe541 in FOXP2) with residues Trp97 and Tyr102 at the N-terminus (Arg504 and Tyr509 in FOXP2) and Trp146 (Trp548 in FOXP2) in the recognition helix (H3). In further support of the role of the hinge in DNA binding specificity, the FOXK1a structure in the presence of DNA provides the first direct evidence of the hinge region in contact with DNA (Tsai *et al.* 2006).

The EMSAs (Figures 25 and 26) indicate that DNA binding of the Y540F variant is slightly diminished compared to the wild-type but not completely abrogated. This suggests that although the Tyr540 residue does not play a direct role in DNA binding in the FOXP family (Stroud *et al.* 2006; Bandukwala *et al.* 2011), it may be implicated in stabilising the protein-DNA interaction. Intrinsic fluorescence studies show that there are no major structural changes to the protein upon binding to DNA (Figure 23) as there is no shift in the emission maximum wavelength between bound and unbound protein. The fluorescence intensity, however, increases for both proteins upon binding DNA. Furthermore, the fluorescence intensity of the DNA-bound Y540F variant is enhanced approximately four-fold compared to the DNA-bound WT FOXP2 FHD. The enhanced fluorescence intensity is unusual as no tryptophan residues are located in the DNA binding site itself. The nearest tryptophan residue, Trp533, is located in close proximity to the recognition helix (H3) (Figure 31) but is not directly involved in the DNA binding interaction. This suggests that the microenvironment of the tryptophan residues in the Y540F variant is different compared to the WT FOXP2 FHD. In the crystal structure of the FOXP2 FHD monomer the Tyr540 residue is located in close proximity to the DNA backbone (Figure 31) which may allow it to participate in stabilising van der Waals contacts. These findings show that the hinge region is important in the interaction of the FOXP2 FHD with DNA.

The residues in the hinge region are also shown to play an important role in dimerisation of the FOXP FHD and may account for the difference in stability between the FOXP2 and FOXP3 DSDs. The importance of the FOXP hinge-loop region in the stability of formation of DSDs has been shown by mutations in this region that lead to disruption of DSD formation such as the A539P mutation in FOXP2 (Stroud *et al.* 2006) and the IPEX mutation (F373A) in FOXP3 (Bennett *et al.* 2001). The Ala539 residue is conserved as a proline in other members of the FOX superfamily and these members have monomeric FHDs. Indeed, the A539P mutation disrupts domain swapping by forming exclusively monomeric species (Stroud *et al.* 2006). Proline residues are well-known helix-breakers (Kim and Kang 1999) and it is therefore proposed that the Pro539 residue disrupts domain-swapping by preventing the extension of the hinge-region (H4) into helix H2 which is required to form the domain-swapped dimer (Stroud *et al.* 2006). The F373A mutation in FOXP3 causes disruption of the domain-swapped dimers resulting in a mixture of monomer and dimer being present at physiological conditions when the wild-type appears to be exclusively dimeric (Bandukwala *et al.* 2011). It is believed that the F373A mutation does this by destabilising the network of hydrophobic residues at the domain-swapping interface.

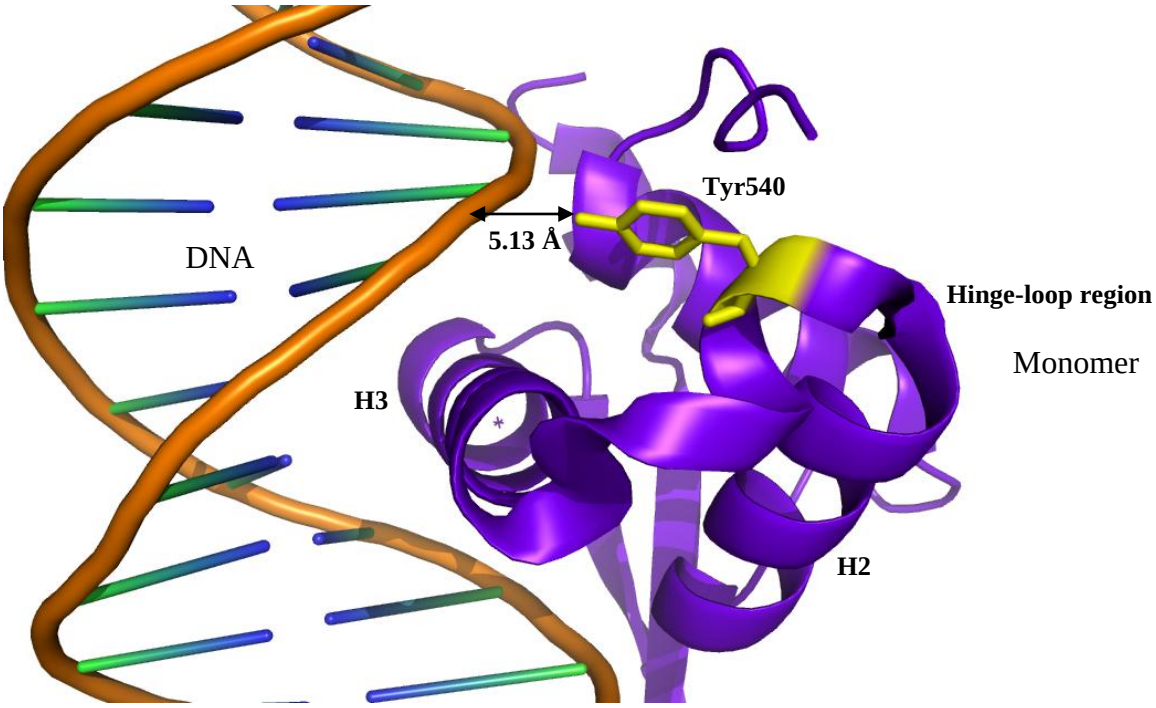


Figure 31. Interaction of the FOXP2 FHD with DNA. Tyr540 is within 5 Angstroms of the DNA backbone which may permit stabilising van der Waals interactions.

