

https://dspace.org/

Structural and functional characterisation of FOXP1 and FOXP2 near-full length variants and their protein-protein interactions

Thulo, Monare

2019

https://hdl.handle.net/10539/29641

Thulo, Monare. (2019). Structural and functional characterisation of FOXP1 and FOXP2 near-full length variants and their protein-protein interactions. University of the Witwatersrand, https://hdl.handle.net/10539/29641

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



STRUCTURAL AND FUNCTIONAL CHARACTERISATION OF FOXP1 AND FOXP2 NEAR-FULL LENGTH VARIANTS AND THEIR PROTEIN-PROTEIN INTERACTIONS

by

Monare Thulo

(1037081)

Thesis

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Dr. Sylvia Fanucchi

June 2019

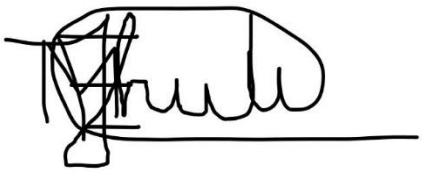
DECLARATION

I, Monare Thulo (Student number: 1037081), am a student registered for the degree of

Doctor of Philosophy (PhD) in the academic year 2019.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.



Monare Thulo

Supervisor: Dr. Sylvia Fanucchi

– To my family –

"Intellect distinguishes between the possible and the impossible; reason distinguishes between the sensible and the senseless. Even the possible can be senseless."

~Max Born

ABSTRACT

FOXP1 and FOXP2 are members of the FOXP subfamily of transcription factors characterised by multiple domains, including a highly conserved C-terminal DNA binding domain (DBD). Additionally, the FOXP proteins are characterised by two functional dimerisation interfaces, comprising of the DBD, the forkhead domain (FHD) and an N-terminal leucine zipper domain. Sequence analysis also shows several regions of intrinsic disorder in the sequence, including the polyglutamine region, the domain linkers and a C-terminal acid rich region, conserved in FOXP1, FOXP2 and FOXP4. FOXP1 and FOXP2 are expressed in several organs including the lungs and brain and are important during embryonic development. FOXP1 is associated with cognition and retardation while FOXP2 has been implicated in language development. The similarities in structure and function between FOXP1 and FOXP2 and the partial overlap that has been reported in their expression patterns in brain regions important for language may indicate a possible interaction between them. More recently, disruptions in *FOXP1* have been implicated in speech delay, intellectual disability, and autism, suggesting that just like FOXP2, FOXP1 has a significant involvement in speech and language development. Because of multiple domains, the expression and purification of the full length FOXP protein is relatively difficult to achieve *in vitro*. In this study, the structural and functional characteristics of FOXP1 and FOXP2 variants were determined using biophysical techniques. Four variants, two of FOXP1 and two of FOXP2 were constructed containing the FHD and the leucine zipper domain and differing in whether or not they additionally contained the C-terminal acid rich tail. The four variants were expressed in the soluble *Escherichia coli* cell fraction and were purified to relative homogeneity using chromatography. Although it has been suggested that FOXP1 and FOXP2 regulate transcription as dimers, we cannot rule out the possibility of both proteins performing their functions as higher oligomers. Indeed, the formation of higher order oligomers was observed in this study through SEC, DLS and native PAGE measurements. Higher order oligomers-DNA complexes were observed in EMSA, therefore suggesting that formation of oligomers may also occur during DNA binding. The significance of the disordered C-terminal tail, ART, to structural stability was studied in the presence of a simple alcohol, ethanol. Simple alcohols are enough to destabilise the tertiary structure of a protein and to stabilise the helical structure, leading to a possible quaternary conformational state change. FOXP1 and FOXP2 variants exist in solution predominantly as higher order oligomers at different concentrations. However, for FOXP1 and FOXP2 LZ-FHD, decreases in solvent dielectric result in dissociation of the higher oligomers to form a mixture of monomer and dimer. Although there is also a decrease in quaternary state in the presence of ethanol for FOXP1 and FOXP2 LZ-End, the ART seem to be enough to maintain dimeric and higher oligomeric state. In agreement with other studies, the use of ethanol was able to destabilise the

tertiary structure, shown by a red shift in fluorescence emission for each variant, and to a characteristic increase in helical content. The folding of FOXP1 and FOXP2 LZ-End in mixtures of water with a simple alcohol directly relate to decrease in the dielectric constant of the solution. The hetero-association of FOXP1 with FOXP2 was defined using pull down assays on purified FOXP1 and FOXP2 truncated variants encompassing the leucine zipper, FHD and ART and fluorescence anisotropy to study the binding affinity of FOXP2 variants for FOXP1 variants *in vitro*. Size exclusion chromatography showed that the hetero-associated proteins exist mainly as dimers and fluorescence anisotropy revealed relatively weak association, both in the presence and absence of DNA in comparison to the isolated FHD of both proteins. The loose assembly of FOXP1 and FOXP2 near full-length variants native structure suggest that this is a mechanism that is needed for both interaction with binding partners such as the other FOXP proteins during heterodimerisation and during DNA binding. These results also suggest a possibility of regulation by a dynamic equilibrium between different states which leads to a 'partial occupancy' upon DNA binding. The existence of different quaternary states and conformation suggests that both FOXP1 and FOXP2 might also control transcription as components of supramolecular regulatory complexes stabilised by different subdomains, including the C-terminal acid rich region. This, therefore, suggests that local folding of the proteins must be frequently coupled to DNA binding. Consequently, hetero-oligomerisation of the full-length protein could be a transient event, that occurs for limited time.

CONFERENCES AND SYMPOSIA

2018: *Poster presentation*:

Molecular Biosciences Research Thrust research day, University of the Witwatersrand, Johannesburg; FOXP1 and FOXP2 associate in transcriptional regulation through multiple and distinct dimerisation domains, 29 November 2018

2016: *Poster presentation*:

Molecular Biosciences Research Thrust research day, University of the Witwatersrand, Johannesburg; Structural characterization of near full length FOXP1 and FOXP2, 08 December 2016

2016: *Oral presentation:*

South African Society of Biochemistry and Molecular Biology congress, East London Convention Center, East London (SA); Expression, purification and biophysical characterization of FOXP1 and FOXP2, 10 - 14 July 2016

2016: *Poster presentation:*

Cross Faculty Symposium, University of the Witwatersrand, Johannesburg; Biophysical characterization of near full length FOXP1 and FOXP2, 01 - 02 March 2016

2015: *Poster presentation*:

Molecular Biosciences Research Thrust research day, University of the Witwatersrand, Johannesburg; Expression, purification and biophysical characterization of near full length FOXP1 and FOXP2, 03 December 2015

WORKSHOPS

2018: 5th European Crystallographic School: Stellenbosch South Africa 08 – 14 July 2018

2017: EMBL Global Exchange practical training course: Structural and biophysical methods for biological macromolecules in solution, Singapore, Singapore, 06 – 14 December 2017

2015: LSMS workshop (Mass spectrometry), Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa (Presented by Alek Dooley and Volker Kruft), 12 – 16 October 2015

ACKNOWLEDGEMENTS

To my supervisor, Dr Sylvia Fanucchi, I would like to thank you for your guidance and support throughout my postgraduate endeavours and for taking a chance with me when I was clueless in Biochemistry. I really enjoyed working with you. The amount of guidance and freedom you provided really allowed me the time to learn and develop as a scientist. You have been a great inspiration.

I have also enjoyed working with everyone in the PSFRU, past and present. I never thought I could enjoy science and life as such as I did with you. Special thanks to Dr Alison Williams, Donald Mhlanga, Zanele Molaudzi and Roland Worth. Prof. Heini Dirr, Prof. Yasien Sayed and Dr Ikechukwu Achilonu, I would like to thank you for your dedication to the training and development of young biochemists.

To the FOXy team, I will not forget the times spent together optimising this, figuring out that and the excitement of new discoveries that came with it. I will miss our long discussions on nothing and on science during our French toast meetings. Major thanks to Dr Gavin Morris for being my inspiration in what I strive to be as a scientist. Special thanks to Dr Kershia Perumal, Ashleigh Blane, Naadira Pahad and Heather Donald.

I am grateful to Dr Sylvia Fanucchi, Prof. Heini Dirr, the University of the Witwatersrand, and the South African National Research Foundation for their generous financial support that enabled my studies and this research.

Finally, I would like to thank my parents and sister, for their support during my studies. Thank you for giving me strength when I felt adrift and thought of giving up. You provided comfort when I was struggling in my research, when experiments were frustrating me and all I wanted to do was cry.

TABLE OF CONTENTS

DECLARATION i

ABSTRACT iii

CONFERENCES AND SYMPOSIA v

WORKSHOPS v

CHAPTER 1 1

INTRODUCTION 1

1.1. CLASSIFICATION OF TRANSCRIPTION FACTORS 2

1.1.1. Helix-turn-helix transcription factors 3

1.1.2. Leucine zipper transcription factors 6

1.1.3. Helix-loop-helix transcription factors 7

1.1.4. Zinc finger transcription factors 8

1.1.5. β-sheet transcription factors 10

1.2. FOX PROTEINS 11

1.2.1. FOXP subfamily 12

1.2.2. FOXP structure 13

1.2.3. FOXP dimerisation interfaces 16

1.2.4. FOXP DNA binding 20

1.3. PROTEIN-PROTEIN INTERACTIONS 22

1.3.1. Protein-protein interactions summary 22

1.3.2. Transcription factor oligomerisation 25

1.3.3. Transcription factor oligomerisation domains 26

1.4. FOXP INTERACTIONS 28

1.5. Interaction between FOXP1 and FOXP2 30

1.6. IDENTIFICATION OF PROBLEM 32

1.7. AIM AND OBJECTIVES 33

CHAPTER 2 34

EXPERIMENTAL 34

2.1. MATERIALS 34

2.2. METHODS 35

2.2.1. FOXP1 and FOXP2 constructs 35

2.2.2. Transformation 36

2.2.3. Expression and purification 37

2.2.4. SDS-PAGE 39

2.2.5. Assessment of protein purity and determination of concentration 40

2.2.6. Circular dichroism 41
2.2.7. Fluorescence spectroscopy 42
2.2.8. Size exclusion chromatography (SEC) 43
2.2.9. Dynamic light scattering 45
2.2.10. Native polyacrylamide gel electrophoresis (BN-PAGE) 46
2.2.11. Disorder prediction using DISOPRED server 47
2.2.12. Electrophoretic mobility shift assay (EMSA) 47
2.2.13. Immobilised metal ion affinity chromatography pull-down assay (IMAC-PDA) 48
2.2.14. Fluorescence anisotropy (FA) 49
2.2.15. Isothermal titration calorimetry (ITC) 51
CHAPTER 3 53
RESULTS 53
3.1. EXPRESSION 53
3.2. PROTEIN PURIFICATION 57
3.3. DNA BINDING AND NATIVE FOLDING 60
3.4. STRUCTURAL CHARACTERISATION 65
3.5. CONFORMATIONAL STUDIES 78
3.6. PROTEIN-PROTEIN INTERACTIONS 95
CHAPTER 4 113
DISCUSSION 113
CONCLUSION 120
CHAPTER 5 121
BIBLIOGRAPHY 121

LIST OF ABBREVIATIONS

ART	acid rich tail
BRET	bioluminescence resonance energy transfer
bHLH	basic Helix-Loop-Helix
dn-HLH	dominant-negative Helix-Loop-Helix
DBD	DNA binding domain
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
FA	Fluorescence anisotropy
FHD	Forkhead domain
HLH	Helix-loop-helix
HPLC	High performance liquid chromatography
HTH	Helix-turn-helix
IMAC	Immobilised metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_D	Dissociation constant
LZ	Leucine zipper
mFoxp2	mouse Foxp2
MW	Molecular weight
PDA	Pull down assay
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TEMED	Tetramethyl ethylenediamine
TF	transcription factors
UV	Ultra Violet

LIST OF FIGURES

Figure 1:The helix-turn-helix motif.. 4
Figure 2: The winged helix/forkhead domain 5
Figure 3: Leucine zipper domain. 7
Figure 4: The helix loop helix domain is characterised by two amphipathic alpha helices separated by an unconserved loop 8
Figure 5: The zinc finger motif 9
Figure 6: The beta-sheet motif of two or more antiparallel hydrogen-bonded β-strands linked by a loop region 10
Figure 7: General FOXP protein architecture. 13
Figure 8: Sequence alignment of FOXP proteins. 15
Figure 9: A) NMR structure of the FOXP1 FHD (PDB: 2kiu), B) crystal structure of the FOXP2 FHD (PDB: 2a07). This is the only member of the family for which there are FHD structures available in both the monomeric (insert) and dimeric form. C) The FOXP3 domain swapped dimer FHD (PDB: 3qrf). D) The alignment of the FOXP1 and FOXP2 FHD monomers. 17
Figure 10: FOXP leucine zipper dimerisation interface. 19
Figure 11: FOXP DNA interactions. 21
Figure 12: FOXP interactions. 28
Figure 13: Architecture of FOXP1 and FOXP2 variants, showing the location of the leucine zipper domain and the FHD followed by the C-terminal acid rich tail. 35
Figure 14: The amino acids sequence that makes up FOXP2 FHD-End variant. 36
Figure 15: SDS-PAGE gel showing representative expression of FOXP1 and FOXP2 variants. 55
Figure 16: The calibration curve for the SDS-PAGE gel shown in Figure 15. 56
Figure 17: SDS-PAGE gels showing purification of FOXP1 and FOXP2 variants. 59
Figure 18: Protein−DNA complex formation of the FOXP variants as determined by electrophoretic mobility shift assay. 61
Figure 19: DNA binding isotherms for the FOXP1 and FOXP2 variants as determined by fluorescence anisotropy (FA). 64
Figure 20: SE-HPLC chromatogram of FOXP1 and FOXP2 variants performed at increasing protein concentrations. 67
Figure 21: Native PAGE was used to determine the native conformation of the FOXP1 and FOXP2 variants. 69
Figure 22: Far-UV circular dichroism spectra of FOXP1 and FOXP2 variants using ~8 µM of FOXP1 variants and ~5 µM FOXP2 variants. 71
Figure 23: Intrinsic tryptophan fluorescence spectra of FOXP1 and FOXP2 variants using 5 µM protein. 74

Figure 24: Thermal melting curves of the FOXP1 and FOXP2 variants obtained by monitoring the far-UV CD absorbance at 222 nm while incrementally increasing temperature from 20 ℃ to 80 ℃. 76
Figure 25: pH dependent study of the quaternary structure of FOXP1 and FOXP2 variants using ~20 µM protein. 80
Figure 26: Size distribution of FOXP1 and FOXP2 near full-length variants at different pH (5.5-9).....82
Figure 27: Dynamic light scattering measurements for FOXP1 and FOXP2 variants obtained from Malvern Zetasizer Nano S at different pH (5.5-9) using ~15 µM protein. 84
Figure 28: Hydrodynamic size distribution of ~15 µM FOXP1 and FOXP2 variants at different pH (5.5-9) determined using dynamic light scattering.....85
Figure 29: Disorder simulation of FOXP1 and FOXP2 variants**.** 87
Figure 30: Secondary structure characteristics of FOXP1 variants as measured by CD spectra in the presence of increasing concentrations of ethanol (0 %-60 %).....89
Figure 31: Tertiary structure characteristics of FOXP1 variants in the presence of increasing ethanol concentrations (0-60%) as measured by intrinsic tryptophan fluorescence.....91
Figure 32: Apparent quaternary structure characterisation of FOXP1 and FOXP2 variants..93
Figure 33: Homo-association of FOXP1 variants and FOXP2 variants following a pull-down assay.....96
Figure 34: Heterotypic interactions of FOXP1 and FOXP2 variants as detected on a 12 % tris-glycine polyacrylamide gel following pull down assays with increasing concentrations of the prey protein (~5 µM - ~50 µM). 99
Figure 35: Size exclusion chromatography of FOXP1 and FOXP2 heterotypic combinations following pull down assays.....101
Figure 36: Secondary and tertiary structural characterisation of FOXP1 and FOXP2 variants following hetero-association using pull down assays.....103
Figure 37: Heterotypic binding isotherms of FOXP1 and FOXP2 truncated variants as determined by fluorescence anisotropy spectroscopy.....105
Figure 38: Comparison of FOXP1 and FOXP2 variant association binding affinity K_Ds as measured using fluorescence anisotropy in the presence and absence of DNA. 107
Figure 39: Comparison of the DNA binding affinity of the hetero-associated proteins with the DNA binding affinity of the homo-associated proteins. 109
Figure 40: Titrations of FOXP1 LZ-End with FOXP2 LZ-End. The experiments were performed in 10 mM HEPES, pH 7.4, 100 mM sodium chloride, at 25 ℃. 111
Figure 41: Models of the FOXP1/FOXP2 hetero-associated protein complex. 119

LIST OF TABLES

Table 1: Buffers used to study the secondary and tertiary structures of FOXP1 and FOXP2 variants 42
Table 2: Different buffers used for determination of quaternary structure 45
Table 3: Summary of melting points of FOXP1 and FOXP2 variants 75
Table 4: Changes in apparent oligomer size in the presence of ethanol 94
Table 5: Summary of ITC thermodynamic parameters obtained for the FOXP1/FOXP2 hetero-oligomerisation, FOXP1 LZ-End and FOXP2 LZ-End dissociation 112

CHAPTER 1

INTRODUCTION

Transcription is the initial step in gene expression and depends on the binding of transcription factors (TFs) to DNA response elements (RE) and chromatin-modifying enzymes within the promoter region (a region of DNA that initiates transcription of a particular gene) as well as the recruitment of the RNA Polymerase II complex to the transcription start site (located 3′ of the promoter) (Gao, Foat and Bussemaker, 2004; Pan *et al.*, 2010).

Transcription factors are DNA-binding, regulatory proteins that control the rate of transcriptional initiation. They bind to regulatory sequences, usually in the 5′ upstream promoter region of target genes (or they can bind to more distal control elements such as enhancers (Funnell and Crossley, 2012), in order to increase (or occasionally decrease) the rate of gene transcription (Barnes, 2006). They thus influence the fate of cells by interpreting the regulatory DNA within a genome. Transcription factors are recruited to transcription start sites by small segments of DNA- typically a few base pairs in length called enhancers- through short, specific DNA sequences (motifs) in order to regulate transcription (see review: Spitz and Furlong, 2012).

Several factors and mechanisms influence the binding of transcription factors to DNA and thereby impact gene expression. The intricate three-dimensional interactions between transcription factors and gene promoters lead to the formation of large multicomponent complexes. Analysis of the structure and interactions of these complexes provides us with information on the specificity and regulation of transcription factor DNA binding. Specificity refers to the observation that many transcription factors preferentially bind to specific DNA sequences over other sequences. The preference that a transcription factor has for a given nucleotide at a specific position in the DNA sequence is mainly determined by the physical interactions between the amino acid side chains of the transcription factor and accessible edges of the base pairs that are contacted. These contacts include direct hydrogen bonds, water-mediated hydrogen bonds and hydrophobic contacts (Slattery *et al.*, 2014). The location of a specific transcription factor on a specific DNA sequence, its interaction with other regulatory

proteins and the restriction of these interactions to a certain position in space and time contributes to the regulation of gene expression at the level of transcription.

1.1. CLASSIFICATION OF TRANSCRIPTION FACTORS

Transcription factors (TF) are categorised into two broad classes: 'general' TF, which are part of basal transcription machinery organised around RNA polymerases; and 'specific' TF that react to various biological stimuli and control the expression of appropriate target genes by binding to their *cis*-regulatory sequence in transcription activation or repression (Georges *et al.*, 2010; Benayoun, Caburet and Veitia, 2011).

Several families of TF exist, and members of each family may share structural characteristics, most commonly a highly conserved DNA binding domain and it is, therefore, common for TFs within a family to recognise either the same, or very similar, consensus DNA target sequences. Transcription factor families are classified into four predominant groups based on their DNA binding domains: helix-turn-helix, helix-loop-helix, zinc finger and basic protein/leucine zipper (Luscombe *et al.*, 2000; Gray *et al.*, 2004; Adcock and Caramori, 2009; Vaquerizas *et al.*, 2009) as well as a less defined β-sheet motif (Barnes, 2006; Adcock and Caramori, 2009). The combination of one or more of these DNA-binding motifs in a transcription factor contributes to the specificity of transcription factor binding and its interactions.

There are a number of advantages to combining multiple DNA-binding domains within a transcription factor. These include: **(1)** the chance for metabolite or substrate channelling; **(2)** the potential to protect unstable intermediates **(3)** the opportunity to allow interaction between sequential proteins leading to allosteric regulatory function and lastly, **(4)** the potential for one- or two- dimensional diffusion of the proteins if they are attached to or are part of a cellular substructure (Hawkins and Lamb, 1995).

Metabolite channelling is a process mostly defined in enzymatic reactions, where formation of multienzyme complexes allows the coordination and regulation of individual enzymes influencing the overall catalytic efficiency and specificity (Baker *et al.*, 2012). Combinations of domains to form complexes can greatly impact the catalysis of successive reactions in metabolic pathways and can also influence the transfer of common intermediates from one active site to another without release to the surrounding environment (Nagradova, 2001).

Substrate channelling also influences the ability of multidomain proteins to protect unstable intermediates. Because of multiple reactions occurring sequentially, labile or toxic intermediates are easily separated and stabilised; therefore preventing formation of unwanted products from competing cellular reactions (Miles, Rhee and Davies, 1999; Huang, Holden and Raushel, 2001; Raushel, Thoden and Holden, 2003). For this to occur, products are not released until each functional domain (active site) has completed its reaction process (Nagradova, 2001).

Furthermore, combining multiple DNA-binding domains within a transcription factor provides the opportunity to allow interaction between sequential proteins leading to allosteric regulatory function (Hawkins and Lamb, 1995). Because of allosteric regulatory function, one site on a protein can modify the function at another spatially distinct site (Cui and Karplus, 2008; Smock and Gierasch, 2009; Reynolds, Mclaughlin and Ranganathan, 2011), a process important for information flow between functional surfaces (Reynolds, Mclaughlin and Ranganathan, 2011), known as cross-talk (Vuzman, Polonsky and Levy, 2010).

Lastly, the combination of multiple domains is important for the one- or two- dimensional diffusion of the proteins if they are attached to or are part of a cellular substructure (Hawkins and Lamb, 1995). Diffusion of proteins is essential during DNA binding and is a mechanism used to search for the correct sequence on DNA (Berg, Winter and von Hippel, 1981; Slutsky and Mirny, 2004; Givaty and Levy, 2009; Vuzman, Polonsky and Levy, 2010; Vuzman and Levy, 2012). The search for a target sequence on DNA is facilitated by a combination of 3D diffusion and other search modes including sliding, hopping and jumping, which are performed at both one- and two-dimensional space (Vuzman and Levy, 2012). Initially, the protein is bound non-specifically to DNA through a single domain, while the other domain search DNA using the different modes of diffusion (Givaty and Levy, 2009; Vuzman, Polonsky and Levy, 2010).

1.1.1.Helix-turn-helix transcription factors

The helix-turn-helix (HTH) motif (Figure 1) is a simple fold comprising of three core helices that form a right handed helical bundle with a partly open configuration interconnected by short random coils (Aravind *et al.*, 2005). Variants of the HTH motif may contain two α-helices that are separated by a β-turn (Struhl, 1989). The loop between the 1st helix and the 2nd helix accounts for the most variability in the structure and may accommodate modifications in the

different classes of the helix-turn-helix motif (Aravind *et al.*, 2005). The characteristic sharp turn (Figure 1B), which is a defining feature of this motif is situated between the 2nd and 3rd helix and is typically well conserved in all variations of the HTH motif (Struhl, 1989; Aravind *et al.*, 2005).

The 3rd helix, termed the recognition helix (figure 1), forms the primary DNA-protein contacts by inserting into the major groove of DNA (Brennan and Matthews, 1989; Otting *et al.*, 1990). The helix turn helix motif binds to a palindromic sequence of DNA as a symmetric dimer that uses both the two fold symmetry of the DNA helical backbone as well as the nucleotide sequence to facilitate binding (Struhl, 1989; Khare *et al.*, 2004; Klug, 2010). Furthermore a basic patch at the N-terminus of helix 1 may form secondary contacts with DNA (Otting *et al.*, 1990) and additionally, the other α-helices can lie across the major groove and also non-specifically interact with DNA (Struhl, 1989).

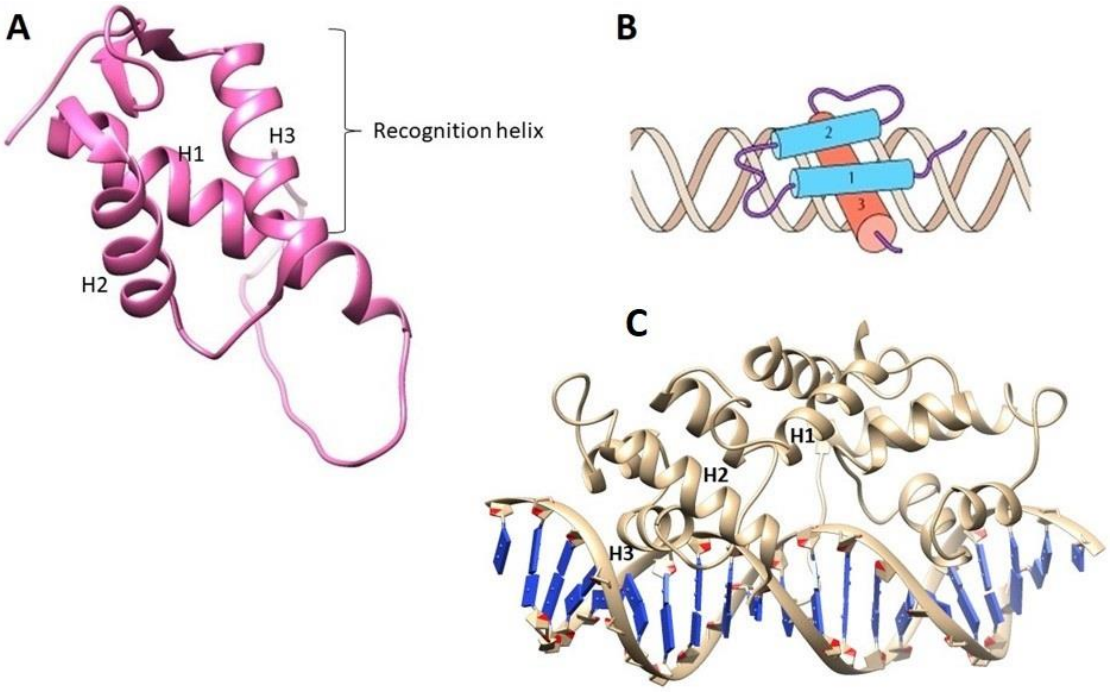


Figure 1:The helix-turn-helix motif. A) A representative HTH motif of HFH-1 (PDB: 1kq8). B) A schematic representation of the helix-turn helix motif (Cooper, 2000). C) The HTH motif binds to DNA as a dimer shown here by the HTH of lambda repressor-operator complex (PDB: 1lmb). Helix 3 forms contacts with the major groove of DNA. Model was rendered in USCF, Chimera, alpha version 1.13 (Build 41666) (Pettersen *et al.*, 2004).

An example of a transcription factor domain that contains the HTH motif is the winged helix domain (Figure 2) (Clark *et al*, 1993). The winged helix domain is a compact α/β structure (Figure 2A) consisting of two wings (W1 and W2), three α helices (H1, H2 and H3) and three β strands (S1, S2 and S3) (Barnes *et al.*, 2000). In the conventional HTH, the helices are connected by short random coils or β-turns, while in the winged helix motif, the loops connecting the helices are extended to form large random coils or 2 stranded β-sheet wings (Aravind *et al.*, 2005). The winged helix motif of the FOX transcription factors is characterised by a ~100 amino acid long, monomeric DNA binding domain (Kaestner *et al.*, 2000) called the forkhead domain (FHD), first identified in the *Drosophila melanogaster Forkhead* gene mutants and in the mammalian HNF-3α TF (FOXA1) (Weigel and Jäckle, 1990). Like all other HTH domains, the forkhead domain makes contact with DNA via helix 3, by inserting into the major groove of DNA. This interaction allows not only helix 3 to form interactions via the major groove but the loop, wing 1 and wing2 regions are also able to interact with the more distal parts of the DNA target site (Brennan, 1993).

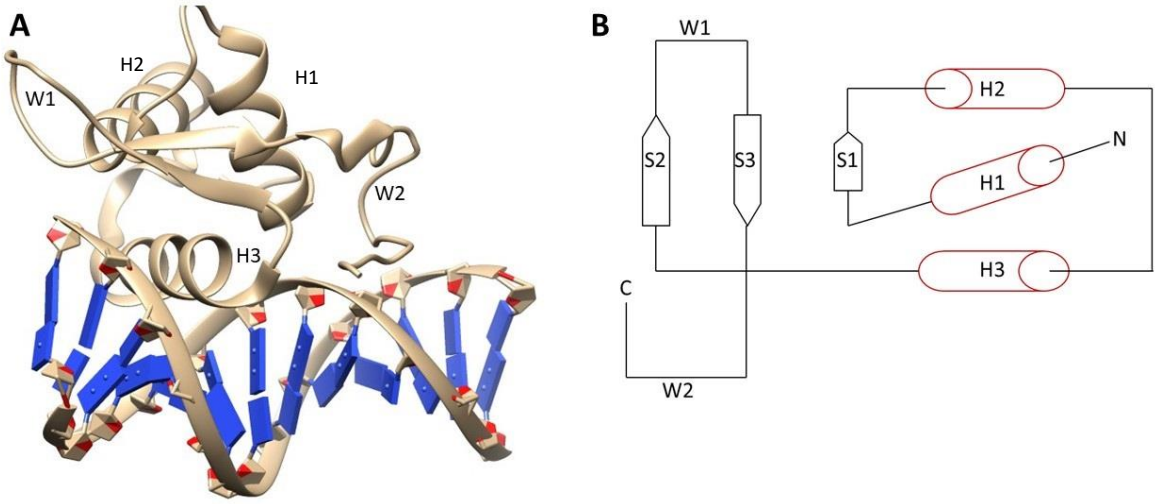


Figure 2: The winged helix/forkhead domain. A) The winged helix/forkhead domain of HNF-3 (PDB: 1vtn) was one of the first modified HTH domains to be described. As with all HTH domains, helix 3 insets into the major groove of DNA. B) Winged helix-turn-helix consists of two wings: W1 and W2, three α-helices (H1, H2 and H3) and three ß-strands (S1, S2 and S3) (Al-khayyat, 2015). Models were rendered in USCF, Chimera, alpha version 1.13 (Build 41666) (Pettersen *et al.*, 2004).

1.1.2.Leucine zipper transcription factors

The leucine zipper domain (Figure 3) contains four or five leucine residues that are spaced exactly seven residues apart and hence could be viewed as being repeated every two turns of an α-helix (Struhl, 1989). Indeed, the polypeptide segment that contains these periodic repeats of leucine residues does form a helical conformation. The leucine side chains extend from the α-helix of one polypeptide chain and can interact with those of a second polypeptide chain, thus facilitating dimerisation (Hakoshima, 2005). The α-helices in the leucine zipper dimer wind around each other forming a coiled-coil or supercoiled arrangement that helps maximise interactions among the hydrophobic side chains of the leucine residues (Figure 3A and B). This supercoiled structure gives the motif its zipper-like appearance. The hydrophobic core of the leucine zipper dimer is stabilised via van der Waals forces in addition to the hydrophobic interactions (O'Shea *et al.*, 1989; Hjalt, 2004; Hakoshima, 2005). The dimeric leucine zipper region alone does not interact with DNA; however, in some instances the leucine zipper may contain a segment of ~20 amino acids that is rich in basic residues (Figure 3B) at its C-terminus. It is this sequence that facilitates non-specific DNA binding by interacting with the negatively charged phosphate groups of DNA (Hakoshima, 2005). These basic regions stick out from the coiled coil of the zip region to interact with DNA (Figure 3B).

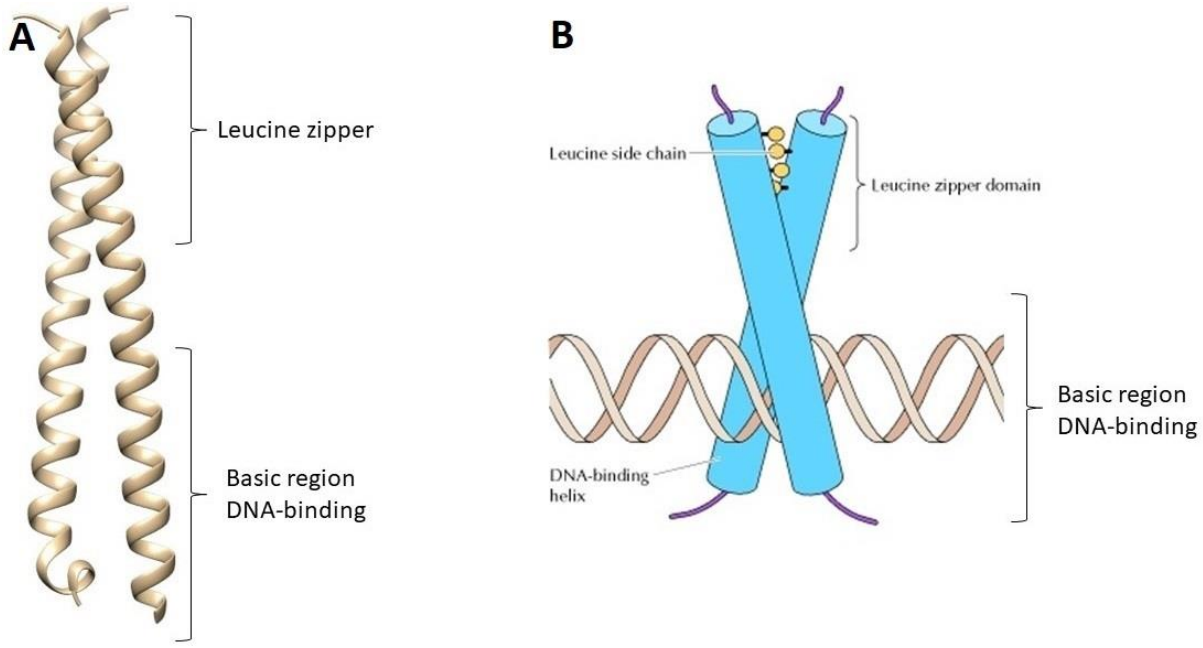


Figure 3: Leucine zipper domain. A) The leucine zipper domain of PKG1-α (PDB: 4r4l). B) A schematic representation of the leucine zipper with a basic region interacting with DNA. Model was rendered in USCF, Chimera, alpha version 1.13 (Build 41666) (Pettersen *et al.*, 2004). The schematic was taken from (Cooper, 2000).

1.1.3.Helix-loop-helix transcription factors

The helix-loop-helix (HLH) motif, shown in Figure 4, forms two amphipathic alpha helices, containing alternating hydrophobic residues, such as isoleucine or leucine in such a way that they are present on one side of the helix, creating a hydrophobic surface (Hjalt, 2004). Although the HLH motif consists of an extended alpha helical structure similar to that present in the leucine zipper domain, unlike the leucine zipper, it also has a loop inserted into the middle of the helix (Kadesch, 1992; Massari and Murre, 2000) (Figure 4A). As with the leucine zipper domain, DNA binding to the HLH motif requires an additional cluster of basic amino acids positioned immediately to the C-terminal side of the motif (Figure 4B). Part of both the interhelical loop and the second helix of each polypeptide chain in the dimer can also interact with DNA (Hjalt, 2004).

There are three different classes of the HLH motif. The first is the basic HLH motif (bHLH), described above and shown in Figure 4. It has two conserved and functionally distinct domains which together make up a region of approximately 60 amino acid residues (Jones, 2004). The basic amino acids cluster at the N-terminus on helix 1 and form the DNA binding site (Kadesch,

1992). The second class of HLH is denoted by the presence of a leucine zipper adjacent (C-terminal) to the HLH motif. These zippers help stabilise HLH mediated protein-protein interactions and function to restrict the dimerisation specificities of the protein (Beckmann, Su and Kadesch, 1990). The last class of HLH is the dominant negative HLH (dn-HLH). This family possess an HLH motif but does not contain the adjacent stretch of basic amino acids required for DNA binding.

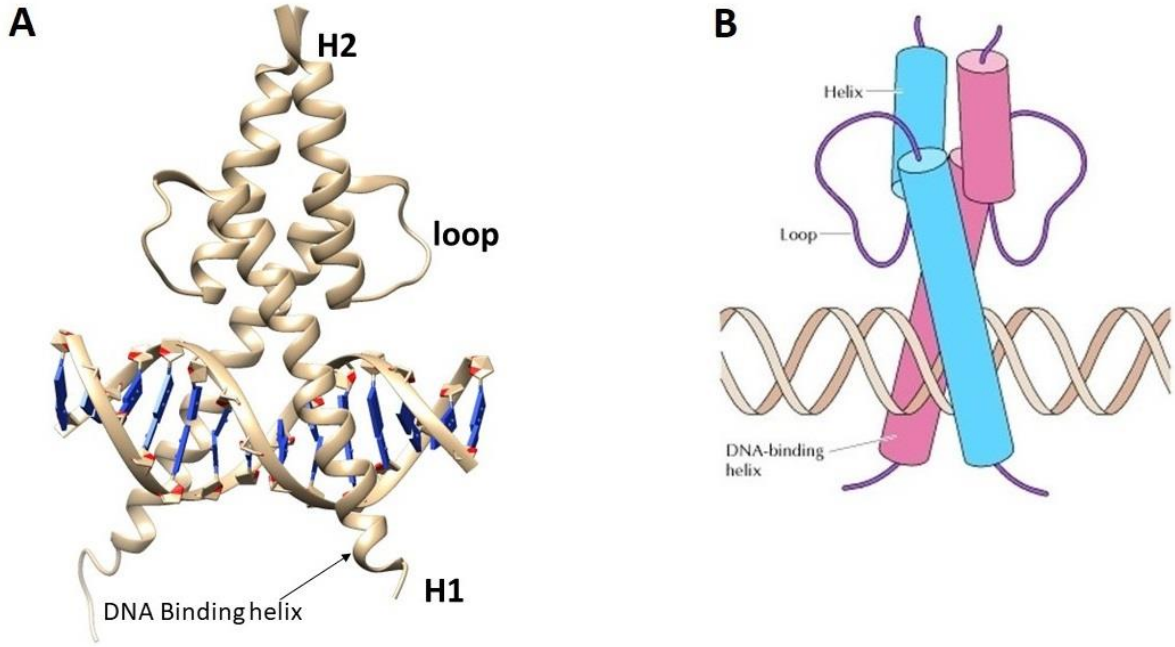


Figure 4: The helix-loop-helix domain is characterised by two amphipathic alpha helices separated by an unconserved loop A) Representation of a helix-loop-helix domain using the MyoD bHLH structure (PDB: 1mdy). B) a general schematic showing the helix-loop-helix domain interacting with DNA via H1. Model was rendered in USCF, Chimera, alpha version 1.13 (Build 41666) (Pettersen *et al.*, 2004). The schematic for the model was taken from (Cooper, 2000).

1.1.4.Zinc finger transcription factors

The zinc finger motif (Figure 5) was first described in the TFIIIA protein of *Xenopus* oocytes (Miller, McLachlan and Klug, 1985; Foster *et al.*, 1997) which contains 7-11 zinc atoms per molecule and nine repeating units of approximately 30 amino acid residues (Miller, McLachlan and Klug, 1985). In the C_2H_2 zinc finger model, the most common zinc finger, cysteine and histidine residue pairs serve as a tetrahedral coordination site for a single zinc ion and the amino acids between these coordination sites project out as fingers (Miller, McLachlan and Klug,

1985). The zinc-coordination complex maintains a scaffold that allows the exposure of amino acids sidechains along the only alpha helix in the domain (Klug, 2010). In a different model which is analogous with structures of other metalloproteins, an anti-parallel β-sheet and α-helix exist between the coordination sites (Berg, 1990). Figure 5C represents the topology of various zinc finger domains models.

The zinc finger binds to DNA tandemly in a linear, polar fashion and can recognise DNA sequences of different lengths, allowing several zinc fingers to be strung together in order to recognise DNA sequences of variable length (Klug, 2010). The α-helix (Figure 5A) binds to the major groove of DNA through specific hydrogen-bond interactions between the amino acids at helical positions 1, 3 and 6 and three successive bases on one strand of the DNA (Pavletich and Pabo, 1991).

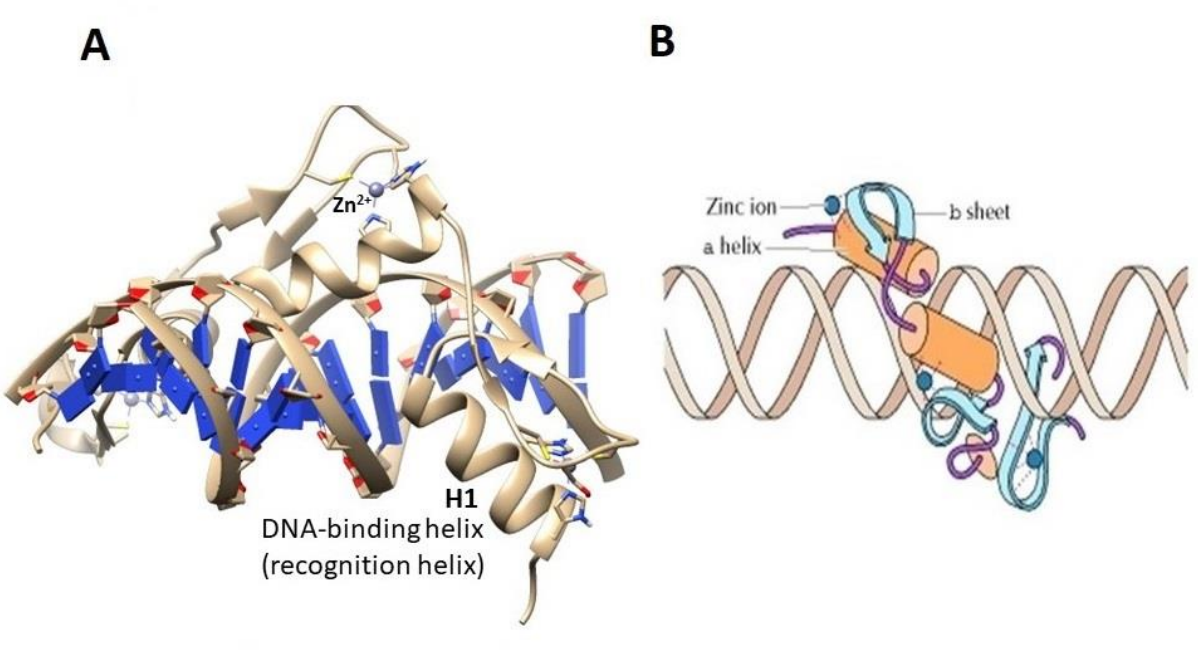


Figure 5: The zinc finger motif. A) the NMR structure of the zinc finger motif of Sp1 (PDB: 1VA3), coordinated by a zinc ion, modelled from the structure of the TFIIIA zinc finger motif. Helix 1 forms contacts with DNA. The linear fashion in which the zinc finger binds allows several zinc fingers to be strung together in order to recognise DNA sequences of variable length Model was rendered in USCF, Chimera, alpha version 1.13 (Build 41666) (Pettersen *et al.*, 2004). B) A general schematic of a zinc finger motif interacting with DNA. Helix 1 is shown interacting with the major groove of DNA, taken from (Cooper, 2000).

1.1.5. β-sheet transcription factors

The β-sheet motif, shown in Figure 6, is the least studied of all the transcription factor classes. The main roadblock to their study is the fact that isolated fragments of β-structure tend to aggregate in solution (Richardson and Richardson, 2002). The β-hairpin is the simplest β-sheet motif. A β-hairpin consists of two antiparallel hydrogen-bonded β-strands linked by a loop region (Pantoja-Uceda, Santiveri and Jiménez, 2006). β-Hairpin motifs differ in the length and shape of the loop and are classified according to the number of residues in the turn and the number of inter-strand hydrogen bonds between the residues flanking the turn (Pantoja-Uceda, Santiveri and Jiménez, 2006). When the β-hairpin loops are short (detailed description: (Pantoja-Uceda, Santiveri and Jiménez, 2006)), the loop conformation corresponds to β-turns with geometries adequate for the characteristic right-handed twist of antiparallel β-sheets (Pantoja-Uceda *et al.*, 2006).

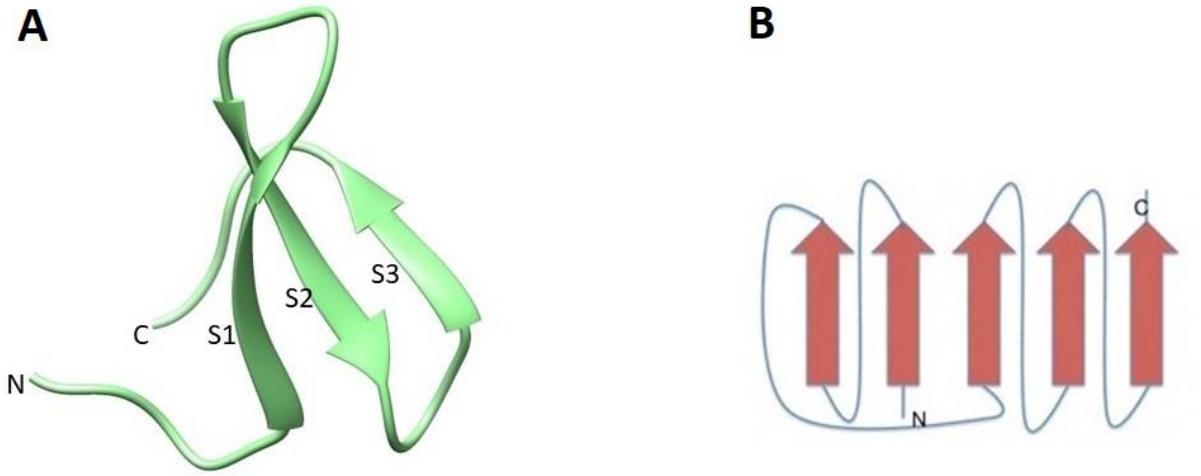


Figure 6: The beta-sheet motif of two or more antiparallel hydrogen-bonded β-strands linked by a loop region. A) The beta-sheet motif of ARRDC3 (PDB:4n7h) showing the antiparallel β-strands and the linking loops. Model was rendered in USCF, Chimera, alpha version 1.13 (Build 41666) (Pettersen *et al.*, 2004) B) A schematic representation of a general β-sheet motif, showing the N- and C-termini, and the general arrangement of β-strands.