Size-exclusion chromatography was used to investigate the role of Tyr540, and hence the hinge-loop region in dimerisation by monitoring the monomer-dimer ratio with increasing protein concentration. Some proteins exhibit formation of domain-swapping oligomers only under extremely high protein concentrations which are used for crystallisation. Figure 28A shows that the WT FOXP2 FHD does not form a significant amount of dimer up to 100 μM which falls within the physiological concentration range (Hiller and Weber 1978; Anderson and Anderson 2002; Beck *et al.* 2011). The proportion of dimer in the Y540F variant, on the other hand, increases with increasing protein concentration. This suggests that having a phenylalanine at position 540 increases the propensity of the protein to dimerise.

Phe373 in the FOXP3 DSD is suggested to promote dimerisation by stabilising the extensive hydrophobic network at the dimer interface (Stroud *et al.* 2006; Bandukwala *et al.* 2011). This study shows that the WT FOXP2 FHD does not form a significant amount of dimer up to 100 μM which is in contrast to the monomer-dimer mixture of the WT FOXP2 FHD previously reported by Stroud *et al.* 2006. However, the WT FOXP2 DSD may be observed only at extremely high protein concentrations which are used for protein crystallisation. Another member of the FOXP-subfamily, the FOXP1 FHD, is also shown to exist in a mixture of monomer and dimer species (Chu *et al.* 2011). The FOXP1 FHD exhibits a significantly higher amount of dimer than monomer at 86 μM (Chu *et al.* 2011) compared to the mainly monomeric WT FOXP2 FHD at 100 μM. Even though the WT FOXP2 FHD is mainly monomeric at physiological concentrations, the full-length protein contains the zinc finger/leucine zipper motif which may aid in bringing the two monomers together in close proximity in order to dimerise. Since the size-exclusion work was performed in the absence of DNA, it cannot be ruled out that DNA may be required for dimerisation as seen in other transcription factors, Oct-1(Remenyi *et al.* 2001) and SgrAI (Park *et al.* 2010b).

In the FOXP2 FHD crystal structure, the dimer is shown to form fewer contacts with DNA and thus a lower affinity for DNA than its monomeric counterpart (Figure 32) (Stroud *et al.* 2006). This study has shown that the Y540F variant has an increased propensity to dimerise which is accompanied by slightly diminished DNA binding compared to the WT FOXP2 FHD. Based on these findings, we suggest that dimerisation of the FOXP2 FHD may be a DNA binding regulatory mechanism *in vivo*. In the cell, the full-length FOXP2 binds to DNA to regulate a specific subset of target genes. It is thought that above a certain concentration threshold, the FOXP2 protein forms dimers which results in decreased affinity for that particular DNA sequence. The FOXP2 dimer may be able to then bind another specific DNA sequence with high affinity to regulate a different subset of target genes.

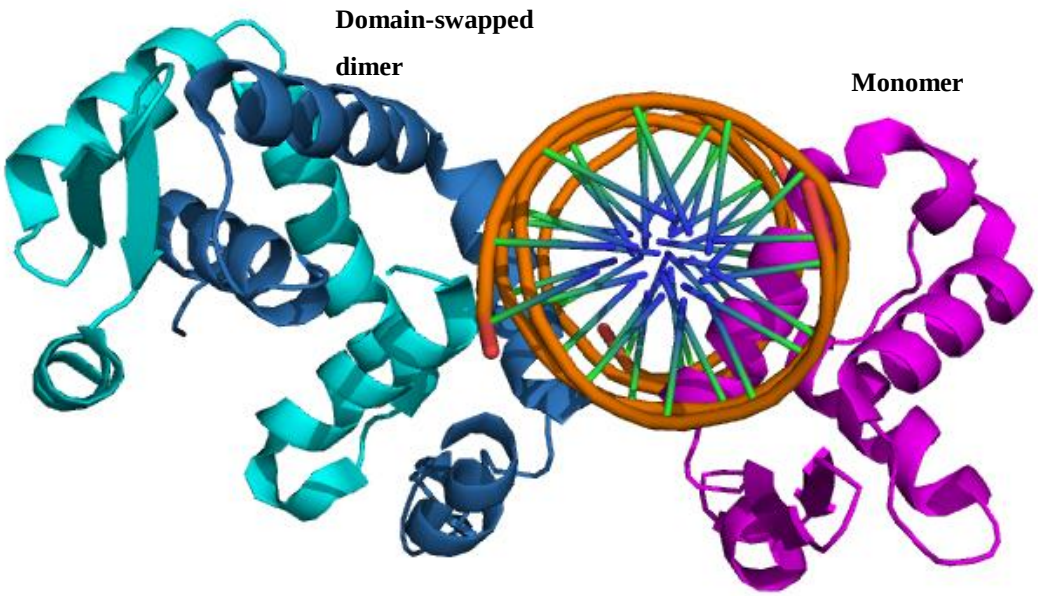


Figure 32. Interaction of the WT FOXP2 FHD dimer and monomer with DNA. The recognition helix of the monomer (magenta) is fully inserted into the major groove of DNA. The recognition helix (H3) of the dimer (cyan and blue) interacts with fewer contacts with DNA than the monomer. This figure was generated using PyMol.

4.2 Physiological role of the FOXP domain-swapped dimer

Intracellular protein concentrations are important in individuals that are heterozygous for a particular gene that contains a disease-causing mutation. This is because only half of the protein produced would be functional. In the case of the FOXP proteins, the concentration would be insufficiently low for dimers to form. The dimers therefore would not be able to carry out their unique function which results in disease. Mutation of this phenylalanine residue (F373A) to an alanine is implicated in the IPEX syndrome (Bennett *et al.* 2001) by disrupting the formation of domain-swapped dimers (Bandukwala *et al.* 2011). In FOXP3, the DSD is shown to be important for the protein's suppressive function by bringing together two distal DNA strands in close proximity to facilitate higher-order transcription complexes as shown in the crystal structure of the FOXP3 FHD (Bandukwala *et al.* 2011). (Bandukwala *et al.* 2011) but the role of the DSD in FOXP2 is still unclear.

The significance of our work shows that replacing Tyr540 with a phenylalanine as in the FOXP3 FHD significantly stabilises the dimer. Thus one single residue in the hinge region has a large effect on the domain-swapping propensity of the FHD. Perhaps the WT FOXP2 FHD may bind as a monomer *in vivo* as size-exclusion studies show here that it is almost exclusively monomeric at concentrations

lower than 100 μM. And perhaps the formation of domain-swapped dimers is to regulate DNA binding under conditions when the protein concentration is increased.

The presence of other transcription factors or binding partners can also influence the dimerisation or domain-swapping propensity. This is exemplified in the crystal structures of the FOXP3 and FOXP2 FHDs in the presence of DNA. The FOXP3 FHD was only obtained in the presence of NFAT1, a transcription factor, and cognate DNA. There is no crystal structure of the FOXP3 FHD alone in the presence of DNA. The structure of FOXP2 obtained in the presence of NFAT1 and cognate DNA only exhibited monomeric species (Wu *et al.* 2006).

4.3 Conclusion

The largely diverse conformations of the hinge region among members of the FOX superfamily are suggested to be important in DNA binding. This study uses a conservative mutation (Y540F) in the hinge region of the FOXP2 FHD to highlight the relevance of Tyr540 in stabilising DNA binding and in dimerisation. The residue corresponding to Tyr540 in FOXP2 is responsible for the IPEX syndrome upon mutation (F373A). This mutation causes disease by disrupting the formation of domain-swapped dimers and hence affects the suppressive function of the FOXP3 protein. In this study, replacing Tyr540 with a phenylalanine is shown to promote dimerisation and it is believed to do so by stabilising the core of the hydrophobic network at the dimer interface. DNA binding is diminished in the Y540F variant but not completely abrogated which shows that the hinge region plays important roles in two non-mutually exclusive processes, DNA binding and dimerisation. Although, the proposed physiological relevance of the DSD is that it can bring together two DNA molecules in close proximity to facilitate the formation of higher-order complexes during transcription, we found that the WT FOXP2 FHD is almost exclusively monomeric at physiological concentrations. We are aware that the full-length FOXP2 protein contains a zinc finger/leucine zipper dimerisation domain which may bring the two monomers together to aid domain swapping of the FHD, however. Alternatively, we see that the FOXP FHD can bind to DNA as a monomer. This suggests that domain swapping may be a possible DNA binding regulatory mechanism. However, it cannot be ruled out that different binding partners, such as other transcription factors, and the protein/DNA ratio can alter the DNA binding and dimerisation properties of the FOXP FHD.

The mechanism, kinetics and thermodynamics of domain-swapping of the FOXP2 FHD have yet to be established. The 3-D crystal structure of the Y540F variant in the presence of DNA has yet to be solved to determine if any structural changes occurred at atomic resolution compared to the 3-D structure of the WT FOXP2 FHD bound to DNA. This will help us understand how the phenylalanine alone increased the propensity to dimerise. Future work entails further mutations being created in the highly variable regions that flank the hinge region as well as mutations of the FOXP2 FHD region that corresponds to the secondary interface in the FOXP3 DSD. These mutations will help us understand the role of these residues in stabilising the DSD by using techniques such as size-exclusion chromatography, X-ray crystallography, circular dichroism and fluorescence spectroscopy and compare it to wild-type. We also wish to obtain K_d values for DNA binding of the monomer and dimer separately to quantify the affinity of each species for DNA using fluorescence spectroscopy and electrophoretic mobility shift assays. In order to do so, a monomeric variant and a dimeric variant will be created as the dimer may have higher affinity or specificity for another DNA sequence. This will provide further insight into understanding the physiological relevance of the DSD.

CHAPTER 5: REFERENCES

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

Amoutzias, G., A. Veron, J. Weiner, M. Robinson-Rechavi, E. Bornberg-Bauer, S. Oliver and D. Robertson (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, S. G. Oliver and E. Bornberg-Bauer (2004). Convergent evolution of gene networks by single-gene duplications in higher eukaryotes. *EMBO Rep* 5(3): 274-279.

Anderson, N. L. and N. G. Anderson (2002). The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1(11): 845-867.

Anthony-Cahill, S. J., P. A. Benfield, R. Fairman, Z. R. Wasserman, S. L. Brenner, W. d. Stafford, C. Altenbach, W. L. Hubbell and W. F. DeGrado (1992). Molecular characterization of helix-loop-helix peptides. *Science* 255(5047): 979-983.

Arnau, J., C. Lauritzen, G. E. Petersen and J. Pedersen (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr Purif* 48(1): 1-13.

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, M. Feuerer, Y. Chen, B. Barboza, S. Ghosh, J. C. Stroud, C. Benoist, D. Mathis, A. Rao and L. Chen (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, E. Campo, P. L. Fernandez, C. Fidler, K. Gatter, M. Jones, D. Y. Mason, J. E. Prime, P. Trougouboff, K. Wood and J. L. Cordell (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.

Beavis, R. C., B. T. Chait and H. Fales (1989). Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Communications in Mass Spectrometry* 3(12): 432-435.