1.2. FOX PROTEINS

The forkhead box (FOX) family of transcription factors currently has over 2000 members identified in 108 species of animals and fungi (Benayoun, Caburet and Veitia, 2011). Each subfamily is denoted by a letter (A-S) (Benayoun, Caburet and Veitia, 2011). The FOX proteins are characterised by highly conserved DNA-binding domain (DBD) called the forkhead domain (FHD), which was initially identified in the *Drosophila melanogaster Forkhead* gene (Weigel and Jäckle, 1990; Kaestner, Kno and Martı, 2000; Benayoun, Caburet and Veitia, 2011). The DNA binding domain folds into a variation of the helix-turn-helix motif (Figure 1B) termed a winged helix (Figure. 3) and is made up of three α-helices (H1, H2 and H3), two characteristic large loops or 'wings' (W1 and W2) which are extended and three β-strands (S1, S2 and S3) (Clark *et al.*, 1993; Kaestner, Kno and Martı, 2000; Huffman and Brennan, 2001). The FOX winged HTH is different from the conventional HTH in that there are two extended C-terminal loops that form a two-stranded anti-parallel β-sheet that make up the first wing (Clark *et al.*, 1993). The second wing is usually made up of the a fourth α-helix or an extended loop following the antiparallel β-sheets (Clark *et al.*, 1993).

The FOX transcription factors participate in an extensive range of biological functions ranging from embryogenesis to adult life. These include development, growth, stress resistance apoptosis, cell cycle, immunity, metabolism, reproduction and ageing (Burgering and Kops, 2002; Giannakou and Partridge, 2004; Katoh and Katoh, 2004; Arden, 2007; Carter and Brunet, 2007; Peng, 2007; Tuteja, Kaestner and Gene, 2007a, 2007b; Van Der Horst and Burgering, 2007; Partridge and Brüning, 2008). Their roles also include the regulation of gastrulation (Ang and Rossant, 1994; Weinstein *et al.*, 1994) and stem cell niche maintenance (Sackett *et al.*, 2009; Aoki *et al.*, 2016). Furthermore, members of the FOX family are essential for the normal specification, differentiation, maintenance and/or function of several tissues (Zhu, 2016). In humans, mutations in or the abnormal regulation of FOX genes are linked to several developmental disorders and diseases such as cancer (look in ref: Myatt and Lam, 2007; Golson and Kaestner, 2016), Parkinson's disease (Kittappa *et al.*, 2007), autism spectrum disorder (Bowers and Konopka, 2012), ocular abnormalities (Acharya *et al.*, 2011), defects in immune regulation and function (Mercer and Unutmaz, 2009) and deficiencies in language acquisition (Takahashi *et al.*, 2009).

1.2.1. FOXP subfamily

There are four members in the FOXP subfamily (FOXP1-4). They are classified as such because they share homology in a larger proportion of their sequence than just the forkhead domain alone.

FOXP1 functions as a transcription repressor and is widely expressed in the developing and mature brain as a regulator of neural connectivity (Kurz *et al.*, 2010). It has been reported to have expression patterns within the striatal projection neurons. It is associated with cognition and retardation (Kurz *et al.*, 2010). It is also an essential factor in pro/pre-B-cell development and is expressed in a variety of B-cell lymphomas (Sagardoy *et al.*, 2013; De Smedt *et al.*, 2015). FOXP1 is differentially expressed in a number of malignant cell types, including gastrointestinal, lung, head, neck, genitourinary malignancies and breast cancer. FOXP1 is thus a potential tumour suppressor gene (Koon *et al.*, 2007). Studies have also shown the possibility that FOXP1 acts as an oncogene in a number of malignancies, with elevated expression of FOXP1 being an indicator of worse prognosis (Koon *et al.*, 2007). Several mutations in the FOXP1 gene have been implicated in several autism spectrum disorders, mental disorders (schizophrenia) and cognitive disabilities (neurodevelopmental delay and specific language impairment) (Horn *et al.*, 2010; Sollis *et al.*, 2016; Meerschaut *et al.*, 2017).

FOXP2 is highly expressed in the central nervous system (CNS) (Bowers and Konopka, 2012). Its expression is not limited to the CNS, it is also expressed in the lungs, kidneys, intestines, spleen and skeletal muscles (Bowers and Konopka, 2012). Mutations in FOXP2 has been implicated in a severe speech and language disorder (Lai *et al.*, 2001; Teramitsu *et al.*, 2004). Mutations of FOXP2 are rare (Fisher and Scharff, 2009), but have been reported to lead to neuropsychiatric disorder including schizophrenia (Lai *et al.*, 2001; Tomblin *et al.*, 2010; Bowers and Konopka, 2012) and impaired expressive and receptive language (Fisher and Scharff, 2009). Downregulation of FOXP2 has been shown to triggers tumour initiation (Cuiffo *et al.*, 2014; Yan *et al.*, 2015; Jia *et al.*, 2016; Chen *et al.*, 2018), suggesting that FOXP2 can function as a putative tumour/metastasis suppressor in breast cancer (Cuiffo and Karnoub, 2016).

FOXP3 is expressed in developing regulatory T-cells. It acts as a transcription repressor when expressed in either non-lymphoid cells or T-cell lines where it inhibits activation induced cytokine expression (Lopes *et al.*, 2006). FOXP3 is implicated in human immune

dysregulation, polyendocrinopathy enteropathy, X-linked syndrome (IPEX syndrome) (Bennett *et al.*, 2001; Lopes *et al.*, 2006; Torgerson and Ochs, 2007; Konopka *et al.*, 2009) which is a systematic autoimmune disease affecting bowel, skin, endocrine organs and blood (Torgerson and Ochs, 2007).

FOXP4 is the least studied of all the FOXPs. In mice, Foxp4 has been found to be expressed in adult tissues of the heart, brain, lungs, liver, kidney and testis. Like FOXP1 and FOXP2, it is also expressed in the developing lung and gut (Chatila *et al.*, 2000; Shu *et al.*, 2001; Lu *et al.*, 2002; Lai *et al.*, 2003; Konopka *et al.*, 2009). It was found to be located at the chromosome region 6p21, which is a region, linked to prostate cancer (Teufel *et al.*, 2003).

1.2.2. FOXP structure

The four members of the FOXXP subfamily, FOXP1-4 contain 677, 740, 431, and 680 amino acid residues respectively (Chu *et al.*, 2011). Uniquely, the FOXP proteins have a C-terminal FHD (Figure 7) unlike the other FOX family members which have their FHD at the N-terminal side of the protein. The structural architecture of the FOXP transcription factors is thus different to that of the other FOX proteins.

FOXP transcription factors are categorised together due to the conservation of their sequence and structure. In addition to the conserved FHD, these proteins share an N-terminal polyglutamine-rich tract (or proline-rich tract in the case of FOXP3 (Xie *et al.*, 2015)), a zinc finger motif and a leucine zipper domain (Lai *et al.*, 2001; Shu *et al.*, 2001) (figure 7).

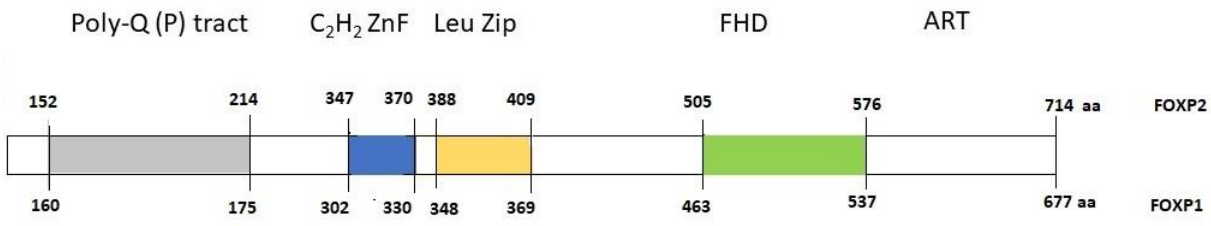


Figure 7: General FOXP protein architecture showing the centrally located domains, the C_2H_2 zinc finger motif, the leucine zipper domain (LeuZip) and the forkhead domain (FHD). FOXP1, FOXP2 and FOXP4 also possess a C-terminal acid rich tail (ART).

The polyglutamine tract of FOXP proteins may be important in protein-protein interactions (Lai *et al.*, 2001). The FOXP proteins also have a Cys_2/His_2 zinc finger domain that is

implicated in both dimerisation (Fairall *et al.*, 1993; Klug, 2010). The α-helix of the zinc-finger is directly linked to that of the leucine zipper producing an extended single long helix (Song *et al.*, 2012). The FOXP leucine zipper is an antiparallel loosely packed α-helical structure, important in dimerisation (Li *et al.*, 2004; Song *et al.*, 2012). This putative leucine zipper consists of a valine residue at some heptad positions instead of a leucine residue. The FOXP proteins are also characterised by an acidic C-terminus which is speculated to be involved in DNA binding and stability (MacDermot *et al.*, 2005; Vernes *et al.*, 2006; Hamdan *et al.*, 2010; Takeuchi *et al.*, 2010). FOXP3 does not have an acid-rich tail and its FHD ends only 11 amino acids from the C terminus of the protein (Schubert *et al.*, 2001), as seen in the sequence alignment (Figure 8).

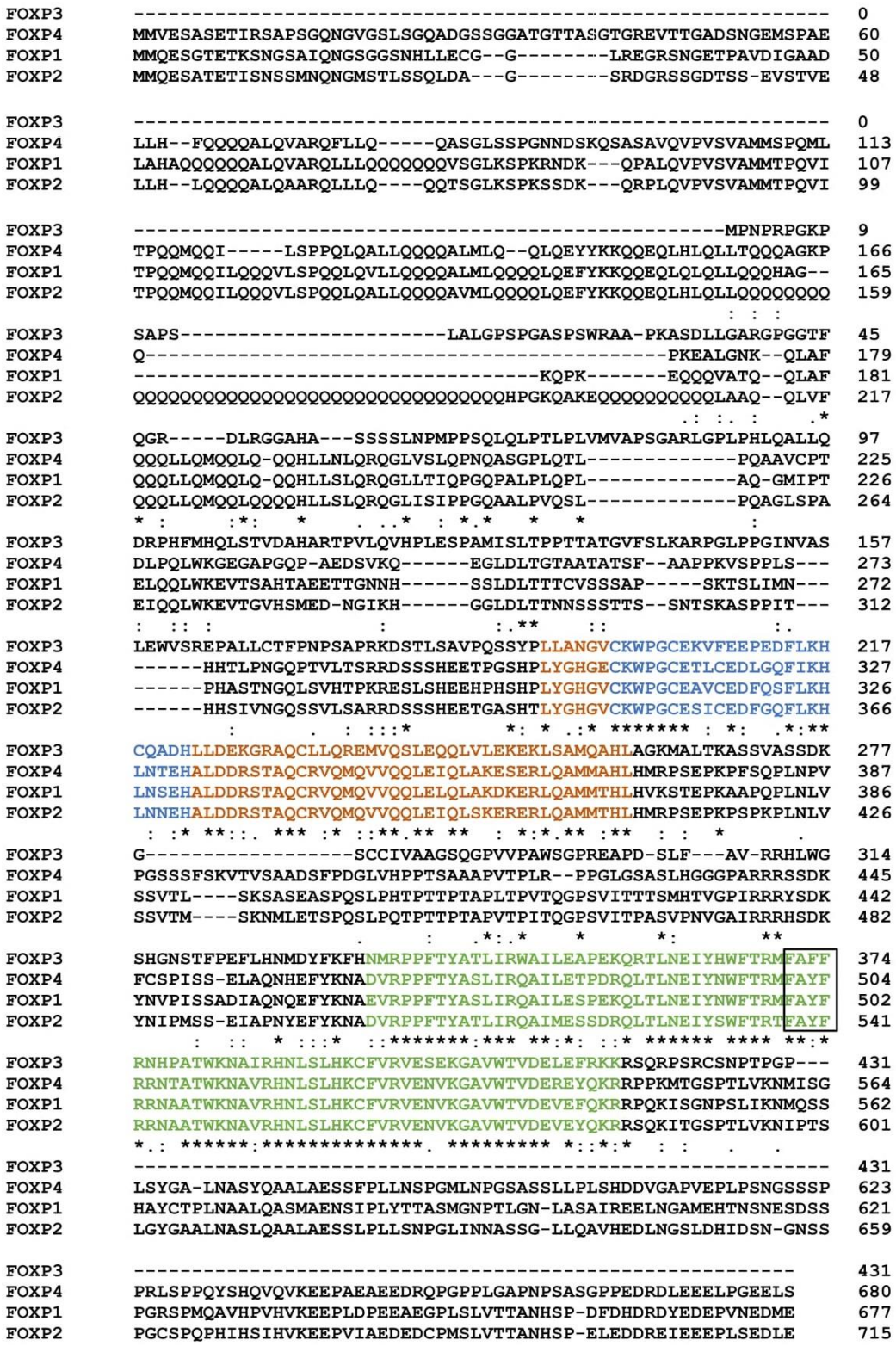


Figure 8: Sequence alignment of FOXP proteins. Each domain of these multiple domain proteins is coloured differently; zinc finger motif (blue), leucine zipper domain (brown) and forkhead-box domain (FHD) (green). The C-terminal acid rich tail extends from the FHD to the end of the sequence in three of the four subfamily members. Multiple sequence alignment performed using CLUSTAL O(1.2.4).

1.2.3.FOXP dimerisation interfaces

The FOXP subfamily of forkhead proteins is remarkable in its ability to dimerise. This is unlike any other FOX proteins studied to date. There are two distinct regions of the protein that have been shown to form a dimer interface: the leucine zipper (Li, Weidenfeld and Morrisey, 2004) and the forkhead domain (Stroud *et al.*, 2006). The NMR structure of the FOXP1 FHD (Chu *et al.*, 2011) and the crystal structures of the FOXP2 (Stroud *et al.*, 2006) and FOXP3 (Bandukwala *et al.*, 2011; Chen *et al.*, 2015) forkhead domains are shown in Figure 9. The crystal structures of the FOXP2 FHD (Figure 9B) and the FOXP3 FHD (Figure 9C) show that both these FOXP FHDs can form domain swapped dimers (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011; Chen *et al.*, 2015). There are major variations in the propensity of the FHD to form dimers between the FOXP subfamily members where the isolated FOXP3 FHD exists almost exclusively as a homodimer, while the FOXP1 and FOXP2 FHDs tend to exist as a mixture of monomer and dimer in solution under similar conditions (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011; Chu *et al.*, 2011; Medina *et al.*, 2016). These differences are highlighted by the fact that a single residue substitution in the FOXP2 FHD, Y540F, to mimic the natural substitution present at a similar position in the FOXP3 FHD, was shown to be sufficient to increase the propensity of the FOXP2 FHD for dimerisation quite substantially (Perumal, Dirr and Fanucchi, 2015). Interestingly, although crystallisation studies have shown that the FOXP FHD proteins can dimerise, the FOXP2 FHD has been shown to exist predominantly as a monomer at concentrations as high as 300 μM (Blane and Fanucchi, 2015; Perumal, Dirr and Fanucchi, 2015; Morris and Fanucchi, 2016). However, it is possible that dimerisation of the FOXP2 FHD does occur at 300 μM but the resultant dimer is highly unstable, as shown by the instability of a disulfide-linked dimeric mutant in some studies (Morris and Fanucchi, 2016).

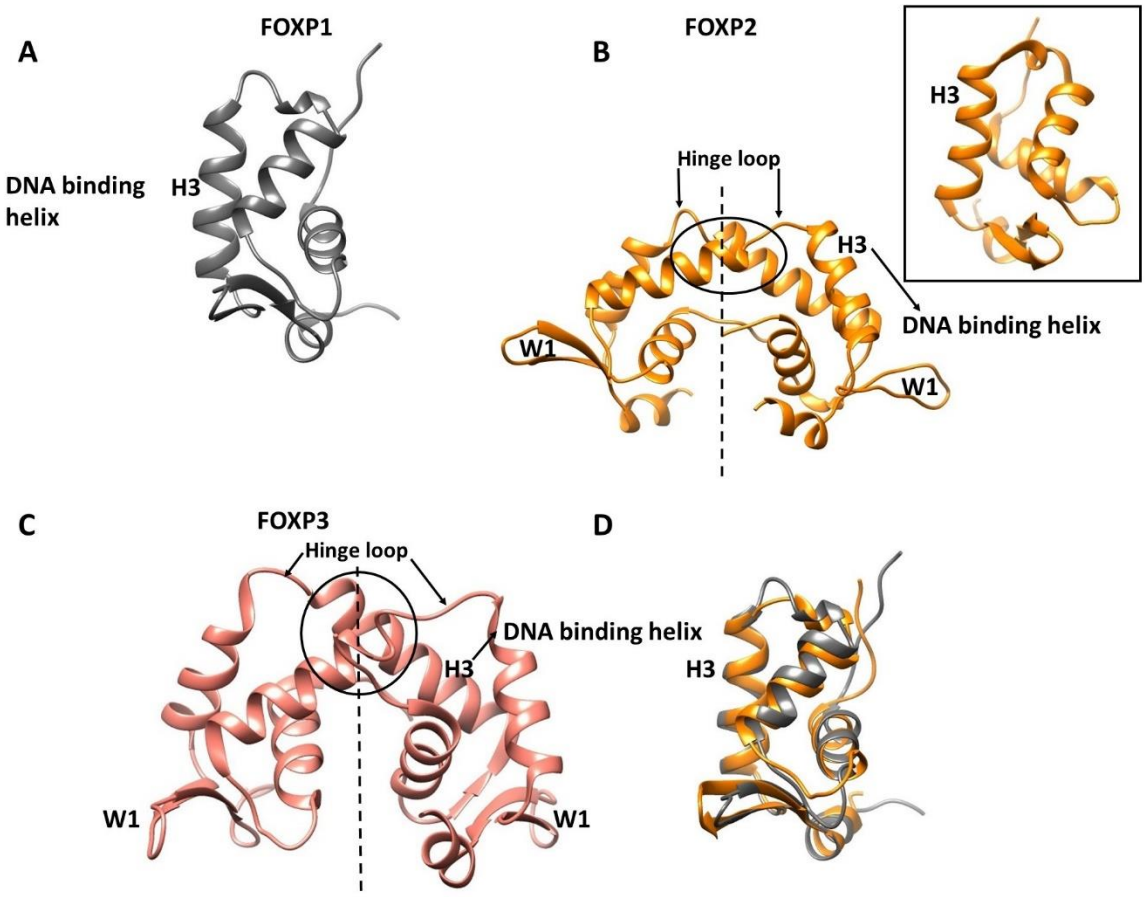


Figure 9: A) NMR structure of the FOXP1 FHD (PDB: 2kiu), B) crystal structure of the FOXP2 FHD (PDB: 2a07). This is the only member of the family for which there are FHD structures available in both the monomeric (insert) and dimeric form. C) The FOXP3 domain swapped dimer FHD (PDB: 3qrf) D) The alignment of the FOXP1 and FOXP2 FHD monomers. The domain swapping interface is shown (circular). The hinge loop region connecting the exchanged domains is highly conserved in FOX proteins allows extension of helix 1 and helix 2 during domain swapping. Models were rendered in USCF, Chimera, alpha version 1.13 (Build 41666) (Pettersen *et al.*, 2004).

Domain swapping is a unique trait observed in the FHD of FOXP proteins and is not seen in the FHD of other FOX proteins which remains monomeric. In domain swapping, specific regions of proteins are swapped over to form an intertwined dimer (Schlunegger, Bennett and Eisenberg, 1997). Domain swapped regions can differ in size from primary structural elements including α-helices or β-sheets (Khazanovich *et al.*, 1996) to whole or many domains (Bennett, Schlunegger and Eisenberg, 1995; Liu *et al.*, 1998, 2002). The only structural difference between monomer and the subunits of a domain-swapped dimer is the hinge-loop region joining the two subunits (Ogihara *et al.*, 2001; Rousseau *et al.*, 2004).

In the FOXP FHD, helix 1 (H1) and helix 2 (H2) of one monomer are exchanged upon formation of the domain swapped dimer due to extension of helix 2 (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011; Chen *et al.*, 2015). The hinge loop region connecting the exchanged domains is highly conserved in FOX proteins. However, in the FOXP subfamily, an alanine residue replaces the conserved proline that is found in this region in other FOX proteins. Proline is cyclic and as such, its structure is likely to prevent the extension of helix 2 and helix 4, thus hindering domain swapping. Indeed, mutation of the alanine to a proline residue in the FOXP1 and FOXP2 hinge loop region prevents domain swapping and renders the FOXP2 FHD exclusively monomeric (Stroud *et al.*, 2006; Chu *et al.*, 2011; Morris and Fanucchi, 2016). Although domain swapping has been observed in the crystal structures of both the FOXP2 and FOXP3 FHDs (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011; Chen *et al.*, 2015), the physiological relevance of this event is yet to be elucidated and it is uncertain whether it is not simply an artefact of crystallisation. However, there is evidence to support that the hinge loop region, thought to be responsible for domain swapping is important in regulating the mechanism of DNA binding of the FHD since it allows for flexibility of the FHD in FOXPs (Morris *et al.*, 2018). Indeed, this dimerisation seems to be crucial for the FOXP3 FHD and, therefore, FOXP3 binding since the domain swapped dimer can bridge DNA (Bandukwala *et al.*, 2011; Chen *et al.*, 2015).

The FOXP leucine zipper domain is the other significant dimerisation interface through which the full length FOXP proteins can form dimers (Li *et al.*, 2004; Song *et al.*, 2012). A crystal structure of the FOXP3 leucine zipper domain has been solved (Song *et al.*, 2012) and is shown in Figure 10.

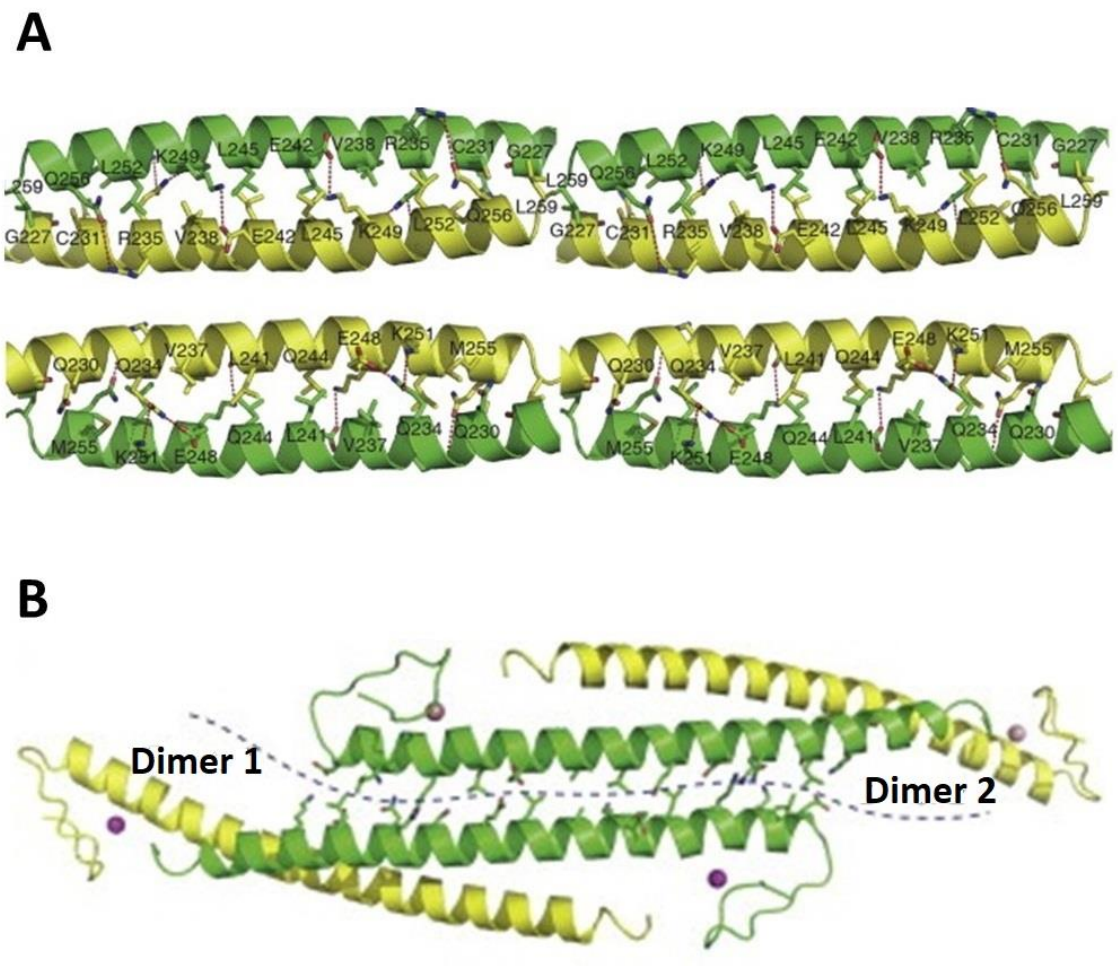


Figure 10: FOXP leucine zipper dimerisation interface. A) Amino acids involved in the FOXP leucine zipper domain, and their respective interactions with each other. B) Homo-oligomerisation of FOXP3 leucine zipper domain. The diagram was obtained from (Song *et al.*, 2012) (PDB 4i1l).

The FOXP3 leucine zipper is a loosely packed anti-parallel coiled coil (Song *et al.*, 2012). This leucine zipper domain is able to facilitate a type of high-order packing between dimers (Song *et al.*, 2012) (Figure 10B) due to the extended hydrophobic stretch on the coiled-coil surface, a trait observed in other leucine zipper domains (Bresnick and Felsenfeld, 1994; Zeng, Herndon and Hu, 1997; West *et al.*, 2004; Taylor and Keating, 2008). Because of high sequence similarities amongst the FOXP proteins, the fact that the FOXP3 leucine zipper can mediate not only dimer formation but also formation of higher oligomers, implies that the other FOXP leucine zippers are also likely to be involved in higher oligomer formation. Mutations in the FOXP leucine zipper domain that disrupt dimerisation have also been shown to affect DNA

binding in the full-length protein (Wang *et al.*, 2003; Li *et al.*, 2004; Song *et al.*, 2012), implying that dimerisation regulates DNA binding to some degree. Furthermore, dimerisation has also been shown to be important in regulating and fine tuning transcriptional activity (Sin *et al.*, 2014). The full-length FOXP protein has a zinc finger motif adjacent to the leucine zipper domain. The FOXP zinc finger motif is not believed to be directly involved in dimerisation (Song *et al.*, 2012) and its role in the protein is yet to be elucidated.

1.2.4. FOXP DNA binding

The forkhead domain of most of the FOX family members binds to DNA as an obligate monomer (Benayoun, Caburet and Veitia, 2011). However, the FOXP subfamily members are an exception and the FOXP FHD is believed to be able to interact with DNA as a dimer due to a proline to alanine substitution in the hinge loop region which allows the FHD to exist in a domain swapped dimeric form. Indeed, the crystal structure of the FOXP2 FHD shows both dimer and monomer species interacting with DNA (Stroud *et al.*, 2006).

As with all HTH motifs, DNA binding by FOXP proteins is mediated through insertion of the recognition helix, (helix 3) into the major groove of DNA irrespective of the dimerisation state (whether monomer or domain-swapped dimer) (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011; Chen *et al.*, 2015). Although they vary in their affinity of binding, the FOXP family of transcription factors are all capable of binding to the same DNA sequence (Li, Weidenfeld and Morrisey, 2004; Webb *et al.*, 2017). The dominant contact between the FOXP forkhead domain and DNA is non-specific with only a few base specific contacts being made (Stroud *et al.*, 2006). Despite this, specific interactions such as ionic interactions, hydrogen bonding, have still been shown to play an important role in DNA binding (Morris *et al.*, 2018).

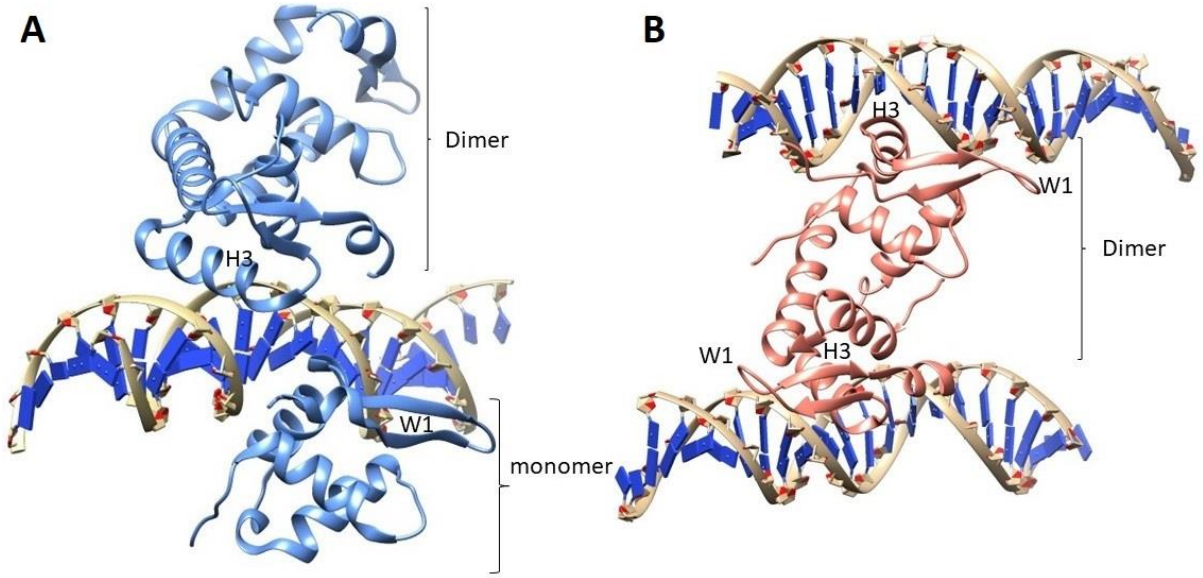


Figure 11: FOXP DNA interactions. A) FOXP2 FHD interacts with DNA as both a monomer and a dimer in the crystal structure. B) FOXP3 FHD interacts with DNA as a dimer, capable of bridging two DNA helices. Model was rendered in USCF, Chimera, alpha version 1.13 (Build 41666) (Pettersen *et al.*, 2004) using the PDB codes: 2a07 (FOXP2 FHD) and 3qrf (FOXP3 FHD), respectively.

Although the alanine to proline mutation in the FOXP FHD results in a larger DNA binding surface compared to the wild type, there is evidence to suggest that interactions between the wild type FOXP2 FHD and DNA are highly controlled by ionic contacts, facilitating greater affinity for DNA than the alanine to proline mutant (Morris *et al.*, 2018). Furthermore, binding of the FOXP2 FHD to DNA results in a decrease in the flexibility of the protein as it becomes more alpha-helical, especially in the presence of high affinity sequences (Webb *et al.*, 2017). These changes upon DNA binding can also be induced by changes in p*H* (Blane and Fanucchi, 2015). Indeed, according to the crystal structure, the FOXP2 FHD mediates a p*H* dependent ionic contact through its His554 which forms an H-bond with the N4 amino group of Cyt11 (Stroud *et al.*, 2006; Blane and Fanucchi, 2015).

1.3. PROTEIN-PROTEIN INTERACTIONS

1.3.1. Protein-protein interactions summary

Proteins play a vital role in almost all biological functions and processes, but seldom act by themselves. Most of the molecular processes rely on molecular machines which are made up of many different proteins which bind to each other via direct physical protein-protein interactions. Such interactions occur through protein interfaces. Protein interfaces (or binding sites) are certain functional patches on the surface of each protein through which contact between the interacting partners occur (Jones and Thornton, 1996). The average patch size of an interface lies between 1600 and 4000 $Å^2$ (Conte, Chothia and Janin, 1999). Easy access to binding patches is a central requirement for the interaction to form (Ma *et al.*, 2003). Three-dimensional protein structures obtained through several techniques including x-ray crystallography show that protein interfaces mostly consist of completely buried cores, surrounded by partially accessible rims (Bogan and Thorn, 1998; Chakrabarti and Janin, 2002). The major features of the interaction vary substantially among proteins; however, certain amino acid types (leucine residues, isoleucine residue and valine residue, etc.) tend to be more common than others. Furthermore, there are differences in amino acid composition between the core and rim regions of the interface (Jones and Thornton, 1996; Bogan and Thorn, 1998; Chakrabarti and Janin, 2002; Guharoy and Chakrabarti, 2005). Much of the binding affinity at the interface resides in small, conserved, independent and highly packed regions which are called "hot spots" (Delano, 2002; Ma *et al.*, 2003).

Protein–protein interactions can be divided into different classes according to their composition, affinity and life time (Nooren and Thornton, 2003a; Park *et al.*, 2009; Khan *et al.*, 2011) as: (i) homo- and hetero-oligomeric complexes (Jones and Thornton, 1996) , (ii) non-obligate and obligate complexes (Nooren and Thornton, 2003a; Perkins *et al.*, 2010) and (iii) transient and permanent complexes (Ozbabacan *et al.*, 2011).

Homo-complexes are frequently observed as permanent complexes. This is because homo-complexes tend to be symmetric and provide a good scaffold for stable macromolecules (Ozbabacan *et al.*, 2011). Therefore, their interfaces resemble protein cores (Tsai *et al.*, 1997; Tsai, Xu and Nussinov, 1997). Homo-interfaces are typically large, hydrophobic (as measured by high values of nonpolar buried surface areas) and display good complementarity between the two chains (Keskin *et al.*, 2008). In contrast, in hetero-complexes, chains differ from each other and the interactions are largely non-permanent (Jones and Thornton, 1996; Keskin *et al.*,

2008; Cardinale *et al.*, 2010). The stability of hetero-complexes can vary substantially and often hetero-complexes form the base to which different proteins that associate in order to form a macromolecular complex (Ozbabacan *et al.*, 2011). Hetero-complexes may be formed and deformed based on the cellular environment and/or external factors (Jones and Thornton, 1996; Cardinale *et al.*, 2010).

The key point for differentiation between obligate and non-obligate protein-protein complexes is affinity (Nooren and Thornton, 2003a; Perkins *et al.*, 2010; Ozbabacan *et al.*, 2011). Components of an obligate complex are unstable on their own *in vivo* and can only exist when forming part of the complex, while non-obligate components are usually capable of existing independently and stably on their own, outside of the complex system (Ozbabacan *et al.*, 2011). Therefore, obligate interactions are usually permanent (Nooren and Thornton, 2003a) whereas non-obligate interactions are mostly transient (Janin, Bahadur and Chakrabarti, 2008). Furthermore, the interfaces of non-obligate interactions tend to be smaller, less tightly packed, more polar, less conserved, and overall more similar to protein surfaces in terms of amino acid composition than those of obligate interactions (Zhu *et al.*, 2006). Obligatory associations are in general tighter, with a stronger hydrophobic component, better packing, and fewer structural water molecules trapped between the monomers, and they manifest better shape complementarity (Keskin *et al.*, 2008; Khan *et al.*, 2011; Ozbabacan *et al.*, 2011). Homodimers provide a nice example of obligatory complexes; however, many other proteins consisting of hetero-oligomers may also fall into this category (Khan *et al.*, 2011; Ozbabacan *et al.*, 2011).

Transient and permanent protein-protein complexes are characterised based on the lifetime (or stability) of the complex (Ozbabacan *et al.*, 2011). Permanent interactions are usually very stable and irreversible (Nooren and Thornton, 2003b). By contrast, formation of transient complexes depends on the functional state of the partners (Ozbabacan *et al.*, 2011). The components of the transient interactions associate and dissociate temporarily *in vivo* (Mintseris and Weng, 2003; Nooren and Thornton, 2003b, 2003a; Block *et al.*, 2006; Janin, Bahadur and Chakrabarti, 2008; Levy and Pereira-Leal, 2008). The interfaces in the transient complexes are widely less extensive and have more polar/charged amino acids, and the surfaces of the interacting proteins at their interface are not as optimised, leading to weaker associations with the exception of some enzyme-inhibitor complexes (Block *et al.*, 2006; Janin *et al.*, 2006; Winter *et al.*, 2006). Non-obligate interactions are predominantly transient; however, there are a few examples of permanent, non-obligate interactions (Janin, Bahadur and Chakrabarti, 2008). Obligate interactions are permanent in nature (Nooren and Thornton, 2003b).

It is commonly understood that proteins will associate through hydrophobic patches on their surfaces (Jones and Thornton, 1996). While folding is largely driven by the hydrophobic effect, there is actually great diversity in the manner in which proteins associate with each other (Jones and Thornton, 1996; Stites, 1996; Conte, Chothia and Janin, 1999). There are, however, other forces that are crucial to protein-protein interactions. These forces involved in packing of oligomeric proteins are surface complementarity, van der Waals interactions and electrostatic interactions, such as H-bonds and salt bridges (Nakamura, 1996; Tsai *et al.*, 1997). Protein-protein interactions seem to follow an 'anything is possible' approach, indicating that there might not be one general mechanism governing protein-protein interactions as there is for protein folding (Lendel, Dogan and Härd, 2006; Keskin *et al.*, 2008; Khan *et al.*, 2011).

In general, the highest hydrophobicity of protein-protein interfaces exists in-between protein cores and exposed surfaces, with the interfaces of homo-complexes being slightly more hydrophobic than those of hetero-complexes (Janin, Miller and Chothia, 1988; Jones and Thornton, 1996; Keskin *et al.*, 2008). Indeed, there are often a number of aromatic residues located at protein-protein interfaces implying that hydrophobic contributions do play a significant role (Young, Jernigan and Covell, 1994; Tsai *et al.*, 1997; Tsai, Xu and Nussinov, 1997; Jones, 2004). Furthermore, many, although not all, protein-protein complexes have relatively large changes in heat capacities ($\Delta C_p°$) for dissociation, which is as expected for a process driven by the hydrophobic effect (Stites, 1996).

Despite the prevalence of hydrophobic contacts at the interface, polar interactions between proteins are also common (Stites, 1996; Conte, Chothia and Janin, 1999; Keskin *et al.*, 2008; Khan *et al.*, 2011). Arginine residues, for example, are commonly found at these interfaces, indicating that electrostatic interactions are also important (Xu, Lin and Nussinov, 1997; Sheinerman, Norel and Honig, 2000; Sheinerman and Honig, 2002; Norel *et al.*, 2008). Further support for the prominence of electrostatic forces at protein-protein interfaces comes from indications that the desolvation free energy of polar atoms is lower in interfaces than in protein cores (Xu, Lin and Nussinov, 1997; Sheinerman, Norel and Honig, 2000) and from the effect of long-range electrostatic interactions on the on-rates of complex formation (Schreiber and Fersht, 1996).

1.3.2. Transcription factor oligomerisation

Transcription factors often form oligomers because they either cannot bind to DNA as monomers, their monomers are unstable, or the oligomers have a higher affinity or specificity for DNA-binding than the monomers (Funnell and Crossley, 2012). The formation of transcription factor homo- and hetero-oligomers is, therefore, important for transcriptional regulation as it modulates their interactions with specific DNA sequences (Zaret and Carroll, 2011). Oligomerisation events result in either the inhibition or the enhancement of transcriptional activity at a site distinct from the consensus target site for a transcription factor (McCarty *et al.*, 2003; Guharoy and Chakrabarti, 2005; Maiese, 2010). Proper dimerisation is thus a requirement for cooperative binding to multiple binding sites and synergistic transactivation of targets (Payre *et al.*, 1997).

Transcription factor oligomerisation is favoured because the DNA-binding domain (DBD) of a monomeric protein only recognises a relatively short DNA sequence unit and cannot bind DNA as stably (Spitz and Furlong, 2012). The binding of a dimer to two DNA sites can be much tighter than the binding of either monomer (Schleif, 2013) due to increased number of interactions. Furthermore, oligomer formation increases the selectivity of protein-DNA interactions by causing the proteins to bind to a more complex sequence consisting of a combination of the short units recognised by each DBD (Funnell and Crossley, 2012). When the transcription factors and their interacting partners are expressed in overlapping spatial domains, the combinatorial binding that results, causes discrete and precise patterns of transcriptional activity (Spitz and Furlong, 2012). Many transcription factors are homo-oligomeric and simultaneously utilise two DNA binding domains to contact two similar DNA sites (Schleif, 2013). This allows DNA binding affinity to be controlled by regulating the separation and/or relative orientation of the domains (Schleif, 2013). Hence, if dimerisation or oligomerisation is controlled, then DNA binding can be controlled. The ability of these transcription factors to activate transcription is also regulated by protein-protein interactions among the monomers rather than between the oligomer and DNA (Kadonaga, 2004; Vinson, Acharya and Taparowsky, 2006; Hashikawa, Yamamoto and Sakurai, 2007).

Transcription factors homo- and hetero-oligomerisation can occur via different mechanisms: **(1)** An unbound transcription factor can be synergistically attracted to a bound transcription factor thus enabling it to bind to neighbouring transcription factor binding sites; **(2)** interaction of a monomer with DNA induces allosteric changes in the protein which in turn increases its

affinity for an incoming monomer; or binding of a monomer to DNA may lead to conformational changes in the neighbouring binding site, thereby increasing its affinity for another monomer; **(3)** homodimerisation can occur between monomers in solution with or without intervention from promoters (Review: Georges, 2010). However, sometimes dimers are not pre-assembled before DNA binding, but are formed during the recognition event (Georges *et al.* 2010).

1.3.3. Transcription factor oligomerisation domains

Some major transcription factor domains, such as the basic leucine zipper domain, the basic helix-loop-helix (bHLH), as well as other less defined domains including the MADS box and the Rel homology domain, require noncovalent dimerisation in order to bind DNA (See full review: Funnell and Crossley, 2012). Dimerisation is one of the major control mechanisms in basic HLH transcription factor activity regulation (Hjalt, 2004). For example, protein-protein interactions between factors belonging to two classes of transcription factor, the homeodomain (HD) and the basic helix-loop-helix (bHLH) class, have been shown to form the basis for cell-specific transcription of pituitary pro-opiomelanocortin (POMC) gene expression by forming dimerisation interactions via their HD and bHLH domains (Poulin *et al.*, 2000). The corticotrophic specificity of POMC transcription depends on the interaction between the following three transcription factors: Pitx1 (an HD containing protein) Pan1 (a bHLH containing protein), and NeuroD1 dimers which also contain a bHLH motif (Lamonerie *et al.*, 1996; Poulin, Turgeon and Drouin, 1997). Pitx1, physically associates with Pan1 through the basic helix-loop-helix domain (Poulin *et al.*, 2000). Pan1 interacts with NeuroD1, and their heterodimer can also interact with Pitx1 through both the basic helix-loop-helix domain and the homeodomain (Poulin *et al.*, 2000).

Several transcription factors have been shown to associate through the leucine zipper domain. The specific arrangement of amino acids within the leucine zipper is responsible for the relative stability and specificity of leucine zipper domains (Krylov, Mikhailenko and Vinson, 1994). While oligomerisation through the leucine zipper domain occurs largely due to the contribution of the hydrophobic effect, the zipper forms through contacts between both hydrophobic residues (mainly leucine) but also polar residues of the corresponding monomer (Krylov and Vinson, 2001; Hakoshima, 2005). Indeed, oligomerisation specificity between two leucine zipper motifs is determined primarily by electrostatic interactions between the two helices

(Hakoshima, 2005). One of the best described transcription factor families that associate through the leucine zipper domain are the FOS and JUN transcription factors, proto-oncogenes involved in the signal transduction pathways (Abate *et al.*, 1990a; Abate *et al.*, 1990b; Abate *et al.*, 1991). FOS and JUN onco-proteins readily hetero-dimerise via their leucine zipper domain and strongly bind to DNA (Shea *et al.*, 1989). JUN can also form DNA-binding homodimers while FOS homodimers are unstable and do not bind to DNA (Abate, Luk, Gentz, *et al.*, 1990). Both homo- and hetero-dimerisation of these proteins, therefore, offers important advantages such as increased stability, regulation of binding site accessibility and increased complexity (Funnell and Crossley, 2012; Matthews and Sunde, 2012).

The zinc finger domain can also mediate protein-protein interactions. Of the zinc fingers, the Cys_2/His_2 zinc finger is the one that is most prominent in mediating protein-protein contacts (Fairall *et al.*, 1993). Oligomerisation occurs through contacting amino acids in the anti-parallel β-sheet that forms at the zinc coordination site (Fairall *et al.*, 1993; Mackay and Crossley, 1998). For instance, the TFIIIA transcription factor has a strong requirement for its 6^{th} Cys_2/His_2 zinc finger domain cluster in order to form dimers (Klug, 2010).

1.4. FOXP INTERACTIONS

The FOXP proteins are ubiquitously expressed, occurring in both embryonic and mature cells and tissue. Because of the vast expression patterns of the FOXP transcription factors, there are major overlaps in expression patterns amongst themselves, with other FOX proteins, and other transcription factors. A number of putative FOXP interactions have been identified in in several surveys, outlining the importance of these proteins in development and maintenance (Li *et al.*, 2004; Ravasi *et al.*, 2010; Sakai *et al.*, 2011; Corominas *et al.*, 2014). Figure 12 below shows the different interactions of the FOXP proteins that are currently known, determined exprerimentally through the use of high throughput pull down assays and through data mining.