Beck, M., A. Schmidt, J. Malmstroem, M. Claassen, A. Ori, A. Szymborska, F. Herzog, O. Rinner, J. Ellenberg and R. Aebersold (2011). The quantitative proteome of a human cell line. *Mol Syst Biol* 7(549.

Bennett, C. L., J. Christie, F. Ramsdell, M. E. Brunkow, P. J. Ferguson, L. Whitesell, T. E. Kelly, F. T. Saulsbury, P. F. Chance and H. D. Ochs (2001). The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet* 27(1): 20-21.

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

Benos, P. V., A. S. Lapedes and G. D. Stormo (2002). Is there a code for protein–DNA recognition? Probab (ilistical) ly…. *Bioessays* 24(5): 466-475.

Bettelli, E., M. Dastrange and M. Oukka (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, L. Chong, P. Elias and T. de Lange (1997). TRF1 is a dimer and bends telomeric DNA. *EMBO J* 16(7): 1785-1794.

Bierzynski, A., P. S. Kim and R. L. Baldwin (1982). A salt bridge stabilizes the helix formed by isolated C-peptide of RNase A. *Proceedings of the National Academy of Sciences* 79(8): 2470-2474.

Birnboim, H. C. and J. Doly (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7(6): 1513-1523.

Bogan, A. A. and K. S. Thorn (1998). Anatomy of hot spots in protein interfaces. *J Mol Biol* 280(1): 1-9.

Bogdarina, I., D. G. Fox and G. G. Kneale (1998). Equilibrium and kinetic binding analysis of the N-terminal domain of the Pf1 gene 5 protein and its interaction with single-stranded DNA. *J Mol Biol* 275(3): 443-452.

Bonk, T. and A. Humeny (2001). MALDI-TOF-MS analysis of protein and DNA. *Neuroscientist* 7(1): 6-12.

Boura, E., L. Rezabkova, J. Brynda, V. Obsilova and T. Obsil (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.

Braman, J., C. Papworth and A. Greener (1996). Site-directed mutagenesis using double-stranded plasmid DNA templates. *Methods Mol Biol* 57(31-44.

Braz, V. A. and K. J. Howard (2009). Separation of protein oligomers by blue native gel electrophoresis. *Anal Biochem* 388(1): 170-172.

Brent, M. M., R. Anand and R. Marmorstein (2008). Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. *Structure* 16(9): 1407-1416.

Byeon, I. J., J. M. Louis and A. M. Gronenborn (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 and A. M. Gronenborn (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, P. J. Brown, R. J. Launchbury, P. Bignone, J. Chi, G. Roncador, C. H. Lawrie, K. C. Gatter, R. Kusec and A. H. Banham (2010). Aberrant expression of the neuronal transcription factor FOXP2 in neoplastic plasma cells. *Br J Haematol* 149(2): 221-230.

Carroll, S. F., J. T. Barbieri and R. J. Collier (1986). Dimeric form of diphtheria toxin: purification and characterization. *Biochemistry* 25(9): 2425-2430.

Carson, B. D., J. E. Lopes, D. M. Soper and S. F. Ziegler (2006). Insights into transcriptional regulation by FOXP3. *Front Biosci* 11(1607-1619.

Carson, M., D. H. Johnson, H. McDonald, C. Brouillette and L. J. DeLucas (2007). His-tag impact on structure. *Acta Crystallographica Section D: Biological Crystallography* 63(3): 295-301.

Castanet, M., S. M. Park, A. Smith, M. Bost, J. Leger, S. Lyonnet, A. Pelet, P. Czernichow, K. Chatterjee and M. Polak (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.

Chacko, S., E. Silverton, L. Kam-Morgan, S. Smith-Gill, G. Cohen and D. Davis (1995). Structure of an antibody–lysozyme complex unexpected effect of a conservative mutation. *J Mol Biol* 245(3): 261-274.

Chae, W. J., O. Henegariu, S. K. Lee and A. L. Bothwell (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 and R. L. Miesfeld (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.

Chayen, N. E. (2004). Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol* 14(5): 577-583.

Chayen, N. E. (2005). Methods for separating nucleation and growth in protein crystallisation. *Prog Biophys Mol Biol* 88(3): 329-337.

Choo, Y. and A. Klug (1997). Physical basis of a protein-DNA recognition code. *Curr Opin Struct Biol* 7(1): 117-125.

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.

Chu, Y. P., C. H. Chang, J. H. Shiu, Y. T. Chang, C. Y. Chen and W. J. Chuang (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.

Clark, K. L., E. D. Halay, E. Lai and S. K. Burley (1993a). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5.

Clark, K. L., E. D. Halay, E. Lai and S. K. Burley (1993b). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364(6436): 412-420.

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.

Collins-Racie, L. A., J. M. McColgan, K. L. Grant, E. A. DiBlasio-Smith, J. M. McCoy and E. R. LaVallie (1995). Production of recombinant bovine enterokinase catalytic subunit in Escherichia coli using the novel secretory fusion partner DsbA. *Biotechnology (N Y)* 13(9): 982-987.

Cowan-Jacob, S. W., G. Fendrich, P. W. Manley, W. Jahnke, D. Fabbro, J. Liebetanz and T. Meyer (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.

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 and W. Wahli (2006). Transcriptional regulation of metabolism. *Physiological reviews* 86(2): 465-514.

Drenth, J. (1999). Principles of Protein X-Ray Crystallography, Springer.

Ercole, C., F. Avitabile, P. Del Vecchio, O. Crescenzi, T. Tancredi and D. Picone (2003). Role of the hinge peptide and the intersubunit interface in the swapping of N-termini in dimeric bovine seminal RNase. *Eur J Biochem* 270(23): 4729-4735.

Fersht, A. R. (1987). The hydrogen bond in molecular recognition. *Trends Biochem Sci* 12(301-304.