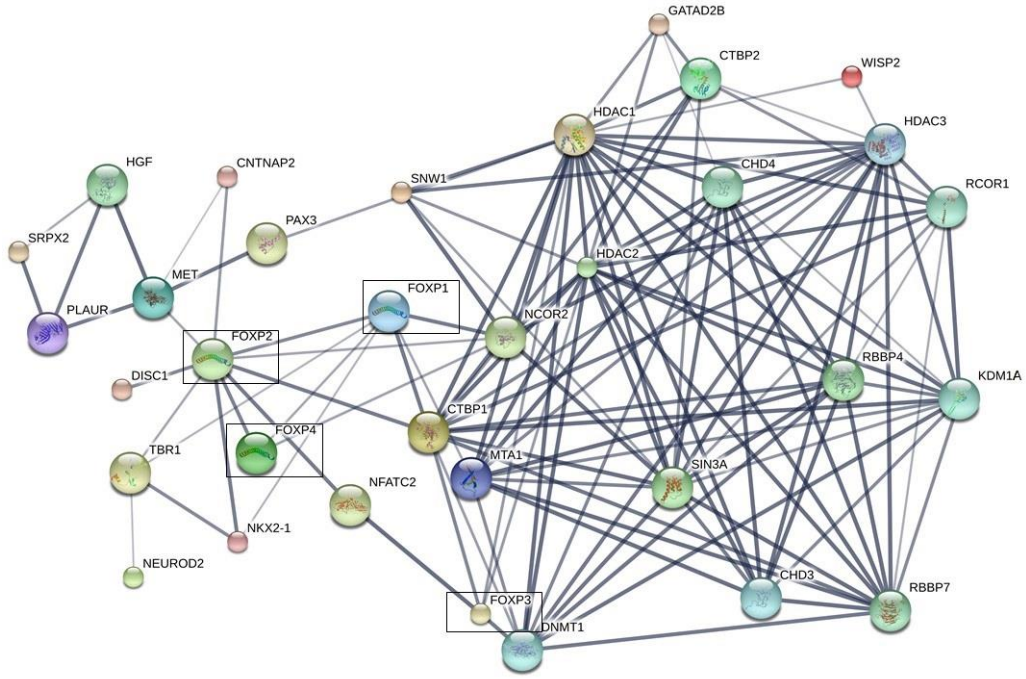


Figure 12: FOXP interactions. The confidence level was set at 40%. Each line represents interactions between the proteins labelled. The thickness of each line corresponds to the level of confidence, with thick bands corresponding to high confidence level and thin (faint) bands at lower confidence level. Some interactions have not been determined experimentally; therefore, were determined from text mining by the online tool. Image made using STITCH online protein-protein interaction tool (Szklarczyk *et al.*, 2016). Proteins with solved crystal structures are shown in bigger bubbles.

Several proteins partners have been identified which associate with the FOXPs to regulate transcription. The C-terminal binding proteins (CTBP) 1 and 2, which function as co-repressors

for several transcription factors (Li *et al.*, 2004; Estruch *et al.*, 2016) interact with FOXP1 and FOXP2 but not with FOXP4. Evidence of the interaction between CTBP1 and CTBP2 and FOXP2 has come from multiple independent yeast two hybrid screens (Sakai *et al.*, 2011; Corominas *et al.*, 2014; Rolland *et al.*, 2014) and a bioluminescence resonance energy transfer (BRET) assay (Estruch *et al.*, 2016). Another co-factor that has been shown to interact with FOXP1, FOXP2 and FOXP4 is GATAD2B, a component of the NuRD chromatin- remodelling complex (Chokas *et al.*, 2010). Luciferase assay confirmed the interaction of GATAD2B with FOXP1 and FOXP4 suggesting that GATAD2B cooperates with FOXP1 and FOXP4 in repression. However, such interactions were not observed between GATAD2B and FOXP2. Further interactions with other components of the NuRD complex were shown for FOXP1 and FOXP4; MTA1 interacts with FOXP1 and HDAC1/2 interacts with FOXP1 and FOXP4 (Deriziotis *et al.*, 2014a; Estruch *et al.*, 2018). FOXP1 also interacts with the co-repressor NCOR2 (Jepsen *et al.*, 2008). It has been demonstrated that FOXP1 and NCOR2 cooperatively repress the expression of common target genes to promote cardiac growth and regulate macrophage differentiation (Jepsen *et al.*, 2008). An association between FOXP2 and POT1, a nuclear protein involved in telomere maintenance and DNA repair mechanisms has also been reported (Tanabe *et al.*, 2011). The interaction between mFoxp2 and NKX2-1 was identified in mouse lung cell lines using a mammalian two-hybrid assay and confirmed with coimmunoprecipitation experiments (Zhou *et al.*, 2008). NKX2-1 is a transcription factor with important roles in lung, brain and thyroid development (Sussel *et al.*, 1999). These interactions of FOXP1, FOXP2 and FOXP4 were confirmed using coimmunoprecipitation assays (Estruch *et al.*, 2018).

Another transcription factor known to interact with FOXP2 is NFATC2. This was discovered in a study in which the FOXP2 forkhead domain was co-crystallised with NFATC2 (Wu *et al.*, 2006). NFATC2 is expressed in several organs including the brain (Vihma *et al.*, 2016), hence it is possible that it interacts with the FOXPs in neural sites of co-expression to regulate brain development. However, to date the functional consequences of the interaction between FOXP and NFATC2 have only been studied for FOXP3 in the context of the immune system (Wu *et al.*, 2006).

The neuron-specific transcription factor TBR1 has been suggested as a putative interactor of FOXP2 (Sakai *et al.*, 2011) and the interaction was confirmed using a BRET assay (Deriziotis *et al.*, 2014b). The interaction involves the FOXP2 region encompassing amino acid residues 122-258, and the DNA-binding domain (T-box) of TBR1 (Deriziotis *et al.*, 2014b). These two

encoded proteins are expressed in overlapping areas of the brain, and, therefore, the suggestion is that the interaction between FOXP2 and TBR1 may be important for brain development and may also have implications in autism spectrum disorder (Deriziotis *et al.*, 2014b).

1.5. Interaction between FOXP1 and FOXP2

The members of the FOXP subclass share a highly conserved DNA binding domain and have high sequence similarities. FOXP1 and FOXP2 are the most related of the FOXP proteins, sharing 64% total protein sequence identity, and 89% in the forkhead domain (Bacon and Rappold, 2012). In humans, members of the FOXP family, especially FOXP1 and FOXP2, have generated considerable interest because of their roles in regulating cognitive developmental processes such as speech acquisition. FOXP2, more specifically, was the first member of the FOXP subclass to be associated with language deficits (Lai *et al.*, 2001). However, more recently, FOXP1 has also been implicated in speech development (Hamdan *et al.*, 2010; Horn *et al.*, 2010).

FOXP1 and FOXP2 are both expressed in neural tissue during development (Shu *et al.*, 2001). Both proteins have two highly homologous repression domains (Li *et al.*, 2004). An expression study of FOXP2 and FOXP1 in developing and adult mice brains and in human foetal brain tissue found that while both genes were expressed in the brain, they each had their own distinct expression pattern with regards to different brain regions, with the exception of the basal ganglia and both genes were expressed in the cortex, albeit in different layers (Ferland *et al.*, 2003). Interestingly, a study of FOXP2 and FOXP1 expression in human and zebra finch (a songbird) brains found that both genes were expressed in brain circuits related to song learning and production (Teramitsu *et al.*, 2004). Expression patterns of both genes in human and zebra finch brain appeared to be highly similar (Teramitsu *et al.*, 2004). The structural and functional similarities between the two genes and the partial overlap in expression patterns in brain regions important for language may suggest a possible interaction between them. More recently, disruptions in FOXP1 have been implicated in speech delay, intellectual disability, and autism (Hamdan *et al.*, 2010; Pariani *et al.*, 2010; O'Roak *et al.*, 2011), suggesting that just like FOXP2, FOXP1 has a significant involvement in speech and language development. FOXP1 and FOXP2 heterodimerisation may have different transcriptional outcomes than their homodimers (Mohd *et al.*, 2017). In a situation where levels of FOXP2 are low,

heterodimerisation with FOXP1 could repress transcription, but as FOXP2 increases in amount, competition between FOXP2 homodimers and endogenous FOXP1 can lead to transcriptional activation (Spiteri *et al.*, 2007). The fact that FOXP1 and FOXP2 co-operate in the regulation of non-neural developmental processes (Shu *et al.*, 2007) and have overlapping expression patterns in a number of tissues and brain (Li *et al.*, 2004), suggests that these two proteins might have overlapping functions and thus may be partners in speech development. Indeed FOXP1/FOXP2/FOXP4 have been shown to hetero-associate and regulate tissue-specific gene transcription (Sin, Li and Crawford, 2014) and tissue-specific gene transcription (Li, Weidenfeld and Morrisey, 2004). Homodimerisation of FOXP1 and FOXP2 occur through the leucine zipper domain and the FHD (Wang *et al.*, 2003; Li, Weidenfeld and Morrisey, 2004; Stroud *et al.*, 2006; Chu *et al.*, 2011) which also mediate heterodimerisation (Shu *et al.*, 2001; Li, Weidenfeld and Morrisey, 2004; Sin, Li and Crawford, 2014), suggesting that the mechanism of association is identical.

Mutations in both FOXP1 and FOXP2, more especially mutations that disrupt the dimer interface or that disrupt DNA binding, are associated with cognitive dysfunctions, such as autism spectrum disorder (ASD) and intellectual disability (Bacon and Rappold, 2012). However, disorders associated with the two genes do not have completely overlapping symptoms, indicating that the two transcription factors might regulate different but related brain functions (Golson and Kaestner, 2016). For instance, language deficits caused by variations in mFoxp2 are generally accompanied by deficits in orofacial movements, whereas impairments in language acquisition associated with mFoxp1 variants, are always accompanied by another cognitive impairment (Bacon and Rappold, 2012).

1.6. IDENTIFICATION OF PROBLEM

Protein-protein interactions are important in both metabolic and regulatory processes. There is considerable variation in the types of complexes that are observed: large macromolecular complexes, such as the ribosome, are highly stable and permanent while dynamic and transient interactions control signalling and regulatory networks (Bhattacharyya *et al.*, 2006; Stein *et al.*, 2009; Bashor *et al.*, 2010).

The fact that the FOXP subfamily has evolved the ability to form homo- and hetero-dimers which is unusual in the FOX superfamily, might be of significance in the way they bind DNA and regulate transcription. Indeed, it is likely that the combination of multiple domains and multimeric FOXP complexes can regulate transcription through the spatial and temporal regulation of complex formation. Thus, knowing the structure, behaviour and protein-protein interactions of FOXP proteins is significant. This will enable us to determine the role of each structural feature of FOXP proteins in their transcriptional function.

Because of the importance of both FOXP1 and FOXP2 in development and regulation, cognitive and developmental disorders and their increasing prevalence in cancer, and because of the unusual ability of these proteins to form both homo- and hetero-dimers, it is of interest to determine whether oligomerisation of FOXP1 and FOXP2 contributes to DNA binding and hence transcriptional regulation. This better understanding of the mechanism of transcriptional regulation employed by these proteins could lead to the development of novel therapeutic strategies. Although the biophysical features of the individual dimerisation interfaces have been investigated in isolation, structural and biophysical data of the intact dimerisation interface, encompassing both the FHD and leucine zipper domains is required in order to uncover the exact mechanisms of protein-protein interactions.

1.7. AIM AND OBJECTIVES

Knowledge of the structure of proteins is critical in order to determine their functions and how they perform such functions. The FOXP proteins are believed to form dimers through two interfaces: the leucine zipper domain and forkhead box domain (FHD). The mechanisms underlying dimerisation and protein-protein interactions are, therefore, of great importance in understanding the function of FOXP proteins since they can provide necessary information on plausible reasons why the FOXP proteins can dimerise while other FOX proteins cannot. As a result, it is important to determine the structural behaviour of these transcription factors and their DNA binding characteristics. By studying the behaviour of the isolated variant of the FOXP protein containing both the FHD and leucine zipper domains, we can obtain important insight into dimerisation.

The main aim of this research is to characterise the structures of FOXP1 and FOXP2 focusing on their dimerisation interfaces so as to identify the most critical structural features and elements. This will facilitate the investigation into the mechanisms of both homo- and hetero-typic protein-protein interaction of FOXP1 and FOXP2.

To this end, the specific objectives of this research were to:

1. express and purify recombinant FOXP1 and FOXP2 variants encompassing both the leucine zipper and forkhead box domain
2. characterise the secondary and tertiary structural conformations of FOXP1 and FOXP2 variants
3. assess the native quaternary conformation of FOXP1 and FOXP2 variants using chromatographic techniques
4. study the protein-protein interaction of FOXP1 and FOXP2 including self-association and hetero-complex formation
5. investigate binding of both proteins to their consensus DNA sequences and the effects of protein-protein interactions on DNA binding

CHAPTER 2

EXPERIMENTAL

2.1. MATERIALS

Thrombin from human plasma (20 units) was supplied by Sigma-Aldrich (St. Louis, USA). Competent cells were purchased from New England Biolabs (Ontario, Canada). The chaperone plasmid kit used in protein expression was supplied by Clontech Laboratories, Inc (Beijing, China). Isopropyl β-D-1-thiogalactopyranoside (IPTG) and Dithiothreitol (DTT) were obtained from Inqaba Biotech (Pretoria, South Africa). SDS-PAGE molecular weight markers were purchased from Amersham® Biosciences (Buckinghamshire, UK). L-Histidine, 4-(2-hydroxyethyl)-1-piperazineethanasulfonic acid (HEPES) and Coomassie Brilliant Blue G250 were obtained from Sigma-Aldrich (St. Louis, USA). All proteins were purified with fast protein liquid chromatography using an ÄKTAprime plus liquid handling system and the following columns were used: 5 mL Histrap^{TM} Purification column, 5 mL desalting column and 5 mL Q-sepharose FF ion-exchange column, all supplied by GE Healthcare (Buckinghamshire, UK). Size exclusion chromatography columns, Superdex 200 Increase 10/300 GL size exclusion column and Yarra™ 3u SEC-2000, LC Column were purchased from GE Healthcare (Buckinghamshire, UK) and Phenomenex (Torrance, USA), respectively. Size exclusion chromatography standards were purchased from Sigma Aldrich (St. Louis, USA). FOXP Cognate DNA sequence 3′-GATACTTTCATTTGTGGATT-5′ (Nelson *et al.*, 2013) was synthesised and supplied by IDT, Whitehead Scientific (Cape Town, South Africa). All other chemicals used were of standard analytical grade. All solutions were filtered before spectroscopic studies using a 0.22 μm acetate filter (Osmonics).

2.2. METHODS

2.2.1. FOXP1 and FOXP2 constructs

The full length FOXP protein is made up of a number of different domains all N-terminal to the forkhead domain (FHD) (Shu *et al.*, 2001; Lai *et al.*, 2003). These domains all have different functions and can influence the expression of the full length FOXP protein in different ways when using *Escherichia coli* expression machinery. For instance, the polyglutamine tract is highly prone to aggregation (Scherzinger *et al.*, 1997; Nagai *et al.*, 2007), making poly-glutamine containing proteins highly unstable. Similarly, the zinc finger motif requires the addition of high concentration of zinc to growth culture during its expression which can inhibit the growth of *Escherichia coli* cells (Yao *et al.*, 2005), thus reducing the acquired protein yield. This study is focused on oligomerisation, particularly the behaviour of the FHD and leucine zipper domain. Therefore, near full-length protein, truncated N-terminal of the leucine zipper so as to remove the polyglutamine tract and the zinc finger but to include the FHD and the leucine zipper domain were used (figure 13).

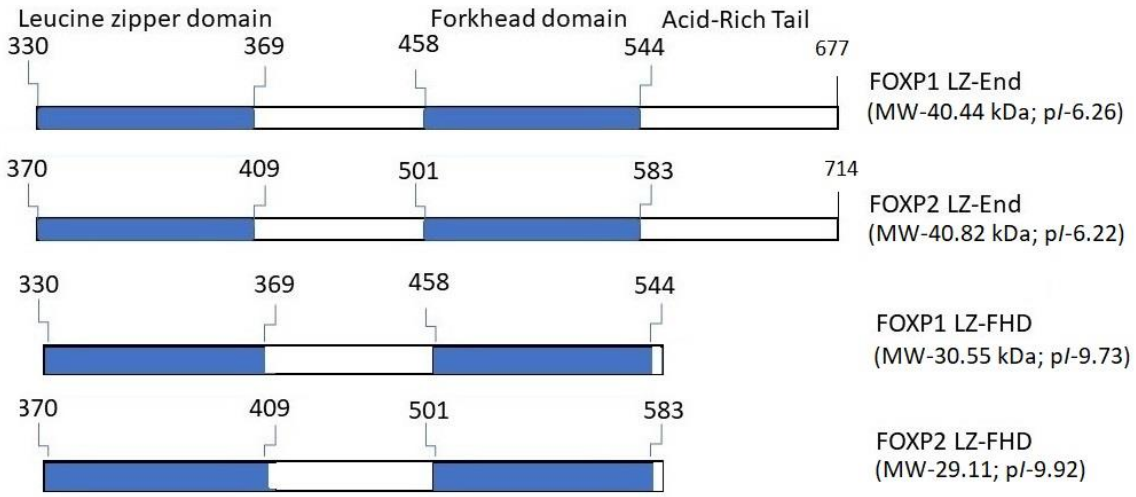


Figure 13: Architecture of FOXP1 and FOXP2 variants, showing the location of the leucine zipper domain and the FHD followed by the C-terminal acid rich tail. The amino acid sequence numbers encompassing the leucine zipper domain and the FHD are shown. The physical parameters (MW and p*I*) are given in the figure.

Each of the 4 variants used in this study contain both the leucine zipper domain and the FHD. Two of the variants include the amino acid sequence from just after the zinc finger domain to the end of the amino acid sequence (amino acid 330-677 (40.44 kDa) for FOXP1 and amino acid 370-714 (40.82 kDa) for FOXP2), these are termed FOXP1 LZ-End and FOXP2 LZ-End, respectively. The other two variants encompass amino acid 330-544 and 370-583 for FOXP1 and FOXP2, respectively and are truncated C-terminal of the FHD. They are termed FOXP1 LZ-FHD and FOXP2 LZ-FHD, respectively. These two variants were used to study dimerisation through the FHD and the leucine zipper domain without interference from the C-terminal acid rich region. Another variant, incorporating only the FHD and the C-terminal acid rich region was also designed (amino acid 501-714) called FOXP2 FHD-End (Figure 14). This variant was used to determine if the FHD is involved in heterodimerisation.

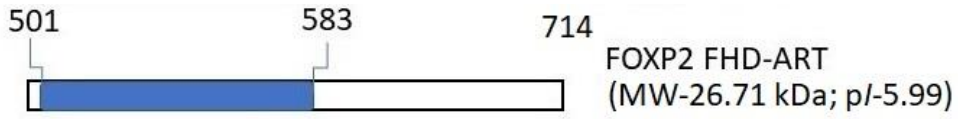


Figure 14: The amino acids sequence that makes up FOXP2 FHD-End variant. This variant was used to assess the involvement of specifically the FHD in hetero-association.

The genes encoding FOXP1 and FOXP2 variants were codon-optimised for expression in *Escherichia coli* bacterial cells. The genes were, respectively, cloned into the pET28a plasmid that has kanamycin resistance and pET11a with ampicillin resistance by GenScript (USA). The genes have a C-terminal hexa-His-tag preceded by a thrombin cleavage site, which is used to remove the tag.

2.2.2. Transformation

Two pET28a vectors, each encoding one of the FOXP1 variants (FOXP1 LZ-FHD and FOXP1 LZ-End) were used to transform BL21 (DE3) *Escherichia coli* cells for expression of the FOXP1 protein variants. The cells that were transformed with pET28a for expression of FOXP1 LZ-FHD were further transformed using the pKJE7 vector encoding *Dna*J-*Dna*K-*Grp*E to aid in correct folding of the protein during expression and thereby to assist with soluble expression. Three pET11a vectors, each encoding one of the FOXP2 variants (FOXP2 FHD-End, FOXP2 LZ-FHD and FOXP2 LZ-End) were used to transform T7 express *pLys*S *E. coli* cells for over-expression and subsequent purification of the FOXP2 protein variants.

A one-step transformation of BL21 (DE3) *E. coli* cells was performed as follows: 50 µL of cells were left to thaw on ice. Following this, 5 µL of plasmid DNA (and 1 µL of chaperone plasmid in the case of FOXP1 LZ-FHD) were added to the cell mixture and left for 30 minutes so as to stabilise the lipid membranes of the cells. The cells were heat-shocked at 42°C on a heating block for 90 seconds for pET28a transformation and for 45 seconds for pET11a transformation. The heat changes the state of the fluid membrane of the cell, increasing its permeability and allowing the DNA to enter the cell. The mixture was immediately placed on ice for 5 minutes to cool down and allow cell recovery. Thereafter, 950 µL of SOC media [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 10% (w/v) NaCl, 0.02 M glucose] was added to the transformed cells to provide nutrients and allow for cell recovery from heat-shock stress. The cells were then incubated at 37 °C for an hour with shaking at 230 rpm. The cells were then plated onto LB-agar [1% tryptone, 0.5% yeast, 1% NaCl, 1.5% agar] plates supplemented with 30 µg/ml kanamycin and 50 µg/ml chloramphenicol for selection of pET28a and pKJE7 transformants or with 100 µg/mL ampicillin to select for pET11a transformants. The plated cells were then incubated at 37 °C overnight, to allow for growth of successfully transformed cells.

2.2.3. Expression and purification

Glycerol stocks of transformed bacterial cells were added to flasks containing 100 mL LB media supplemented with either 30 µg/mL kanamycin or 100 µg/mL ampicillin, for subsequent expression of FOXP1 or FOXP2, respectively. The cells were left to grow overnight for 16 hours at 37 °C with 230 rpm shaking.

FOXP1 and FOXP2 variants were purified as required from four to six litre batch cultures of transformed BL21 or T7 express *Escherichia coli* cells, respectively. Fresh media was inoculated with the overnight culture with a 1:100 dilution. In the case of FOXP1 LZ-FHD, 0.5 mg/ml of L-arabinose was additionally added to the media during inoculation so as to induce expression of chaperones which are required to assist in the correct folding of FOXP1 LZ-FHD during over-expression. Prior to induction of expression, the culture was allowed to grow at 37 ℃ with shaking at 230 rpm to mid-log phase (OD_{600}=0.6) which took about 2 hours.

Overexpression of all variants was induced with 0.5 mM final concentration of isopropyl β-D-galactopyranoside (IPTG). FOXP1 LZ-End, FOXP1 LZ-FHD and FOXP2 LZ-FHD were all expressed for 22 hours at 37 ℃ with 230 rpm shaking following induction, while FOXP2 LZ-

End was left to overexpress for 4 hours post induction. FOXP2 FHD-End was expressed as outlined in Stroud *et al*. (2006) and Morris and Fanucchi (2016) (Stroud *et al.*, 2006; Morris and Fanucchi, 2016).

Following over-expression, the cells were harvested by centrifugation at 6500xg at 4 ℃ and resuspended in HisTrap^{TM} Equilibration Buffer [20 mM Tris-Cl, pH 7.6, 500 mM NaCl and 50 mM imidazole]. Then, 10 µg/mL lysozyme and 1 µg/mL DNase were added to the resuspended cells. And the cells were then stored at -20^{o}C. Lysozyme assists in cell lysis by weakening the cell wall of bacterial cells (Birdsell and Cota-Robles, 1967; Hunter, Muir and Thirkell, 1973) while DNase fragments bacterial DNA, thereby minimising DNA contamination of proteins (Tetz, Artemenko and Tetz, 2009; Tetz and Tetz, 2010).

Lysis of transformed BL21 and T7 express cells containing FOXP1 and FOXP2 variants was done using sonication. In this method of cell lysis, a sonicator probe is submerged into the resuspended cell solution and high frequency sound waves are generated for around 30-60 seconds. The sound waves force the cells to break via shearing force and cavitation. Sonication is usually done on ice to reduce protein denaturation since some heat is produced during the process (Sambrook and Russel 2006).

Cells were allowed to defrost at 20 ℃ for about 2 hours. The cells were then lysed on ice by sonication for 5 rounds of 5x6 second pulses, using a power output of 15 with a Sonicator Ultrasonic Processor purchased from Misonix Incorporates (New York, USA). The cell debris was then, separated from the soluble fraction by centrifugation at 6500xg for 20 minutes at 4^{o}C.

Once the soluble fraction containing the protein of interest had been separated from the cell debris and verified using SDS-PAGE, it was subsequently loaded onto a 5 mL nickel pre-charged HisTrap purification column (GE Healthcare, UK). The column was first equilibrated with HisTrap Equilibration Buffer [20 mM Tris-Cl, pH 7.6, 500 mM NaCl, 50 mM imidazole]. Purification was performed using the ÄKTA fast protein liquid chromatography (FPLC) purification system (GE Healthcare, UK). The hexa-histidine tagged protein bound non-covalently to the nickel ions on the HisTrap column. This was followed by a high salt wash [20 mM Tris-Cl, pH 7.6, 1.5 M NaCl, 50 mM imidazole, 1% Triton-X100, 0.5 % Tween 20]. The high salt buffer helps to get rid of cellular DNA fragments that may be bound to the protein. The column was then washed with 10 column volumes of the HisTrap Imidazole Wash Buffer [20 mM Tris-Cl, pH 7.6, 500 mM NaCl, 100 mM imidazole]. This helped to get rid of

contaminants that might bind to the column less tightly than the tagged protein. A one-step elution was used to elute FOXP1 or FOXP2 variants using a HisTrap Elution Buffer [20 mM Tris-Cl, pH 7.6, 500 mM NaCl, 500 mM imidazole]. At high concentrations, imidazole out-competes the tagged protein for binding to the nickel ions in the column and thus the protein is eluted.

Chaperones such as *dnaK* and *dnaJ* usually co-purify with the protein of interest, as they have a tendency of interacting with the thrombin cleavage site (Rial and Ceccarelli, 2002). Chaperone contamination was observed following IMAC purification of FOXP1 LZ-FHD. Ion exchange chromatography was thus used as a purification step after IMAC in the purification of FOXP1 LZ-FHD to remove the extra contaminants (chaperones) that came off the Ni^{2+} HisTrap™ column during FOXP1 LZ-FHD purification. The sample fractions collected from IMAC were pooled and loaded onto a 5 mL desalting column connected to a Q-Sepharose FF HiTrap™ (GE Healthcare, UK) anion exchange column pre-equilibrated with a 20 mM Tris-HCl buffer (pH 7.5). At pH 7.5, FOXP1 LZ-FHD (p*I* 9.73) does not bind the negatively charged resin and passes through the column while the chaperones, which have a p*I* below 6 do bind. The bound chaperones were then eluted by increasing the salt concentration of the buffer.

The purified protein was then dialysed into Storage Buffer [20 mM Tris-Cl, pH 7.5, 500 mM NaCl, 0.5 mM EDTA, 2 mM DTT]. High salt concentrations were used as they aid in reducing protein aggregation.

2.2.4. SDS-PAGE

Sodium dodecyl sulfate (SDS) is an anionic detergent, with a negative charge over a wide pH range. SDS binds proteins in proportion to their relative mass, confers a net negative charge and denatures the protein into individual polypeptide units (Roy and Kumar, 2011). Electrophoresis is based on the movement of charged particles in an electric field. Electrophoresis on a discontinuous polyacrylamide gel used as a support medium was performed when using the SDS-PAGE method according to the Laemmli (Laemmli, 1970) to separate proteins. To retain the reduced state of proteins, a reducing agent, such as β-mercaptoethanol is added to the sample buffer. Polyacrylamide gel electrophoresis in the presence of SDS provides an easy way for rapid and simple estimation of molecular weight of proteins and their subunits (Shapiro *et al.*, 1967).

SDS-PAGE gels consisted of 4% acrylamide stacking gels [4% (w/v) acrylamide, 0.36% (w/v) bis-acrylamide, 50 mM Tris-Cl pH 6.8, 0.1% (w/v) SDS, 0.005% (w/v) ammonium per sulfate, 0.2% (v/v) TEMED] and 12.5% acrylamide separating gels [12.5% (w/v) acrylamide, 1.08% (w/v) bis-acrylamide, 250 mM Tris-Cl, pH 8.8, 0.1% (w/v) SDS, 0.05% (w/v) ammonium per sulphate, 0.2% TEMED]. The samples were all diluted with equal volume of SDS reducing sample buffer [125 mM Tris-HCl, pH 6.8, 4% (w/v) SDS, 20% (v/v) glycerol, 10% β-mercaptoethanol, 3.5 µg/mL bromophenol blue]. The samples were then boiled at 95ºC for 5 minutes to further denature the protein before being loaded on the polymerised gel. The gels were subjected to electrophoresis using a tank buffer (running buffer) [250 mM Tris-Cl, *p*H 8.3, 192 mM glycine, 0.1% (w/v) SDS] at 165 V for about 2 hours to resolve the proteins. A molecular weight standard containing the following proteins was used: β-galactosidase (116 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), lactate dehydrogenase (35 kDa), and restriction endonuclease Bsp981 (25 kDa), β-lactoglobulin (18.4 kDa) and lysozyme (14.4 kDa). The gels were then stained using a Coomassie stain solution [0.1% (w/v) Coomassie Brilliant Blue R-250 dye in 1:5:4 acetic acid-methanol-water solution] for about 1 hour. This was followed by de-staining overnight using a de-staining solution [1: 5: 4 acetic acid: methanol: water].

2.2.5. Assessment of protein purity and determination of concentration

Samples of the purified protein were resolved on a 12.5% discontinuous polyacrylamide gel with a glycine buffer system and stained with Coomassie Brilliant Blue R-250 to assess the degree of purity. Only protein samples with a purity of at least 95%, determined by densitometry using a ChemiDoc™ XRS+ Imaging System (Bio-Rad, USA), were pooled and used for subsequent experiments. DNA contamination levels were assessed using the ratio of 280 nm light absorbance to 260 nm light absorbance (A280/A260) by the protein sample. Absorbance measurements were performed on a Jasco V-630 UV/Vis absorbance spectrophotometer in scanning mode. Sample concentrations were adjusted to give an absorbance at 280 nm in the range of 0.5 to 1.

Protein concentration was determined by absorbance at 280 nm wavelength light using the Beer-Lambert law ($A=\varepsilon cl$) with extinction coefficients (FOXP1 LZ-FHD: 27055 M^{-1} cm^{-1}; FOXP2 LZ-FHD: 28545 M^{-1} cm^{-1}; FOXP1 LZ-End: 30035 M^{-1} cm^{-1}; FOXP2 LZ-End:

28670 M^{-1} cm^{-1}) determined from the amino acid sequence using ProtParam embedded in the ExPASy server (Gasteiger *et al.*, 2005). Measurements were taken on a Jasco V630 UV/Vis absorbance spectrophotometer in single wavelength mode. Before all experiments were performed, protein samples were centrifuged for 5 minutes at 12 000 x g to remove aggregates. The absorbance of the undiluted solution was determined by extrapolation of a linear regression fitted to the absorbance values of five serial dilutions. The determination of concentration was repeated in triplicate and the average taken as the final concentration of the sample for further experiments.

2.2.6. Circular dichroism

Circular dichroism (CD) measures the differences in light absorbed between left-circularly polarised light and right-circularly polarised light (Woody, 1995). CD is displayed by chiral molecules because these molecules have the ability to interact with polarised light (Kelly, Jess and Price, 2005). The amide chromophore of the peptide bond has two low energy electronic transitions, $n \rightarrow \Pi^*$ at 215-230 nm and $\Pi_o \rightarrow \Pi^*$ 185-200 nm (Woody, 1995; Corrêa and Ramos, 2009). In the far-UV CD range, proteins show characteristic behaviour depending on the secondary structural elements of the polypeptide backbone (Woody, 1995; Kelly, Jess and Price, 2005). Therefore, far-UV CD is an excellent probe to use study the secondary structure of proteins.

For far-UV CD measurements, ~8 µM of FOXP1 variants and ~5 µM FOXP2 variants was used. More protein was used for FOXP1 variants because at low concentration (below 5 µM) the signal was noisy. FOXP1 and FOXP2 variants were filtered using a 0.2 µm filter to remove aggregates or dust particles that may interfere with polarisation of light. Far-UV circular dichroism measurements were performed at different pH 5-9 (Table 1). The effects of ethanol on secondary structure of the FOXP1 and FOXP2 variants was also tested using far-UV CD. Circular dichroism measurements were done, using a buffer containing 200 mM NaCl, 10 mM HEPES pH 7.5, 2 mM DTT and increasing concentrations of ethanol (0-60 %). The spectra were collected on a Jasco J-1500 spectropolarimeter at 20 ℃ with a data pitch of 1 nm, scanning speed of 200 nm/min and a band width of 0.5 nm. An average of five spectral accumulations was obtained for each of the samples. The average spectrum was corrected by subtracting the buffer spectrum from the average sample spectrum.

Table 1: Buffers used to study the secondary and tertiary structures of FOXP1 and FOXP2 variants

	Buffer	Components
Fluorescence & Far- UV circular dichroism	pH 5 buffer	10 mM sodium acetate, 200 mM NaCl, 2 mM DTT at pH 5
	pH 6 buffer	10 mM sodium acetate, 200 mM NaCl, 2 mM DTT at pH 6
	pH 7.5 buffer	10 mM HEPES, 200 mM NaCl, 2 mM DTT at pH 7.5
	pH 8 buffer	10 mM borate, 200 mM NaCl, 2 mM DTT at pH 8
	pH 9 buffer	10 mM borate, 200 mM NaCl, 2 mM DTT at pH 9

The data was obtained in millidegrees and was converted to mean residue ellipticity (MRE) (deg.cm^2.dmol^{-1}) using the following equation:

$$[\theta]MRE = \frac{100\theta}{Cnl}$$

Where n is the number of amino acids, θ is the observed ellipticity in degrees, l is the path length in cm and C is the concentration (mmol).

Thermal unfolding of the FOXP1 LZ-FHD, FOXP1 LZ-End, FOXP2 LZ-FHD and FOXP2 LZ-End were monitored with far-UV circular dichroism. Measurements were taken on a Jasco J-1500 circular dichroism spectropolarimeter following the 222 nm signal, a wavelength that gives an indication of the α-helical content of a protein, over a temperature range of 20-80 ℃ with a gradient of 1 ℃/min.

2.2.7. Fluorescence spectroscopy

Fluorescence is a spectrochemical method of analysis based on excitation and emission of electromagnetic radiation (Lakowicz, 2002). In fluorescence experiments, molecules of the analyte are excited by irradiation at a certain wavelength and the emission of radiation is monitored. Fluorescence corresponds to the relaxation of a molecule from a singlet excited state to a singlet ground state with emission of light that is short-lived (lifetime of about 10^{-8} seconds) (Povrozin and Barbieri, 2016). Protein molecules contain three aromatic amino acid residues, tryptophan, tyrosine and phenylalanine, which contribute to their intrinsic fluorescence (Chen and Barkley, 1998; Vivian and Callis, 2001). However, most intrinsic

fluorescence measurements of protein in the near-UV range focus on tryptophan and changes in its environment, since the fluorescence intensity of all the other residues is usually swamped by that of tryptophan (Chen and Barkley, 1998). When tryptophan is buried in the nonpolar core of globular proteins it results in a characteristic blue shift of its fluorescence spectrum (λ_{ex}= 295 nm) (Chen and Barkley, 1998; Vivian and Callis, 2001; Lakowicz, 2002; Lakowicz and Masters, 2008). The position of the tryptophan fluorescence spectrum, being sensitive to the polarity of the microenvironment, allows us to study protein folding (Stryer, 1968) or other conformational transitions accompanied by the change of tryptophan solvatation, i.e. processes leading to considerable changes in the compactness of the protein molecule (Uversky *et al.*, 1997). Therefore, fluorescence is an ideal technique to study the tertiary structure of FOXP1 and FOXP2 variants.

The tryptophan residues in both FOXP1 and FOXP2 variants are localised within the forkhead domain (FHD) (for FOXP1: Trp494, Trp509 and Trp534 and for FOXP2: Trp533, Trp548 and Trp573). Therefore, fluorescence is useful as a local probe to monitor any conformational changes within the DNA binding domain (FHD). Fluorescence measurements were performed on purified FOXP1 and FOXP2 variants (5 µM) at different pH (5-9) (Table 1) using a 10 mm path length quartz cuvette on a Jasco FP -8300 spectrofluorometer. The effects of ethanol on tertiary structure of the FHD was also tested using intrinsic tryptophan fluorescence. Fluorescence measurements were done using 200 mM NaCl, 10 mM HEPES pH 7.5, 2 mM DTT and increasing concentrations of ethanol (0-60%). The emission was recorded from 300 nm to 400 nm using a data pitch of 0.5 nm, excitation bandwidth of 5 nm and emission band width of 5 nm. Three spectral accumulations were collected for each sample, averaged and corrected by subtracting the buffer spectrum.

2.2.8. Size exclusion chromatography (SEC)

Size exclusion chromatography (SEC) is a technique that involves separation of molecules according to their size in solution (Hagnauer, 1981). It is based on the observation that small molecules could be excluded from the small pores of zeolites as a function of their molecular size (Hong *et al.*, 2012). SEC provides a means to determine the oligomeric state of proteins by measuring the hydrodynamic volume of all the different species (if there are multiple), in native solution conditions (Ramsey, Daugherty and Kelm, 2006). Size exclusion chromatography is also based on the assumption that proteins are globular and is often

erroneous with asymmetric particles (Rambo, 2017). The hydrodynamic volume of asymmetric protein particles is usually larger than the corresponding hydrodynamic volume of a globular protein of the same size. This anomalous behaviour in SEC, therefore, makes it difficult to obtain accurate sizes. However, knowledge of the hydrodynamic volume of a protein from SEC can be used to provide a qualitative indication of the oligomeric state of the protein. Thus, SEC was used to assess the hydrodynamic volumes, and therefore the oligomeric state, of FOXP1 and FOXP2 variants.

Size exclusion-High Performance Liquid Chromatography was used as an analytical technique to determine changes in quaternary structure that may be induced by increasing protein concentrations. Different concentrations of purified FOXP1 and FOXP2 variants (5 μM–100 μM) were separated on a Yarra™ 3u SEC-3000 (Phenomenex, USA), LC column pre-equilibrated with HPLC Equilibration Buffer [10 mM HEPES *p*H 7.5, 500 mM NaCl, 2 mM DTT] attached to a SHIMADZU SPD20A HPLC machine. The column was calibrated using Bio-Rad Standards 1.3 kDa – 670 kDa (Bio-Rad, USA).

The native quaternary conformation at equilibrium following incubation at 20 ℃ for 20 hours was investigated using ~ 20 μM purified FOXP1 and FOXP2 variants. The column was calibrated using Gel Filtration Markers Kit for Protein Molecular Weights 12 kDa – 200 kDa (Sigma-Aldrich, USA). Changes in the native conformation were then determined at different buffer *p*H (5-9) (table 2). Because ethanol has an impact on hydrophobic interactions, its effect on oligomerisation was also investigated. This was done to attempt to understand the strength of the hydrophobic interactions in the dimer interface. The effects of ethanol were tested on FOXP1 and FOXP2 variants using ethanol buffer [500 mM NaCl, 10 mM HEPES, pH 7.5, 2 mM DTT, 15% ethanol]. using size exclusion chromatography (SEC) on a Superdex® 200 10/300 GL column (GE Healthcare, USA). The column was calibrated using Gel Filtration Markers Kit for Protein Molecular Weights 12 kDa – 200 kDa (Sigma Aldrich, USA).

Table 2: Different buffers used for determination of quaternary structure and oligomeric state

	Buffer	Components
Size exclusion chromatography & Dynamic Light Scattering	pH 5 SEC buffer	10 mM sodium acetate, 500 mM NaCl, 2 mM DTT at pH 5
	pH 6 SEC buffer	10 mM sodium acetate, 500 mM NaCl, 2 mM DTT at pH 6
	pH 7.5 SEC buffer	10 mM HEPES, 500 mM NaCl, 2 mM DTT at pH 7.5
	pH 8 SEC buffer	10 mM borate, 500 mM NaCl, 2 mM DTT at pH 8
	pH 9 SEC buffer	10 mM borate, 500 mM NaCl, 2 mM DTT at pH 9

2.2.9. Dynamic light scattering

Dynamic light scattering (DLS is a non-invasive technique for measuring the translational diffusion coefficient (DT) of a macromolecule undergoing Brownian motion in solution (Schmitz and Phillies, 1991; Ferré-D'Amaré and Burley, 1994). Samples, generally at equilibrium, are illuminated with monochromatic light, scattering the light based on concentration and density fluctuations; the scattered light acquires a power spectrum mirroring the temporal evolution of the fluctuations (Schmitz and Phillies, 1991). By scattering light from small particles, their geometrical structure and their state of motion can be measured (Goldburg, 1999). Hydrodynamic methods such as DLS can be applied to characterise the dynamic properties and particle size of proteins in solution.

The hydrodynamic volume and protein size of FOXP1 and FOXP2 variants were determined using dynamic light scattering on a Malvern Zetasizer Nano S. A 10 mm path length glass cuvette was loaded with ~ 15 µM of filtered (0.02 µm) FOXP1 and FOXP2 variants, individually at different *p*H (5-9) (Table 2). Each protein sample was allowed to equilibrate for 120 seconds in the machine at 20 ℃ before each measurement was carried out. Each measurement was done in triplicate, averaged and normalised against buffer measurements. Measurements for two standard proteins (conalbumin, 75 kDa; diameter of ~8.72 ±0.37 nm and aldolase, 158 kDa; diameter of ~ 10.3 nm) were carried out and their sizes used as a reference point for the sizes of FOXP1 and FOXP2 variants.

2.2.10. Blue-Native polyacrylamide gel electrophoresis (BN-PAGE)

Blue native polyacrylamide gel electrophoresis (BN-PAGE) was first developed for membrane proteins by Schagger and von Jagow (1991) (Schägger and von Jagow, 1991). BN-PAGE is a discontinuous native protein gel electrophoresis system that allows the separation of proteins according to their size, oligomeric state, and shape (Niepmann and Zheng, 2006). The technique follows the same principles as SDS-PAGE; however, proteins are separated without being denatured. The anionic dye Coomassie Blue G-250 is added to the sample before electrophoresis to add a negative charges to the proteins (Wittig, Braun and Schägger, 2006). Native proteins and complexes migrate as blue bands through BN gels facilitating visualisation and excision of specific bands (Wittig, Braun and Schägger, 2006). It is ideal for the determination of the oligomeric state of proteins and in determination of complex formation. Usually, a second dimension (such as denaturing SDS-PAGE) is added for confirmation of oligomeric states and complexes. The system is ideal for biological species as it works at a *p*H near 7.5 (Schägger and von Jagow, 1991).

In this study, BN-PAGE was used to confirm the oligomeric state of each of FOXP1 and FOXP2 variants and to monitor complex formation with increasing protein concentration. Electrophoresis on a discontinuous polyacrylamide gel used as a support medium was performed on an 8% gel prepared according to the blue native PAGE protocol (Niepmann & Zheng, 2006; Wittig *et al.*, 2006). The native-PAGE gels consisted of 4% acrylamide stacking gels [4% (w/v) acrylamide, 0.36% (w/v) bis-acrylamide, 50 mM Tris-Cl pH 6.8, 0.005% (w/v) ammonium persulfate, 0.2% (v/v) TEMED] and 8% acrylamide separating gels [8% (w/v) acrylamide, 1.08% (w/v) bis-acrylamide, 250 mM Tris-Cl, *p*H 8.8, 0.05% (w/v) ammonium per sulfate, 0.2% TEMED]. The samples were all diluted with equal volume of sample buffer [40% (v/v) glycerol, 0.5% (w/v) Coomassie Blue G250]. Different concentrations 5-50 µM of FOXP1 and FOXP2 were used. Prior to electrophoresis, FOXP1 and FOXP2 variants were first dialysed into native-PAGE buffer [10 mM HEPES, pH 7.5, 100 mM NaCl]. The gels were then subjected to electrophoresis using a cathode buffer [100 mM L-Histidine, titrated to pH 8.0 using 1 M Tris base] and an anode buffer [200 mM Tris-Cl, pH 8.8] at 100 V for about 3 hours at 4 ℃ to resolve the proteins. The separated protein species were compared to standard proteins: ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa) and ovalbumin (44 kDa). Pictures of the gels were obtained immediately following electrophoresis using the ChemiDoc™ XRS+ Imaging System (Bio-Rad, USA).

2.2.11. Disorder prediction using DISOPRED server

The DISOPRED server uses a knowledge-based method to predict dynamically disordered regions from the amino acid sequence (Ward *et al.*, 2004). Single letter amino acid sequences are submitted to the DISOPRED server and the results are delivered to the user by email. To predict disorder, DISOPRED2 initially conducts a PSI-BLAST search over a filtered sequence database (Altschul *et al.*, 1997). Each residue is then encoded by the profile for a window of 15 positions in the sequence and classified using a neural network (Ward *et al.*, 2004). The classifier is trained using a support vector machine learning algorithm and outputs a probability estimate of the residue being disordered (Ward *et al.*, 2004). The PSIPRED protein structure prediction server incorporates three recently developed methods, GenTHREADER (Jones, 1999a), PSIPRED (Jones, 1999b) and MEMSAT 2 (Jones, Taylor and Thornton, 1994) for predicting secondary structural information about a protein from its amino acid sequence alone (McGuffin, Bryson and Jones, 2000). DISOPRED2 and PSIPRED were used to predict disorder and secondary structural elements of FOXP1 and FOXP2 variants based on their amino acid sequences.

2.2.12. Electrophoretic mobility shift assay (EMSA)

Gel electrophoresis mobility shift assay (EMSA) is used to detect protein complexes with nucleic acids (Hellman and Fried, 2007b). Its based on the observation that the electrophoretic mobility of a protein-nucleic acid complex is typically less than that of the free nucleic acid (Fried, 1989). The potential to separate complexes of differing stoichiometry or conformation is a major advantage of the gel method over other common assays and like other gel assays, electrophoretic mobility shift assay (EMSA) is a rapid and sensitive method in the detection of protein-nucleic acid interactions. The assay is easy to use and can accommodate a wide range of binding conditions making it excellent to study DNA-binding affinities (Fried, 1989; Hellman and Fried, 2007b). The binding of proteins to DNA is an important step in many cellular functions, including DNA replication, recombination and repair, transcription and viral assembly, and therefore it is crucial to understand formation of such protein-DNA complexes, in order to fully dissect these biological processes.

EMSA was used to monitor the ability of FOXP1 and FOXP2 variants to form protein-DNA complexes. The confirmation of the DNA-binding function of these proteins also served to confirm that each variant folded into its native 3D structure. Protein was first dialysed into buffer containing 10 mM HEPES, pH 7.5, 100 mM NaCl. A duplex cognate DNA oligonucleotide containing a single binding site as determined by (Nelson *et al.*, 2013), 5′-TTAGG**TGTTTA**CTTTCATAG-3′ was used for EMSA analysis. The DNA and protein samples for analysis were mixed in the ratio 1:0.5, 1:1, 1:2, 1:5 1:10 and 1:20 (DNA:protein) and a binding buffer containing 10 mM HEPES, pH 7.5, 1 mM $MgCl_2$, 100 mM KCl, 10% (w/v) glycerol and 0.1 mg/ml BSA was added. BSA is added to the binding buffer to minimise non-specific losses of binding proteins during solution handling (Hellman and Fried, 2007b). The samples were then incubated on ice for 30 minutes to allow for complex formation and then loaded on an 8% continuous EMSA gel [8% (w/v) acrylamide, 0.69% (w/v) bis-acrylamide, 5xTBE (450 mM Tris, 450 mM Boric acid and 13 mM $Na_2EDTA.H_2O$), 0.05% (w/v) ammonium persulfate and 0.2% TEMED]. The gel was subjected to electrophoresis at 4 ℃ for 4 hours at 160 V in 1xTBE buffer [90 mM Tris, 90 mM Boric acid and 2.6 mM $Na_2EDTA.H_2O$]. The complexed samples were subjected to electrophoresis in conjunction with free DNA samples. The gels were then stained using SYBR® Gold and viewed under UV-light using a ChemiDoc™ XRS+ Imaging System (Bio-Rad, USA).