Fiala, G. J., W. W. Schamel and B. Blumenthal (2011). Blue native polyacrylamide gel electrophoresis (BN-PAGE) for analysis of multiprotein complexes from cellular lysates. *J Vis Exp* 48 2164-2171.

Fisher, S. E. and C. Scharff (2009). FOXP2 as a molecular window into speech and language. *Trends Genet* 25(4): 166-177.

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. and A. Revzin (1986). The use of gel electrophoresis to detect and study nucleic acid-protein interactions. *Trends Biochem Sci* 11(395-396.

Garvie, C. W., J. Hagman and C. Wolberger (2001). Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* 8(6): 1267-1276.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31(13): 3784-3788.

Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook, Humana Press**:** 571-607.

Georgiou, C. D., K. Grintzalis, G. Zervoudakis and I. Papapostolou (2008). Mechanism of Coomassie brilliant blue G-250 binding to proteins: a hydrophobic assay for nanogram quantities of proteins. *Anal Bioanal Chem* 391(1): 391-403.

Glaser, F., D. M. Steinberg, I. A. Vakser and N. Ben-Tal (2001). Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins* 43(2): 89-102.

Glover, J. M. and S. C. Harrison (1995). Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA. *Nature* 373, 257-261

Gong, X., M. Jia, Y. Ruan, M. Shuang, J. Liu, S. Wu, Y. Guo, J. Yang, Y. Ling, X. Yang and D. Zhang (2004). Association between the FOXP2 gene and autistic disorder in Chinese population. *Am J Med Genet B Neuropsychiatr Genet* 127B(1): 113-116.

Gotte, G., M. Bertoldi and M. Libonati (1999). Structural versatility of bovine ribonuclease A. Distinct conformers of trimeric and tetrameric aggregates of the enzyme. *Eur J Biochem* 265(2): 680-687.

Grandori, C., S. M. Cowley, L. P. James and R. N. Eisenman (2000). The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annual review of cell and developmental biology* 16(1): 653-699.

Graslund, S., P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H. W. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S. H. Kim, Z. Rao, Y. Shi, T. C. Terwilliger, C. Y. Kim, L. W. Hung, G. S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J. L. Sussman, R. C. Stevens, S. A. Lesley, I. A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M. I. Donnelly, W. H. Eschenfeldt, Y. Kim, L. Stols, R.

Wu, M. Zhou, S. K. Burley, J. S. Emtage, J. M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S. C. Almo, J. B. Bonanno, A. Fiser, S. Swaminathan, F. W. Studier, M. R. Chance, A. Sali, T. B. Acton, R. Xiao, L. Zhao, L. C. Ma, J. F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C. K. Ho, D. Wang, H. Wang, M. Jiang, G. T. Montelione, D. I. Stuart, R. J. Owens, S. Daenke, A. Schutz, U. Heinemann, S. Yokoyama, K. Bussow and K. C. Gunsalus (2008). Protein production and purification. *Nat Methods* 5(2): 135-146.

Grönholm, J., S. Vanhatupa, D. Ungureanu, J. Väliaho, T. Laitinen, J. Valjakka and O. Silvennoinen (2012). Structure-function analysis indicates that sumoylation modulates DNA-binding activity of STAT1. *BMC biochemistry* 13(1): 20.

Gusev, A. I., W. R. Wilkinson, A. Proctor and D. M. Hercules (1995). Improvement of signal reproducibility and matrix/comatrix effects in MALDI analysis. *Anal Chem* 67(6): 1034-1041.

Halazonetis, T. D., K. Georgopoulos, M. E. Greenberg and P. Leder (1988). c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. *Cell* 55(5): 917-924.

Harris, L. F., M. R. Sullivan and D. F. Hickok (1993). Conservation of genetic information: a code for site-specific DNA recognition. *Proceedings of the National Academy of Sciences* 90(12): 5534-5538.

Harrison, S. (1991). A structural taxonomy of DNA-binding domains. *Nature* 353(715-719.

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.

Hiller, G. and K. Weber (1978). Radioimmunoassay for tubulin: a quantitative comparison of the tubulin content of different established tissue culture cells and tissues. *Cell* 14(4): 795-804.

Hollis, T. (2007). Crystallization of protein-DNA complexes. Macromolecular Crystallography Protocols, Springer**:** 225-237.

Holmes, D., A. Jones, R. Reed and J. Weyers (2003). Practical Skills in Biomolecular Sciences England, Pearson Educaton.

Huang, Y., H. Cao and Z. Liu (2012). Three-dimensional domain swapping in the protein structure space. *Proteins* 80(6): 1610-1619.

Janin, J. (1997). The kinetics of protein-protein recognition. *Proteins: Structure, Function, and Bioinformatics* 28(2): 153-161.

Janowski, R., M. Kozak, E. Jankowska, Z. Grzonka, A. Grubb, M. Abrahamson and M. Jaskolski (2001). Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nat Struct Biol* 8(4): 316-320.

Jayaram, B. and T. Jain (2004). The role of water in protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* 33(343-361.

Jee, J., I. J. Byeon, J. M. Louis and A. M. Gronenborn (2008). The point mutation A34F causes dimerization of GB1. *Proteins* 71(3): 1420-1431.

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.

Jordan, S. R. and C. O. Pabo (1988). Structure of the lambda complex at 2.5 A resolution: details of the repressor-operator interactions. *Science* 242(4880): 893-899.

Kaufmann, E. and W. Knochel (1996). Five years on the wings of fork head. *Mech Dev* 57(1): 3-20.

Kaufmann, R. (1995). Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: a novel analytical tool in molecular biology and biotechnology. *J Biotechnol* 41(2): 155-175.

Khazanovich, N., K. Bateman, M. Chernaia, M. Michalak and M. James (1996). Crystal structure of the yeast cell-cycle control protein, p13suc1, in a strand-exchanged dimer. *Structure* 4(3): 299-309.