2.2.13. Immobilised metal ion affinity chromatography pull-down assay (IMAC-PDA)

The *in vitro* pull-down assay is a method that can be used to confirm direct binding in protein-protein interactions in cases when an interaction was inferred from other interaction assays (Suzuki *et al.*, 2004). The pull-down method relies on the immobilisation of a protein via an affinity tag (e.g. His-tag or GST-tag) on a solid phase while a secondary protein, usually untagged, is allowed to flow over the bound protein (or incubated) on the solid support (Schechtman, Mochly-rosen and Ron, not dated; Craig, Ciufo and Morgan, 2004). The method consists of first immobilising the tagged protein (bait) on an affinity ligand specific to the tag. This creates an affinity support to capture and purify other proteins (prey) that are capable of interacting with the bait (Louche, Salcedo and Bigot, 2017). The extent of binding is assayed by SDS-PAGE (Craig, Ciufo and Morgan, 2004). Pull- down assays are relatively cheap and easy to perform on recombinant proteins as they do not require expensive machinery (Craig,

Ciufo and Morgan, 2004). The main limitation of this technique is the possibility of a false interaction positive due to either non-specific interaction with the solid support (Nguyen and Goodrich, 2006; Louche, Salcedo and Bigot, 2017). Additives such as salts and detergents can eliminate any non-specific interactions (Emmott and Goodfellow, 2014). In this work, FOXP2 variants were used as the bait while FOXP1 variants were used as the prey, in order to assess if an interaction could be detected between FOXP1 and FOXP2.

Protein-protein interactions between FOXP1 and FOXP2 variants were investigated using IMAC pull-down assays on a HiTrap™ Ni^{2+} IMAC column (GE Healthcare, USA) using purified FOXP1 variants as the prey and FOXP2 variants as the bait. Purified FOXP1 proteins were dialysed into thrombin cleavage buffer [100 mM Tris-Cl, pH 8.0, 2 mM $CaCl_2$ and 100 mM NaCl]. The C-terminal his-tag was removed by incubation of the protein with Thrombin for 12 hours at 20 ℃. The undigested fusion protein isolated his-tag and thrombin were then removed by IMAC followed by benzamidine affinity chromatography. The isolated tag-free FOXP1 protein was then dialysed into IMAC binding buffer [10 mM HEPES, pH 7.5, 100 mM NaCl, 30 mM imidazole]. The column was first equilibrated using IMAC binding buffer and the purified FOXP2 protein was loaded. FOXP1 protein was then loaded onto the column at a flow rate of 1 ml/min. Unbound protein was washed off using the binding buffer [10 mM HEPES, pH 7.5, 500 mM NaCl, 30 mM imidazole] while bound protein was eluted from the column using IMAC elution buffer [10 mM HEPES, pH 7.5, 100 mM NaCl, 500 mM imidazole]. The eluted protein was visualised using SDS-PAGE on a 12% tris-glycine polyacrylamide gel prepared as described above. Two bands on SDS PAGE would be indicative of an interaction.

2.2.14. Fluorescence anisotropy (FA)

Anisotropy of fluorescence is described as the difference in vertical and horizontal emission components with respect to the total fluorescence intensity when vertically polarised excitation is used (LeTilly and Royer, 1993). In the FA, a fluorophore is irradiated with linearly polarised light. The resultant fluorescence intensity is measured through a polarisation filter placed in front of the detector and oriented either parallel or perpendicular to the incident polarised light (Zhang, Wu and Berezin, 2015). The polarised emission is influenced by a number of processes, including motions that occur within the lifetime of the excited fluorophore, also known as rotational diffusion (Ghisaidoobe and Chung, 2014). Information on tumbling, rotational diffusion or hydrodynamic aspects of a macromolecular system are useful in

studying binding processes, that is, the association of small ligands with macromolecules or macromolecule-macromolecule interactions (Jameson and Sawyer, 1995). Fluorescence anisotropy (FA) is suitable for the study of protein-DNA complex formation, protein-ligand interactions and protein-protein interactions because its sensitivity allows detection at low protein or DNA concentrations (Deprez *et al.*, 2001). In molecular interactions involving protein biomolecules and DNA ligands, the use of simple anisotropy can provide evidence that interactions have occurred between the species and give the binding constants (Zhang, Lu and Wang, 2011).

Fluorescence anisotropy experiments were conducted on each variant to determine their DNA-binding affinity. A Perkin Elmer LS-50B fluorescence spectrophotometer was used, fitted with an anisotropy filter to monitor the tumbling of DNA as it binds to purified FOXP1 or FOXP2 variants. 5-carboxy-X-rhodamine (ROX)-labelled DNA (500 nM) was incubated with increasing concentrations of each of each variant on ice for 30 min. Fluorescence anisotropy measurements were taken with an emission wavelength of 605 nm following excitation of the ROX dye at a wavelength of 580 nm.

Fluorescence anisotropy experiments were also used to study protein-protein interactions between FOXP1 and FOXP2 variants so as to investigate whether they can form hetero-oligomers. In this case, FOXP1 LZ-End (~50 µM) and FOXP1 LZ-FHD (~45 µM) were individually labelled with nitrotriacetic acid coupled with ATTO 550 (NTA-ATTO 550) which is a fluorescent dye that binds specifically to 6×His-tag and incubated for 60 min at 20 ℃ as outlined in Zhao *et al.*, 2010. Excess fluorescent dye was removed using dialysis into binding buffer [10 mM HEPES, pH 7.5, 100 mM NaCl] for 12 hours. The labelled proteins were then resolved by SEC on a Superdex® 200 10/300 GL column (GE Healthcare, USA) at 500 µl/min. Excluded fractions were collected and monitored by absorbance at 550 nm for NTA-ATTO. The degree of protein labelling (dye/protein) in the pooled fractions was confirmed by comparing the concentration of NTA-ATTO with that of the protein. Labelling efficiency was quite poor; however, the amount of labelled protein was enough to proceed with FA experiments.

Labelled FOX1 LZ-End (5 µM) was incubated with increasing concentrations (0-25 µM) of FOXP2 FHD-End, FOXP2 LZ-FHD and FOXP2 LZ-End for 60 min at 20 ℃. Similarly, labelled FOXP1 LZ-FHD (5 µM) was incubated with increasing concentrations of FOXP2 FHD-End, FOXP2 LZ-FHD and FOXP2 LZ-End for 60 min at 20 ℃. Fluorescence anisotropy

measurements were taken with an emission wavelength of 569 nm following excitation with a wavelength of 554 nm.

Furthermore, fluorescence anisotropy was used to measure the DNA binding affinity of the FOXP1 and FOXP2 heterodimer. This was conducted using labelled FOXP1 LZ-End and labelled FOXP1 LZ-FHD (5 µM) incubated with increasing concentrations (0-20 µM) of FOXP2 FHD-End, FOXP2 LZ-FHD and FOXP2 LZ-End in the presence of 200 nM DNA (cognate DNA: 5′-TTAGG**TGTTTA**CTTTCATAG-3′ (Nelson *et al.*, 2013)) following the protocol outlined in Morris and Fanucchi, 2016. The protein: DNA complex was incubated for 30 minutes prior to fluorescence anisotropy measurements. The NTA-ATTO labelled protein was selectively excited at 554 nm, and the emission was monitored at 569 nm.

All data were obtained in triplicate and were averaged. The data were fit with a single site binding model using SigmaPlot version 13.0, from Systat Software, Inc., San Jose California USA, www.systatsoftware.com. Data analysis was done using a Student t-test in SigmaPlot. The t-test was used to determine the significance of the differences observed in dissociation constants (K_D) of hetero-association and DNA binding.

2.2.15. Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry is used to measure the energetics of biochemical reactions or molecular interactions at constant temperature (Freire *et al.*, 1990). These molecular interactions include ligand binding, enzyme-substrate interactions and protein-protein (or DNA) interactions among components of multi-molecular complexes (Velazquez-Campoy *et al.*, 2015). ITC measures the reversible reactions between biomolecules (Holdgate, 2010) and allows for rapid determination of the binding affinity, number of binding sites (stoichiometry), enthalpy and entropy of the binding reaction (Freire *et al.*, 1990). The reaction is triggered by changing of the chemical composition of the sample by step-wise titration of a required reactant. During this, heat is either absorbed or released and the change is measured by a sensitive microcalorimeter. The thermodynamic analysis of observed heat effects then allows for the quantitative determination of the energetic processes related to the association reaction. A reference cell containing the buffer solution which will account for injection and dilution effects during titration is also present. ITC is the only technique that can resolve the enthalpic and entropic components of binding affinities (Leavitt and Ernesto, 2001).

ITC was used in this study to determine the strength of association of FOXP1 and FOXP2, to characterise the potential multiple binding sites present in both FOXP1 and FOXP2 association interfaces and to characterise the enthalpic and entropic contributions to the Gibbs energy. The binding thermodynamics of the FOXP2 LZ-End to FOXP1 LZ-End was monitored. Protein samples were prepared by extensive dialysis against fresh binding buffer [10 mM HEPES, *p*H 7.5, 100 mM NaCl]. Protein concentration was determined by UV absorbance of 280 nm light using a NanoDrop® ND-1000 UV-Vis Spectrophotometer. Titration experiments were performed on a TA Instruments Nano-ITC by 25×10 μL injections of ~500 μM FOXP2 LZ-End into ~30 μM FOXP1 LZ-End at 25 °C. Titration experiments were also performed using ~480 μM FOXP2 LZ-End and ~100 μM FOXP1 LZ-End titrated into buffer [10 mM HEPES, pH 7.5, 100 mM NaCl] as outlined above. The heats of saturation were averaged, and the value subtracted from all data points before data analysis. Data were fitted using a non-linear least-squares method for multiple site binding independent site binding. Errors were determined as the standard deviation of three averaged independent titrations.

CHAPTER 3

RESULTS

To date, most studies dealing with the FOXP transcription factors are confined to the isolated FHD or, if the full-length protein is under investigation, the studies performed are mostly done *in vivo*. There is very little information available on the behaviour of the isolated proteins *in vitro*. Although work has been done on the isolated FOXP FHD and the isolated FOXP3 leucine zipper, longer variants encompassing more than one domain have not been studied *in vitro.* Furthermore, not much is known about the precise interaction between FOXP1and FOXP2 despite convincing cellular evidence of its existence. In this study a thorough biophysical characterisation of FOXP1 and FOXP2 was performed, focusing on their dimerisation interfaces and I identify the most critical structural features and elements in order to gain insight into their protein-protein interactions. This is done through the creation of FOXP1 and FOXP2 variants containing the FOXP domains in varying proportions and assessing their DNA binding properties, structural characteristics and homo- and hetero-associations.

3.1. EXPRESSION

The presence of multiple domains in the FOXP transcription factors provides challenges for expression and purification of the full-length proteins due to the modular nature and high predicted disorder of these proteins as well as the presence of the aggregation-prone poly-glutamine domain. Although the other domains might also be important in stabilising the full-length protein, it is the leucine zipper domain and the FHD that have thus far been identified as the most important domains in oligomerisation of FOXP proteins. *Escherichia coli* expression machinery was used to over-express all four FOXP1 and FOXP2 variants that contain both the leucine zipper and the FHD. Two of the variants also contain the C-terminal acid rich tail, that could be important in structural stability. Knowledge of the behaviour of these variants serves as a background for inquiry into hetero-association of FOXP1 and FOXP2. The use of *E. coli* for expression provides several advantages. *E. coli* cells multiply

very fast and in glucose-salts media, under optimal environmental conditions, its doubling time is about 20 min (Sezonov, Joseleau-Petit and D'Ari, 2007). Because of the fast growth and doubling time, high cell densities are achieved in optimal growth media (Rosano and Ceccarelli, 2014). Furthermore, it is quite fast and easy to transform *E. coli* cells using exogenous DNA. Some transformation protocols allow plasmid transformation of *E. coli* in as little as 5 min (Pope and Kent, 1996). Because of its advantages, *E. coli* was the expression machinery of choice.

Following expression trials, it was found that three of the four variants (FOXP1 LZ-End, FOXP1 LZ-FHD and FOXP2 LZ-FHD) expressed optimally in *E. coli* cells using 0.5 mM IPTG at 20 ℃ for a period of 20 hours (Fig 15 A, B and C). FOXP2 LZ-End followed a similar protocol and was successfully over-expressed in T7 Express pLysS *E. coli* cells in sufficiently high quantities at 37 ℃ for 4 hours using an IPTG concentration of 0.5 mM (Figure 15D). There was about an equal amount of protein in the soluble and insoluble fractions, however since there was enough quantity protein in the soluble fraction, there was no need for recovery from the insoluble fraction.

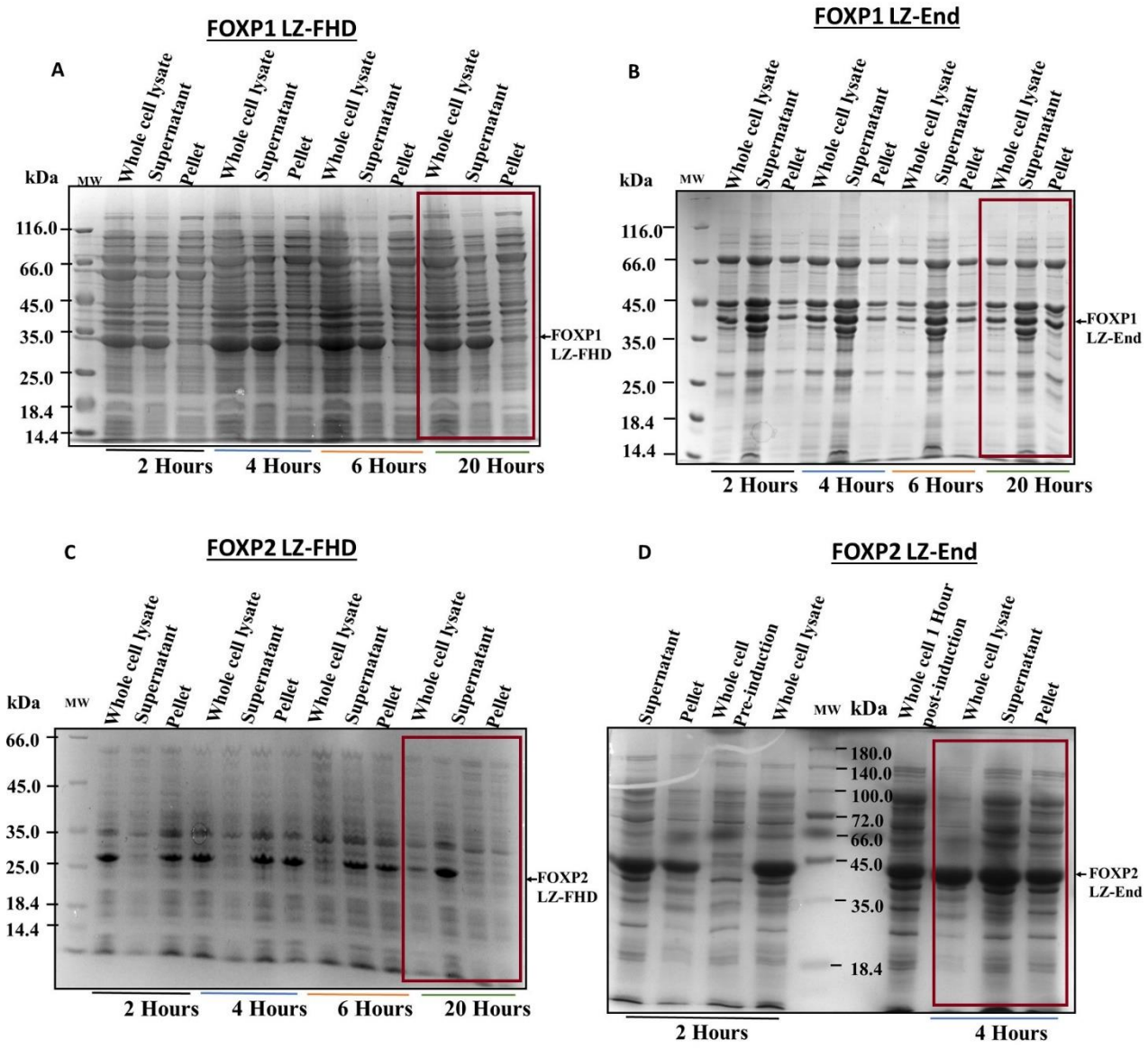


Figure 15: SDS-PAGE gel showing representative expression of FOXP1 and FOXP2 variants. Samples were taken at 4 different time periods post induction with 0.5 mM IPTG (final concentration). (A) FOXP1 LZ-FHD expression. The condition chosen for protein production was induction with 0.5 mM IPTG after 20 hours at 20 ℃ (boxed). (B) Induction trials for FOXP2 LZ-End. This variant expressed at high quantities after 4 hours at 37 ℃, both in the soluble fraction and as inclusion bodies. (C) Expression of FOXP2 LZ-FHD. The condition chosen for protein production was induction with 0.5 mM IPTG after 20 hours (boxed). (D) SDS-PAGE representative of FOXP2 LZ-End expression trials. The condition chosen for protein production was induction with 0.5 mM IPTG after 4 hours (boxed).

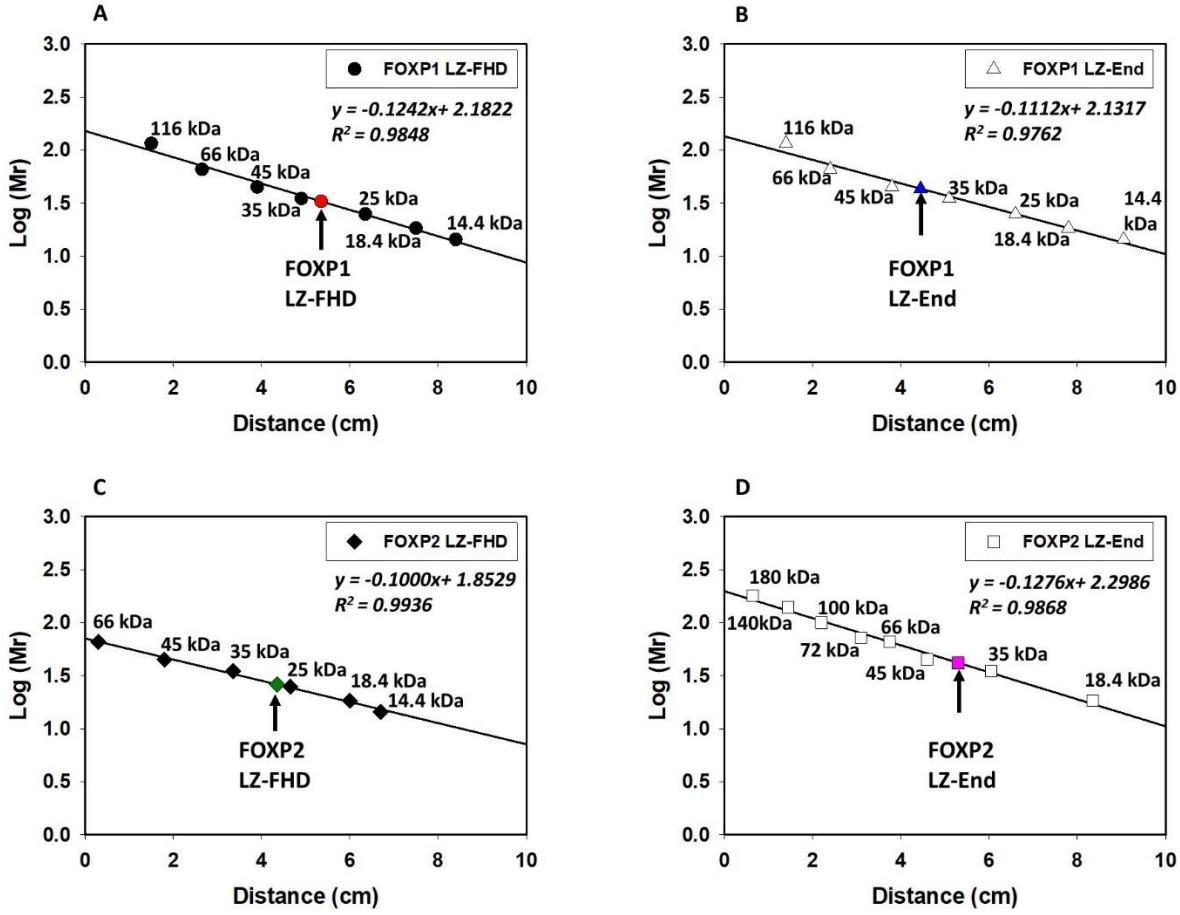


Figure 16: The calibration curve for the SDS-PAGE gel shown in Figure 15. The molecular mass (in kDa) and distances migrated by the standard proteins (shown as MW in gels) were used to construct this curve. (A) FOXP1 LZ-FHD (MW: ~32.9 kDa) (B) FOXP1 LZ-End (MW: ~43.3 kDa). Molecular weight of both FOXP1 LZ-FHD and FOXP1 LZ-End were determined against Unstained Protein Molecular Weight Marker (14.4-116 kDa) (C) Molecular weight of FOXP2 LZ-FHD (MW: ~26.2 kDa) was determined against Unstained Protein MW Marker (Low) (14 - 94 kDa). (D) Molecular weight of FOXP2 LZ-End (MW: ~41.9 kDa) was determined against Bio-Rad Protein Ladder (18.4 - 180 kDa).

FOXP1 LZ-FHD (Figure 15A) was the only one of the four variants that expressed as insoluble inclusion bodies. FOXP1 LZ-FHD was, therefore, the only one of the four variants that was expressed with the assistance of the molecular chaperones *dna*J-*dna*K-*grp*E. These chaperones assist in proper folding of misfolded protein. Under these conditions, FOXP1 LZ-FHD expressed mainly in the soluble fraction of the *Escherichia coli* BL21 (DE3) cells (Figure 15A). The pKJE7 plasmid encoding the chaperones has chloramphenicol; therefore, for successful expression, *E. coli* BL21 (DE3) cells were used for this variant instead of the T7 cells, because they have a *pLysS* plasmid which is used to reduce basal protein expression and encodes chloramphenicol resistance.

3.2. PROTEIN PURIFICATION

The four variants were isolated from other *E. coli* proteins using immobilised metal ion affinity chromatography (IMAC). Each variant has a C-terminal hexa-histidine tag, which exhibits a strong interaction with metals. IMAC interactions occur between a transition metal ion (Co^{2+}, Ni^{2+}, Cu^{2+} or Zn^{2+}) immobilised on a matrix and histidine residue side chains of the hexa-histidine tag (Bornhorst and Falke, 2000).

All four variants were purified using immobilised metal ion affinity chromatography charged with either Ni^{2+} or Co^{2+}. Both FOXP1 variants (FOXP1 LZ-End and FOXP1 LZ-FHD) as well as FOXP2 LZ-FHD were purified using a Ni^{2+}-charged IMAC HiTrap™ column. Because of the high quantities of FOXP2 LZ-End expressed in the soluble fraction of *E. coli* cells (Figure 15), FOXP2 LZ-End was purified using a Co^{2+}-charged IMAC column rather than a Ni^{2+}-charged column. His-tags have a weaker affinity for cobalt compared to Ni^{2+} but higher specificity (Soghoian, 2004; Wu *et al.*, 2013); therefore, because large amounts of FOXP2 LZ-End were produced, the reduced affinity for the cobalt was outweighed by the fact that there were fewer non-specific binding proteins and the purification of the FOXP2 LZ-End variant resulted in a relatively higher degree of purity while still being at an acceptable yield.

For all four variants, a high salt wash (1.5 M NaCl) was done after loading the supernatant onto the IMAC column to remove any unbound proteins and *E. coli* DNA fragments that may be bound to the protein of interest. This step is of importance because both FOXP1 and FOXP2 are transcription factors and have high affinity for DNA. Changes in absorbance were not observed during the salt wash, implying that unbound *E. coli* proteins and unbound fragments

of *E. coli* DNA had already passed through the column together with the non-specific proteins. After the salt wash, a one-step elution using 500 mM imidazole was performed which successfully removed the his-tagged protein from the column.

At this stage of the purification, FOXP1 LZ-End (Figure 17B) and both FOXP2 variants (Figure 17 C & D) were isolated to a relatively high degree of purity (~95%) but FOXP1 LZ-FHD (Figure 17A) co-purified with the chaperones used to aid in its expression. Fusion protein-chaperone co-purification is in fact common in expressions involving *dna*K. This is because *dna*K, more than other chaperones in the group, can bind to the protease restriction sites for thrombin, enterokinase, or Factor Xa (Rial and Ceccarelli, 2002). The fusion proteins used in this study were designed with a thrombin cleavage site; therefore, providing a site for chaperone binding. Several ion exchange chromatography steps were employed in order to further isolate FOXP1 LZ-FHD (Figure 17A). Ion exchange chromatography exploited the differences in isoelectric points of protein samples at a given buffer pH. FOXP1 LZ-FHD has a p*I* of around 9.73 while the p*Is* of the chaperone group *dna*K-*dna*J-*grp*E are 5.1, 8.5 and 4.68, respectively (Zyliczs *et al.*, 1985; Zylicz *et al.*, 1987; Zmijewski *et al.*, 2004). Ion exchange chromatography was performed on a HiTrap™ Q FF column at pH 7.5. At this pH, FOXP1 LZ-FHD and *dna*J, which are both positively charged at pH 7.5, do not bind the positively charged Q-sepharose resin. However, *dna*K interacts strongly with *dna*J at this pH (Suh *et al.*, 1998; Mayer *et al.*, 1999; Noguchi *et al.*, 2014); therefore, regardless of its p*I*, it does not co-elute with FOXP1 LZ-FHD. Following ion exchange chromatography, samples were resolved on a 12% SDS-PAGE gel under denaturing and reducing conditions to assess purity of the isolated FOXP1 LZ-FHD (Figure 17A).

FOXP2 LZ-End migrates as a significantly larger protein than a protein of its typical molecular weight on an SDS-PAGE gel. It typically appears as ~43 kDa protein on the denaturing gel although its molecular weight is ~40.82 kDa. All the other variants, FOXP1 LZ-FHD (Figure 17A), FOXP1 LZ-End (Figure 17B) and FOXP2 LZ-FHD (Figure 17C) are resolved with molecular weight corresponding to their apparent masses (30.55 kDa, 40.44 kDa, 29.11 kDa, respectively) determined from their amino acid sequences. The cause of this anomalous behaviour in the case of FOXP2 LZ-End (Figure 17D) could be because of the highly charged C-terminal tail, though the trait is not observed in FOXP1 LZ-End, which also possess the tail.

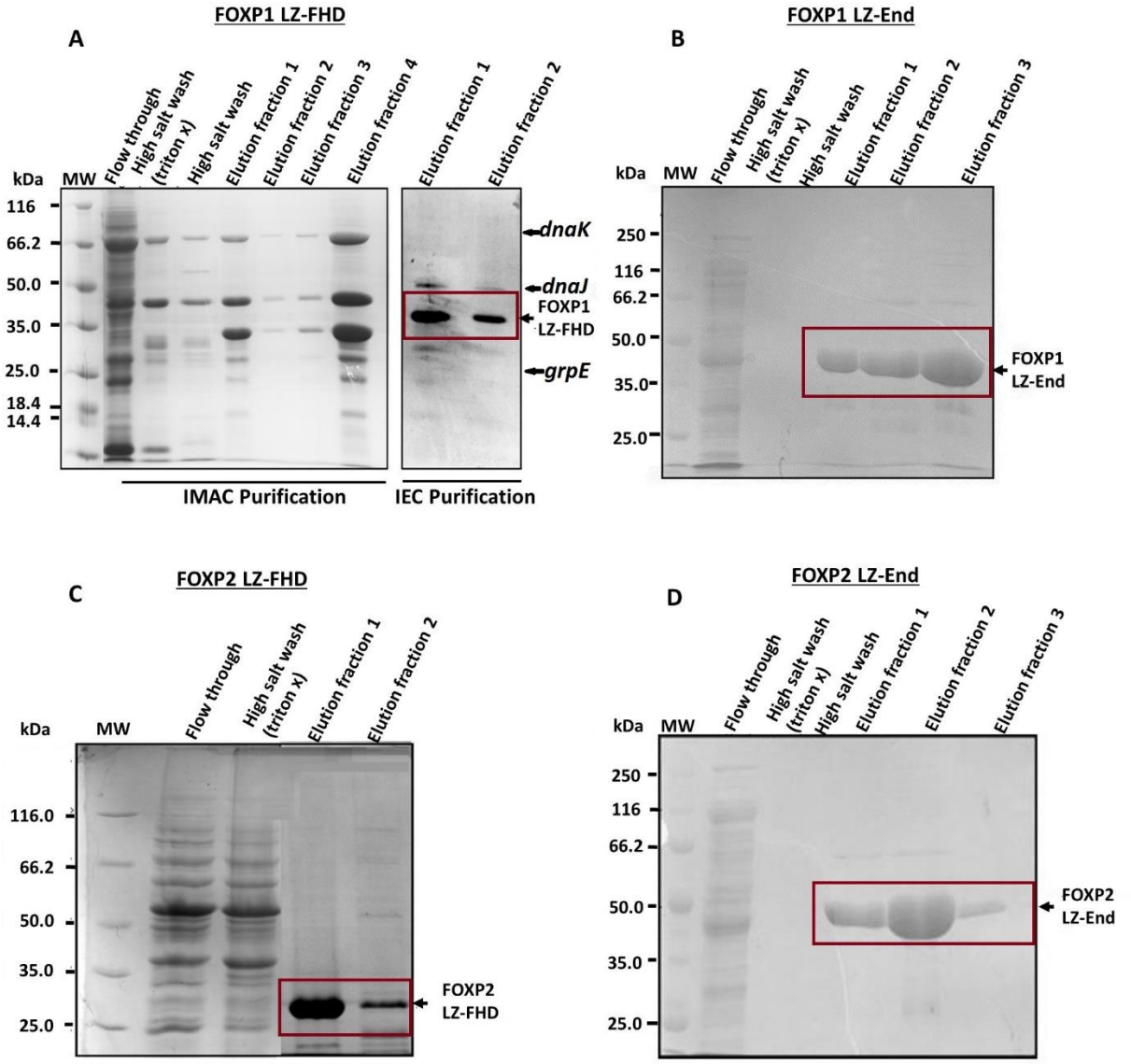


Figure 17: 12.5 % SDS-PAGE gels showing purification of FOXP1 and FOXP2 variants. Bands representing fractions containing pure protein that was used in subsequent experiments are boxed in red. (A) FOXP1 LZ-FHD was purified using a Ni^{2+}-charged IMAC HiTrap™ column followed by ion exchange chromatography (IEC) on a Q FF IEC column in order to remove the co-purified chaperones. (B) FOXP1 LZ-End IMAC purification on a Ni^{2+}-charged column. (C) FOXP2 LZ-FHD purification on a Ni^{2+}-charged IMAC column. (D) FOXP2 LZ-End was purified using IMAC on a Co^{2+}-charged column.

Generally, ~50 µM of protein was purified per litre of culture for all four variants and about 93-95% pure protein was obtained following purification. The protein yield and purity obtained were enough for all subsequent experiments.

The minor bands for FOXP1 and FOXP2 variants were sequenced using LC-Mass Spectrometry. Analysis of mass spectrometry data suggests that the minor bands are isomers of either FOXP1 or FOXP2 with minimum amounts of contamination.

3.3. DNA BINDING AND NATIVE FOLDING

FOXP1 and FOXP2 are transcription factors and, therefore, their main function is to control the expression of genes by binding to specific sequences of DNA. Because the ability of a protein to perform its function can be used as a direct measure of structural integrity, the ability of all four variants to bind to DNA was used as a means of confirming their structural folding. Electrophoretic mobility shift assay (EMSA) was used to determine the ability of FOXP1 and FOXP2 variants to form complexes with DNA. EMSA is based on the observation that the electrophoretic mobility of a protein-nucleic acid complex is typically less than that of the free nucleic acid (Hellman and Fried, 2007a). In EMSA, solutions of protein and nucleic acid are combined, and the resulting mixtures are subjected to electrophoresis under native conditions through polyacrylamide or agarose gel. Therefore, EMSA can also provide an indication of binding of multiple species of different sizes as separation occurs during migration of the sample in the electric field (Fried and Crothers, 1981; Hudson and Fried, 1990). EMSA provides a qualitative overview on the binding capability of protein to its target DNA sequence. Here, EMSA was conducted in order to determine whether all variants were in the correct conformation by assessing their ability to bind DNA. This would imply that folding of the proteins in the *E. coli* expression machinery mimics folding of the proteins in their native cell environment.

Binding of FOXP1 and FOXP2 variants to DNA is indicated by the bands that are observed on the EMSA gels which appear higher than the free DNA bands, suggesting that a protein-DNA complex was formed (Figure 18). Because EMSA is a separation technique, different complexes between protein and DNA will be separated and will show as different bands on the gel. In the case of FOXP1 LZ-End, FOXP1 LZ-FHD and FOXP2 LZ-FHD, a number of bands are observed indicating the likelihood that there are different protein-DNA complexes being

formed and these different complexes are of different sizes. Only a single band representative of a large species, is observed for the FOXP2 LZ-End-DNA complex at the high concentration, shown in Figure 18D. The lanes in all the EMSAs in Figure 18 are fairly smeared. This is likely due to weak binding in the protein-DNA complex which results in the complex breaking and reforming as it migrates through the gel.

The fact that all FOXP1 and FOXP2 variants are capable of binding to DNA (Figure 18) implies that the DNA-binding domain, the FHD, is in its native conformation. Furthermore, a number of studies have shown that impairment of the leucine zipper domain, either in-frame deletions or mutations eliminates homo- or hetero-oligomerisation in the full-length FOXP proteins (Li, Weidenfeld and Morrisey, 2004; Bin Li *et al.*, 2007; Song *et al.*, 2012) and in most cases also DNA binding (Bennett *et al.*, 2001; Brunkow *et al.*, 2002; Li, Weidenfeld and Morrisey, 2004). This, therefore, implies that there is cooperativity between the different domains in all variants and impairment of one domains influences binding of the other. The fact that all FOXP1 and FOXP2 variants bind to DNA through their FHD means that the entire protein is in its native conformation.

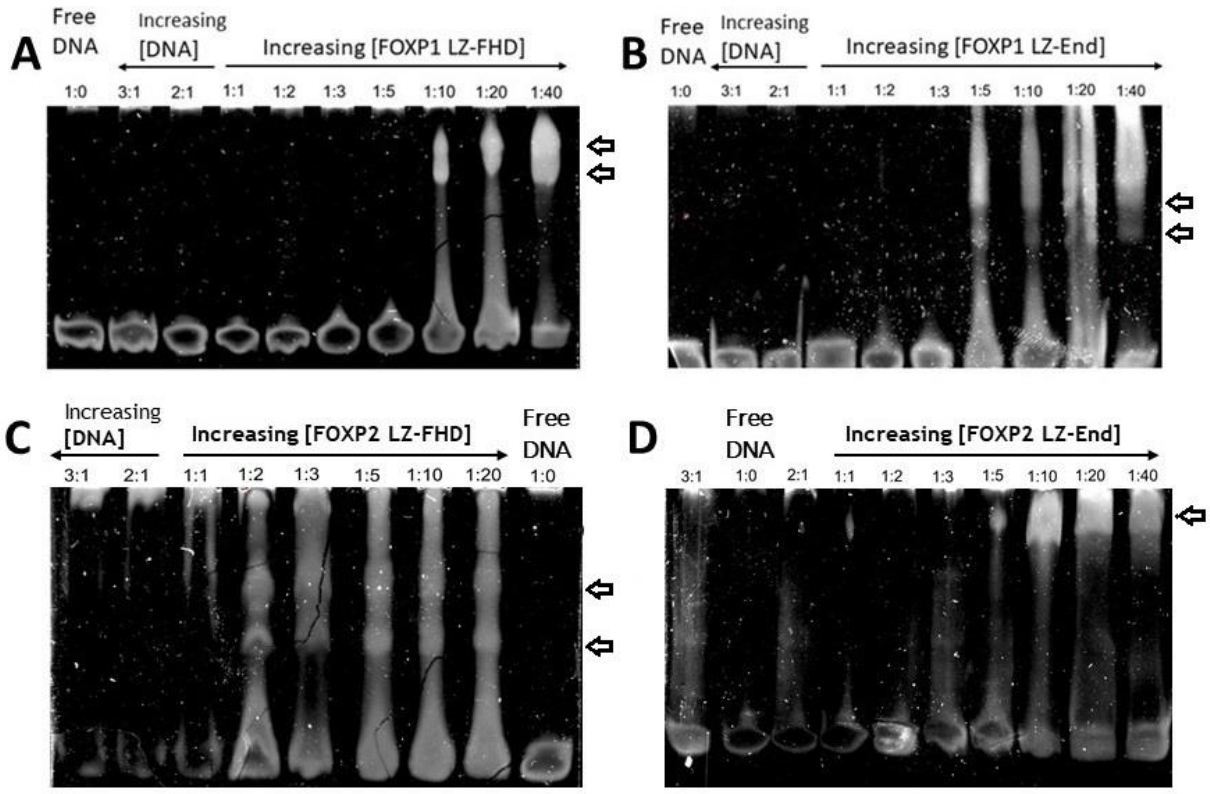


Figure 18: Protein−DNA complex formation of the FOXP variants as determined by electrophoretic mobility shift assay. 500 nM cognate DNA was incubated with increasing concentrations (5−40 µM) of: (A) FOXP1 LZ-FHD, (B) FOXP1 LZ-End, (C) FOXP2 LZ-

FHD, (D) FOXP2 LZ-End. All variants indicated DNA binding capability, suggesting that the FHD (DNA binding domain) is correctly folded. It appears that multiple species form weakly associated (smeared bands) complexes with the DNA as indicated by the arrows.

Matrix-based separation methods, such as electrophoretic mobility shift assays, can easily perturb the reaction equilibrium by pulling reactants (DNA and protein) away from products (complex); thereby, creating a concentration gradient which pushes the reactant towards individual equilibrium (Licata and Wowor, 2008). Therefore, EMSA was used in this study for qualitative purposes only and the affinity of binding could not be quantified from the EMSA gel. Fluorescence anisotropy (FA) was, therefore, used in addition to the EMSA in order to determine the differences in binding of all four variants to DNA by using DNA binding affinities.

Fluorescence anisotropy measures the rotational diffusion of a molecule. So when the reactant binds its partner, the larger product has a lower rotational diffusion coefficient and a higher fluorescence anisotropy (Lakowicz, 2002; Gradinaru, Marushchak and Krull, 2010; Pollard, 2010; Weinreis, Ellis and Cavagnero, 2011). Although the FHD is the main DNA-binding domain of both FOXP1 and FOXP2, the other domains may influence the strength of DNA binding. Furthermore, the oligomeric state of FOXP1 and FOXP2 variants could also influence the strength of DNA-binding. Therefore, in this study, FA was used to assess and compare the strength of binding of FOXP1 and FOXP2 variants to DNA (Figure 19).

Figure 19A and 19B show the DNA binding isotherms of all four variants (FOXP1 in figure 19A and FOXP2 in Figure 19B). These isotherms were fitted to a single-site binding model. K_D values obtained for DNA binding to FOXP1 LZ-End and FOXP1 LZ-FHD were 1281.2 ± 70.4 nM and 765.5 ±120.7 nM, respectively, while the K_D values for DNA binding to the FOXP2 variants was 347.1 ± 55.4 nM and 326.2 ±73.2 nM for the LZ-End and the LZ-FHD variants, respectively. As can be seen in Figure 18 C-F, all FOXP2 variants bind the DNA sequence with a significantly greater affinity than the FOXP1 variants. While there is no statistically significant difference in DNA binding affinity between the two FOXP2 variants, there is a significant difference in the DNA binding affinity of the FOXP1 variants where the longer variant shows significantly weaker affinity, indicating that the presence of the acid rich tail interferes with DNA binding strength in the case of FOXP1 but not FOXP2.

The DNA binding assessment performed using the EMSA (Figure 18) showed that the variants bind to DNA as a number of species of various sizes. Therefore, the binding affinities obtained from fluorescence anisotropy represents the average binding affinity of all the species. Nonetheless, the binding affinity of both FOXP2 variants shows considerable improvement and tighter binding compared to the K_D obtained for the isolated FOXP2 FHD (~892 ± 49.9 nM) using fluorescence anisotropy (Morris and Fanucchi, 2016). Similarly, comparison of the dissociation constants obtained here with those obtained for the isolated FOXP1 FHD monomeric mutants (K_D ~3650 ± 400 nM and ~1510 ± 60 nM) using standard fluorometric titration analysis (Chu *et al.*, 2011), shows that both FOXP1 variants also bind DNA with a stronger affinity than the isolated FOXP1 FHD.

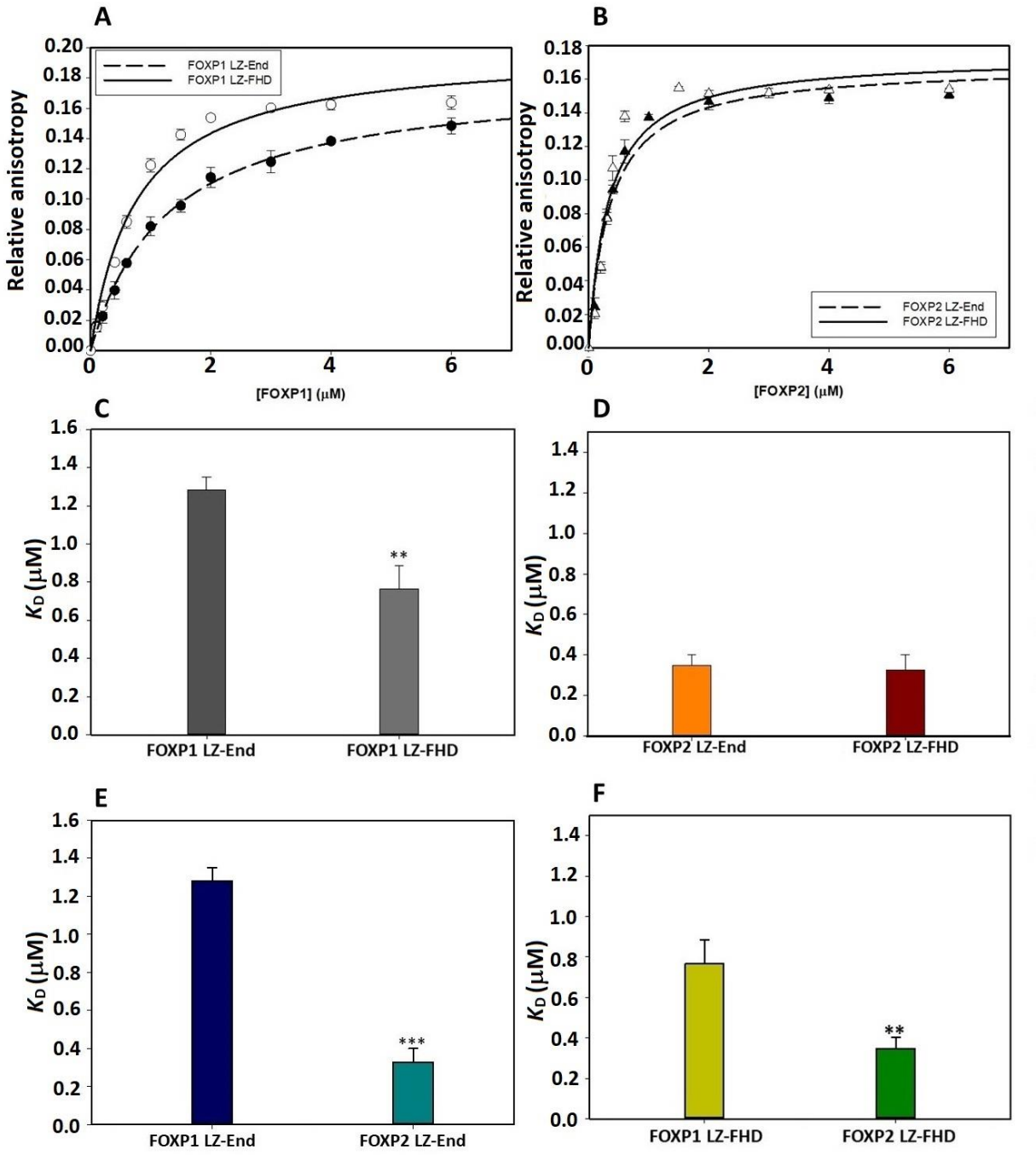


Figure 19: DNA binding isotherms for the FOXP1 and FOXP2 variants as determined by fluorescence anisotropy (FA). For each experiment, 500 nM dsDNA containing a single FOXP2 binding site labelled at the 5′ end with ROX was incubated with increasing concentrations of each protein variant (0 μM-6 μM). The fluorescence anisotropy at each concentration was measured and the data fitted to a single site binding model. (A) The K_D obtained for FOXP1 LZ-FHD and FOXP1 LZ-End DNA binding are 765.5 ± 120.7 nM and 1281.2 ± 70.4 nM, respectively. (B) The DNA binding K_D for FOXP2 LZ-FHD and FOXP2 LZ-End DNA binding are 347.1 ± 55.4 nM and 326.2 ± 73.2 nM respectively. (C-E) Comparison of the K_D for each variant. The statistical significance of each comparison was determined using a student's t-test (*p < 0.05; **p < 0.01; ***p<0.001). Errors indicate the standard deviation of three averaged replicates.

3.4. STRUCTURAL CHARACTERISATION

FOXP proteins interact with DNA as dimers *in vivo* (Wang *et al.*, 2003; Li, Weidenfeld and Morrisey, 2004; Sin, Li and Crawford, 2014), through the leucine zipper domain and the forkhead domain (FHD) (Li, Weidenfeld and Morrisey, 2004; Bin Li *et al.*, 2007). The isolated FHD has been shown to form domain-swapped dimers (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011; Chen *et al.*, 2015) and the FOXP3 leucine zipper domain has been shown to mediate not only dimer formation, but also higher order oligomerisation (Bin Li *et al.*, 2007; Song *et al.*, 2012). Therefore, because of the reported differences in oligomeric state, it is worth investigating the structural characteristics of FOXP1 and FOXP2 variants in isolation, to gain a better understanding of which structural elements are critical to structural integrity, DNA binding and oligomerisation. Both electrophoretic mobility shift assay (EMSA) and fluorescence anisotropy (FA) have shown that all four variants bind to DNA; therefore, implying that the proteins have attained their native fold during expression, allowing for other biophysical characteristics of the proteins to be investigated. Furthermore, a number of bands of varying sizes were observed in EMSA, suggesting that the variants bind to DNA as multiple species of different sizes. Structural characterisation was, therefore, done in order to get a thorough understanding of the structural architecture of the isolated proteins and to make any significant comparisons between the variants. This can lead to an understanding of the dynamics and structure-function of protein-protein interactions and protein-DNA associations.