Kim, J. S. and R. T. Raines (1993). Ribonuclease S-peptide as a carrier in fusion proteins. *Protein Sci* 2(3): 348-356.

Kim, M. K. and Y. K. Kang (1999). Positional preference of proline in α-helices. *Protein Science* 8(7): 1492-1499.

Klose, D. P., B. A. Wallace and R. W. Janes (2010). 2Struc: the secondary structure server. *Bioinformatics* 26(20): 2624-2625.

Knaus, K. J., M. Morillas, W. Swietnicki, M. Malone, W. K. Surewicz and V. C. Yee (2001). Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat Struct Biol* 8(9): 770-774.

Kneale, G. G. (1994). DNA-protein interactions : principles and protocols. Totowa, N.J, Humana Press.

Koh, K. P., M. S. Sundrud and A. Rao (2009). Domain requirements and sequence specificity of DNA binding for the forkhead transcription factor FOXP3. *PLoS One* 4(12): e8109.

Kono, H. and A. Sarai (1999). Structure-based prediction of DNA target sites by regulatory proteins. *Proteins: Structure, Function, and Bioinformatics* 35(1): 114-131.

Kurt, S., S. E. Fisher and G. Ehret (2012). Foxp2 mutations impair auditory-motor association learning. *PLoS One* 7(3): e33130.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259): 680-685.

Lai, C. S., S. E. Fisher, J. A. Hurst, F. Vargha-Khadem and A. P. Monaco (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413(6855): 519-523.

Lakowicz, J. R. (1999). Principles of fluorescence spectroscopy. New York, Plenum Press.

Lane, D., P. Prentki and M. Chandler (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. (2005). Gene regulation : a eukaryotic perspective. New York ; Abingdon, Taylor & Francis.

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 and E. E. Morrisey (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.

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, A. Stein, R. G. Hibbert, T. Heidebrecht, P. Aloy, R. H. Medema and A. Perrakis (2010). Structure of the FoxM1 DNA-recognition domain bound to a promoter sequence. *Nucleic Acids Res* 38(13): 4527-4538.

Liu, Y. and D. Eisenberg (2002). 3D domain swapping: as domains continue to swap. *Protein Sci* 11(6): 1285-1299.

Liu, Y., G. Gotte, M. Libonati and D. Eisenberg (2002). Structures of the two 3D domain-swapped RNase A trimers. *Protein Sci* 11(2): 371-380.

Liu, Y., P. J. Hart, M. P. Schlunegger and D. Eisenberg (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, L. A. Schubert, S. D. Anover, E. L. Ocheltree, H. D. Ochs and S. F. Ziegler (2006). Analysis of FOXP3 reveals multiple domains required for its function as a transcriptional repressor. *J Immunol* 177(5): 3133-3142.

Lu, M. M., S. Li, H. Yang and E. E. Morrisey (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.

Luscombe, N. M., S. E. Austin, H. M. Berman and J. M. Thornton (2000). An overview of the structures of protein-DNA complexes. *Genome biology* 1(1): reviews001.

Luscombe, N. M., R. A. Laskowski and J. M. Thornton (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* 29(13): 2860-2874.

Luscombe, N. M. and J. M. Thornton (2002). Protein–DNA interactions: amino acid conservation and the effects of mutations on binding specificity. *J Mol Biol* 320(5): 991-1009.

Mandel-Gutfreund, Y., O. Schueler and H. Margalit (1995). Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. *J Mol Biol* 253(2): 370-382.

Marianayagam, N. J., M. Sunde and J. M. Matthews (2004). The power of two: protein dimerization in biology. *Trends Biochem Sci* 29(11): 618-625.

Marsden, I., C. Jin and X. Liao (1998). Structural changes in the region directly adjacent to the DNA-binding helix highlight a possible mechanism to explain the observed changes in the sequence-specific binding of winged helix proteins. *J Mol Biol* 278(2): 293-299.

Massari, M. E. and C. Murre (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* 20(2): 429-440.

Matthews, B. W. (1988). No code for recognition. *Nature* 335(294-295.

Mazet, F., J. K. Yu, D. A. Liberles, L. Z. Holland and S. M. Shimeld (2003). Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. *Gene* 316(79-89.

McPherson, A. (2004). Introduction to protein crystallization. *Methods* 34(3): 254-265.

Mecklenburg, L., M. Nakamura, J. P. Sundberg and R. Paus (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.

Mirny, L. A. a. G., M. S (2002). Structural analysis of conserved base pairs in protein-DNA complexes. *Nucleic Acids Research* 30(7): 1704-1711.

Moreira, I. S., P. A. Fernandes and M. J. Ramos (2007). Hot spots--a review of the protein-protein interface determinant amino-acid residues. *Proteins* 68(4): 803-812.

Morris, G. (2013). Honors Research Report. *University of the Witwatersrand*.

Newcomer, M. E. (2002). Protein folding and three-dimensional domain swapping: a strained relationship? *Curr Opin Struct Biol* 12(1): 48-53.

Niepmann, M. and J. Zheng (2006). Discontinuous native protein gel electrophoresis. *Electrophoresis* 27(20): 3949-3951.

Novagen (2011). pET system manual.

Ogihara, N. L., G. Ghirlanda, J. W. Bryson, M. Gingery, W. F. DeGrado and D. Eisenberg (2001). Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc Natl Acad Sci U S A* 98(4): 1404-1409.

Olson, W. K., A. A. Gorin, X.-J. Lu, L. M. Hock and V. B. Zhurkin (1998). DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proceedings of the National Academy of Sciences* 95(19): 11163-11168.

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 L. Nekludova (2000). Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *J Mol Biol* 301(3): 597-624.