3.4.1. QUATERNARY STRUCTURE

3.4.1.1. Quaternary state using size exclusion chromatography (SEC)

SEC provides a means to determine the oligomeric state of proteins by measuring the hydrodynamic volume of all the different species (if there are multiple species), when under native solution conditions (Ramsey, Daugherty and Kelm, 2006). The hydrodynamic volume of a protein gives an indication of the oligomeric state of the protein, and thus SE-HPLC was used to assess the hydrodynamic volumes, and therefore the oligomeric state, of FOXP1 and FOXP2 variants. Because the electrophoretic mobility shift assay (Figure 18) indicated that each of the four variants can likely form complexes of different sizes with DNA, SE-HPLC was used here in order to determine whether the different species exist in the absence of DNA.

The quaternary state of the FOXP1 and FOXP2 variants was determined using size exclusion high performance liquid chromatography, as indicated in Figure 20. Interactions of both the leucine zipper domain and the FHD are predominantly mediated by hydrophobic associations which are usually quite stable (Clark *et al.*, 1993; Krylov and Vinson, 2001; Stroud *et al.*, 2006). Therefore, any quaternary conformation detected using this method will likely be the result of hydrophobic interactions between polypeptide chains.

The HPLC column was equilibrated using High Molecular Weight Markers (GE Healthcare, USA). Given the apparent MW of the variants; 28.8 kDa (FOXP2 LZ-FHD), 30. 2 kDa (FOXP1 LZ-FHD), 38.8 kDa (FOXP1 LZ-End) and 44.5 kDa (FOXP2 LZ-End) on SDS-PAGE (Figure 16), the large hydrodynamic volumes reported from size exclusion chromatography, indicate that all the purified proteins form higher order oligomers in solution and that this is the most preferred conformation. However, because of several unstructured and highly flexible regions, there is generally an overestimate of the size based on the hydrodynamic volume and the sizes predicted based on the retention times do not necessarily correspond to the actual MW or size (Figure 19). Therefore, although we can speculate on the sizes and hence quaternary structure of the proteins based on their apparent molecular weights, it is also highly likely that the large sizes predicted by size exclusion chromatography could be due to the large hydrodynamic volume that is likely to be occupied by these proteins since they are predicted to have large unstructured domains which could be highly disordered. Therefore, we cannot conclusively size the FOXP1 and FOXP2 variants using SE-HPLC, but rather we can use the results as a qualitative measure of relative size and number of the various protein species present amongst the four variants.

The longer variants of both FOXP1 and FOXP2 tend to elute from the size exclusion column as a single large species with apparent molecular weights corresponding to that of a tetramer for FOXP1 LZ-End and an octamer for FOXP2 LZ-end (Figure 20B & 20D, respectively). The shorter variants elute from the column as two species. FOXP1 LZ-FHD (Figure 20A) elutes with apparent molecular weights equivalent to a mixture of monomer and dimer at concentrations lower than 20 µM and a mixture of dimer and tetramer at concentrations greater than 20 µM, while FOXP2 LZ-FHD (Figure 20C) elutes as two species with apparent molecular weights corresponding to that of a hexamer and a trimer at all concentrations. The lack of the acid rich tail (ART) region in the shorter variants (FOXP1 LZ-FHD and FOXP2 LZ-FHD) thus seems to imply this region plays a critical role in the oligomeric state and

stability of the FOXP variants. This therefore suggests that this region could be important for stability and can influence dimer formation.

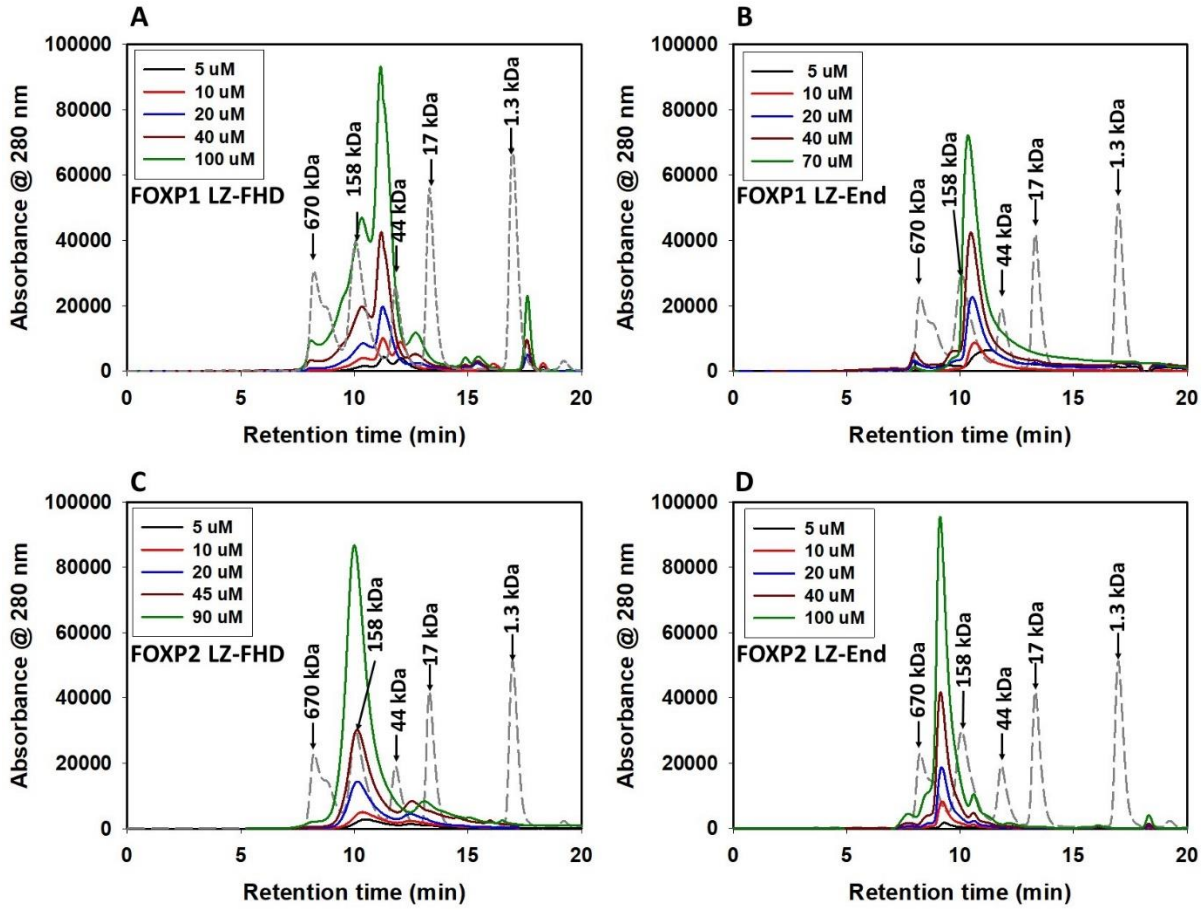


Figure 20: SE-HPLC chromatogram of FOXP1 and FOXP2 variants performed at increasing protein concentrations in 10 mM HEPES buffer, pH 7.4 containing 500 mM NaCl and 2 mM DTT. (A) Elution profile of (5 – 100 µM) FOXP1 LZ-FHD. At low concentrations (< 20 µM), FOXP1 LZ-FHD exists as a mixture of apparent monomer and dimer, and at concentrations greater than 20 µM it exists as a mixture of apparent dimer and tetramer. (B) Elution profile showing the quaternary state of (5 – 70 µM) FOXP1 LZ-End. FOXP1 LZ-End exists as an apparent tetramer at all concentrations. (C) Elution profile of 5-100 µM FOXP2 LZ-FHD. The variant exists as a mixture of apparent monomer and hexamer at concentrations greater than 20 µM. (D) The elution profile of 5-100 µM FOXP2 LZ-End. FOXP2 LZ-End exist as an octamer at all concentrations. Molecular weight standards are indicated with grey dashed lines in each figure. For SE-HPLC, Bio-Rad's Gel Filtration Standard, 1.35 kDa-670 kDa (Bio-Rad, USA) were used.

3.4.1.2. Quaternary state using native polyacrylamide gel electrophoresis (BN-PAGE)

The quaternary states of FOXP1 and FOXP2 variants were further studied using Blue Native PAGE, as shown in Figure 21 below. Although separation techniques such as EMSA and BN-PAGE can easily perturb the reaction equilibrium by pulling reactants (DNA and protein) away from products (complex), they provide a qualitative indication of an interaction (Eubel, Braun and Millar, 2005). Therefore, in this study, BN-PAGE was used to complement and confirm the quaternary state of all four variants that was observed in size exclusion chromatography.

As with the SE-HPLC results, the native gels indicate that the variants exist as a mixture of several species. This is seen by the number of bands on the gel for each variant at each concentration resolved. The longer variants, both FOXP1 and FOXP2 appeared to form larger higher order aggregates than the two smaller variants. This result is consistent with the SE-HPLC result. A difference is that there was predominantly one (very large) species for both FOXP1 LZ-End and FOXP2 LZ-End when using SE-HPLC but the BN-PAGE (Figure 21) indicates that FOXP1 LZ-End exists as a mixture of species. The short variants appeared as two or three species on the BN-PAGE which is in agreement with the SE-HPLC result although the exact proportion of each species present tends to differ between the two techniques.

Both variants of FOXP1 are capable of existing in solution as different concentration dependent complexes, a trait that has been observed for the isolated FOXP1 FHD and the isolated FOXP3 leucine zipper domain (Bin Li *et al.*, 2007; Chu *et al.*, 2011; Song *et al.*, 2012; Medina *et al.*, 2016). Comparatively, both FOXP1 variants seem to favour formation of tetramer in solution, regardless of the C-terminal acid rich tail which is present in FOXP1 LZ-End and absent in FOXP1 LZ-FHD. This also suggests that the C-terminal acid rich tail has minimal effect on the quaternary structure under native conditions, pH 7.4 and 20 ℃. The association of different subunits is supported by the fact that the FOXP leucine zipper domain surface is highly hydrophobic; therefore, favouring clustering, which supports formation of higher oligomers, not just dimers (Song *et al.*, 2012).

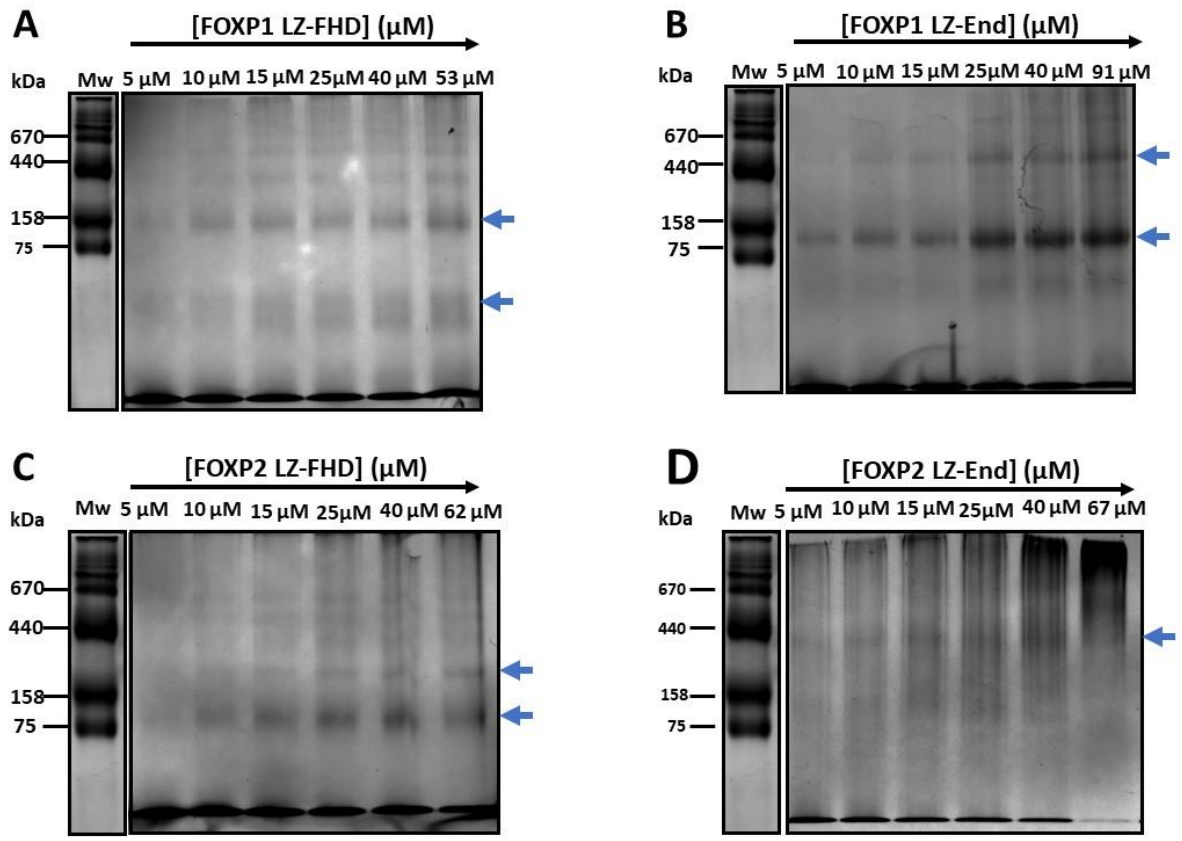


Figure 21: Native PAGE was used to determine the native conformation of the FOXP1 and FOXP2 variants. Different concentrations (between 5 and 90 µM) of FOXP1 and FOXP2 variants were subjected to electrophoresis under non-denaturing conditions on an 8 % polyacrylamide gel. (A) FOXP1 LZ-FHD (B) FOXP1 LZ-End (C) FOXP2 LZ-FHD (D) FOXP2 LZ-End. The bands observed in each lane are fairly smeared. This is due to a constant dissociation and association; therefore, creating a concentration gradient which pushes reactants toward their individual equilibrium. However, some bands (indicated by arrows) were observed, for each quaternary state. Because of inconsistencies in standard protein bands, it is fairly difficult to accurately determine the sizes of each species observed and this method is used as a qualitative technique only.

According to native-PAGE analysis, FOXP2 LZ-FHD exist almost exclusively as a trimer at concentration lower than 20 µM, but as a mixture of trimer and hexamer at higher concentration (Figure 21C). Formation of higher order oligomers is also observed for FOXP2 LZ-FHD at high concentrations of protein. Although SE-HPLC and native-PAGE agree about formation of a hexamer, the hexameric form is the most prominent conformation for FOXP2 LZ-FHD in SE-HPLC, while the trimer is more prevalent in native-PAGE. The differences observed in both techniques could be attributed to the shape of the protein, which ultimately influences the tumbling volume and hydrodynamic diameter (Niepmann and Zheng, 2006).

FOXP2 LZ-End (Figure 20D) exists predominantly as a higher oligomer both in size exclusion chromatography and Native-PAGE. A relatively defined band is observed at most FOXP2 LZ-End concentrations, corresponding to an apparent molecular weight of ~326 kDa, corresponding to an octamer, both in SEC and Native-PAGE.

Higher oligomers are mostly favoured by both FOXP2 variants (Figure 20C & D), though FOXP2 LZ-FHD exists mainly as a mixture of trimer and hexamer in solution, compared to FOXP2 LZ-End which exists predominantly in a quaternary state resembling an octamer. Different quaternary conformations have also been observed for the isolated FOXP2 FHD, which exists in solution as a mixture of monomer and dimer (Stroud *et al.*, 2006).

Although the general architecture of FOXP proteins is similar, there are differences in folding which make each protein unique in comparison to the others. Both SE-HPLC and native-PAGE (Figure 20 & 21) suggest that the FOXP2 variants could exist predominantly as extended proteins, while the FOXP1variants favour a more compact conformation. Nonetheless, formation of higher order oligomers is a trait that seems to be prevalent in FOXP proteins. It has also been observed in studies focusing on FOXP3 (Li *et al.*, 2007(a); Li *et al.*, 2007(b); Song *et al.*, 2012).

3.4.2. Secondary structure

The far-UV CD spectra of FOXP1 and FOXP2 variants is shown in Figure 22. The shapes of the spectra are indicative of predominantly α-helical content, intrinsic disorder and a minor β-sheet component for all the variants. From the solution structure of the FOXP1 FHD determined using NMR spectroscopy (Chu *et al.*, 2011) and the crystal structure of the FOXP2 FHD (Stroud *et al.*, 2006), it is clear that the FHD is predominantly helical. Furthermore, the leucine zipper domain is an antiparallel coiled coil structure mainly made up of two intertwined α-helices (Landschulz, Johnson and Mcknight, 1988; O'Shea, Rutkowski and Kim, 1989; Alber, 1992; Krylov and Vinson, 2001; Song *et al.*, 2012). As the FHD and the leucine zipper make up a 25 % proportion of the structures of the long variants and a 37 % proportion of the structure of the short variants, the CD results are believable.

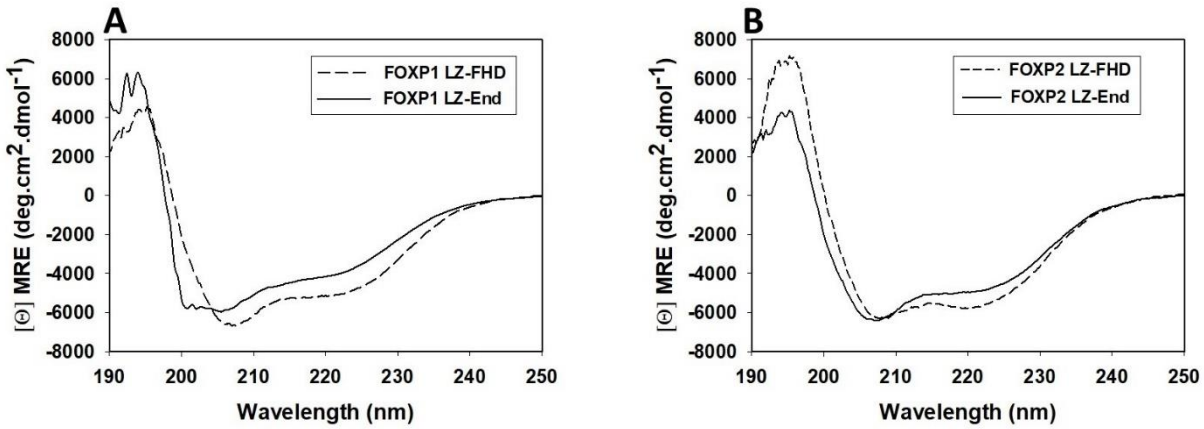


Figure 22: Far-UV circular dichroism spectra of FOXP1 and FOXP2 variants using ~8 μM of FOXP1 variants and ~5 μM FOXP2 variants. (A) The secondary structure of FOXP1 variants is mostly α-helical as seen by troughs at 222 nm and 208 nm, with a minor presence of β-turn/sheets usually seen as a trough at 218 nm. There are slight differences in α-helical content and disorder observed for both FOXP1 variants. (B) FOXP2 variants exhibits a predominantly α-helical secondary structure. There is not much of a change in secondary structure upon truncation of the C-terminal acid rich tail in FOXP2 variants, as shown by superimposition of the far-UV CD spectra (ellipticity).

Amino acid sequence analysis using the disorder prediction tool DISOPRED2 Disorder Prediction Server and PSIPRED (Jones, 1999b; Buchan *et al.*, 2013) indicates that the regions of highest disorder in the sequences of all four variants consists of the C-terminal acid rich tail and an N-terminal linker between the leucine zipper domain and the forkhead domain while the leucine zipper domain and the FHD have the least amount of disorder Ffigure 29). The regions of high disorder contribute to the disorder in the FOXP1 and FOXP2 variants' secondary structure detected in the CD spectra. Although all far-UV CD spectra of FOXP1 species show typical α- helical behaviour, FOXP1 LZ-End also shows the most pronounced disorder of the variants, indicated by the trough at ~203 nm. The pronounced disorder in FOXP1 LZ-End compared to FOXP1 LZ-FHD is due to the extended C-terminal acid rich tail, which is absent in FOXP1 LZ-FHD (see Figure 29A & B).

There are only slight differences in far-UV CD spectra of FOXP2 variants (Figure 22B). Both variants show a predominantly α-helical secondary structure, with no evidence of intrinsic disorder. However, disorder prediction using the amino acid sequences shows that like FOXP1 variants, the C-terminal acid rich tail and the domain linker are regions of major intrinsic disorder. Nonetheless, the lack of differences in CD spectra suggest that the C-terminal acid rich tail does not contribute much to observable disorder, indicating that the tail might fold upon the FOXP2 FHD.

Based on the crystal and solution structures, the FOXP1 and FOXP2 FHDs have less than 50% α-helical content, and far-UV CD spectrum of the wild type FOXP2 FHD indicates that it has around 30% α-helical content and about 30% disorder (Blane and Fanucchi, 2015; Perumal *et al.*, 2015), with the dimer showing a slight increase in ordering of the polypeptide backbone (Perumal, Dirr and Fanucchi, 2015; Morris and Fanucchi, 2016). Therefore, the increase in α-helical content to ~50 to 60% upon inclusion of the leucine zipper domain in the near full-length constructs indicates that the leucine zipper is responsible for increased helical content and hence increased structural stability of FOXP1 and FOXP2.

3.4.3. Tertiary structure

The tertiary structure of all four FOXP1 and FOXP2 variants was determined using intrinsic tryptophan fluorescence. All four variants in the current study have three tryptophan residues and they are all located in the DNA binding domain, the FHD. The other domains and

subdomains are devoid of tryptophan residues; therefore, the intrinsic tryptophan fluorescence probes the changes in the microenvironment of the FHD alone.

The fluorescence spectra of the FOXP1 variants both have a maximum emission at 330 nm and 339 nm (Figure 23A). The double peaks observed in fluorescence emission are due to different environments of the different tryptophan residues in the FOXP1 FHD. According to the crystal structure, in the FOXP2 FHD dimer, the tryptophan residue on position 509 on helix 3 and the tryptophan residue 494 are buried in the hydrophobic core while Trp534 on the wing is in a relatively exposed environment. Because these tryptophan residues are in three different environments, they, therefore, contribute to the multiple fluorescence peaks. The differences in secondary structure, though slight, also influence the tertiary structure of FOXP1 variants. This is shown by the differences in quantum yield for both variants. The FOXP2 variants exhibit fluorescence emission maxima at 331 nm and 338 nm (Figure 23B). The maximum fluorescence emission at 330 nm is indicative of buried tryptophan residues while the red shifted emission, 338 nm and 339 nm indicate partially exposed tryptophan residues (Lakowicz and Masters, 2008).

Comparison of the fluorescence emission maxima of FOXP1 and FOXP2 variants (Figure 23) indicates that there are minimal differences in tertiary structure between the long and short variants of both FOXP1 and FOXP2.

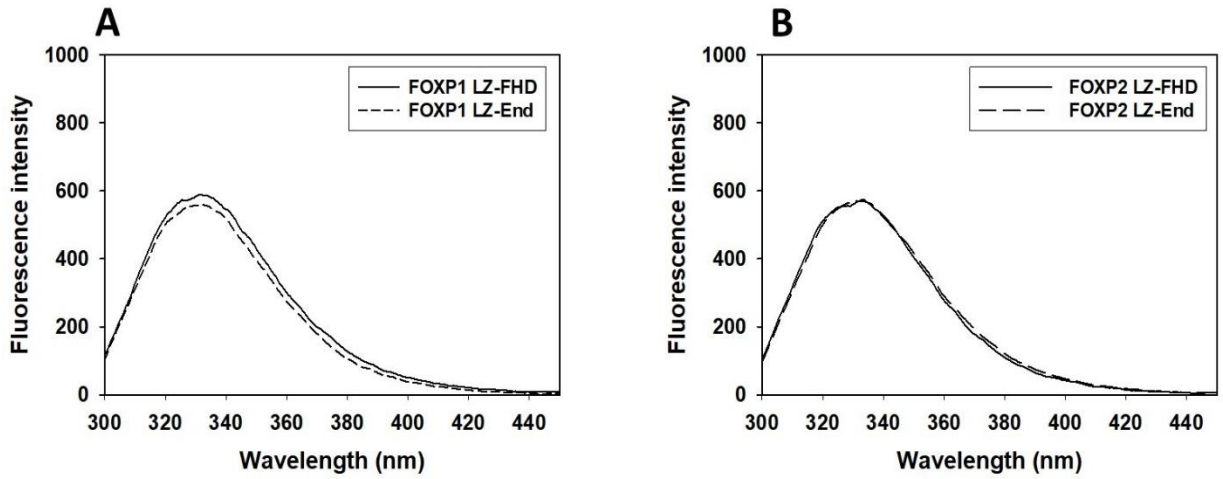


Figure 23: Intrinsic tryptophan fluorescence spectra of FOXP1 and FOXP2 variants using 5 µM protein. (A) Both FOXP1 variants show emission maxima with peaks at 330 nm. (B) FOXP2 variants show a shoulder at 325 nm and a maximum emission at 330 nm. There is no difference is fluorescence spectra for both FOXP2 variants, indicating that changes in architecture have no effect on micro-environment of the three (3) tryptophan residues in the FOXP2 FHD.

3.4.4. Structural stability

Most proteins are characterised by well-defined three-dimensional structures, which are environmentally sensitive and generally exist within narrow limits of specific environmental conditions. Outside these conditions, proteins exhibit denatured and structurally unfolded states. In order to obtain some idea of the stability of the proteins in the absence of DNA, thermal unfolding of FOXP1 and FOXP2 variants was monitored using CD as a probe. (figure 24). The thermal behaviour of the four variants will help give an indication of the role of the C-terminal acid rich tail in structural stability.

The thermal melts for all four variants were not reversible upon cooling and all the variants showed considerable precipitation at low and high temperature following thermal denaturation. FOXP1 LZ-End has a midpoint melting temperature of ~59.4 ± 3.5 ℃ (Figure 24B) and unfolds in a linear manner starting at a temperature below 20 °C up to 60 °C similar to that seen in the DNA-binding domains of other transcription factors (Dragan *et al.*, 2004). In contrast, the FOXP1 LZ-FHD unfolds in a sigmoidal fashion with a midpoint melting temperature of ~46.6 ± 0.7 °C. In contrast both FOXP2 variants unfold in a relatively more sigmoidal fashion (Figure 24B). However, FOXP2 LZ-FHD has a more pronounced pre-transition region and a less pronounce post-transition region in comparison to FOXP2 LZ-End, indicating the unfolding cooperativity of the variants. FOXP2 LZ-End has a midpoint melting temperature of ~58.0 ± 1.1 ℃ while FOXP2 LZ-FHD has a midpoint melting temperature of ~52.3 ± 0.5 ℃, therefore suggesting that FOXP2 LZ-End has a relatively more stable conformation than FOXP2 LZ-FHD (Summarised in Table 3). In general, FOXP1 LZ-End is more stable than all the other variants and FOXP1 LZ-FHD is the least stable.

Table 3: Summary of melting points of FOXP1 and FOXP2 variants

Variants	Temperature (T_M) (℃)
FOXP1 LZ-FHD	46.6 ± 0.7
FOXP1 LZ-End	59.4 ± 3.5
FOXP2 LZ-FHD	52.3 ± 0.5
FOXP2 LZ-End	58.0 ± 1.1

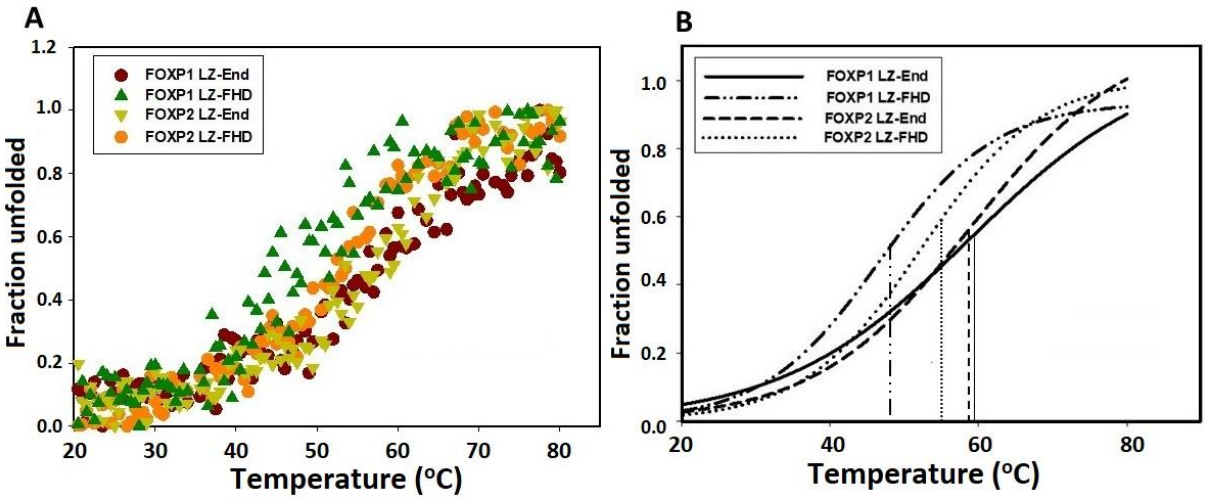


Figure 24: Thermal melting curves of the FOXP1 and FOXP2 variants obtained by monitoring the far-UV CD absorbance at 222 nm while incrementally increasing temperature from 20 ℃ to 80 ℃. A) The thermal unfolding profiles of FOXP1 and FOXP2 variants. B) Thermal unfolding data was fitted using the global curve fitting wizard in SigmaPlot from Systat Software, Inc., San Jose California USA. FOXP1 LZ-END (T_M = ~59.42 ± 3.54 ℃) unfolds in an almost linear fashion from 20 ℃ to 80 ℃. The protein remains mostly stable at low temperature 20-50 ℃. FOXP1 LZ-FHD (T_M = ~ 46.59 ± 0.67 ℃) unfolds in a sigmoidal fashion, showing a pre-transition region, a transition region and a post-transition region where the protein is fully unfolded or denatured. FOXP1 LZ-End shows very low cooperativity compared to FOXP1 LZ-FHD Both FOXP2 LZ-FHD (T_M = ~ 52.34 ± 0.45 ℃) and FOXP2 LZ-End (T_M = ~ 58.02 ± 1.15 ℃) unfold in a sigmoidal fashion although FOXP2 LZ-FHD is more susceptible to temperature variations. Both show relative stability at low temperature 20 – 40 ℃, before unfolding at higher temperature. Thermal unfolding was done at pH 7.5 in 20 mM Tris-Cl and 200 mM NaCl. The lines indicate the midpoint melting temperature of each protein.

Furthermore, both FOXP1 and FOXP2 LZ-End variants show less cooperativity compared to the LZ-FHD variants (Figure 24B). The lack of cooperativity could in part be attributed to the presence of the C-terminal acid rich tail in the LZ-End variants, a region of intrinsic disorder, shown in Figure 30. An interesting feature of intrinsically disordered proteins or regions is their extraordinary resilience, where such proteins or regions can sustain exposure to extremely harsh environmental conditions, being able either to keep its core structure and functionality under these extreme conditions or to rapidly regain it after returning to normal conditions (Uversky, 2017). Because of the multidomain nature of the variants under investigation, further uncooperative behaviour can be attributed to unfolding of the different domains. Each domain behaves differently; therefore, influencing the thermal stability of the protein. For instance, the FOXP2 FHD also shows uncooperative thermal unfolding (Morris and Fanucchi, 2016), while the leucine zipper domain usually shows a classical two-state unfolding (Weiss, 1990). This, therefore, suggests that the uncooperative thermal unfolding of FOXP1 LZ-End and FOXP2 LZ-End (figure 24B) is a direct feature of the intrinsically disordered regions and the FHD. Furthermore, the behaviour of FOXP1 LZ-FHD and FOXP2 LZ-FHD is reminiscent of a classical two state unfolding pattern; therefore, suggesting that the stabilising effect of the C-terminal acid rich tail is almost completely lost. This, therefore, highlights the significance of this subdomain to the structural stability of both FOXP1 and FOXP2. Nonetheless, the fact that all FOXP1 and FOXP2 variants under investigation exist as higher order oligomers in solution at pH 7.5, suggests that oligomeric state of a protein also plays an important role in counteracting effect of thermal denaturation.

3.5. CONFORMATIONAL STUDIES

The folding of protein into its proper 3D conformation is the most fundamental and universal example of biological self-assembly; understanding this complex process, therefore, provides a unique insight into the way in which evolutionary selection has influenced the properties of a molecular system to its functional advantage. The stability of the native fold of a protein is a function of its surrounding environmental variables such as pH, temperature, ionic strength, and solvent composition as they affect various intramolecular bonds responsible for stability and integration of the protein (Dubey and Jagannadham, 2003).

3.5.1. EFFECTS OF pH ON PROTEIN STRUCTURE

3.5.1.1. Effects of pH on the hydrodynamic radius and quaternary structure

Biophysical techniques were used to characterise and study the effect of pH on the structure of the variants at equilibrium after incubation for 20 hours at 20 ℃. Since all four of the variants contain the two major dimerisation interfaces on both FOXP1 and FOXP2 (i.e. the leucine zipper and the FHD), and since we have established that all four variants exist in some form of oligomeric state at pH 7.4, size exclusion chromatography (Figure 25) was used to investigate the quaternary conformation of each variant at each pH tested. This was to determine whether the oligomerisation interfaces are sensitive to pH as this will give an indication into how oligomerisation occurs. By comparing each of the variants' response to pH, we will also be able to establish whether or not the acid rich tail plays a role in pH-dependent oligomerisation.

All size exclusion chromatography studies were conducted under reducing conditions and high salt concentrations in order to eliminate disulfide bond formation between the different monomers, as well as to eliminate non-specific and ionic interactions.

There is a clear *p*H dependent transition in the elution size of both FOXP1 LZ-FHD and FOXP2 LZ-FHD variants (Figure 25). They both appear to have a larger size at low pH (less than pH 7.5), while at pH greater than 7.5, they appear to adapt a more compact conformation. Unlike the shorter variants, the two longer variants do not show a complete shift in hydrodynamic volume to a more compact species at high pH but rather a partial shift, where the majority of the species remain large but at *p*H greater than 7.5. A second, more compact species is also

detected. In the case of FOXP2 FHD, this smaller species is apparent throughout the pH range even at low *p*H albeit at low relative concentrations.

The trend observed that larger species are present at low pH and smaller species are more prominent at high pH suggests one of two things. Either oligomerisation is promoted at low pH, or the protein becomes less compact and more loosely packed at low pH. The presence of the acid rich tail seems to resist the transition to the more compact, smaller species at high pH. FOXP1 and FOXP2 both display similar behaviour and trends although the changes are more pronounced with FOXP1. Also, notably FOXP2 tends to form larger oligomers than FOXP1 with the largest FOXP2 species detected being in the order of octamers while the large FOXP1 species are in the order of tetramers.

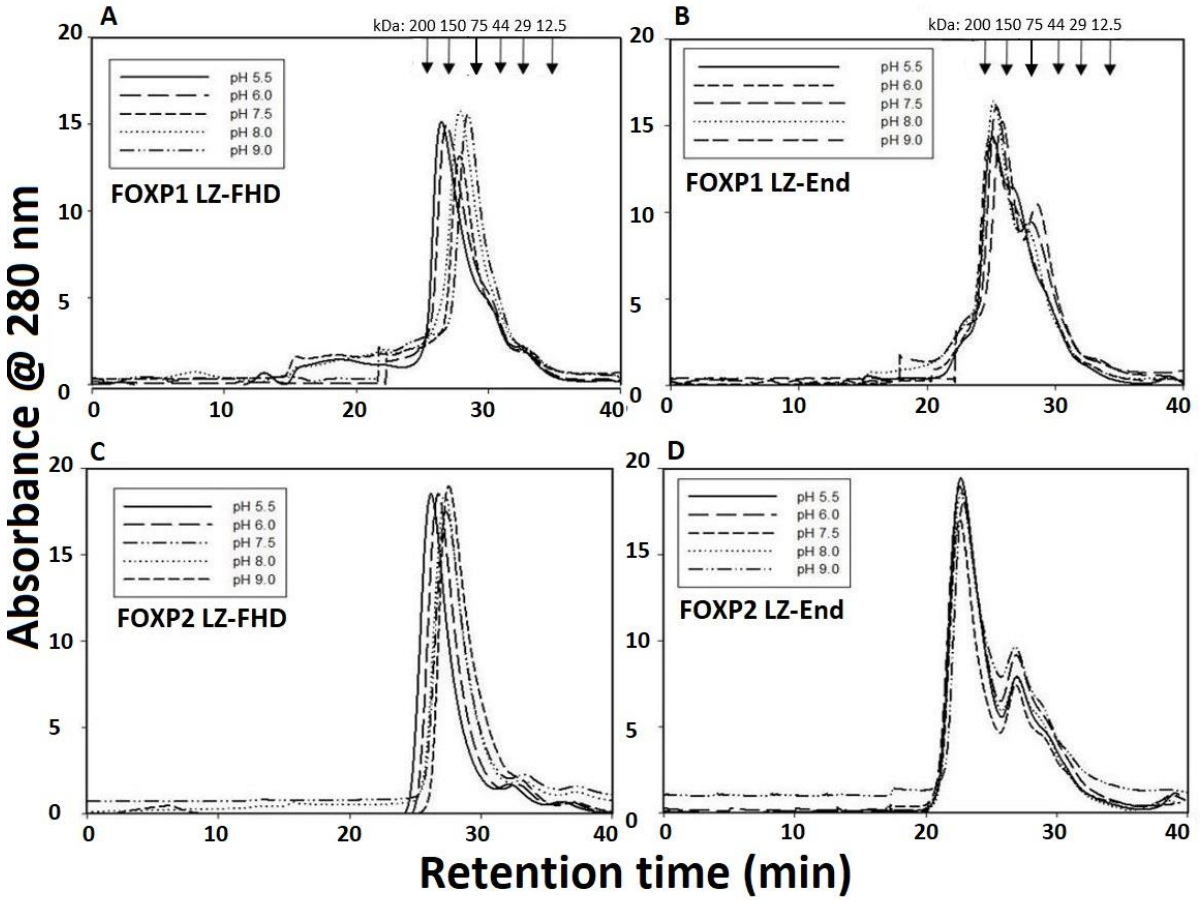


Figure 25: pH dependent study of the quaternary structure of FOXP1 and FOXP2 variants using ~20 µM protein. Size exclusion chromatography was performed on a Superdex® 200 10/300 GL column at different pH in the range 5.5 – 9. The column was calibrated using High Molecular Weight Gel Filtration standards (12.5-200 kDa) from Sigma-Aldrich, USA. All protein samples were incubated for 24 hours at 20 ℃ prior to gel filtrations to allow for maximum equilibration of the protein species. Gel filtration profile for (A) FOXP1 LZ-FHD, (B) FOXP1 LZ-End, (C) FOXP2 LZ-FHD and (D) FOXP2 LZ-End show *p*H-dependent shifts in size to smaller, more compact species at high pH.

As indicated in Figure 24, the existence of the C-terminal acid tail in FOXP2 LZ-End contributes to its structural stability. It is clearly seen, that there was little or no change in the quaternary state between pH 9.0 and 5.0. This is because intrinsically disordered protein regions (IDPR) induce a ‘‘turned out’’ response to changes in pH (Uversky, 2009). On the other hand, FOXP2 LZ-FHD experiences changes in MW in the pH range (Figure 25C, 26B). At *p*H≥7.5, FOXP2 LZ-FHD becomes more compact, eluting from the SEC column with a much smaller apparent MW. Similar behaviour has been observed for the isolated FOXP2 FHD, which also undergoes compaction at high *p*H (Blane and Fanucchi, 2015).

The trend for all variants is a decrease in apparent size with an increase in pH. This is shown convincingly in the plot of the size of each species as a function of *p*H (Figure 26). Figure 25C shows the ratio of the different species of FOXP1 and FOXP2 LZ-End variants, indicating that the relative proportion of smaller “dimeric” species increases with an increase in pH. It is clear that although the trend is there for all variants, both the FOXP1 variants undergo more pronounced changes in quaternary state with respect to pH compared to the FOXP2 variants.

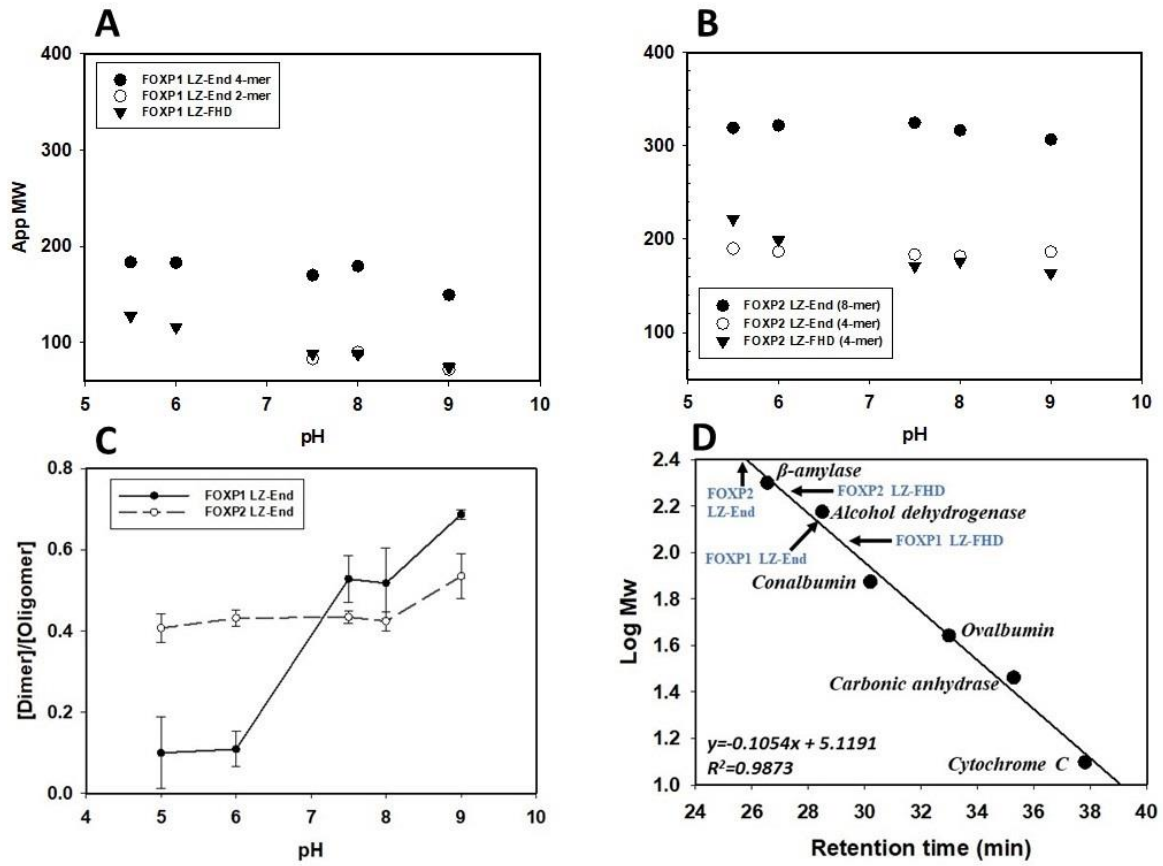


Figure 26: Size distribution of FOXP1 and FOXP2 near full-length variants at different pH (5.5-9). (A) Estimated apparent molecular weight of each of the species of the FOXP1 variants with respect to pH (B) Estimated apparent molecular weight of each of the species of the FOXP2 variants with respect to pH. (C) Changes in species distribution for FOXP1 LZ-End and FOXP2 LZ-End with changes in pH indicating increased formation of the lower oligomer at higher pH (pH≥7.5). (D) The apparent size of each prominent protein peak for each variant determined using High Molecular Weight standards (Sigma-Aldrich, USA), ranging from 12.5 kDa to 200 kDa in size.

The results from the size exclusion chromatography show that protonation and deprotonation of a protein can influence its oligomeric state to an extent (Figure 26). The study was thus extended by examining how the hydrodynamic volume occupied by the protein variants is influenced by changes in the environment. This was done by detecting the dynamic light scattering (DLS) of each of the variants monitored at different pH (Figure 27). All variants display hydrodynamic diameter ranging from 9.0 nm to 14 nm at all pH (Figure 28).

For all variants, the changes in hydrodynamic volume (diameter) resemble changes in apparent size observed in size exclusion chromatography. The trend is a decrease in hydrodynamic diameter with increasing pH (above *p*H 7.5) (Figure 28) with minimum change in polydispersity of each variant (Figure 28E). These changes again suggest that high pH increases the compactness of the variants, while lowering pH leads to opening of the structure. The sizes of all variants were compared to the sizes of standard proteins, conalbumin (75 kDa; diameter of ~8.72 ±0.37 nm) and aldolase (158 kDa; diameter ~10.3 nm) (Figure 27F). The hydrodynamic radius of both conalbumin and aldolase determined by DLS corresponds to those estimated in other studies (Miller *et al*., 2011; Takeuchi *et al.*, 2014; Tonkin *et al.*, 2015). Because of the correlation between the sizes of the standard proteins and our variants, it can be concluded that FOXP1 and FOXP2 are not only capable of forming dimers but can form an "oligomer-of-dimers" assembly.

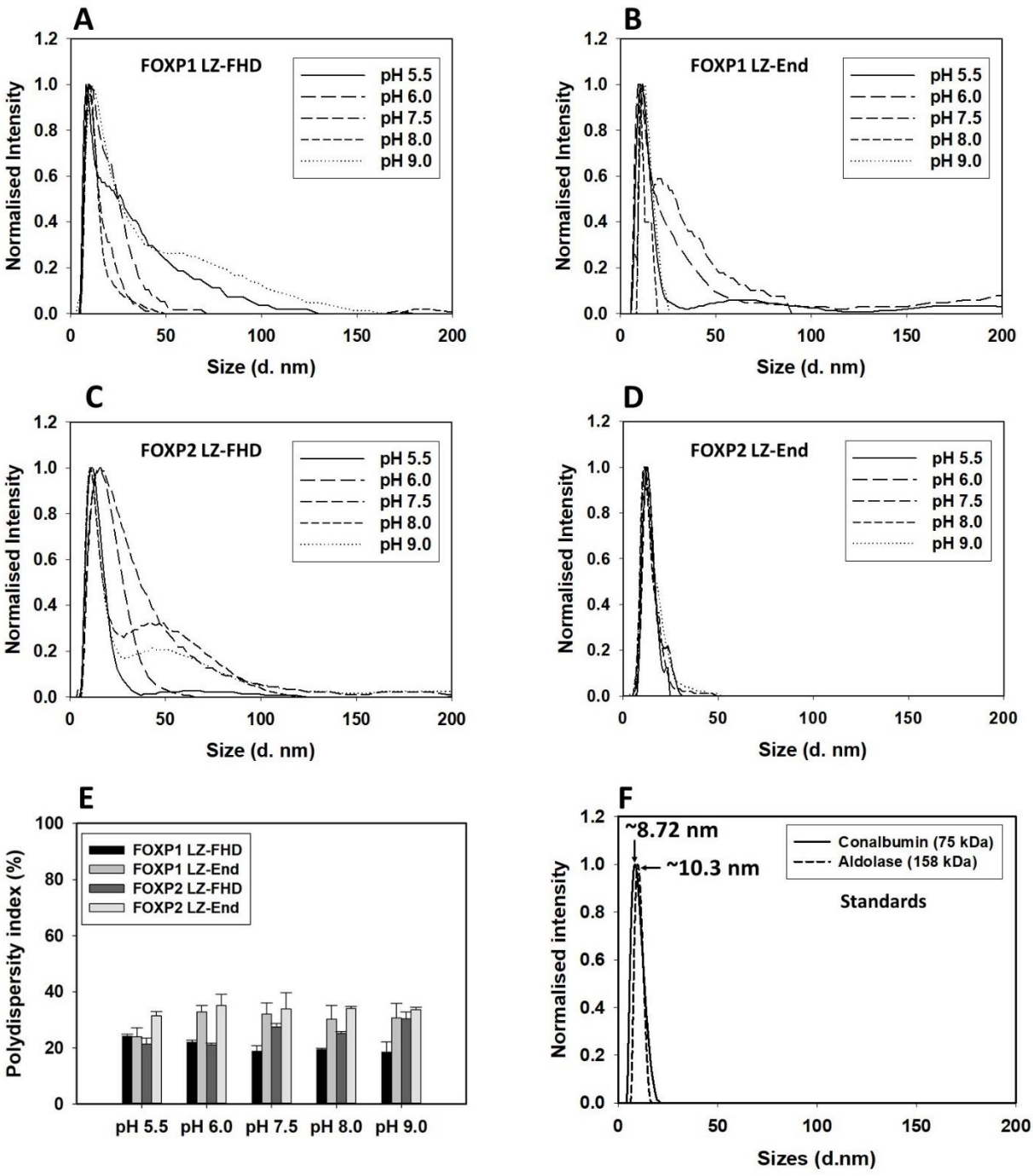


Figure 27: Dynamic light scattering measurements for FOXP1 and FOXP2 variants obtained from Malvern Zetasizer Nano S at different *p*H (5.5-9) using ~15 µM protein. (A) FOXP1 LZ-FHD (B) FOXP1 LZ-End (C) FOXP2 LZ-FHD (D) FOXP2 LZ-End. (E) The polydispersity index for all four variants at different *p*H values. In general, the dispersity ranges from 20% to ~35%, indicating predominantly monodispersed samples. (F) Sizes of proteins (Conalbumin and Aldolase) of known molecular weight were used to compare the sizes obtained for all four variants.

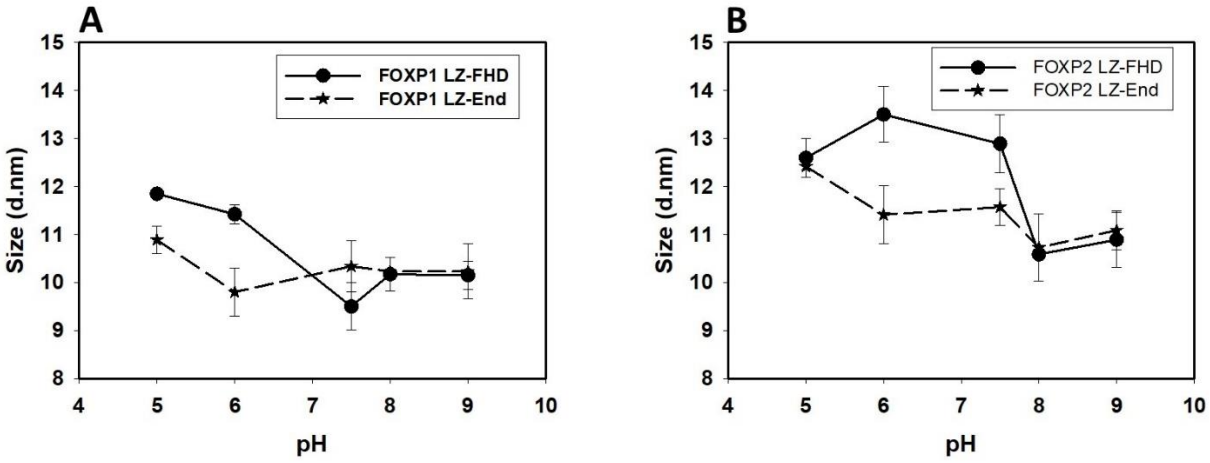


Figure 28: Hydrodynamic size distribution of ~15 µM FOXP1 and FOXP2 variants at different pH (5.5-9) determined using dynamic light scattering (in diameter). (A) pH-induced changes in hydrodynamic radii of FOXP1 variants (B) pH-induced changes in hydrodynamic radii of FOXP2 variants. The trend is a decrease in average hydrodynamic radius with increasing pH. For each variant, ~15 µM protein was used.

3.5.2. EFFECTS OF ORGANIC SOLVENT ON PROTEIN STRUCTURE

Organic solvents such as alcohols are known to cause the destruction of rigid native structure, induce formation of alpha-helical secondary structure and dissolve peptide aggregates (Hirota, Mizuno and Goto, 1998). In solvents of low polarity, such as alcohols, hydrophobic interactions stabilising the native structure or the protein aggregate are weakened (Hirota, Mizuno and Goto, 1998). Furthermore, simple alcohols, such as ethanol, can induce changes in quartenary structure by influencing the formation of secondary structure elements in intrinsically disorderd proteins and regions (Munishkina *et al.*, 2003; Martin *et al.*, 2008). The leucine zipper domain and the FHD association and structures are stabilised by hydrophobic interactions (Clark *et al.*, 1993; Krylov and Vinson, 2001; Stroud *et al.*, 2006; Song *et al.*, 2012; Blane and Fanucchi, 2015). In this study, ethanol was used to disrupt the hydrophobic interactions that stabilise FOXP1 and FOXP2 dimerisation interface. This was done in order to understand the most crucial interactions that stabilise the FOXP1/FOXP2 oligomers.

3.5.2.1. Effects of ethanol on the secondary structure of FOXP1 and FOXP2 variants

The structure of FOXP1 and FOXP2 variants comprises of areas of intrinsic disorder. Intrinsic disorder predictions were done for each of the four variants, based on amino acid sequence using DISOPRED and PSIPRED (Figure 29). The C-terminal region is predicted to be highly disordered. Because alcohols can promote the formation of secondary structure elements in intrinsically disordered regions, it is worth investigating the influence of ethanol on the secondary structure of FOXP1 and FOXP2 variants. This information can provide an idea of the role of this disordered C-terminal region on the stability of FOXP1 and FOXP2, and insight into the folding mechanism that the C-terminal acid rich region might follow on upon DNA binding. Far-UV CD was used to monitor changes that occur in the FOXP1 and FOXP2 variants secondary structure in the presence of ethanol.

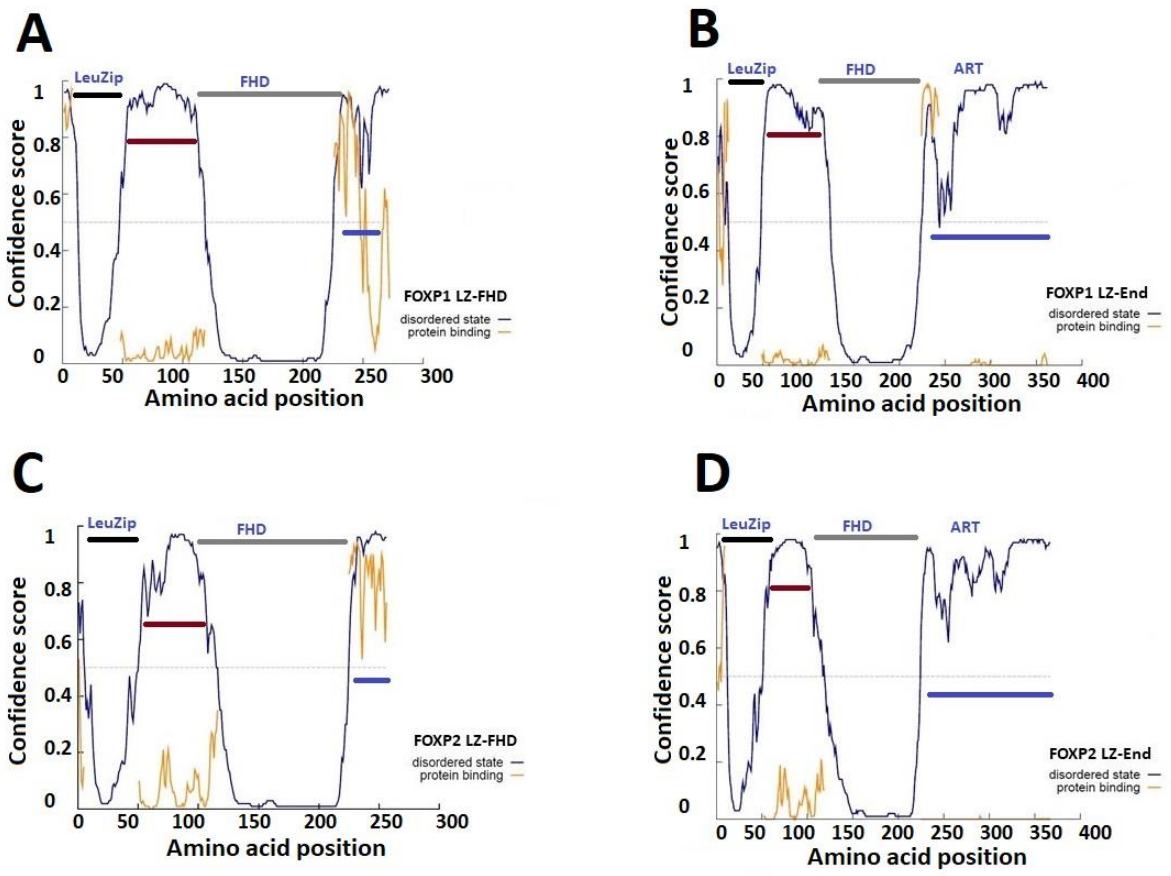


Figure 29: Disorder simulation of FOXP1 and FOXP2 variants using the DISOPRED2 Disorder Prediction Server (Buchan *et al.*, 2013) and PSIPRED (Jones, 1999b; Buchan *et al.*, 2013) online servers. (A) FOXP1 LZ-End (B) FOXP1 LZ-FHD. (C) FOXP2 LZ-End. (D) FOXP2 LZ-FHD. The blue line (—) represents regions of highest disorder. Amino acids in the input sequence are considered disordered when the confidence score is higher than 0.5. The orange (—) line shows the confidence of disordered protein binding residue predictions. The different domains are indicated; leucine zipper domain (—), forkhead-box domain (FHD) (—) and the C-terminal acid rich tail (—).

A plot of the CD value at 222 nm (that primarily reflects α-helicity) as a function of alcohol concentration at 20 ℃ is shown in Figure 30 for FOXP1 and FOXP2 respectively. For both proteins, the long variants show the most sensitivity to increased ethanol by becoming more helical. This suggests that the acid rich tail is probably fairly disordered in solution but does have helical propensity particularly when exposed to an apolar environment. The shorter variants do not really respond to increases in ethanol (0-60%). The decreased ellipticity at high concentrations may be due to denaturation.

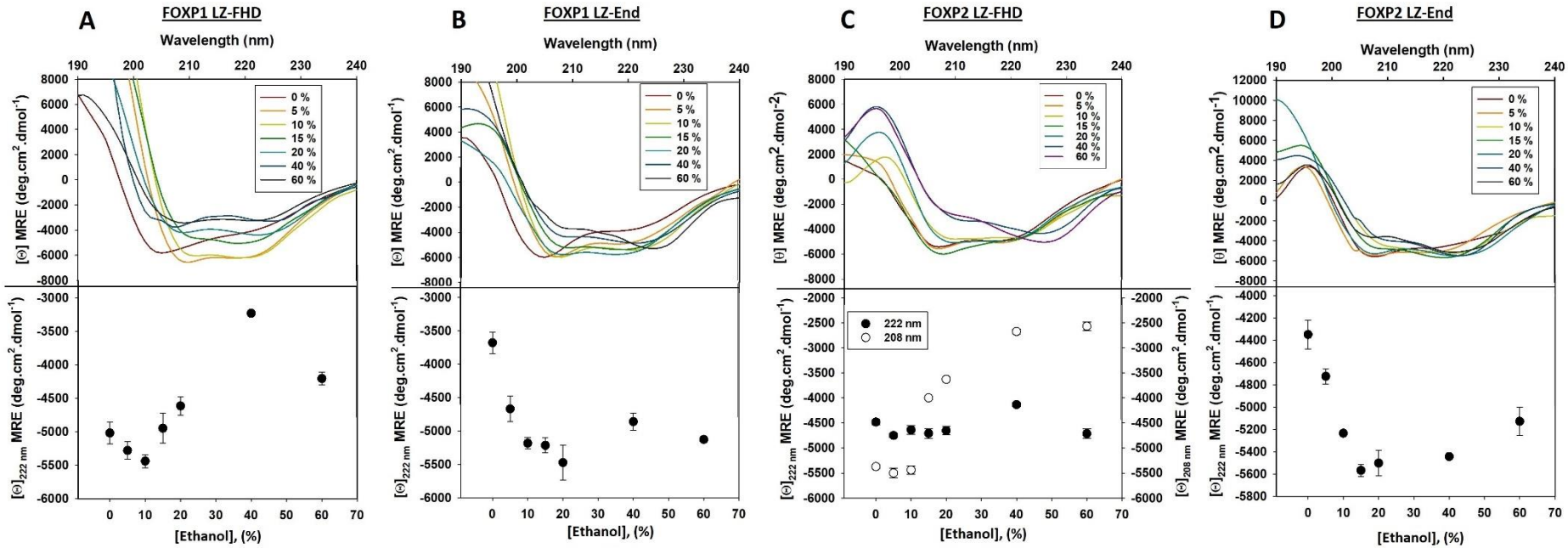


Figure 30: Secondary structure characteristics of FOXP1 variants as measured by CD spectra in the presence of increasing concentrations of ethanol (0%-60%). (A) FOXP1 LZ-FHD. There is minimal change in α-helical content at low ethanol concentration for FOXP1 LZ-FHD. However, at 20 % ethanol and greater, the alpha helicity of the protein is reduced. (B) FOXP1 LZ-End. Changes in concentration of ethanol lead to an increase in α-helical content, shown by MRE at 222 nm. (C) FOXP2 LZ-FHD. There is minimal change in MRE at 222 nm. Changes in secondary structure are shown by decrease in the depth of the trough at 208 nm. (D) FOXP2 LZ-End. The mean residue ellipticity (MRE) of FOXP2 LZ-End obtained at 222 nm is also shown. As the ethanol concentration increases, there is also an increase in α-helical content. Mean residue ellipticity (MRE) of FOXP1 LZ-End obtained at 222 nm.

3.4.2.2. Effects of organic solvents on FOXP1 and FOXP2 tertiary structure

Intrinsic tryptophan fluorescence was used to monitor the influence of ethanol on tertiary structure of FOXP1 and FOXP2 variants, and ensure that the alcohol has not introduced total disruption of the tertiary structure. Because all four variants have tryptophan residues only in the FHD, intrinsic tryptophan fluorescence was used as a local probe to investigate the influence of ethanol on the FHD.

For both FOXP1 and FOXP2 variants (Figure 31), increases in ethanol concentration induces a red shift from 330 nm at 0 % ethanol to 339 nm (Figure 31A and 31D), 342 nm (Figure 31B) and 341 nm (Figure 31C). A red shift in the wavelength maximum and an increase in fluorescence intensity observed may be attributed to unfolding of the protein in the presence of ethanol, reflecting the considerable increase of the accessibility of these residues to the solvent. Changes in the polarity of the solvent as a result of increasing ethanol concentration results in exposure of the tryptophan residues from the hydrophobic core.

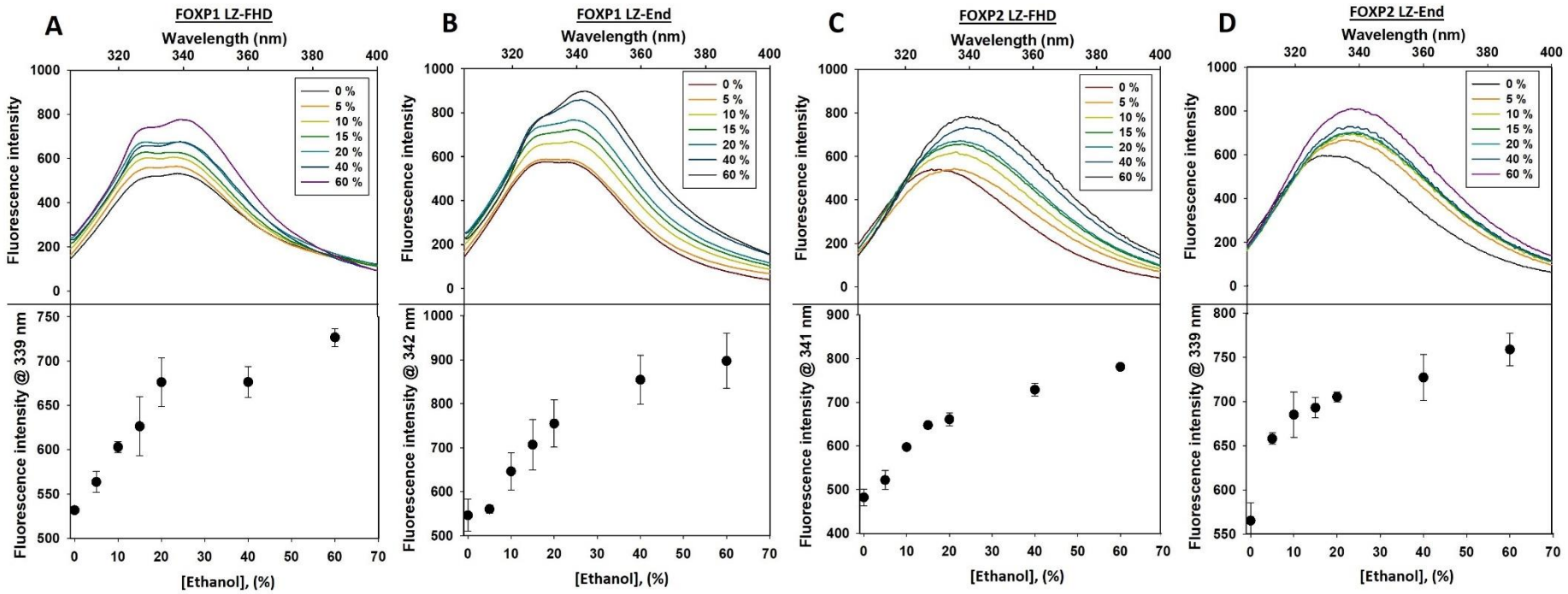


Figure 31: Tertiary structure characteristics of FOXP1 variants in the presence of increasing ethanol concentrations (0-60%) as measured by intrinsic tryptophan fluorescence. (A) FOXP1 LZ-FHD. Changes in tryptophan fluorescence with changes in ethanol concentration for FOXP1 LZ-FHD. (B) FOXP1 LZ-End. Changes in tryptophan fluorescence at maximum emission wavelength with changes in ethanol concentration for FOXP1 LZ-End are shown. (C) FOXP2 LZ-FHD. Maximum intensity was observed at 341 nm for FOXP1 LZ-FHD when ethanol concentration was 60%. (D) FOXP2 LZ-End. At maximum ethanol concentration (60% ethanol), the fluorescence intensity had a maximum wavelength of 339 nm, from 330 nm at 0% ethanol concentration. There are only slight changes in fluorescence intensity at maximum wavelength regardless of changes in ethanol concentration. The trend is an increase in fluorescence emission with increasing concentration of ethanol.

3.5.2.3. Changes in apparent quaternary state in the presence of ethanol

Simple alcohols can induce conformational transitions of a protein structure into a conformational state resembling that of a molten globule, by altering hydrophobic interactions and hydrogen bonds (Hirota, Mizuno and Goto, 1998). Hydrophobic interactions that stabilise the native structure or the protein oligomers are weakened, and simultaneously the local hydrogen bonds are strengthened, resulting in denaturation or dissolution, and stabilisation of extended β-helical structures (Hirota, Mizuno and Goto, 1998). In this study, the effect of alcohol on the quaternary structure of FOXP1 and FOXP2 variants was monitored using SE-HPLC. Ethanol was used to try to disrupt the hydrophobic dimerisation interface, made up of the leucine zipper and the FHD.

Both FOXP1 and FOXP2 variants exist in solution in different quaternary states, ranging from monomer to dimer to tetramer to higher order oligomers (Figure 20). Changes in solvent hydrophobic properties; therefore, dielectrics, can also induce conformational changes that could allow dissociation of higher order oligomeric states into lower oligomers, reminiscent of the functional conformation. These experiments were conducted using 15% ethanol. Only 15% ethanol was used because analysis of intrinsic tryptophan fluorescence and far-UV CD spectra, revealed that at this concentration, the FOXP1and FOXP2 variants have not yet been denatured (Figures 30 and 31).

Figure 32 shows changes in apparent quaternary state of FOXP1 and FOXP2 variants in the presence of ethanol. Figure 32E and 32 F shows differences in retention time of size exclusion standards in the presence and absence of ethanol. There is minimum difference in retention time of standards in 15% ethanol compared to 0% ethanol buffer. Therefore changes observed in apparent quaternary state can be attributed to changes in solvent polarity. In general, all variants appear to be smaller/more compact in the presence of 15% ethanol. Table 4 shows the changes in apparent molecular weight of all variants in the presence of 15% ethanol. FOXP2 LZ-End remains larger than FOXP1 LZ-End and both FOXP1 LZ-FHD and FOXP2 LZ-FHD form a mixture of species compared to the longer variants which are a single species. The decrease in apparent molecular weight might be because of dissociation of the native quaternary state.

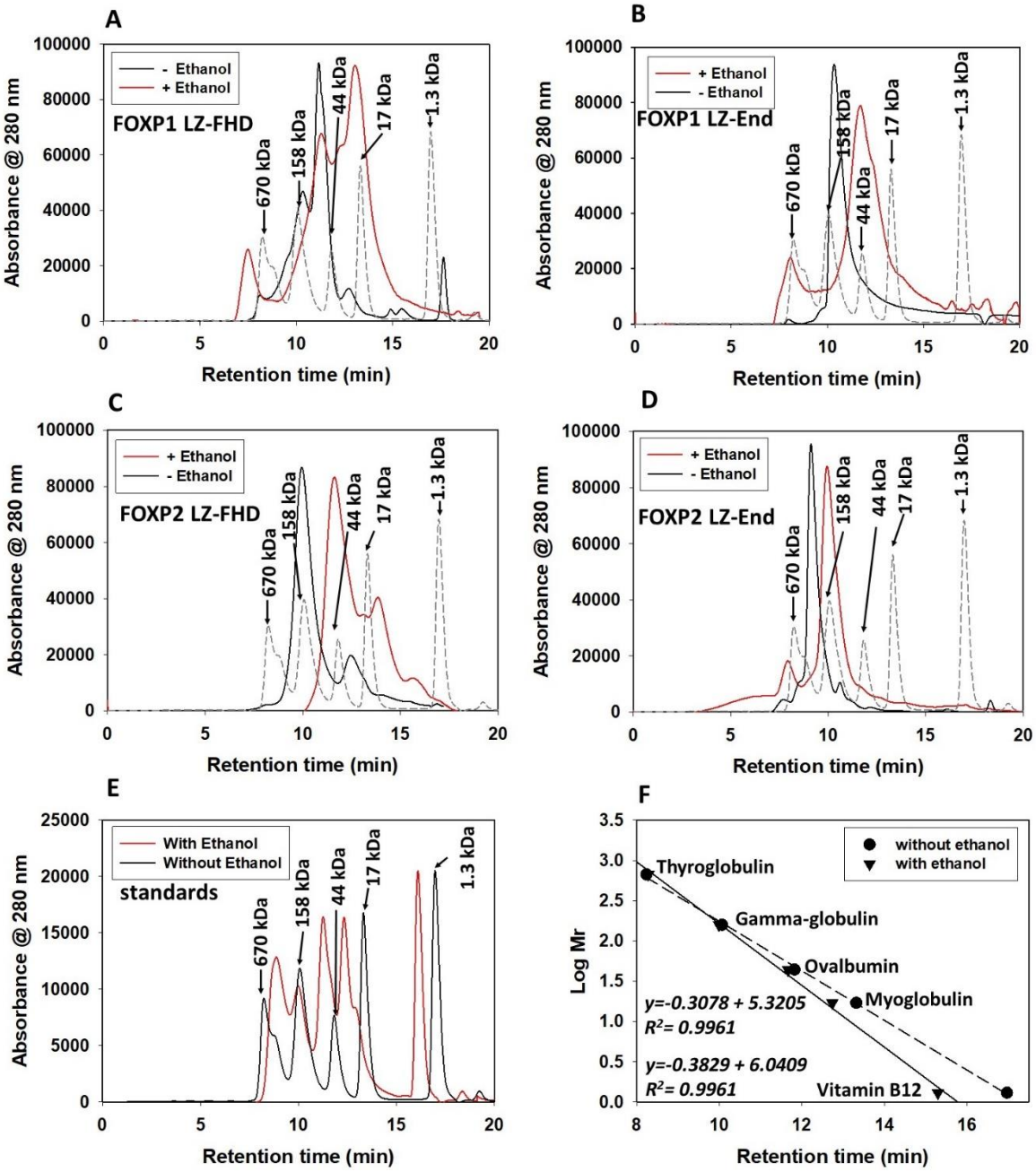


Figure 32: Apparent quaternary structure characterisation of FOXP1 and FOXP2 variants. The elution profile was resolved by SE-HPLC using 10 mM HEPES pH 7.4, 500 mM NaCl, 2 mM DTT and either 0% (-)(—) or 15 % (+)(—) ethanol. The column was calibrated using Bio-Rad Gel Filtration Standards (---) 1.3 kDa - 670 kDa (Bio-Rad, USA). (A) Elution profiles of FOXP1 LZ-FHD in the presence and absence of ethanol. FOXP1 LZ-FHD exists as a mixture of dimer and tetramer in the absence of ethanol and as a mixture of monomer and dimer in the presence of ethanol. (B) Elution profiles of FOXP1 LZ-End with and without ethanol show single peaks with apparent MW of ~76 kDa (dimer) and ~146 kDa (tetramer), respectively. (C) Elution profile of FOXP2 LZ-

FHD with and without ethanol. FOXP2 LZ-FHD exists as an apparent hexamer in the absence of ethanol and as a mixture of monomer and trimer in the presence of 15% ethanol. (D) Elution profiles of FOXP2 LZ-End in the presence and absence of ethanol. FOXP2 LZ-End exists as an apparent octamer in the absence of ethanol and an apparent hexamer in the presence of ethanol. (E) Bio-Rad standards were used to calibrate the SE-HPLC column in the presence and absence of ethanol. Minor differences were observed. (F) The calibration curve for the SE-HPLC standards shown in E. The molecular mass (in kDa) and retention times by the standard proteins were used to construct this curve.

Table 4: Changes in apparent oligomer size in the presence of ethanol

Variants	0 %Ethanol		15% Ethanol	
FOXP1 LZ-FHD	~76 kDa	~146 kDa	~30 kDa	~76 kDa
FOXP1 LZ-End	~146 kDa		~76 kDa	
FOXP2 LZ-FHD	~180 kDa		~26 kDa	~76 kDa
FOXP2 LZ-End	~326 kDa		~249 kDa	

3.6. PROTEIN-PROTEIN INTERACTIONS

One of the most important aspects of transcriptional regulation is the ability of transcription factors to physically interact with each other to form homo- or hetero-oligomers, resulting in inhibition or enhancement of transcriptional activity at a site distinct from the consensus target for a transcription factor (Manna, Dyson and Stocco, 2016). It has been shown that FOXP1/2/4 are likely capable of interacting with each other *in vivo* to form hetero-oligomers (Li *et al.*, 2004; Sin *et al.*, 2014). FOXP1 has also been shown to associate with FOXP3 through the leucine zipper to form hetero-oligomers (Song *et al.*, 2012). It is, therefore, evident that a neuro-molecular complex consisting of multiple proteins and nucleic acids contributes to the regulation of transcription by these factors, likely in a spatially and temporally controlled manner. Knowledge on the mechanism of this regulation will be empowering in understanding how these transcription factors function and how misfunction leads to disease. Using pull-down assays, size exclusion chromatography, fluorescence anisotropy and isothermal titration calorimetry, homotypic and heterotypic associations of FOXP1 and FOXP2 were studied.

3.6.1. Homo-oligomerisation of FOXP1 and FOXP2

It has been shown at length in this thesis up to this point that all four FOXP variants are able to homo-associate to form higher order oligomeric structures ranging from apparent dimers to apparent octamers. Furthermore, it has been shown by others that *in vivo* FOXP2 homodimers are functionally important for DNA biding (Li *et al*., 2004; Wang *et al*., 2003).

In this study, the truncated FOXP1 and FOXP2 variants (FOXP1 LZ-FHD and FOXP2 LZ-FHD) were used to verify the occurrence of homo-association by visualising their interaction with their respective longer variant (i.e. FOXP1 LZ-FHD with FOXP1 LZ-End and FOXP2 LZ-FHD with FOXP2 LZ-End) on an SDS-PAGE gel following a pull-down assay on a Ni^{2+}-charged affinity chromatography column. Because both the truncated variants do not possess the acid rich tail (ART) C-terminal to the FHD, all the truncated variants can be distinguished from the longer variants based on size; therefore, allowing us to identify whether or not homo-association occurred. The 6XHis-tagged FOXP1 LZ-End species was able to homo-associate with the untagged FOXP1 LZ-FHD at all concentrations of FOXP1 LZ-FHD (10-40 µM) (Figure 33A). Similar experiments performed with FOXP2 (figure 33B) also showed homo-

association. These findings therefore suggest that the leucine zipper domain and FHD are enough to mediate homo-oligomerisation of both FOXP1 and FOXP2.

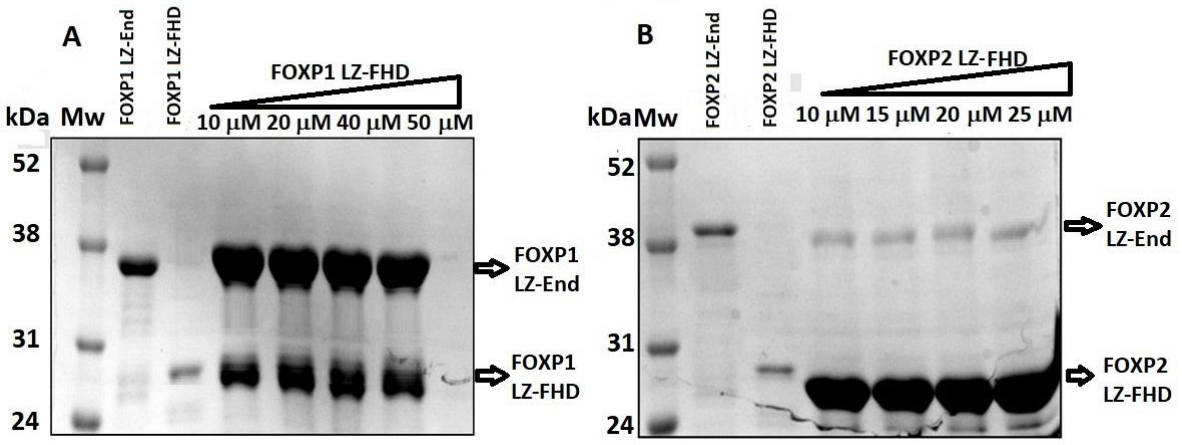


Figure 33: Homo-association of FOXP1 variants and FOXP2 variants following a pull-down assay. (A) SDS-PAGE gel showing homo-association of FOXP1 variants following a pull-down assay using 6xHis tagged FOXP1 LZ-End (~39 kDa) as prey and FOXP1 LZ-FHD (~30 kDa) as bait. (B) SDS-PAGE gel showing homodimerisation of FOXP2 variants following pull down assay of 6x His tagged FOXP2 LZ-End (~45 kDa) as prey and FOXP2 LZ-FHD (~29 kDa) as bait. Gel electrophoresis was performed on a 12.5% tris-glycine polyacrylamide gel. Lane 1 contains molecular weight marker, lanes 2 and 3 in both gels contain the respective purified proteins while lanes 4-7 on both gels show the eluent after the pull-down assays using increasing concentrations of the shorter variant.

3.6.2. Hetero-oligomerisation of FOXP1 and FOXP2

FOXP1 and FOXP2 have widely overlapping expression patterns in both developing and mature brain (Teramitsu *et al.*, 2004; Bowers and Konopka, 2012; Fong *et al.*, 2018). Because FOXP1 and FOXP2 have been shown to regulate the same upstream and downstream genes, it is likely that they hetero-associate *in vivo* so as to regulate transcription (Sin, Li and Crawford, 2014). In order to demonstrate whether this hetero-association can be detected *in vitro* in an isolated system, the hetero-association of FOXP1 and FOXP2 variants containing the intact leucine zipper domain and FHD was investigated in this study.

Although the crystal structure of the isolated FOXP2 FHD shows that the FHD can exist as a domain-swapped dimer (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011), and it is generally accepted that all FOXP FHDs do dimerise (Chu *et al.*, 2011; Chen *et al.*, 2015), there is evidence supporting the fact that the FOXP2 FHD is not exclusively dimeric and indeed exists as a monomer at concentrations as high as 300 µM (Morris and Fanucchi, 2016). Furthermore, the monomeric FHD is capable of binding to DNA in isolation (Morris and Fanucchi, 2016). The leucine zipper domain is likely to have a very strong propensity to dimerise (Massari and Murre, 2000). Indeed, the FOXP3 leucine zipper has been shown to form both dimers and higher order oligomers (Song *et al.*, 2012) which corresponds with the work presented in this thesis on FOXP1 and FOXP2. Although the leucine zipper is not believed to directly interact with DNA, mutated leucine zipper domains have been shown to influence DNA binding (Song *et al.*, 2012). It is likely that if FOXP1 and FOXP2 do indeed hetero-associate in isolation, this interaction occurs through the leucine zipper domain (Li, Weidenfeld and Morrisey, 2004) but we cannot discount the presence of the FHD in the dimer and this is why the various variants were used in this study.

In order to thoroughly account for all possible FOXP dimer interfaces, a further variant was used in this study: a truncated form of FOXP2 containing only the FHD and acid rich tail (amino acids 501-714). This variant was used to determine the contribution of the FHD to hetero-association in the absence of the leucine zipper.

All three of the FOXP2 variants (FOXP2 FHD-End, FOXP2 LZ-FHD and FOXP2 LZ-End) were used as the bait in the pull-down assays. The hexa-histidine tag was left intact on all the FOXP2 variants, in other to allow then to interact with the Ni^{2+} ion used to charge the IMAC column. Alternatively, thrombin cleavage to remove the hexa-histidine tag was performed on both FOXP1 variants (FOXP1 LZ-FHD and FOXP1 LZ-End). To eliminate any chance of false

interactions, the tag free FOXP1 variants were first loaded onto the Ni^{2+}-charged IMAC column using experimental conditions so as to ensure that the prey protein does not interact non-specifically with the column. In this control experiment (Figure 34G) both FOXP1 variants (prey protein) did not remain bound to the column and eluted at high salt concentration (500 M NaCl).

For the pull-down assays, each of the FOXP2 variants (bait protein) was separately loaded onto the IMAC column and each bait protein was used to pull down each of the FOXP1 variants in turn which were used as prey. This resulted in six separate pull-down assays being performed (Figure 34). Any interaction between the FOXP1 and FOXP2 variants would result in the FOXP1 variant binding to the immobilised FOXP2 on the column. The eluents for each pull-down assay experiment were resolved using SDS-PAGE. Each gel (Figure 34A-34F) shows the interaction between FOXP1 and FOXP2 variants. The presence of two bands in each of the eluents of the pull-down assays means that an interaction between the two proteins was detected. At all concentrations tested (5- 50 µM), all six heterotypic combinations exhibited association during the pull-down assays (Figure 34). The ability of FOXP1 and FOXP2 variants to hetero-associate (including the FOXP2 FHD-end variant), suggests that not only is the leucine zipper necessary for hetero-association, but the FHD also contributes significantly to this event.

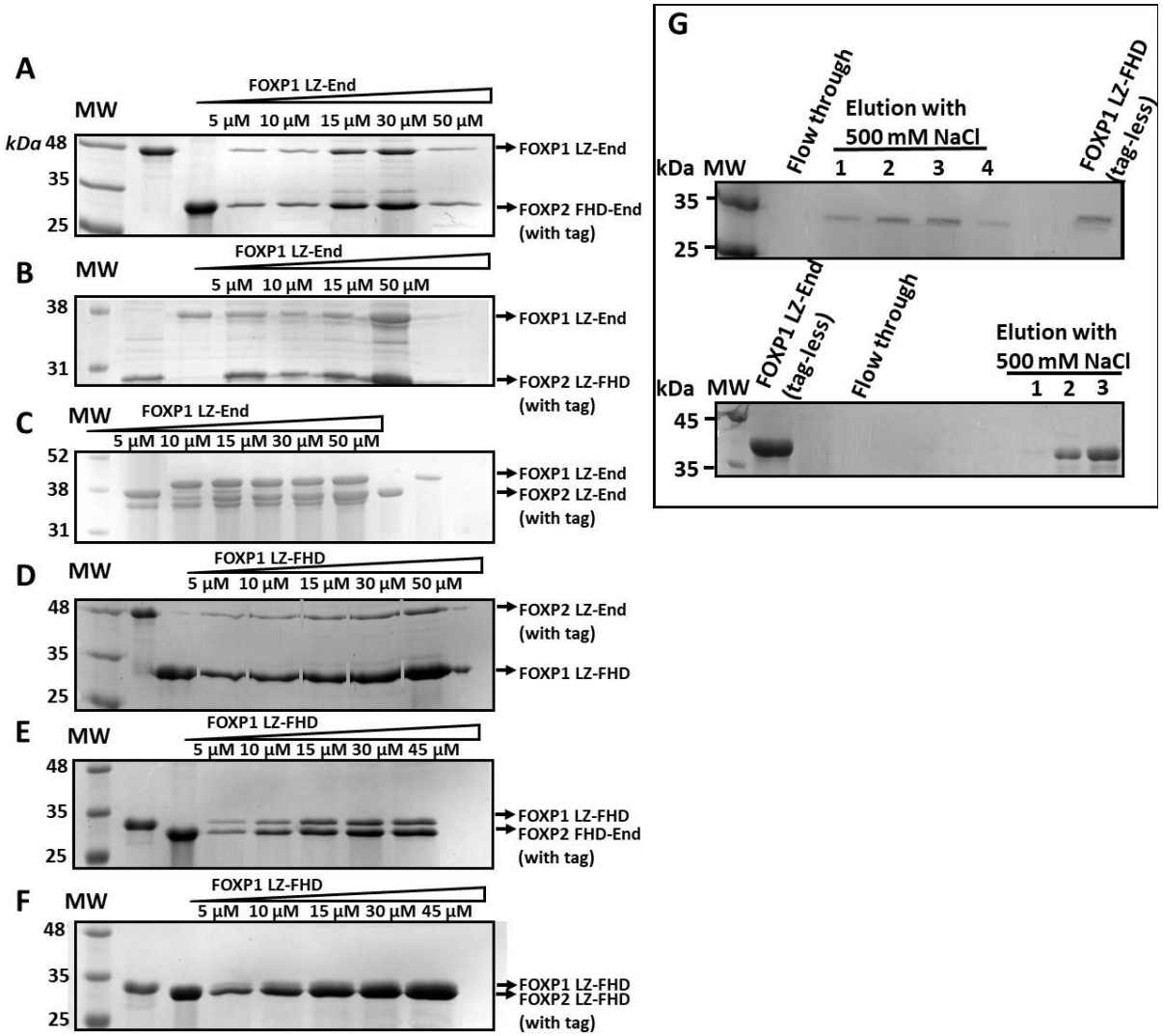


Figure 34: Heterotypic interactions of FOXP1 and FOXP2 variants as detected on a 12% tris-glycine polyacrylamide gel following pull down assays with increasing concentrations of the prey protein (~5 µM - ~50 µM). (A) The interaction between FOXP1 LZ-End (prey) and FOXP2 FHD-End (bait) (B) The interaction of FOXP1 LZ-End (prey) and FOXP2 LZ-FHD (bait) (C) The interaction between FOXP1 LZ-End (prey) and FOXP2 LZ-End (bait) (D) The interaction between FOXP1 LZ-FHD (prey) and FOXP2 LZ-End (bait). (E) The interaction between FOXP1 LZ-FHD (prey) and FOXP2 FHD-End (bait). (F) The interaction between FOXP1 LZ-FHD (prey) and FOXP2 LZ-FHD (bait). Gel electrophoresis was performed on a, following pull down assays. The purified proteins prior to heterotypic interaction are found in the second and third lane of each gel except gel C where they are in the 8th and 9th lanes and gel D where they are in the 2nd and 9th lanes. The presence of two bands in each of the eluents of the pull-down assays means that an interaction between the two proteins was detected.

3.6.3. Quaternary structure characterisation of hetero-oligomers

Size exclusion chromatography was used to further characterise hetero-oligomerisation of the FOXP1 and FOXP2 variants following the pull-down assay using the eluents from the IMAC column. Although the purified variants eluted mostly as single peaks, the heterotypic combinations of the variants eluted with traits of both the proteins combined (Figure 35). For the most part FOXP1 LZ-End + FOXP2 FHD-End (Figure 35A) and FOXP1 LZ-FHD + FOXP2 LZ-End (Figure 35D) eluted as a single peak. For the other truncated FOXP1 and FOXP2 variants (Figure 35B, 35C, 35E & 35F), the heterotypic association eluted as more than one distinct peak, characterised by shoulders on the elution profile. This suggests that even though hetero-association occurs, the affinity of the proteins for each other is not as high and there is also dissociation to form lower order hetero-oligomers.

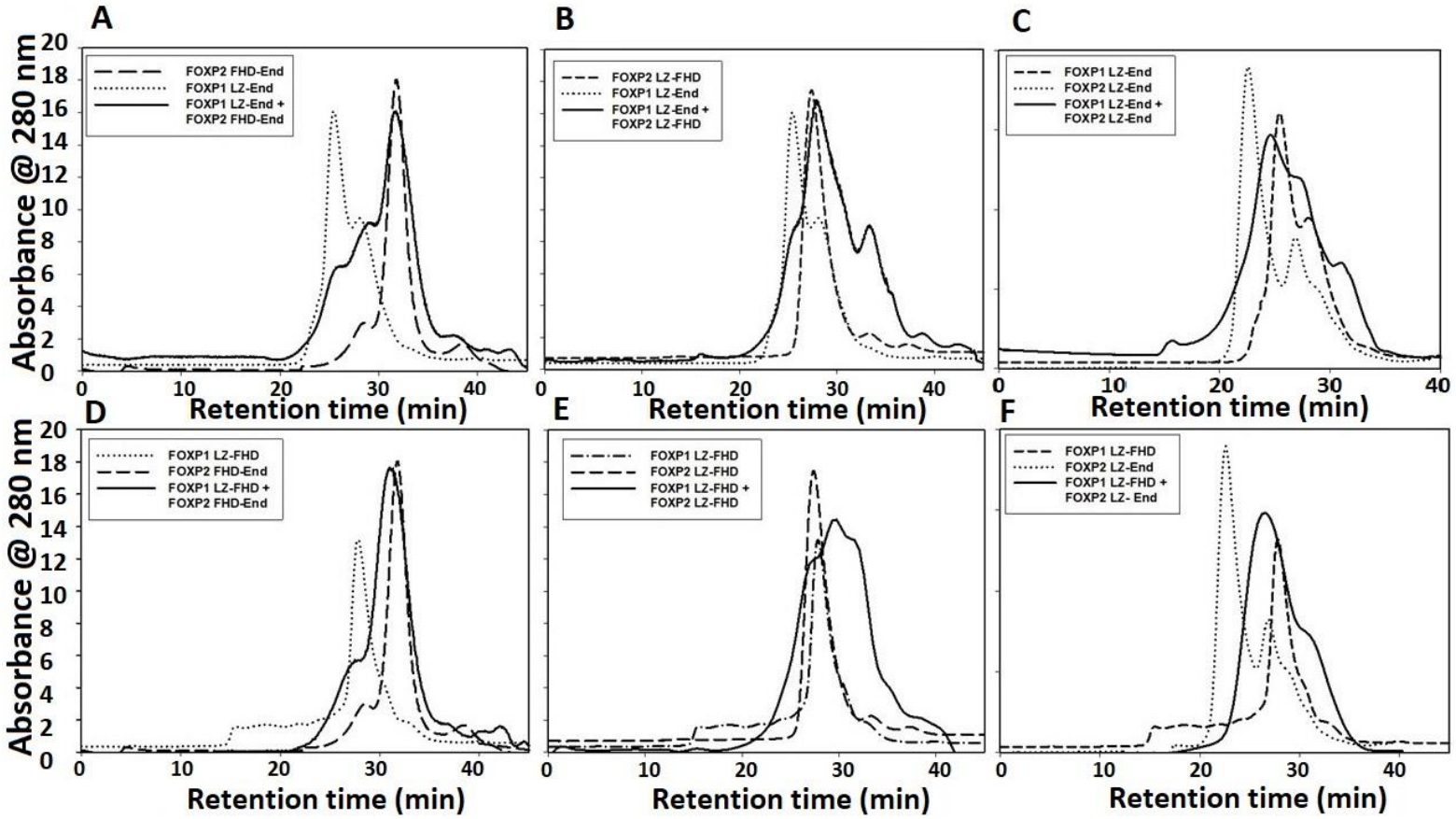


Figure 35: Size exclusion chromatography of FOXP1 and FOXP2 heterotypic combinations following pull down assays. (A) FOXP1 LZ-End + FOXP2 FHD-End (B) FOXP1 LZ-End + FOXP2 LZ-FHD (C) FOXP1 LZ-End + FOXP2 LZ-End (D) FOXP1 LZ-FHD + FOXP2 FHD-End (E) FOXP1 LZ-FHD + FOXP2 LZ-FHD (F) FOXP1 LZ-FHD + FOXP2 LZ-End. Size exclusion chromatography was performed on a Superdex® 200 10/300 GL column to characterise the quaternary structure of the hetero-association between FOXP1 and FOXP2 variants.