Pabo, C. O. and R. T. Sauer (1984). Protein-DNA recognition. *Annu Rev Biochem* 53(293-321.

Pace, C. N., B. A. Shirley, M. McNutt and K. Gajiwala (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, A. Agrawal, M. I. Ghare, E. J. Little, P. W. Dunten, J. Bitinaite and N. C. Horton (2010a). Domain swapping in allosteric modulation of DNA specificity. *PLoS Biol* 8(12): e1000554.

Park, C. K., A. P. Stiteler, S. Shah, M. I. Ghare, J. Bitinaite and N. C. Horton (2010b). Activation of DNA cleavage by oligomerization of DNA-bound SgrAI. *Biochemistry* 49(41): 8818-8830.

Partridge, L. and J. C. Bruning (2008). Forkhead transcription factors and ageing. *Oncogene* 27(16): 2351-2363.

Picone, D., A. Di Fiore, C. Ercole, M. Franzese, F. Sica, S. Tomaselli and L. Mazzarella (2005). The role of the hinge loop in domain swapping. The special case of bovine seminal ribonuclease. *J Biol Chem* 280(14): 13771-13778.

Pitt-Rivers, R. and F. S. Impiombato (1968). The binding of sodium dodecyl sulphate to various proteins. *Biochem J* 109(5): 825-830.

Potschka, M. (1987). Universal calibration of gel permeation chromatography and determination of molecular shape in solution. *Anal Biochem* 162(1): 47-64.

Preneta, A. Z. (1993). Separation on the basis of size: gel permeation chromatography. Protein purification methods: a practical approach. E. L. V. a. A. Harris, S. New York, Oxford University Press**:** 293-306.

Ptashne, M. and A. Gann (2002). Genes & signals, Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY:.

Raines, R. T., M. McCormick, T. R. Van Oosbree and R. C. Mierendorf (2000). [23] The S· tag fusion system for protein purification. *Methods Enzymol* 326(362-376.

Rausa, F. M., Y. Tan and R. H. Costa (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.

Reddy, C. K., A. Das and B. Jayaram (2001). Do water molecules mediate protein-DNA recognition? *J Mol Biol* 314(3): 619-632.

Remenyi, A., A. Tomilin, E. Pohl, K. Lins, A. Philippsen, R. Reinbold, H. R. Scholer and M. Wilmanns (2001). Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Mol Cell* 8(3): 569-580.

Rivas, G., J. A. Fernandez and A. P. Minton (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.

Rohs, R., S. M. West, A. Sosinsky, P. Liu, R. S. Mann and B. Honig (2009). The role of DNA shape in protein–DNA recognition. *Nature* 461(7268): 1248-1253.

Rousseau, F., J. W. Schymkowitz and L. S. Itzhaki (2003). The unfolding story of three-dimensional domain swapping. *Structure* 11(3): 243-251.

Rousseau, F., J. W. Schymkowitz, H. R. Wilkinson and L. S. Itzhaki (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.

Saleem, R. A., S. Banerjee-Basu, F. B. Berry, A. D. Baxevanis and M. A. Walter (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, T. C. Murphy, A. Baxevanis and M. A. Walter (2004). Essential structural and functional determinants within the forkhead domain of FOXC1. *Nucleic Acids Res* 32(14): 4182-4193.

Samanta, U., D. Pal and P. Chakrabarti (2000). Environment of tryptophan side chains in proteins. *Proteins: Structure, Function, and Bioinformatics* 38(3): 288-300.

Sauter, C., J. D. Ng, B. Lorber, G. Keith, P. Brion, M. Wais Hosseini, J.-M. Lehn and R. Giegé (1999). Additives for the crystallization of proteins and nucleic acids. *Journal of crystal growth* 196(2): 365-376.

Schagger, H. (2006). Tricine-SDS-PAGE. *Nat Protoc* 1(1): 16-22.

Schagger, H., W. A. Cramer and G. von Jagow (1994). Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal Biochem* 217(2): 220-230.

Schleif, R. F. (2013). Modulation of DNA Binding by Gene-Specific Transcription Factors. *Biochemistry* 52 (39): 6755-6765.

Schlunegger, M. P., M. J. Bennett and D. Eisenberg (1997). Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv Protein Chem* 50(61-122.

Schmid, F. X. (1997). Optical spectroscopy to characterise protein conformation. Protein Structure - a practical approach. T. E. Creighton. UK, Oxford University Press**:** 261-297.

Schymkowitz, J. W., F. Rousseau, H. R. Wilkinson, A. Friedler and L. S. Itzhaki (2001). Observation of signal transduction in three-dimensional domain swapping. *Nat Struct Biol* 8(10): 888-892.

Seeliger, M. W., H. K. Biesalski, B. Wissinger, H. Gollnick, S. Gielen, J. Frank, S. Beck and E. Zrenner (1999). Phenotype in retinol deficiency due to a hereditary defect in retinol binding protein synthesis. *Invest Ophthalmol Vis Sci* 40(1): 3-11.

Seeman, N. C., J. M. Rosenberg and A. Rich (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proceedings of the National Academy of Sciences* 73(3): 804-808.

Shameer, K., P. N. Shingate, S. C. Manjunath, M. Karthika, G. Pugalenthi and R. Sowdhamini (2011). 3DSwap: curated knowledgebase of proteins involved in 3D domain swapping. *Database (Oxford)* 2011(bar042.

Sheinerman, F. B., R. Norel and B. Honig (2000). Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* 10(2): 153-159.

Sheng, W., M. Rance and X. Liao (2002). Structure comparison of two conserved HNF-3/fkh proteins HFH-1 and genesis indicates the existence of folding differences in their complexes with a DNA binding sequence. *Biochemistry* 41(10): 3286-3293.