3.6.4. Secondary and tertiary structural characterisation of the hetero-oligomers

Far-UV circular dichroism and intrinsic tryptophan fluorescence spectroscopy were used to characterise the secondary and tertiary structure, respectively, of the FOXP1 and FOXP2 heterotypic associations in comparison to the parent proteins. This was done so as to give an indication of any structural changes that may have been induced by hetero-association. All far-UV CD spectra (Figure 36A) are indicative of predominantly α-helical structure with a minor β- sheet component and suggest that the homo-associated and hetero-associated variants are almost identical in secondary structure; therefore, suggesting that no major conformational changes occur upon hetero-association. Similarly, there is little difference in the fluorescence spectra (Figure 36B) between the parent proteins and the hetero-associated proteins (under reducing conditions). The maximum emission wavelength of 330 nm observed for the isolated FOXP2 FHD (Morris and Fanucchi, 2016), indicative of buried tryptophan residues, was also observed for all FOXP1 and FOXP2 variants regardless of homo-association or hetero-association. The FHD is a good model for comparison because the only tryptophan residues are located in the FHD and so fluorescence gives information on the local environment of the FHD in all the variants.

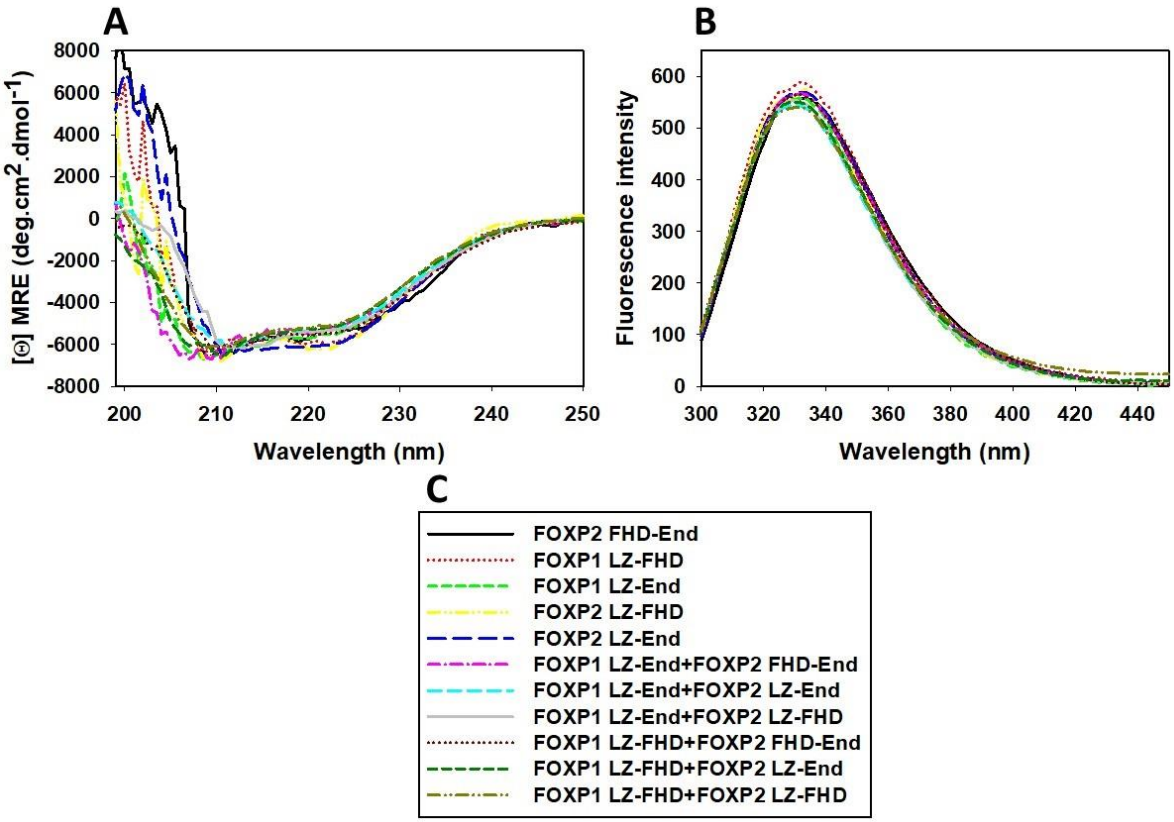


Figure 36: Secondary and tertiary structural characterisation of FOXP1 and FOXP2 variants following hetero-association using pull-down assays. (A) Far-UV circular dichroism spectra of FOXP1 and FOXP2 heterotypic association in comparison to the parent proteins. (B) Intrinsic tryptophan fluorescence spectra of FOXP1 and FOXP2 variants, both homo-associations and hetero-associations. (C) Guide to identifying the different combination for homo- and hetero-associations.

3.6.5 Association affinity studies for FOXP1/FOXP2 oligomers

Since this work established that all the variants can hetero associate, the question was whether this hetero-association happened with the same or different affinity and also whether or not it was affected by the presence of DNA. Fluorescence anisotropy (FA) (Figure 37) was used to determine the hetero-association affinity of FOXP1 and FOXP2 variants for each other, both in the presence and absence of cognate DNA (Nelson *et al.*, 2013). The binding isotherms for the FOXP1 and FOXP2 hetero-associations all fit well to a single-site binding model, and all the variants appear to have similar binding affinities for each other with all K_Ds being in the low micromolar range (Figure 37 and Figure 38). Each variant has the same binding affinity to any of the other variants and there is no significant difference in binding affinity between any of the variants in the absence of DNA. In the presence of DNA, however, the shorter FOXP1 variant, shows improved affinity for all the FOXP2 variants.

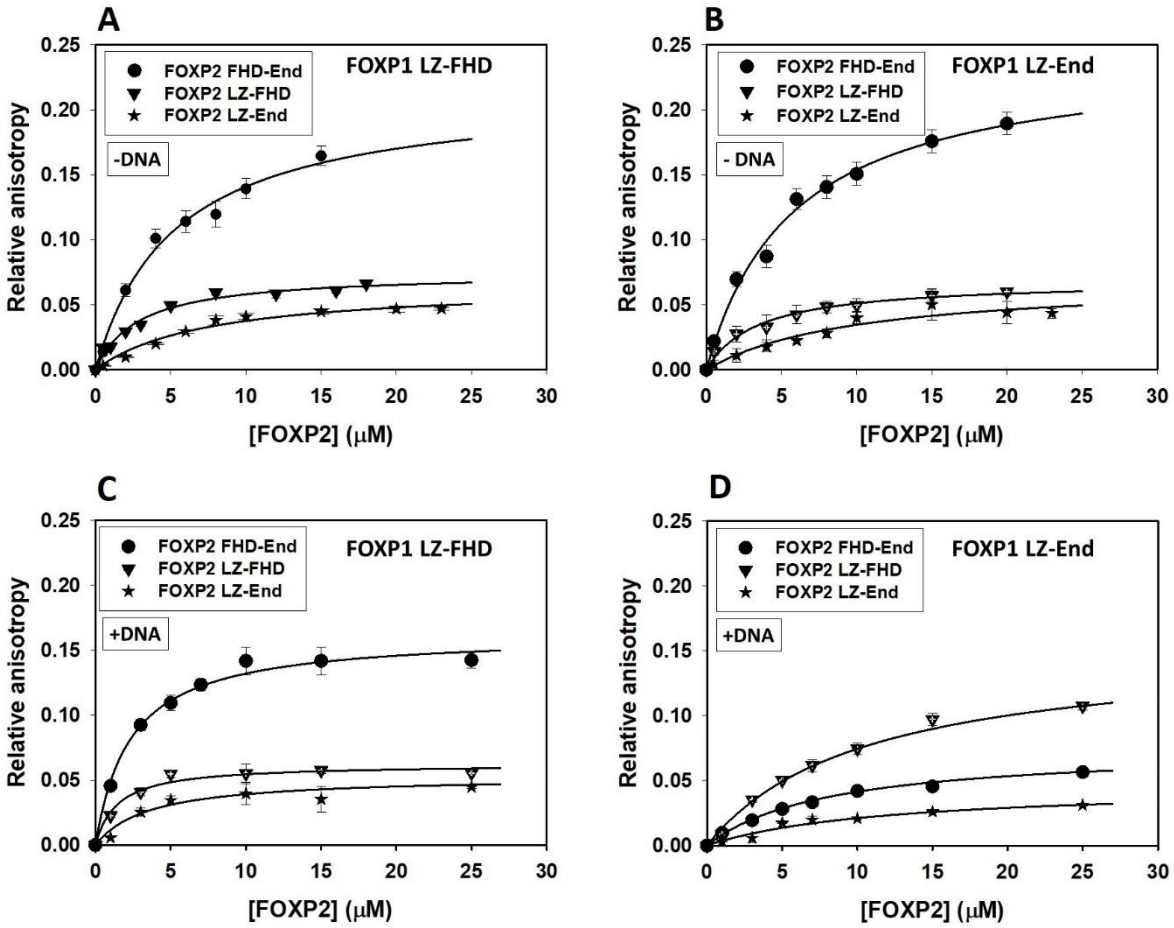


Figure 37: Heterotypic binding isotherms of FOXP1 and FOXP2 truncated variants as determined by fluorescence anisotropy spectroscopy. For each experiment, 5 µM of either labelled (labelled with NTA-ATTO 550) FOXP1 LZ-FHD or labelled FOXP1 LZ-End was incubated with increasing concentrations of unlabelled FOXP2 FHD-End, FOXP2 LZ-End and FOXP2 LZ-FHD. The fluorescence anisotropy at each FOXP2 concentration was measured and the data fitted with a single-site binding model. (A) FOXP1 LZ-End binding isotherms with each of the FOXP2 variants. The dissociation constants are: 10.1 ± 3.5 µM for the interaction with FOXP2 FHD-End, 5.8 ± 0.7 µM for the interaction with FOXP2 LZ-FHD and 3.5 ± 0.6 µM for the interaction with FOXP2 LZ-End (B) FOXP1 LZ-FHD binding isotherms with each of the FOXP2 variants. The dissociation constants are 5.2 ± 0.9 µM for the interaction with FOXP2 FHD-End, 7.5 ± 1.5 µM for the interaction with FOXP2 LZ-FHD and 2.9 ± 0.4 µM for the interaction with FOXP2 LZ-End. (C) FOXP1 LZ-End binding isotherms with each of the FOXP2 variants in the presence of DNA. The dissociation constants are: 8.2 ± 0.5 µM for the interaction with FOXP2 FHD-End, 10.5 ± 1.1 µM for the interaction with FOXP2 LZ-FHD and

12.0 ± 4.6 µM for the interaction with FOXP2 LZ-End. (D) FOXP1 LZ-FHD binding isotherms with each of the FOXP2 variants in the presence of DNA. The dissociation constants are: 2.3 ± 0.4 µM for the interaction with FOXP2 FHD-End, 1.4 ± 0.4 µM for the interaction with FOXP2 LZ-FHD and 3.6 ± 1.1 µM for the interaction with FOXP2 LZ-End.

The K_D values for hetero-association are compared in Figure 41 in the presence and absence of DNA. Interestingly, while in the absence of DNA there is no real significant difference in hetero-association binding affinity between FOXP1-End and FOXP1-FHD, when DNA is added, it is clear that FOXP1-FHD makes significantly stronger hetero-associations with all FOXP2 variants implying the significance of the acid rich tail in transcriptional regulation. The presence or absence of the C-terminal acid rich tail, though significant for structural stability, can cause thermodynamic destabilisation or stabilisation of the protein in the absence or presence of DNA, respectively (Crane-Robinson *et al.*, 2006; Privalov *et al.*, 2007; Tóth-Petróczy *et al.*, 2009). Consequently, this tail could interfere with profound interactions between the hetero-association domains; therefore, slight weakening of the interactions.

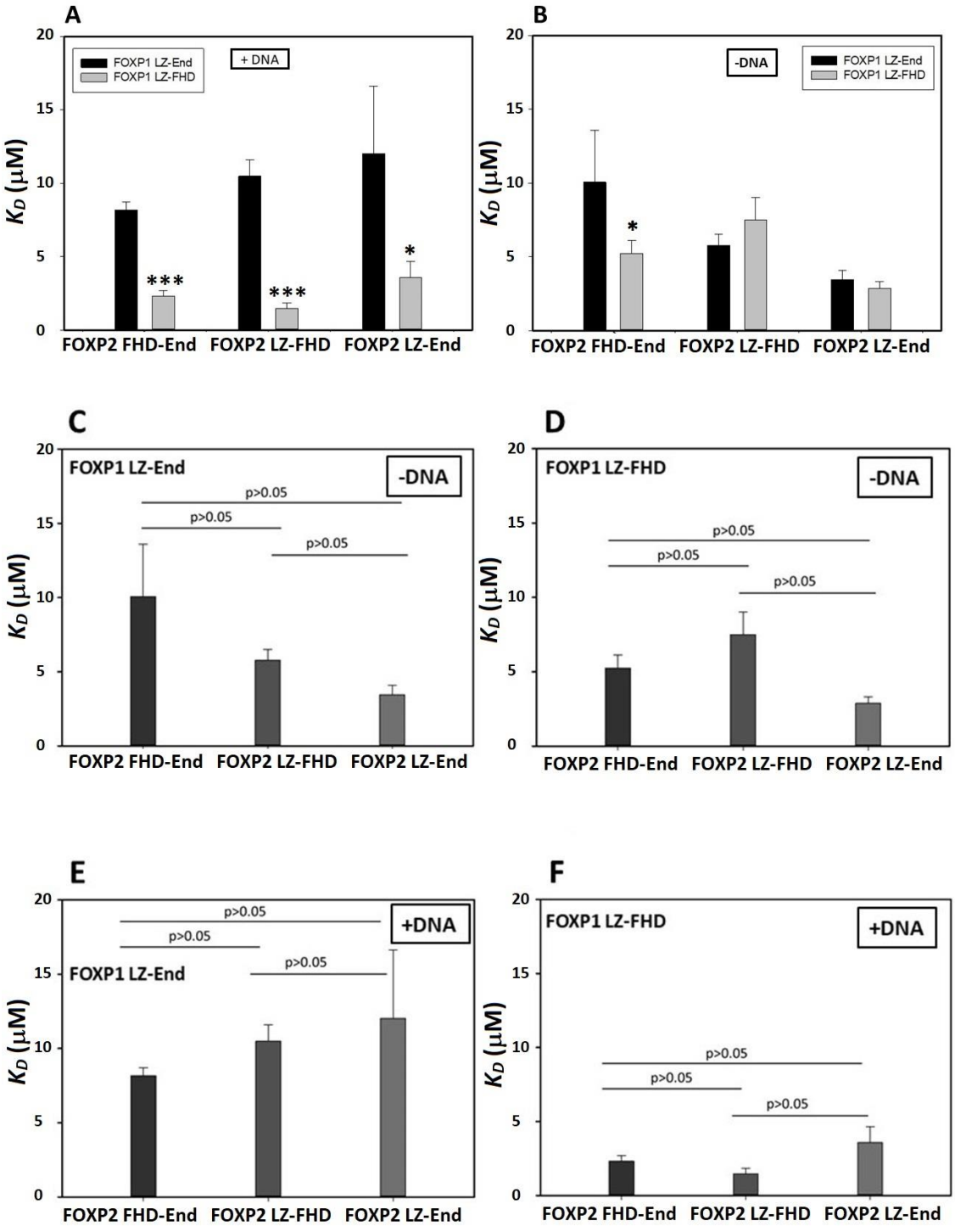


Figure 38: Comparison of FOXP1 and FOXP2 variant association binding affinity K_Ds as measured using fluorescence anisotropy in the presence and absence of DNA. (A) Comparison of the binding affinity for hetero-association of FOXP1 variants with FOXP2 variants when in the presence of DNA. (B) Comparison of the binding affinity for hetero-association of FOXP1 variants with FOXP2 variants when in the absence of DNA. (C) Statistical comparison of the dissociation of constants for FOXP1 LZ-End hetero-

association with all FOXP2 in the absence of DNA. (D) Statistical comparison of the dissociation constants for FOXP1 LZ-FHD hetero-association with all FOXP2 variants in the absence of DNA. (E&F) Statistical comparison of dissociation constants of hetero-association upon DNA-binding. Errors indicate the standard deviation of three averaged replicates. Statistical significance of the difference in affinity between hetero-oligomers was determined using a student *t*-test (*p < 0.05; **p < 0.01; ***p < 0.001).

Figure 38 shows the comparison of FOXP1 and FOXP2 hetero-association dissociation constants (K_D) with that of the purified FOXP1 and FOXP2 variants. In general, there is low affinity for DNA upon hetero-association in comparison to the DNA-binding of the homo-associated FOXP1 and FOXP2 variants. The only exception being the K_D for binding of FOXP1 LZ-FHD/FOXP2 LZ-FHD hetero-association in the presence of DNA (Figure 39B) which is similar to that obtained for homo-associated FOXP1 LZ-FHD in the presence of DNA. Comparatively, hetero-association (Figure 39) leads to decreased binding to the cognate binding site on DNA 5′-GGAAA**ATTTGT**TTCA-3′ (FOXP binding sites are under- lined, bases in bold match with the *in vitro* selected sequence) (Wu *et al.*, 2006).

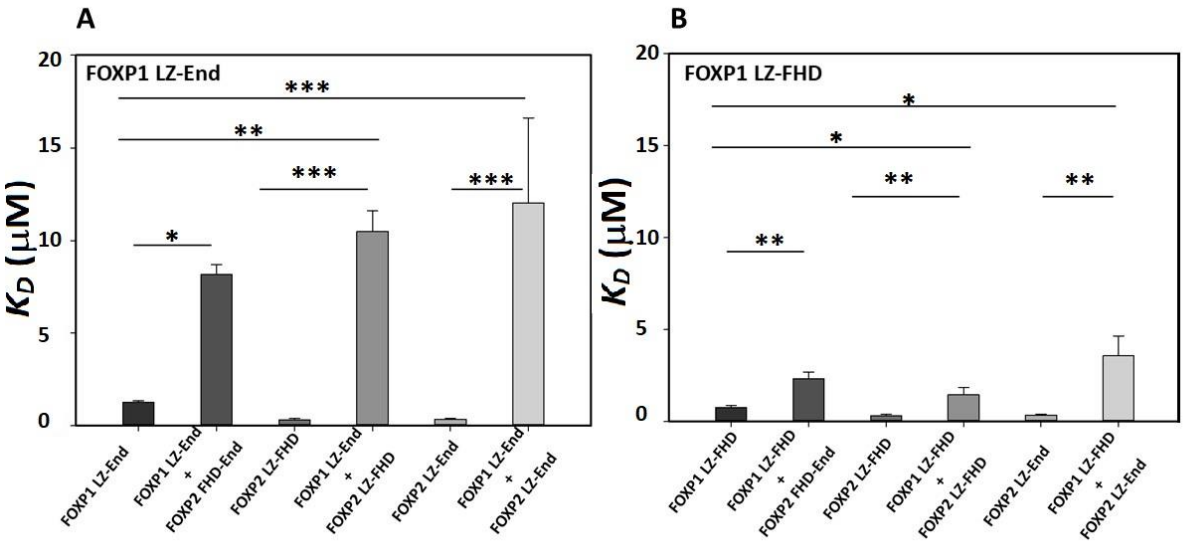


Figure 39: Comparison of the DNA binding affinity of the hetero associated proteins with the DNA binding affinity of the homo-associated proteins. A) FOXP1 LZ-End was used as the fluorescent-tagged FOXP1variant and its association with DNA in the presence of FOXP2 variants was measured. B) FOXP1 LZ-FHD was used as the fluorescent-tagged variant and its association with DNA in the presence of FOXP2 variants was measured. Hetero-complexes do not associate as tightly with DNA as homo-associated proteins. Errors indicate the standard deviation of three averaged replicates. Statistical significance of the difference in affinity between hetero-oligomers was determined using a Student *t*-test (*p < 0.05; **p < 0.01; ***p < 0.001).

In the current study characteristics of FOXP1 and FOXP2 focusing on the dimerisation interface were investigated. All four variants, lack some of the interfaces that are present in the full-length protein; however, FOXP1 LZ-End and FOXP2 LZ-End better represent the full-length protein as all the other domains and subdomains are intact except for the polyglutamine region and the zinc finger motif. Therefore, thermodynamic parameters of binding were obtained for the protein-protein interaction between FOXP1 LZ-end and FOXP2 LZ-End variants using isothermal titration calorimetry (Figure 40) in order to validate and verify the hetero-association observed using fluorescence anisotropy. This was done to investigate the mode of interaction and binding. The isothermal titration data fit to a multiple-site model obtained in the NanoAnalyzer software that is provided by the manufacturer (TA Instruments, USA). Because FOXP proteins have two distinct dimerisation interfaces, the multiple site binding implies that both the leucine zipper and FHD are involved in the heterodimerisation event. However, it cannot be ruled out that both dissociation and association events could be occurring simultaneously. Thermodynamic parameters for FOXP1 and FOXP2 protein-protein interactions are summarised in Table 5. The free energy (ΔG_1=-39.88 ± 0.925 kJ/mol and ΔG_2=-33.91 ± 1.499 kJ/mol) for both events occurring during ITC titrations are negative, indicating spontaneity of heterodimerisation at pH 7.4 and room temperature (25 ℃) (Table 5). The dissociation constant K_D 3.67 ± 0.037 µM obtained for FOXP1/FOXP2 LZ-End association in ITC (Table 4) corresponds to that obtained for FOXP1/FOXP2 LZ-End variants (Figure 40) obtained from fluorescence anisotropy.

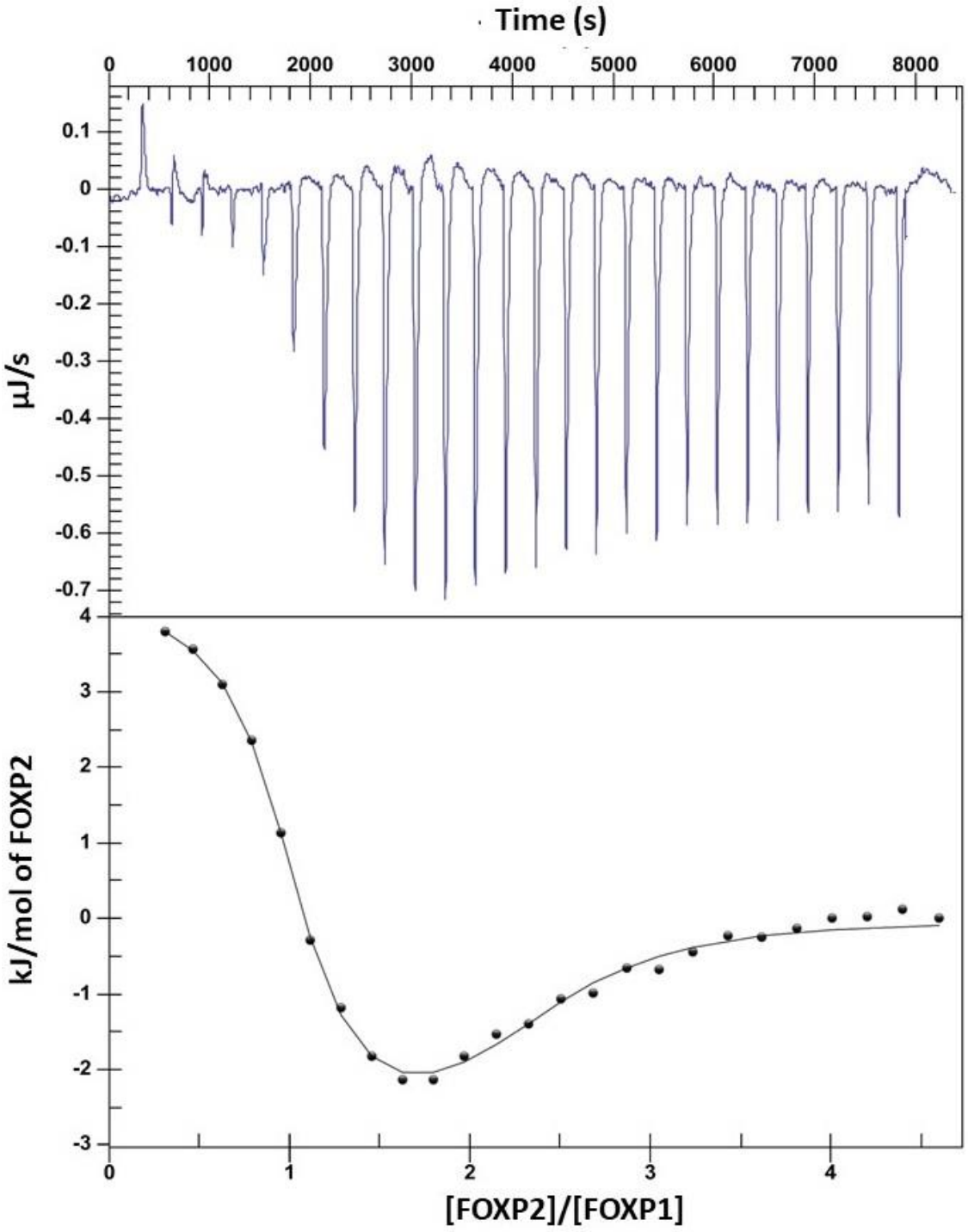


Figure 40: Titrations of FOXP1 LZ-End with FOXP2 LZ-End. The experiments were performed in 10 mM HEPES, pH 7.4, 100 mM sodium chloride, at 25 ℃. The concentrations of reactants are ~32 µM FOXP1 LZ-End (in cell) and ~480 µM FOXP2 LZ-End (in syringe). The isotherms were obtained by multiple 9.86 µL injections of protein into protein or into buffer. Thermodynamic parameters were obtained using a multiple site binding model and are outlined in Table 5 below.

Table 5: Summary of ITC thermodynamic parameters obtained for the FOXP1/FOXP2 hetero-oligomerisation, FOXP1 LZ-End and FOXP2 LZ-End dissociation. Fitting of the raw data was done using the multiple site model for the hetero-oligomer and the independent site model for the dissociation of FOXP1 and FOXP2 LZ-End variants on the NanoAnalyze Software provided by the manufacturer. Errors are the standard deviation of three averaged independent titration experiments.

	N	K_D	ΔH (kJ/mol)	TΔS (kJ/mol)	ΔG (kJ/mol)
FOXP1/FOXP2 LZ-End hetero-oligomer	0.9 ± 0.1	102.4 ± 26.9 nM	4.3 ± 1.6	44.2 ± 2.3	-39.9 ± 0.9
	1.5 ± 0.1	3.7 ± 0.0 nM	-3.3 ± 1.4	30.6 ± 4.8	-33.9 ± 1.5
FOXP2 LZ-End	_	458.0 ± 1.0 nM	-15.0 ± 3.6	20.6 ± 0.7	-35.6 ± 4.3
FOXP1 LZ-End	_	1939 ± 12.3 nM	-53.1 ± 0.4	-21.0 ± 4.0	-32.1 ± 3.6

In summary, it was found that the FHD and leucine zipper domains of FOXP1 and FOXP2 are sufficient to mediate hetero-association, which, therefore, suggests that any mutations that occur could disrupt not only this event, but homodimerisation as well, as observed in human XLAAD/IPEX patients (Li *et al.*, 2007). Just as FOXP3 associates with different supramolecular ensembles in an apparent dynamic manner (Li *et al.*, 2007), it is hypothesised that FOXP1 and FOXP2 can be found in both homo- and hetero-dimeric or tetrameric associations.

CHAPTER 4

DISCUSSION

The idea of a structure-function paradigm predicates on the fact that proteins need a rigid three-dimensional (3D) structure in order to perform their functions. Although this might be the case for small globular proteins, there is extensive evidence that such a phenomenon is not always valid. Many eukaryotic proteins exist in solution and in the cellular environment in disordered conformations (Mittag *et al.*, 2010). These proteins, known as intrinsically disordered protein (IDP) lack secondary structure and tertiary structure conformations, and yet they are still able to perform their functions (Uversky, 2002). This is because, unlike globular proteins, IDP undergo a distinct disorder-to-order transition when in the presence of a suitable ligand (Vuzman and Levy, 2012). Because in more cases than not, these IDP are usually DNA-binding proteins, such a ligand is usually a suitable DNA sequence on the DNA template. Interestingly enough, there are other DNA binding domains that exhibit both globularity and intrinsic disorder in their native conformations. These proteins are usually characterised by multiple domains, linked by flexible linkers, which usually are regions of disorder, characterised by a high random coil content (Vuzman *et al.*, 2010; Van Der Lee *et al.*, 2014). In conjunction with these multiple domains, which cooperatively contribute to function and functional diversity, most multidomain proteins also comprise of N- and C-terminal tails. In transcription factors such as p53, the N-terminal tail is composed of negatively charged amino acids and is usually very involved in transcriptional repression (Laptenko *et al.*, 2016). The C-terminal tail on the other hand is usually very rich in positively charged amino acids, making this subdomain important in mediating DNA binding affinity and specificity, especially considering that it is usually close to the DNA binding domain (DBD) (Vuzman and Levy, 2012).

The major focus of this thesis was to dissect the structure of the FOXP proteins, FOXP1 and FOXP2 with emphasis on oligomerisation and protein-protein interactions. This study also sought to identify structural elements other than the leucine zipper domain and the forkhead-box domain that are important to oligomerisation and stability. In particular, the contribution of the C-terminal acid rich tail, a region of intrinsic disorder to the overall oligomerisation

stability and its contribution to protein-protein interaction affinity, especially during hetero-associations between FOXP1 and FOXP2 were examined. The work presented in this study is novel and the construct used to study homo- and hetero-oligomerisation have not been used in other studies. However, a recent study was published on the oligomerisation of FOXP2 using similar constructs (Häußermann *et al.*, 2019).

The FOXP proteins are not only divergent in the structure of the DNA binding domain to the rest of the FOX proteins (Shu *et al.*, 2001; Wang *et al.*, 2003), but also in how they interact with DNA (Li, Weidenfeld and Morrisey, 2004). Though highly conserved, the FOXP FHD, the DNA-binding domain, is the only FOX DBD that is capable of forming domain-swapped dimers (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011; Chu *et al.*, 2011; Chen *et al.*, 2015). Indeed, the ability of the FOXP proteins to dimerise seem to be unique to these proteins and is likely to be very important to DNA binding. Together with their conserved but divergent FHD, the FOXPs also have a leucine zipper domain, that represents the major and most important dimerisation interface in the full-length protein (Li *et al.*, 2004; Lopes *et al.*, 2006; Li *et al.*, 2007; Song *et al.*, 2012). Indeed, in-frame deletion of the leucine zipper not only affects protein-protein interactions, but also affect DNA binding (Li *et al.*, 2004). Furthermore, the FHD is succeeded by a C-terminal tail, the function of which is yet to be fully examined.

In all variants that have the C-terminal tail (FOXP1 and FOXP2 LZ-End), higher order oligomers are the predominant and only quaternary ensemble, suggesting that this C-terminal tail could be quite important to FOXP1 and FOXP2 homo-associations, and possibly hetero-complex formations. In contrast, both FOXP1 LZ-FHD and FOXP2 LZ-FHD show a variety of quaternary species, resembling to some extent those observed for FOXP3 (Bin Li *et al.*, 2007). The FOXP3 leucine zipper domain, which is similar in all FOXPs, consist of a hydrophobic stretch on the molecular surface which promotes formation of higher oligomers(Song *et al.*, 2012). It could, therefore, be suggested that FOXP proteins could mediate specificity and stability by formation of higher order oligomeric conformations. Indeed, this behaviour seems to be much more prevalent and preferred in most transcription factors (Hinde *et al.*, 2016; Presman *et al.*, 2016). The idea that different oligomeric state in a protein may mediate discrete functions was first proposed for other repressive/activating factors such as p53 (Tarunina and Jenkins, 1993). It is this feature that is probably important in distinguishing FOXP1 and FOXP2 during transcriptional regulation despite high sequence similarities and overlapping expression patterns.

FOXP1 LZ-End and FOXP2 LZ-End exists as higher oligomers at despite changes in the pH of the solution. However, FOXP1 LZ-End is more susceptible pH changes as it dissociates at high pH (pH>7.5). Interestingly, FOXP1 LZ-FHD and FOXP2 LZ-FHD are more affected by changes in pH as they are reduced in size at high pH even though they largely maintain their quaternary state. These slight changes in globularity that occur with increasing pH have been reported for the isolated FOXP2 FHD (Blane and Fanucchi, 2015). This is not surprising considering the fact that both dimerisation interfaces are predominated by hydrophobic interactions (Stroud *et al.*, 2006; Song *et al.*, 2012) and should not be affected by electrostatic interactions. The interaction between the monomers is, therefore, strong as it involves hydrophobic patches. Indeed most transcription factors that homo-associate exhibit their functions as obligate dimers (Van Der Lee *et al.*, 2014). Therefore, it is not surprising that FOXP1 and FOXP2 form obligate associations.

Consequently, changes in the dielectric constant of the buffer affect the oligomeric state of all the variants. It is interesting to see that addition of ethanol results in dissociation of each variant to lower oligomeric states. This is because alcohols can cause unfolding of proteins before denaturation (Singh *et al.*, 2010). It is well understood through unfolding studies that protein dissociation occurs before unfolding (Lencki, Arul and Neufeld, 1992; Lau and Bowie, 1997; Maestro and Sanz, 2007; van der Vegt and Nayar, 2017). Although there is evidence that suggests that decreases in dielectric constant of water due to ethanol enhances electrostatic interactions compared to in water (Yoshikawa *et al.*, 2012a, 2012b), hydrophobic associations would be reduced because the hydrophobic effect will not be as pronounced when the solvent is less polar (van der Vegt and Nayar, 2017). The fact that changes in electrostatic interactions did not affect the oligomeric state of all variants, but the collapse of the hydrophobic effect leads to dissociation of FOXP1 and FOXP2 oligomeric states, could serve as evidence supporting the fact that the FOXP dimerisation interface is mainly mediated by hydrophobic interactions.

All variants exhibit spectroscopic sensitivity in the presence of ethanol. Addition of small amounts of ethanol induces formation of more α-helical content for the C-terminal tail containing variants, while only slight changes are observed in FOXP1 LZ-FHD and FOXP2 LZ-FHD, especially at 222 nm. In the oligomeric helical structure, the hydrophobic surfaces are buried within the oligomer interface, leading to stable quaternary structure. When these variants assumes a dimeric helical structure upon interaction with DNA, the hydrophobic surface is buried, making contact with exposed bases (Rohs *et al.*, 2012; Morris *et al.*, 2018).

It is interesting to note that interactions with DNA leads to formation of α-helices in disordered C-terminal tails (Vuzman, Polonsky and Levy, 2010; Vuzman and Levy, 2012). The present picture in which alcohols stabilize the helical structure by direct hydrophobic interactions is analogous to that of DNA-induced helical formation. Indeed FOXP2 FHD becomes more α-helical upon interaction with DNA (Webb *et al.*, 2017). A huge shift is observed in fluorescence spectra of all variants in increasing ethanol. Since intrinsic tryptophan fluorescence was used to monitor changes in the microenvironment of the FHD, the blue shift observed in spectra is more related to exposure of tryptophan to the non-polar environment. This blue shift is a direct indication of unfolding of the FHD, and probably the entire protein, which consequently leads to dissociation of the oligomeric state. Unlike interactions with DNA, ethanol does not lead to quenching of tryptophan fluorescence for the FHD as seen in other studies (Blane and Fanucchi, 2015), but an increase in quantum yield. This could be because tryptophan residues favour a more non-polar environment (Stryer, 1968; Lakowicz, 2002).

Unsurprisingly, thermal unfolding has shown that FOXP1 LZ-End and FOXP2 LZ-End are generally more stable than both FOXP1 LZ-FHD and FOXP2 LZ-FHD. This is expected since disordered proteins and protein regions are resistant to changes in environment because of the disorder which in most cases resembles unfolded conformations (Vuzman and Levy, 2012). This, therefore, suggest that the C-terminal acid rich tail might be important in structural stability. All variants showed considerable aggregation at temperatures exceeding 60 ℃. This was expected as irreversible aggregation follows the thermal unfolding transitions of multidomain proteins at elevated temperature (Lepock *et al*., 1992; Vogl *et al*., 1997; Fitter and Haber-Pohlmeier, 2004; Duy and Fitter, 2005, 2006), resulting in irreversible unfolding of proteins.

The fact that these variants dissociate in the presence of an organic solvent to oligomeric states that have been reported to interact with DNA, could be a direct indication of the dynamic state of the FOXP coiled coil. Indeed, a flexible conformation was reported for the FOXP3 leucine zipper domain (Song *et al.*, 2012). The lack of tight assembly that exist means that FOXP proteins are able to undergo tightening to form a stable dimer for high-affinity DNA binding upon differential posttranslational modifications or rapid loosening to break apart for hetero-associations.

FOXP1 forms hetero-oligomers with FOXP2 via the same interface involved in homo-oligomerisation. In all combinations, hetero-association of FOXP1 LZ-FHD with all the

FOXP2 variants shows significantly better affinity than all the FOXP1 LZ-End associations. This may be an indication that the C-terminal acid rich tail is not important for hetero-associations despite being important for quaternary stability. Although it is likely that hetero-association occur in the same manner as homo-association, the current study indicates that hetero-association between FOXP1 and FOXP2 are quite weak, both proteins preferring their homo-oligomeric conformations. This, therefore, suggests that FOXP1-FOXP2 hetero-associations may compete with either FOXP1 or FOXP2 homodimerisation and/or formation of higher order oligomers at low concentrations of either, and the dynamic balance of these discrete forms of FOXP1 and FOXP2 complexes may directly affect its repressor activity. Although this might be the case, increases in concentration of one protein may result in increased homodimer affinity. If this is the case, then multiple events that may occur as concentration is increased in FA, could result in a false increase in dissociation constant, due to dissociation of the hetero-mer and reassembly of the homo-mers.

FOXP proteins associate with DNA as dimers in cells (Wang *et al.*, 2003; Li, Weidenfeld and Morrisey, 2004; Sin, Li and Crawford, 2014). However, it has been shown that FOXP1 and FOXP2 exist predominantly as higher order oligomers in isolation. Indeed, this behaviour is seen to be important in the FOXP as FOXP3 has also been shown to form tetramers in solution (Bin Li *et al.*, 2007). Electrophoretic mobility shift assay (EMSA) conducted for both FOXP1 and FOXP2 variants indicates protein-DNA complex formation of varying degrees for all variants. More than one band for each variant and a considerable smearing of the bands was observed. Multiple bands suggest that protein-DNA complexes of varying sizes are formed. The smearing observed indicates a constant dissociation of the FOXP1/2-DNA complexes. If the full-length protein binds to DNA as a dimer, then it is therefore likely that the presence of DNA could induce dissociation of the higher order oligomers observed in this study. Assessment of the DNA binding of both FOXP1 and FOXP2 variant DNA binding affinities using fluorescence anisotropy shows similar binding affinities (K_D) with that of the isolated FOXP1 and FOXP2 FHD (Chu *et al.*, 2011; Morris and Fanucchi, 2016). This, therefore, suggests that the conformational stability and state of both FOXP1 and FOXP2 variants resembles that of the functional protein. In contrast, DNA binding is greatly decrease upon hetero-association compared to homo-association (10-50 magnitudes weaker). The decrease in affinity for DNA could be due to constant dissociation that seem to occur during hetero-association, since homo-oligomerisation is the preferred conformation. This constant association and dissociation can provide an elegant means by which FOXP proteins can

regulate gene transcription. It can be argued that this indicates that regulatory activity of the FOXP through either homo- or hetero-association is controlled in a signalling-dependent manner, depending on the stimulus.

Figure 45A shows a model of FOXP1 and FOXP2 interaction based on the model proposed for FOXP1 and FOXP3 interactions by Song and colleagues (Song *et al.*, 2012). Although in both models heterodimerisation occurs through the leucine zippers of both proteins, it does not take into account the association that occurs through the FHD dimerisation interface. There is not much evidence to support hetero-association through the FHD; however, it has been shown in a crystal structures and through other studies that FOXP1, FOXP2 and FOXP3 form domain swapped dimers (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011; Chen *et al.*, 2015; Medina *et al.*, 2016), therefore this has to be taken into account. In the current study, it has been shown that FOXP1/FOXP2 associates into higher order oligomers. Therefore, its proposed that FOXP1/FOXP2 associations occur mainly through the leucine zipper while the FHD retains domain swapping (Figure 45B). Nonetheless, FOXP1 has also been shown, in this study, to associate with the FOXP2 FHD-ART variant, suggesting that the FHD has a contribution to hetero-association.

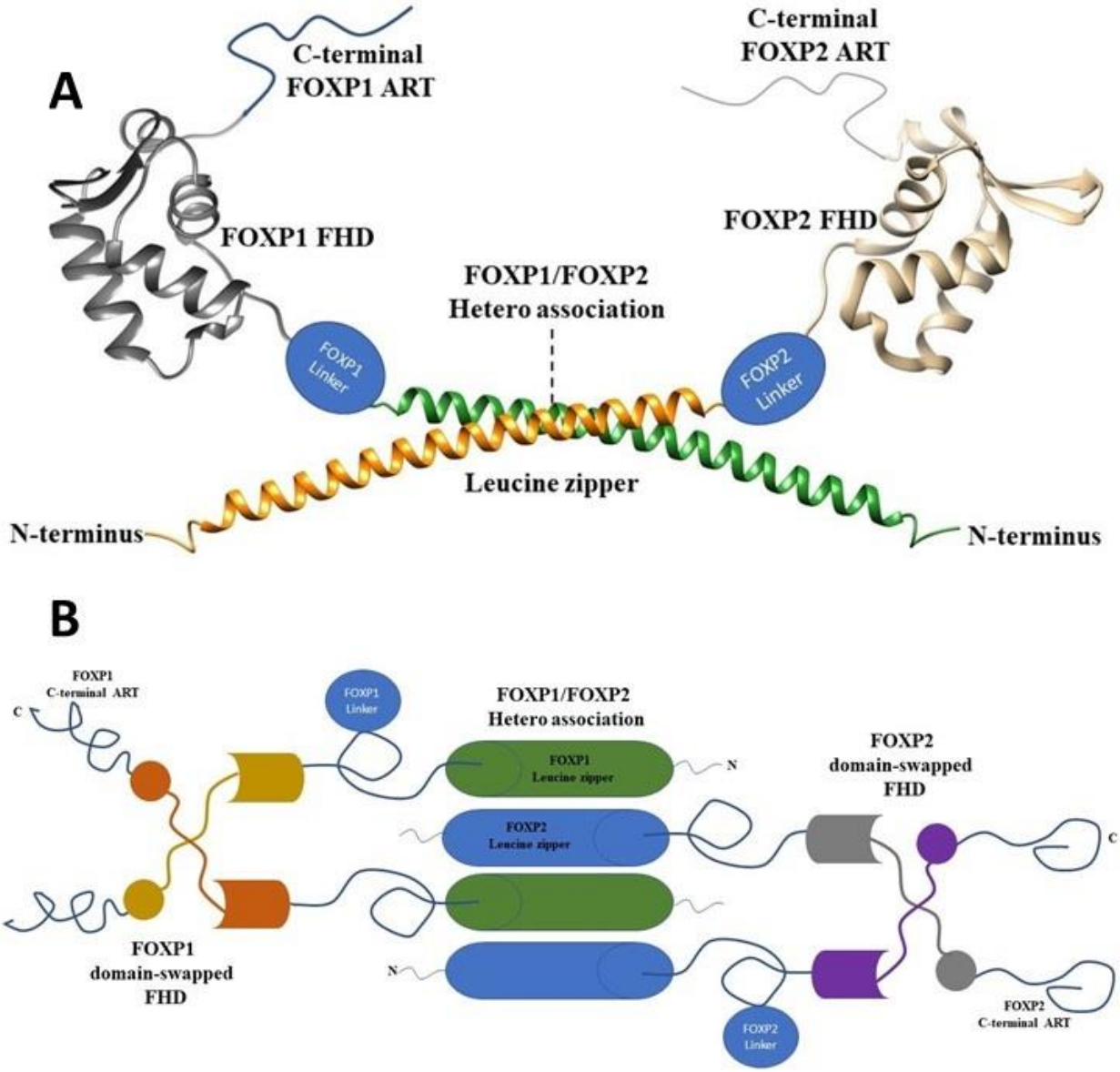


Figure 41: Models of the FOXP1/FOXP2 hetero-associated protein complex. (A) The model was constructed by using the crystal structure of the mFOXP3 coiled-coil (PDB:4i1l), the NMR structure of FOXP1 FHD (PDB:2kiu) and the crystal structure of FOXP2 FHD (PDB:2a07) to emulate that hypothesised for FOXP3 supramolecular complex (Song *et al.*, 2012). Molecular graphics images were produced using the UCSF Chimera package (Pettersen *et al.*, 2004). (B) Models of the FOXP1/FOXP2 hetero-associated protein complex, with domain-swapped FHD. FOXP leucine zipper domain is capable of forming a protein tetramer (Song *et al.*, 2012), a supramolecular complex capable of regulating transcription.

CONCLUSION

Understanding the structure of proteins is crucial for determining their functions and how they perform such functions. Therefore, to be able to study the function of FOXP proteins, it is important to first determine the structural behaviour and DNA binding characteristics of these transcription factors. The main aim of this research was to characterise the structures of FOXP1 and FOXP2 focusing on their dimerisation interfaces, to identify the most crucial structural features and elements. This was done to facilitate the investigation into the mechanisms of both homo- and hetero-typic protein-protein interactions of FOXP1 and FOXP2. DNA binding in FOXP1 and FOXP2 variants has a potential to initiate several events, including a "disorder-to-order" transition of intrinsically disordered regions, oligomer dissociation to form functional dimer and mediate dynamic functional behaviour. Like FOXP3, FOXP1 and FOXP2 exist as a component of supramolecular complex, important for stability and specificity during complex formation. This work shows, for the first time, that hetero-oligomerisation is an important event that occurs at low protein concentration, with relatively low affinity, but which may regulate homo-association and DNA binding.