Shevchenko, A., H. Tomas, J. Havlis, J. V. Olsen and M. Mann (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1(6): 2856-2860.

Shilov, I. V., S. L. Seymour, A. A. Patel, A. Loboda, W. H. Tang, S. P. Keating, C. L. Hunter, L. M. Nuwaysir and D. A. Schaeffer (2007). The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Molecular & Cellular Proteomics* 6(9): 1638-1655.

Sicheri, F., I. Moarefi and J. Kuriyan (1997). Crystal structure of the Src family tyrosine kinase Hck. *Nature* 385(6617): 602-609.

Spaniel, F., J. Horacek, J. Tintera, I. Ibrahim, T. Novak, J. Cermak, M. Klirova and C. Hoschl (2011). Genetic variation in FOXP2 alters grey matter concentrations in schizophrenia patients. *Neurosci Lett* 493(3): 131-135.

Spit, A., R. H. Hyland, E. J. Mellor and L. A. Casselton (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, L. D. Higgins, M. J. Conroy, A. M. Hounslow, R. Jerala, C. J. Craven and J. P. Waltho (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.

Strickler, S. S., A. V. Gribenko, T. R. Keiffer, J. Tomlinson, T. Reihle, V. V. Loladze and G. I. Makhatadze (2006). Protein stability and surface electrostatics: a charged relationship. *Biochemistry* 45(9): 2761-2766.

Stroud, J. C., Y. Wu, D. L. Bates, A. Han, K. Nowick, S. Paabo, H. Tong and L. Chen (2006). Structure of the forkhead domain of FOXP2 bound to DNA. *Structure* 14(1): 159-166.

Su, D. M., S. Navarre, W. J. Oh, B. G. Condie and N. R. Manley (2003). A domain of Foxn1 required for corsstalk-dependent thymic epithelial cell differentiation. *Nature Immunology* 4(1128-1135.

Summers, D. F., J. V. Maizel, Jr. and J. E. Darnell, Jr. (1965). Evidence for virus-specific noncapsid proteins in poliovirus-infected HeLa cells. *Proc Natl Acad Sci U S A* 54(2): 505-513.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 60(5): 523-533.

Thompson, J. D., T. J. Gibson and D. G. Higgins (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* Chapter 2(Unit 2 3.

Tippner, D. and R. Wagner (1995). Fluorescence analysis of the Escherichia coli transcription regulator H-NS reveals two distinguishable complexes dependent on binding to specific or nonspecific DNA sites. *Journal of Biological Chemistry* 270(38): 22243-22247.

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, A. M. Dagnall, M. D. Molto, N. Herrero and R. de Frutos (2010). FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. *BMC Med Genet* 11(114.

Tsai, K. L., C. Y. Huang, C. H. Chang, Y. J. Sun, W. J. Chuang and C. D. Hsiao (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, C. Y. Huang, J. Y. Yang, M. C. Hung and C. D. Hsiao (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, G. Vriend, B. Van der Vinne, G. Venema and V. G. Eijsink (1994). Protein stabilization by hydrophobic interactions at the surface. *Eur J Biochem* 220(3): 981-985.

van Dongen, M. J., A. Cederberg, P. Carlsson, S. Enerbäck and M. Wikström (2000). Solution structure and dynamics of the DNA-binding domain of the adipocyte-transcription factor FREAC-11. *J Mol Biol* 296(2): 351-359.

Wang, B., D. Lin, C. Li and P. Tucker (2003). Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. *J Biol Chem* 278(27): 24259-24268.

Wang, Z. and J. Moult (2001). SNPs, protein structure, and disease. *Hum Mutat* 17(4): 263-270.

Waugh, D. S. (2005). Making the most of affinity tags. *TRENDS in Biotechnology* 23(6): 316-320.

Weigelt, J., I. Climent, K. Dahlman-Wright and M. Wikström (2001). Solution structure of the DNA binding domain of the human forkhead transcription factor AFX (FOXO4). *Biochemistry* 40(20): 5861-5869.
Wittig, I., H. P. Braun and H. Schagger (2006). Blue native PAGE. *Nat Protoc* 1(1): 418-428.
Wittig, I., M. Karas and H. Schagger (2007). High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol Cell Proteomics* 6(7): 1215-1225.
Wittig, I. and H. Schagger (2005). Advantages and limitations of clear-native PAGE. *Proteomics* 5(17): 4338-4346.
Woody, R. W. (1995). Circular dichroism. *Methods Enzymol* 246(34-71.
Wu, Y., M. Borde, V. Heissmeyer, M. Feuerer, A. D. Lapan, J. C. Stroud, D. L. Bates, L. Guo, A. Han and S. F. Ziegler (2006). FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell* 126(2): 375-387.
Xu, D., C. J. Tsai and R. Nussinov (1997). Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng* 10(9): 999-1012.
Yang, S., S. S. Cho, Y. Levy, M. S. Cheung, H. Levine, P. G. Wolynes and J. N. Onuchic (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 and D. G. Covell (1994). A role for surface hydrophobicity in protein-protein recognition. *Protein Sci* 3(5): 717-729.
Zaluzec, E. J., D. A. Gage and J. T. Watson (1995). Matrix-assisted laser desorption ionization mass spectrometry: applications in peptide and protein characterization. *Protein Expr Purif* 6(2): 109-123.
Zhang, J., Q. Wang, B. Barz, Z. He, I. Kosztin, Y. Shang and D. Xu (2010). MUFOLD: A new solution for protein 3D structure prediction. *Proteins: Structure, Function, and Bioinformatics* 78(5): 1137-1152.
Zhou, Z., X. Song, B. Li and M. I. Greene (2008). FOXP3 and its partners: structural and biochemical insights into the regulation of FOXP3 activity. *Immunol Res* 42(1-3): 19-28.