Dysfunction of these proteins leads to disease including, but not limited to, various cancers, an immune disorder, and a rare genetic speech disorder. Given the increasing prevalence of the FOXP in these diseases, it is critical that the mechanism of dimerisation and its role in DNA binding and hence transcriptional regulation are understood if novel therapeutic strategies are to be developed. This study provides insight into the complexities of transcriptional regulation via FOXP proteins.

CHAPTER 5

BIBLIOGRAPHY

Abate, C., Luk, D. and Curran, T. (1991) 'Transcriptional regulation by Fos and Jun in vitro: interaction among multiple activator and regulatory domains.', *Molecular and Cellular Biology*, 11(7), pp. 3624–32. doi: 10.1128/MCB.11.7.3624.Updated.

Abate, C., Luk, D., Gagne, E., Roeder, R. G. and Curran, T. (1990) 'Fos and jun cooperate in transcriptional regulation via heterologous activation domains.', *Molecular and Cellular Biology*, 10(10), pp. 5532–5. doi: 10.1128/MCB.10.10.5532.

Abate, C., Luk, D., Gentz, R., Rauscher, F. J. and Curran, T. (1990) 'Expression and purification of the leucine zipper and DNA-binding domains of Fos and Jun: both Fos and Jun contact DNA directly.', *Proceedings of the National Academy of Sciences of the United States of America*, 87(3), pp. 1032–6. doi: 10.1073/pnas.87.3.1032.

Acharya, M., Huang, L., Fleisch, V. C., Allison, W. T. and Walter, M. A. (2011) 'A complex regulatory network of transcription factors critical for ocular development and disease', *Human Molecular Genetics*, 20(8), pp. 1610–1624. doi: 10.1093/hmg/ddr038.

Adcock, I. M. and Caramori, G. (2009) 'Transcription Factors', in *Asthma and COPD*, pp. 373–380. doi: 10.1016/B978-0-12-374001-4.00031-6.

Al-khayyat, M. Z. S. (2015) 'In silico structural analysis of quorum sensing genes in Vibrio fischeri', *Molecular Biology Research Communications*, 4(3), pp. 115–124.

Alber, T. (1992) 'Structure of the leucine zipper', *Current Opinion in Genetics and Development*, 2(2), pp. 205–210. doi: 10.1016/S0959-437X(05)80275-8.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*, 25(17), pp. 3389–3402. doi: 10.1093/nar/25.17.3389.

Ang, S. L. and Rossant, J. (1994) 'HNF-3β is essential for node and notochord formation in mouse development', *Cell*, 78(4), pp. 561–574. doi: 10.1016/0092-8674(94)90522-3.

Aoki, R., Shoshkes-Carmel, M., Gao, N., Shin, S., May, C. L., Golson, M. L., Zahm, A. M., Ray, M., Wiser, C. L., Wright, C. V. E. and Kaestner, K. H. (2016) 'Foxl1-Expressing Mesenchymal Cells Constitute the Intestinal Stem Cell Niche', *Cellular and Molecular Gastroenterology and Hepatology*. Elsevier Inc, 2(2), pp. 175–188. doi: 10.1016/j.jcmgh.2015.12.004.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. and Iyer, L. M. (2005) 'The many faces of the helix-turn-helix domain: Transcription regulation and beyond', *FEMS Microbiology Reviews*, 29(2), pp. 231–262. doi: 10.1016/j.femsre.2004.12.008.

Arden, K. C. (2007) 'FoxOs in Tumor Suppression and Stem Cell Maintenance', *Cell*, 128, pp. 235–237. doi: 10.1016/j.cell.2007.01.009.

Bacon, C. and Rappold, G. A. (2012) 'The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders', *Human Genetics*, 131(11), pp. 1687–1698. doi: 10.1007/s00439-012-1193-z.

Baker, P., Hillis, C., Carere, J. and Seah, S. Y. K. (2012) 'Protein − Protein Interactions and Substrate Channeling in Orthologous and Chimeric Aldolase − Dehydrogenase Complexes', *Biochemistry*.

Bandukwala, H. S., Wu, Y., Feuerer, M., Chen, Y., Barboza, B., Ghosh, S., Stroud, J. C., Benoist, C., Mathis, D., Rao, A. and Chen, L. (2011) 'Structure of a Domain-Swapped FOXP3 Dimer on DNA and Its Function in Regulatory T Cells', *Immunity*, 34(4), pp. 479–491. doi: 10.1016/j.immuni.2011.02.017.

Barnes, P. J. (2006) 'Transcription factors in airway diseases', *Laboratory Investigation*, 86(9), pp. 867–872. doi: 10.1038/labinvest.3700456.

Barnes, P. J., Gajiwala, K. S. and Burley, S. K. (2000) 'Winged helix proteins', *Current Opinion in Structural Biology*, (10), pp. 110–116. doi: 10.1038/labinvest.3700456.

Bashor, C. J., Horwitz, A. a, Peisajovich, S. G. and Lim, W. a (2010) 'Rewiring Cells: Synthetic biology as a tool to interrogate the organizational principles of living systems', *Annual Review Biophysics*, 39, pp. 515–537. doi: 10.1146/annurev.biophys.050708.133652.Rewiring.

Beckmann, H., Su, L. K. and Kadesch, T. (1990) 'TFE3: A helix-loop-helix protein that activates transcription through the immunoglobulin enhancer uE3 motif', *Genes and Development*, 4(2), pp. 167–179. doi: 10.1101/gad.4.2.167.

Benayoun, B. A., Caburet, S. and Veitia, R. A. (2011) 'Forkhead transcription factors: Key players in health and disease', *Trends in Genetics*. Elsevier Ltd, 27(6), pp. 224–232. doi: 10.1016/j.tig.2011.03.003.

Bennett, C. L., Christie, J., Ramsdell, F., Brunkow, M. E., Ferguson, P. J., Whitesell, L., Kelly, T. E., Saulsbury, F. T., Chance, P. F. and Ochs, H. D. (2001) 'The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3', *Nature Genetics*. Nature America Inc., 27, p. 20. Available at: https://doi.org/10.1038/83713.

Bennett, M. J., Schlunegger, M. P. and Eisenberg, D. (1995) '3D domain swapping: A mechanism for oligomer assembly', *Protein Science*, 4(12), pp. 2455–2468. doi: 10.1002/pro.5560041202.

Berg, J. M. (1990) 'Zinc fingers and other metal-binding domains', *Journal of Biological Chemistry*, 265(12), pp. 6513–6516.

Berg, O. G., Winter, R. B. and von Hippel, P. H. (1981) 'Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory.', *Biochemistry*, 20(24), pp. 6929–6948. doi: 10.1021/bi00527a028.

Bhattacharyya, R. P., Reményi, A., Yeh, B. J. and Lim, W. A. (2006) 'Domains, Motifs, and Scaffolds: The Role of Modular Interactions in the Evolution and Wiring of Cell Signaling Circuits', *Annual Review of Biochemistry*, 75(1), pp. 655–680. doi: 10.1146/annurev.biochem.75.103004.142710.

Birdsell, D. C. and Cota-Robles, E. H. (1967) 'Production and ultrastructure of lysozyme and ethylenediaminetetraacetate-lysozyme spheroplasts of Escherichia coli.', *Journal of Bacteriology*, 93(1), pp. 427–437.

Blane, A. and Fanucchi, S. (2015) 'Effect of pH on the Structure and DNA Binding of the FOXP2 Forkhead Domain', *Biochemistry*, 54(25), pp. 4001–4007. doi: 10.1021/acs.biochem.5b00155.

Block, P., Paern, J., Hullermeier, E., Sanschagrin, P., Sotriffer, C. A. and Klebe, G. (2006) 'Physicochemical Descriptors to Discriminate Protein–Protein Interactions in Permanent and Transient Complexes Selected by Means of Machine Learning Algorithms Peter', *Proteins: Structure, Function and Bioinformatics*, 65, pp. 607–622. doi: 10.1002/prot.

Bogan, A. A. and Thorn, K. S. (1998) ‘Anatomy of hot spots in protein interfaces’, *Journal of Molecular Biology*, 280(1), pp. 1–9. doi: 10.1006/jmbi.1998.1843.

Bornhorst, J. A. and Falke, J. J. (2000) ‘Purification of Proteins Using Polyhistidine Affinity Tags’, *Methods in Enzymology*, 326, pp. 245–254.

Bowers, J. M. and Konopka, G. (2012) ‘The role of the FOXP family of transcription factors in ASD’, *Disease Markers*, 33(5), pp. 251–260. doi: 10.3233/DMA-2012-0919.

Brennan, R. G. (1993) ‘The winged-helix DNA-binding motif: Another helix-turn-helix takeoff’, *Cell*, 74(5), pp. 773–776. doi: 10.1016/0092-8674(93)90456-Z.

Brennan, R. G. and Matthews, B. W. (1989) ‘The Helix-Turn-Helix DNA Binding Motif’, *Journal of Biological Chemistry*, 264(4), pp. 22–25. Available at: http://www.jbc.org/content/264/4.

Bresnick, E. H. and Felsenfeld, G. (1994) ‘The Leucine Zipper Is Necessary for Stabilizing a Dimer of the Helix-Loop-Helix Transcription Factor USF But Not for Maintenance of an Elongated Conformation *’, *Journal of Biological Chemistry*, 269(33), pp. 21110–21116.

Brunkow, M. E., Jeffery, E. W., Hjerrild, K. A., Paeper, B., Clark, L. B., Yasayko, S., Wilkinson, J. E., Galas, D., Ziegler, S. F. and Ramsdell, F. (2002) ‘Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse’, *Nature Genetics*, 27, pp. 68–73. Available at: papers3://publication/uuid/1759DE47-8712-4DB9-94B1-7E65BC931065.

Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K. and Jones, D. T. (2013) ‘Scalable web services for the PSIPRED Protein Analysis Workbench.’, *Nucleic Acids Research*, 41(Web Server issue), pp. 349–357. doi: 10.1093/nar/gkt381.

Burgering, B. M. T. and Kops, G. J. P. L. (2002) ‘Cell cycle and death control : long live Forkheads’, *TRENDS in Biochemical Sciences*, 27(7), pp. 352–360.

Cardinale, D., Salo-Ahen, O., Ferrari, S., Ponterini, G., Cruciani, G., Carosati, E., Tochowicz, A., Mangani, S., Wade, R. and Costi, M. (2010) ‘Homodimeric Enzymes as Drug Targets’, *Current Medicinal Chemistry*, 17(9), pp. 826–846. doi: 10.2174/092986710790712156.

Carter, M. E. and Brunet, A. (2007) ‘FOXO transcription factors’, *Current Biology*, 17(4), pp. 113–114.

Chakrabarti, P. and Janin, J. (2002) 'Dissecting protein-protein recognition sites', *Proteins: Structure, Function and Genetics*, 47(3), pp. 334–343. doi: 10.1002/prot.10085.

Chatila, T. A., Blaeser, F., Ho, N., Lederman, H. M., Voulgaropoulos, C., Helms, C. and Bowcock, A. M. (2000) 'JM2 , encoding a fork head – related protein , is mutated in X-linked autoimmunity – allergic disregulation syndrome Online first', *Journal of Clinical Investigation*, 106, pp. 75–81.

Chen, M. T., Sun, H.-F., Li, L.-D., Zhao, Y., Yang, L.-P., Gao, S.-P. and Jin, W. (2018) 'Downregulation of FOXP2 promotes breast cancer migration and invasion through TGF β / SMAD signaling pathway', *Oncology Letters*, 15, pp. 8582–8588. doi: 10.3892/ol.2018.8402.

Chen, Y. and Barkley, M. D. (1998) 'Toward understanding tryptophan fluorescence in proteins', *Biochemistry*, 37(28), pp. 9976–9982. doi: 10.1021/bi980274n.

Chen, Y., Chen, C., Zhang, Z., Liu, C. C., Johnson, M. E., Espinoza, C. A., Edsall, L. E., Ren, B., Zhou, X. J., Grant, S. F. A., Wells, A. D. and Chen, L. (2015) 'DNA binding by FOXP3 domain-swapped dimer suggests mechanisms of long-range chromosomal interactions', *Nucleic Acids Research*, 43(2), pp. 1268–1282. doi: 10.1093/nar/gku1373.

Chokas, A. L., Trivedi, C. M., Lu, M. M., Tucker, P. W., Li, S., Epstein, J. A. and Morrisey, E. E. (2010) 'Foxp1/2/4-NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of interleukin', *Journal of Biological Chemistry*, 285(17), pp. 13304–13313. doi: 10.1074/jbc.M109.088468.

Chu, Y.-P., Chang, C.-H., Shiu, J.-H., Chang, Y.-T., Chen, C.-Y. and Chuang, W.-J. (2011) 'Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: Insight into its domain swapping and DNA binding', *Protein Science*, 20(5), pp. 908–924. doi: 10.1002/pro.626.

Clark, K. L., Halay, E. D., Lai, E. and Burley, S. K. (1993) 'Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5', *Nature*, 363, pp. 412–420.

Conte, L. Lo, Chothia, C. and Janin, J. (1999) 'The atomic structure of protein-protein recognition sites11Edited by A. R. Fersht', *Journal of Molecular Biology*, 285(5), pp. 2177–2198. doi: 10.1006/jmbi.1998.2439.

Cooper, G. M. (2000) 'The Cell , 2nd edition', in *The Cell: A Molecular Approach*. Second edi. Suderland (MA): Sinauer Associates, pp. 1–14.

Corominas, R., Yang, X., Lin, G. N., Kang, S., Shen, Y., Ghamsari, L., Broly, M., Rodriguez, M., Tam, S., Trigg, S. A., Fan, C., Yi, S., Tasan, M., Lemmens, I., Kuang, X., Zhao, N., Malhotra, D., Michaelson, J. J., Vacic, V., Calderwood, M. A., Roth, F. P., Tavernier, J., Horvath, S., Salehi-Ashtiani, K., Korkin, D., Sebat, J., Hill, D. E., Hao, T., Vidal, M. and Iakoucheva, L. M. (2014) 'Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism', *Nature Communications*. Nature Publishing Group, 5. doi: 10.1038/ncomms4650.

Corrêa, D. H. and Ramos, C. H. I. (2009) 'The use of circular dichroism spectroscopy to study protein folding, form and function', *African Journal of Biochemistry Research*, 3(5), pp. 164–173. doi: not available.

Craig, T. J., Ciufo, L. F. and Morgan, A. (2004) 'A protein – protein binding assay using coated microtitre plates : increased throughput , reproducibility and speed compared to bead-based assays', *Journal of Biochemical and Biophysical Methods*, 60, pp. 49–60. doi: 10.1016/j.jbbm.2004.04.015.

Crane-Robinson, C., Dragan, A. I. and Privalov, P. L. (2006) 'The extended arms of DNA-binding domains: a tale of tails', *Trends in Biochemical Sciences*, 31(10), pp. 547–552. doi: 10.1016/j.tibs.2006.08.006.

Cui, Q. and Karplus, M. (2008) 'Allostery and cooperativity revisited', *Protein Science*, 17, pp. 1295–1307. doi: 10.1110/ps.03259908.coined.

Cuiffo, B. G., Campagne, A., Bell, G. W., Lembo, A., Orso, F., Lien, E. C., Bhasin, M. K., Raimo, M., Hanson, S. E., Marusyk, A., El-Ashry, D., Hematti, P., Polyak, K., Mechta-Grigoriou, F., Mariani, O., Volinia, S., Vincent-Salomon, A., Taverna, D. and Karnoub, A. E. (2014) 'MSC-regulated microRNAs converge on the transcription factor FOXP2 and promote breast cancer metastasis', *Cell Stem Cell*. Elsevier Inc., 15(6), pp. 762–774. doi: 10.1016/j.stem.2014.10.001.

Cuiffo, B. G. and Karnoub, A. E. (2016) 'Silencing FOXP2 in breast cancer cells promotes cancer stem cell traits and metastasis', *Molecular & Cellular Oncology*, 3(3), p. e1019022. doi: 10.1080/23723556.2015.1019022.

Delano, W. L. (2002) 'Unraveling hot spots in binding interfaces DeLano 15', *Current Opinion in Structural Biology*, pp. 14–20. Available at: https://proxy.library.upenn.edu:14097/S0959440X0200283X/1-s2.0-S0959440X0200283X-

main.pdf?_tid=ac695fb6-cbab-11e7-81cd-00000aacb362&acdnat=1510932565_463babb96da8c2d6fa08f27048f5382d.

Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., Hawkins, M. E. and Brochon, J. C. (2001) 'DNA binding induces dissociation of the multimeric form of HIV-1 integrase: a time-resolved fluorescence anisotropy study.', *Proceedings of the National Academy of Sciences of the United States of America*, 98(18), pp. 10090–10095. doi: 10.1073/pnas.181024498.

Deriziotis, P., Graham, S. A., Estruch, S. B. and Fisher, S. E. (2014a) 'Investigating Protein-protein Interactions in Live Cells Using Bioluminescence Resonance Energy Transfer', *Journal of Visualized Experiments*, (87). doi: 10.3791/51438.

Deriziotis, P., Graham, S. A., Estruch, S. B. and Fisher, S. E. (2014b) 'Investigating Protein-protein Interactions in Live Cells Using Bioluminescence Resonance Energy Transfer', *Journal of Visualized Experiments*, (87), pp. 1–15. doi: 10.3791/51438.

Dragan, A. I., Read, C. M., Makeyeva, E. N., Milgotina, E. I., Churchill, M. E. A., Crane-Robinson, C. and Privalov, P. L. (2004) 'DNA binding and bending by HMG boxes: Energetic determinants of specificity', *Journal of Molecular Biology*, 343(2), pp. 371–393. doi: 10.1016/j.jmb.2004.08.035.

Dubey, V. K. and Jagannadham, M. V. (2003) 'Differences in the Unfolding of Procerain Induced by pH, Guanidine Hydrochloride, Urea, and Temperature', *Biochemistry*, 42(42), pp. 12287–12297. doi: 10.1021/bi035047m.

Duy, C. and Fitter, J. (2005) 'Thermostability of irreversible unfolding ??-amylases analyzed by unfolding kinetics', *Journal of Biological Chemistry*, 280(45), pp. 37360–37365. doi: 10.1074/jbc.M507530200.

Duy, C. and Fitter, J. (2006) 'How aggregation and conformational scrambling of unfolded states govern fluorescence emission spectra', *Biophysical Journal*. Elsevier, 90(10), pp. 3704–3711. doi: 10.1529/biophysj.105.078980.

Emmott, E. and Goodfellow, I. (2014) 'Identification of Protein Interaction Partners in Mammalian Cells Using SILAC-immunoprecipitation Quantitative Proteomics', *Journal of Visualized Experiments*, (89), pp. 1–8. doi: 10.3791/51656.

Estruch, S. B., Graham, S. A., Chinnappa, S. M., Deriziotis, P. and Fisher, S. E. (2016)

‘Functional characterization of rare FOXP2 variants in neurodevelopmental disorder’, *Journal of Neurodevelopmental Disorders*. Journal of Neurodevelopmental Disorders, 8(1), pp. 1–19. doi: 10.1186/s11689-016-9177-2.

Estruch, S. B., Graham, S. A., Quevedo, M., Vino, A., Dekkers, D. H. W., Deriziotis, P., Sollis, E., Demmers, J., Poot, R. A. and Fisher, S. E. (2018) ‘Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders’, *Human Molecular Genetics*, 27(7), pp. 1212–1227. doi: 10.1093/hmg/ddy035.

Eubel, H., Braun, H. and Millar, A. H. (2005) ‘Blue-native PAGE in plants : a tool in analysis of protein-protein interactions’, *Plant Methods*, 1(11). doi: 10.1186/1746-4811-1-11.

Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T. and Rhodes, D. (1993) ‘The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition’, *Letters to Nature*, 366(2 December 1993), pp. 461–464.

Ferland, R. J., Cherry, T. J., Preware, P. O., Morrisey, E. E. and Walsh, C. A. (2003) ‘Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain’, *Journal of Comparative Neurology*, 460(2), pp. 266–279. doi: 10.1002/cne.10654.

Ferré-D’Amaré, A. R. and Burley, S. K. (1994) ‘Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies’, *Structure*, 2(5), pp. 357–359. doi: 10.1016/S0969-2126(00)00037-X.

Fisher, S. E. and Scharff, C. (2009) ‘FOXP2 as a molecular window into speech and language’, *Trends in Genetics*, 25(4), pp. 166–177. doi: 10.1016/j.tig.2009.03.002.

Fitter, J. and Haber-Pohlmeier, S. (2004) ‘Structural stability and unfolding properties of thermostable bacterial ??-amylases: A comparative study of homologous enzymes’, *Biochemistry*, 43(30), pp. 9589–9599. doi: 10.1021/bi0493362.

Fong, W. L., Kuo, H. Y., Wu, H. L., Chen, S. Y. and Liu, F. C. (2018) *Differential and Overlapping Pattern of Foxp1 and Foxp2 Expression in the Striatum of Adult Mouse Brain*, *Neuroscience*. IBRO. doi: 10.1016/j.neuroscience.2018.07.017.

Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M. and Wright, P. E. (1997) ‘Domain packing and dynamics in the DNA complex of the N-terminal zinc finger of TFIIIA’, *Nature Structual Biology*, 4(8), pp. 605–608. doi: 10.1007/978-3-319-52548-8.

Freire, E., Mayorga, O. L. and Straume, M. (1990) ‘Isothermal Titration’, *Analytical*

Chemistry, 62(18), pp. 950A-959A. doi: 10.1021/ac00217a002.

Fried, M. G. (1989) 'Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay', *Electrophoresis*, 10(5–6), pp. 366–376. doi: 10.1002/elps.1150100515.

Fried, M. G. and Crothers, D. M. (1981) 'Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis', *Nucleic Acids Research*, 9(23), pp. 6505–6525.

Funnell, A. P. W. and Crossley, M. (2012) 'Homo- and heterodimerization in transcriptional regulation.', *Advances in Experimental Medicine and Biology*, 747, pp. 105–121. doi: 10.1007/978-1-4614-3229-6_7.

Gao, F., Foat, B. C. and Bussemaker, H. J. (2004) 'Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data', *BMC Bioinformatics*, 5(31), pp. 1–10.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', *The Proteomics Protocols Handbook*, pp. 571–607. doi: 10.1385/1592598900.

Georges, A. B., Benayoun, B. A., Caburet, S. and Veitia, R. A. (2010) 'Generic binding sites , generic DNA-binding domains : where does specific promoter recognition come from?', *The FASEB Journal*, 24(2), pp. 346–356. doi: 10.1096/fj.09-142117.

Ghisaidoobe, A. B. T. and Chung, S. J. (2014) 'Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on förster resonance energy transfer techniques', *International Journal of Molecular Sciences*, 15(12), pp. 22518–22538. doi: 10.3390/ijms151222518.

Giannakou, M. E. and Partridge, L. (2004) 'The interaction between FOXO and SIRT1 : tipping the balance towards survival', *TRENDS in Cell Biology*, 14(8), pp. 408–412. doi: 10.1016/j.tcb.2004.07.001.

Givaty, O. and Levy, Y. (2009) 'Protein Sliding along DNA: Dynamics and Structural Characterization', *Journal of Molecular Biology*. Elsevier Ltd, 385(4), pp. 1087–1097. doi: 10.1016/j.jmb.2008.11.016.

Goldburg, W. I. (1999) 'Dynamic light scattering', *American Journal of Physics*, 67(12), pp. 1152–1160. doi: 10.1119/1.19101.

Golson, M. L. and Kaestner, K. H. (2016) 'Fox transcription factors: from development to disease', *Development*, 143(24), pp. 4558–4570. doi: 10.1242/dev.112672.

Gradinaru, C. C., Marushchak, D. O. and Krull, U. J. (2010) 'Fluorescence anisotropy : from single molecules to live cells', *Analyst*, 135, pp. 452–459. doi: 10.1039/b920242k.

Gray, P. A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D., Tsung, E. F., Cai, Z., Alberta, J. A., Cheng, L., Liu, Y., Stenman, J. M., Valerius, M. T., Billings, N., Kim, H. A., Greenberg, M. E., Mcmahon, A. P., Rowitch, D. H., Stiles, C. D. and Ma, Q. (2004) 'Mouse Brain Organization Revealed Through Direct Genome-Scale TF Expression Analysis', *Science Reports*, 306, pp. 2255–2258.

Guharoy, M. and Chakrabarti, P. (2005) 'Conservation and relative importance of residues across protein–protein interfaces', *PNAS*, 102(43), pp. 15447–15452. doi: 10.1073/pnas.0505425102.

Hagnauer, G. L. (1981) 'Size Exclusion Chromatography.', *Analytical Chemistry*, 54(5), pp. 265–276. doi: 10.1021/ac00036a023.

Hakoshima, T. (2005) 'Leucine Zippers', *Encyclopedia of Life Sciences*, pp. 1–5. doi: 10.1038/npg.els.0005049.

Hamdan, F. F., Daoud, H., Rochefort, D., Piton, A., Gauthier, J., Langlois, M., Foomani, G., Dobrzeniecka, S., Krebs, M. O., Joober, R., Lafrenire, R. G., Lacaille, J. C., Mottron, L., Drapeau, P., Beauchamp, M. H., Phillips, M. S., Fombonne, E., Rouleau, G. A. and Michaud, J. L. (2010) 'De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment', *American Journal of Human Genetics*, 87(5), pp. 671–678. doi: 10.1016/j.ajhg.2010.09.017.

Hashikawa, N., Yamamoto, N. and Sakurai, H. (2007) 'Different mechanisms are involved in the transcriptional activation by yeast heat shock transcription factor through two different types of heat shock elements', *Journal of Biological Chemistry*, 282(14), pp. 10333–10340. doi: 10.1074/jbc.M609708200.

Häußermann, K., Young, G., Kukura, P. and Dietz, H. (2019) 'Dissecting FOXP2 oligomerization and DNA binding', *Angewandte Chemie - International Edition*, 131, pp. 7744–7749. doi: 10.1002/anie.201901734.

Hawkins, A. R. and Lamb, H. K. (1995) 'The Molecular Biology of Multidomain Proteins

Selected Examples', *European Journal of Biochemistry*, 232(1), pp. 7–18. doi: 10.1111/j.1432-1033.1995.tb20775.x.

Hellman, L. M. and Fried, M. G. (2007a) 'Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein- Nucleic Acid Interactions', *Nature Protocols*, 2(8), pp. 1849–1861. doi: 10.1038/nprot.2007.249.Electrophoretic.

Hellman, L. M. and Fried, M. G. (2007b) 'Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions', *Nature Protocols*, 2(8), pp. 1849–1861. doi: 10.1038/nprot.2007.249.

Hinde, E., Pand, E., Yang, Z., Ng, I. H. W., Jans, D. A., Bogoyevitch, M. A., Gratton, E. and Gaus, K. (2016) 'Quantifying the dynamics of the oligomeric transcription factor STAT3 by pair correlation of molecular brightness', *Nature Communications*, 7. doi: 10.1038/ncomms11047.

Hirota, N., Mizuno, K. and Goto, Y. (1998) 'Group additive contributions to the alcohol-induced α-helix formation of melittin: Implication for the mechanism of the alcohol effects on proteins', *Journal of Molecular Biology*, 275(2), pp. 365–378. doi: 10.1006/jmbi.1997.1468.

Hjalt, T. (2004) 'Basic helix-loop-helix proteins expressed during early embryonic organogenesis', *International Review of Cytology*, 236, pp. 251–280. doi: 10.1016/S0074-7696(04)36006-7.

Holdgate, G. (2010) 'Ligand-Macromolecular Interactions in Drug Discovery', in *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols, Methods in Molecular Biology*, pp. 101–133. doi: 10.1007/978-1-60761-244-5.

Hong, P., Koza, S., Bouvier, E. S. P. and Corporation, W. (2012) 'A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates', *Journal of Liquid Chromatography & Related Technologies*, 32, pp. 2923–2950. doi: 10.1080/10826076.2012.743724.

Horn, D., Kapeller, J., Rivera-Brugués, N., Moog, U., Lorenz-Depiereux, B., Eck, S., Hempel, M., Wagenstaller, J., Gawthrope, A., Monaco, A. P., Bonin, M., Riess, O., Wohlleber, E., Illig, T., Bezzina, C. R., Franke, A., Spranger, S., Villavicencio-Lorini, P., Seifert, W., Rosenfeld, J., Klopocki, E., Rappold, G. A. and Strom, T. M. (2010) 'Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language

deficits', *Human Mutation*, 31(11). doi: 10.1002/humu.21362.

Van Der Horst, A. and Burgering, B. M. T. (2007) 'Stressing the role of FoxO proteins in lifespan and disease', *Nature Reviews Molecular Cell Biology*, 8(6), pp. 440–450. doi: 10.1038/nrm2190.

Huang, X., Holden, H. M. and Raushel, F. M. (2001) 'Channeling of substrates and intermediates in enzumer-catalyszed reactions', *Annual Review of Biochemistry*, 70, pp. 149–180.

Hudson, J. M. and Fried, M. G. (1990) 'Co-operative Interactions Between the Catabolite Gene Activator Protein and the Zac Repressor at the Lactose Promoter', *Journal of Molecular Biology*, 214, pp. 381–396.

Huffman, J. L. and Brennan, R. G. (2001) 'Prokaryotic transcription regulators', *Current Opinion in Structural Biology*, pp. 98–106. Available at: https://ac.els-cdn.com/S0959440X02002956/1-s2.0-S0959440X02002956-main.pdf?_tid=195744fc-3286-4878-993a-3e3a1fc78f87&acdnat=1524963742_98c8e58e00e437f3a7b1d0b365cc79e6.

Hunter, M. I. S., Muir, D. D. M. and Thirkell, D. (1973) 'Effect of lysozyme treatment on cell wall ultrastructure in Sarcina flava', *Journal of Bacteriology*, 116(1), pp. 483–487.

Jameson, D. M. and Sawyer, W. H. (1995) 'Fluorescence anisotropy applied to biomolecular interactions', *Methods in Enzymology*, 246(C), pp. 283–300. doi: 10.1016/0076-6879(95)46014-4.

Janin, J., Bahadur, R. P. and Chakrabarti, P. (2008) 'Protein-protein interaction and quaternary structure', *Quarterly Reviews of Biophysics*, 41(2), pp. 133–180. doi: 10.1017/S0033583508004708.

Janin, J., Miller, S. and Chothia, C. (1988) 'Surface, subunit interfaces and interior of oligomeric proteins', *Journal of Molecular Biology*, 204(1), pp. 155–164. doi: 10.1016/0022-2836(88)90606-7.

Janin, J., Rodier, F., Chakrabarti, P. and Bahadur, R. P. (2006) 'Macromolecular recognition in the Protein Data Bank', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 63(1), pp. 1–8. doi: 10.1107/S090744490603575X.

Jepsen, K., Gleiberman, A. S., Shi, C., Simon, D. I. and Rosenfeld, M. G. (2008) 'Cooperative regulation in development by SMRT and FOXP1 service Cooperative regulation in

development by SMRT and FOXP1', *Genes & Development*, pp. 740–745. doi: 10.1101/gad.1637108.

Jia, W. Z., Yu, T., An, Q., Yang, H., Zhang, Z., Liu, X. and Xiao, G. (2016) 'MicroRNA-190 regulates FOXP2 genes in human gastric cancer', *OncoTargets and Therapy*, 9, pp. 3643–3651. doi: 10.2147/OTT.S103682.

Jones, D. T. (1999a) 'GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences', *Journal of Molecular Biology*, 287(4), pp. 797–815. doi: 10.1006/jmbi.1999.2583.

Jones, D. T. (1999b) 'Protein secondary structure prediction based on position-specific scoring matrices', *Journal of Molecular Biology*, 292(2), pp. 195–202. doi: 10.1006/jmbi.1999.3091.

Jones, D. T., Taylor, W. R. and Thornton, J. M. (1994) 'A Model Recognition Approach to the Prediction of All-Helical Membrane Protein Structure and Topology', *Biochemistry*, 33(10), pp. 3038–3049. doi: 10.1021/bi00176a037.

Jones, S. (2004) 'An overview of the basic helix-loop-helix proteins', *Genome Biology*, 5(6), p. 226. doi: 10.1186/gb-2004-5-6-226.

Jones, S. and Thornton, J. M. (1996) 'Principles of protein-protein interactions', *Proceedings of the National Academy of Sciences, USA*, 93(January), pp. 13–20. doi: 10.1007/BF00569836.

Kadesch, T. (1992) 'Helix-loop-helix proteins in the regulation of immunoglobulin gene transcription', *Immunol Today*, 13(1), pp. 31–36. Available at: papers2://publication/uuid/71A118B1-C4DE-42BB-B9CB-324B535AA098.

Kadonaga, J. T. (2004) 'Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors', *Cell*, 116(2), pp. 247–257. doi: 10.1016/S0092-8674(03)01078-X.

Kaestner, K. H., Kno, W. and Martı, D. E. (2000) 'Unified nomenclature for the winged helix / forkhead transcription factors', *Genes & Development*, (14), pp. 142–146.

Katoh, Masuko and Katoh, Masaru (2004) 'Human Fox Gene Family', *International Journal of Oncology*, 25, pp. 1495–1500.

Kelly, S. M., Jess, T. J. and Price, N. C. (2005) 'How to study proteins by circular dichroism', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1751(2), pp. 119–139. doi: 10.1016/j.bbapap.2005.06.005.

Keskin, O., Gursoy, A., Ma, B., Nussinov, R. and Aviv, T. (2008) 'Principles of Protein − Protein Interactions What are the Preferred Ways For Proteins To Interact.pdf', *Chemical Reviews*, 108(4), pp. 1225–1243.

Khan, S. H., Ahmad, F., Ahmad, N., Flynn, D. C. and Kumar, R. (2011) 'Protein-Protein Interactions : Principles , Techniques , and their Potential Role in New Drug Development Protein-Protein Interactions : Principles , Techniques ', *Journal of Biomolecular Structure and Dynamics*, 28(6), pp. 929–938.

Khare, D., Ziegelin, G., Lanka, E. and Heinemann, U. (2004) 'Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB', *Nature Structural and Molecular Biology*, 11(7), pp. 656–663. doi: 10.1038/nsmb773.

Khazanovich, N., Bateman, K. S., Chernaia, M., Michalak, M. and James, M. N. G. (1996) 'Crystal structure of the yeast cell-cycle control protein , p13 suc1 , in a strand-exchanged dimer', *Structure*, 4(3), pp. 299–309.

Kittappa, R., Chang, W. W., Awatramani, R. B. and McKay, R. D. G. (2007) 'The foxa2 gene controls the birth and spontaneous degeneration of dopamine neurons in old age', *PLoS Biology*, 5(12), pp. 2875–2884. doi: 10.1371/journal.pbio.0050325.

Klug, A. (2010) 'The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation', *Annual Review of Biochemistry*, 79(1), pp. 213–231. doi: 10.1146/annurev-biochem-010909-095056.

Konopka, G., Bomar, J. M., Winden, K., Coppola, G., Jonsson, Z. O., Gao, F., Peng, S., Preuss, T. M., James, A. and Geschwind, D. H. (2009) 'Human-Specific Transcriptional Regulation of CNS Development Genes by FOXP2', *Nature*, 462(7270), pp. 213–217. doi: 10.1038/nature08549.Human-Specific.

Koon, H. B., Ippolito, G. C., Banham, A. H. and Tucker, P. W. (2007) 'FOXP1 : a potential therapeutic', *Expert opinion on therapeutic targets*, 11(7), pp. 955–966.

Krylov, D., Mikhailenko, I. and Vinson, C. (1994) 'A thermodynamic scale for leucine zipper stability and dimerization specificity : e and g interhelical interactions Basic region', *The EMBO Journal*, 13(12), pp. 2849–2861.

Krylov, D. and Vinson, C. R. (2001) 'Leucine zipper', *Encyclopedia of Life Sciences*, pp. 1–7. doi: 10.1038/npg.els.0003001.

Kurz, A., Wöhr, M., Walter, M., Bonin, M., Auburger, G., Gispert, S. and Schwarting, R. K. W. (2010) 'Alpha-synuclein deficiency affects brain Foxp1 expression and ultrasonic vocalization', *Neuroscience*. Elsevier Inc., 166(3), pp. 785–795. doi: 10.1016/j.neuroscience.2009.12.054.

Laemmli, U. K. (1970) 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature*, 227(5259), pp. 680–685. doi: 10.1038/227680a0.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-khadem, F. and Monaco, A. P. (2001) 'A forkhead-domain gene is mutated in a sever speech and language disorder', *Nature*, 413, pp. 519–523.

Lai, C. S. L., Gerrelli, D., Monaco, A. P., Fisher, S. E. and Copp, A. J. (2003) 'FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder', *Brain*, 126(11), pp. 2455–2462. doi: 10.1093/brain/awg247.

Lakowicz, J. R. (2002) *Topics in Fluorescence Spectroscopy*. Edited by J. R. Lakowicz. New York: Kluwer Academic Publishers. doi: 10.1007/b112911.

Lakowicz, J. R. and Masters, B. R. (2008) 'Principles of Fluorescence Spectroscopy, Third Edition', *Journal of Biomedical Optics*, 13(2), p. 029901. doi: 10.1117/1.2904580.

Lamonerie, T., Tremblay, J. J., Lanctbt, C. and Gauthier, Y. (1996) 'transcription factor involved in transcribtion of the', *Genes & Development*, pp. 1284–1295.

Landschulz, W. H., Johnson, P. F. and Mcknight, S. L. (1988) 'The Leucine Zipper: A hypothetical structure common to a new class of DNA binding proteins', *Science*, 240, pp. 1759–1764.

Laptenko, O., Tong, D. R., Manfredi, J. and Prives, C. (2016) 'The Tail That Wags the Dog: How the Disordered C-Terminal Domain Controls the Transcriptional Activities of the p53 Tumor-Suppressor Protein', *Trends in Biochemical Sciences*. Elsevier Ltd, 41(12), pp. 1022–1034. doi: 10.1016/j.tibs.2016.08.011.

Lau, F. W. and Bowie, J. U. (1997) 'A Method for Assessing the Stability of a Membrane Protein', *Biochemistry*, 36, pp. 5884–5892.

Leavitt, S. and Ernesto, F. (2001) 'Direct measurement of protein binding energetics by isothermal titration', *Current Opinion in Structural Biology*, 11, pp. 560–566. doi: 10.1016/j.jana.2005.09.004.

Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E. and Babu, M. M. (2014) 'Classification of intrinsically disordered regions and proteins', *Chemical Reviews*, 114(13), pp. 6589–6631. doi: 10.1021/cr400525m.

Lencki, R. W., Arul, J. and Neufeld, R. J. (1992) 'Effect of Subunit Dissociation , Denaturation , Aggregation , Coagulation , and Decomposition on Enzyme Inactivation Kinetics : II . Biphasic and Grace Period Behavior', *Biotechnology and Bioengineering*, 40, pp. 1427–1434.

Lendel, C., Dogan, J. and Härd, T. (2006) 'Structural Basis for Molecular Recognition in an Affibody:Affibody Complex', *Journal of Molecular Biology*, 359(5), pp. 1293–1304. doi: 10.1016/j.jmb.2006.04.043.

Lepock, J. R., Ritchie, K. P., Kolios, M. C., Michael Rodahl, A., Heinz, K. A. and Kruuv, J. (1992) 'Influence of Transition Rates and Scan Rate on Kinetic Simulations of Differential Scanning Calorimetry Profiles of Reversible and Irreversible Protein Denaturation', *Biochemistry*, 31(50), pp. 12706–12712. doi: 10.1021/bi00165a023.

LeTilly, V. and Royer, C. A. (1993) 'Fluorescence Anisotropy Assays Implicate Protein-Protein Interactions in Regulating trp Repressor DNA Binding', *Biochemistry*, 32(30), pp. 7753–7758. doi: 10.1021/bi00081a021.

Levy, E. D. and Pereira-Leal, J. B. (2008) 'Evolution and dynamics of protein interactions and networks', *Current Opinion in Structural Biology*, 18(3), pp. 349–357. doi: 10.1016/j.sbi.2008.03.003.

Li, B., Samanta, A., Song, X., Iacono, K. T., Bembas, K., Tao, R., Basu, S., Riley, J. L., Hancock, W. W., Shen, Y., Saouaf, S. J. and Greene, M. I. (2007) 'FOXP3 interactions with histone acetyltransferase and class II histone deacetylases are required for repression', *Proceedings of the National Academy of Sciences*, 104(11), pp. 4571–4576. doi: 10.1073/pnas.0700298104.

Li, Bin, Samanta, A., Song, X., Iacono, K. T., Brennan, P., Chatila, T. A., Roncador, G., Banham, A. H., Riley, J. L., Wang, Q., Shen, Y., Saouaf, S. J. and Greene, M. I. (2007) 'FOXP3 is a homo-oligomer and a component of a supramolecular regulatory complex disabled in the human XLAAD/IPEX autoimmune disease', *International Immunology*, 19(7), pp. 825–835. doi: 10.1093/intimm/dxm043.

Li, S., Weidenfeld, J. and Morrisey, E. E. (2004) 'Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions.', *Molecular and Cellular Biology*, 24(2), pp. 809–22. doi: 10.1128/MCB.24.2.809.

Licata, V. J. and Wowor, A. J. (2008) 'Applications of Fluorescence Anisotropy to the Study of Protein – DNA Interactions', *Methods in CellBiology*, 84, pp. 243–262. doi: 10.1016/S0091-679X(07)84009-X.

Liu, Y., Gotte, G., Libonati, M. and Eisenberg, D. (2002) 'Structures of the two 3D domain-swapped RNase A trimers', *Protein Science*, 11, pp. 371–380. doi: 10.1110/ps.36602.versa.

Liu, Y., Hart, P. J., Schlunegger, M. P. and Eisenberg, D. (1998) 'The crystal structure of a 3D domain-swapped dimer of RNase A at a 2 . 1-Å resolution', *Proceedings of the National Academy of Sciences of the United States of America*, 95, pp. 3437–3442.

Lopes, J. E., Torgerson, T. R., Schubert, L. A., Anover, S. D., Ocheltree, E. L., Ochs, H. D. and Ziegler, S. F. (2006) 'Analysis of FOXP3 Reveals Multiple Domains Required for Its Function as a Transcriptional Repressor', *The Journal of Immunology*, 177(5), pp. 3133–3142. doi: 10.4049/jimmunol.177.5.3133.

Louche, A., Salcedo, S. P. and Bigot, S. (2017) 'Protein–Protein Interactions: Pull-Down Assays', in Journet, L. and Cascales, E. (eds) *Bacterial Protein Secretion Systems: Methods and Protocols*, *Methods in Molecular Biology*. Springer Science+Business Media LLC, pp. 257–275. doi: 10.1007/978-1-4939-7033-9.

Lu, M. M., Li, S., Yang, H. and Morrisey, E. E. (2002) 'Foxp4 : a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues', *Gene Expression Patterns*, 2, pp. 223–228. doi: 10.1016/S1.

Luscombe, N. M., Austin, S. E., Berman, H. M. and Thornton, J. M. (2000) 'An overview of the structures of protein-DNA complexes', pp. 1–37.

Ma, B., Elkayam, T., Wolfson, H. and Nussinov, R. (2003) 'Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces', *Proceedings of the National Academy of Sciences*, 100(10), pp. 5772–5777. doi: 10.1073/pnas.1030237100.

MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A.-M., Lai, C. S. L., Vernes, S. C., Vargha-Khadem, F., McKenzie, F., Smith, R. L., Monaco, A. P. and Fisher, S. E. (2005) 'Identification

of FOXP2 Truncation as a Novel Cause of Developmental Speech and Language Deficits', *The American Journal of Human Genetics*, 76(6), pp. 1074–1080. doi: 10.1086/430841.

Mackay, J. P. and Crossley, M. (1998) 'Zinc fingers are sticking together', *Trends in Biochemical Sciences*, 23(1), pp. 1–4. doi: 10.1016/S0968-0004(97)01168-7.

Maestro, B. and Sanz, J. M. (2007) 'Extensive unfolding of the C-LytA choline-binding module by submicellar concentrations of sodium dodecyl sulphate', *FEBS Letters*, 581, pp. 375–381. doi: 10.1016/j.febslet.2006.12.042.

Maiese, K (2010) *Forkhead Transcription Factors: Vital Elements in Biology and Medicine, 2010*. Edited by Kenneth Maiese. Springer Science+Business Media. Available at: http://books.google.com/books?hl=en&amp%5Cnlr=&amp%5Cnid=3wIJmjjMc4YC&amp%5Cnoi=fnd&amp%5Cnpg=PR3&amp%5Cndq=Forkhead+Transcription+Factors+Vital+Elements+in+Biology+and+Medicine&amp%5Cnots=hDWBXKa4pC&amp%5Cnsig=E0AN2cav4J6g4i3o0nFYXA_vn4Q%5Cnpapers3://pub.

Manna, P. R., Dyson, M. T. and Stocco, D. M. (2016) 'Role of basic leucine zipper proteins in transcriptional regulation of the steroidogenic acute regulatory protein gene', *Molecular and Cellular Endocrinology*, 302(1), pp. 1–11. doi: 10.1177/0333102415576222.Is.

Martin, S. R., Esposito, V., De Los Rios, P., Pastore, A. and Temussi, P. A. (2008) 'Cold denaturation of yeast frataxin offers the clue to understand the effect of alcohols on protein stability', *Journal of the American Chemical Society*, 130(30), pp. 9963–9970. doi: 10.1021/ja803280e.

Massari, M. E. and Murre, C. (2000) 'MINIREVIEW Helix-Loop-Helix Proteins : Regulators of Transcription in Eucaryotic Organisms', *Molecular and Cellular Biology*, 20(2), pp. 429–440.

Matthews, J. M. and Sunde, M. (2012) 'DIMERS, OLIGOMERS, EVERYWHERE Jacqueline', in *Protein Dimerization and Oligomerization in Biology*, pp. 1–2. doi: 10.1007/978-1-4020-6682-5.

Mayer, M. P., Laufen, T., Paal, K., McCarty, J. S. and Bukau, B. (1999) 'Investigation of the interaction between DnaK and DnaJ by surface plasmon resonance spectroscopy', *Journal of Molecular Biology*, 289(4), pp. 1131–1144. doi: 10.1006/jmbi.1999.2844.

McCarty, A. S., Kleiger, G., Eisenberg, D. and Smale, S. T. (2003) 'Selective dimerization of

a C2H2 zinc finger subfamily', *Molecular Cell*, 11(2), pp. 459–470. doi: 10.1016/S1097-2765(03)00043-1.

McGuffin, L. J., Bryson, K. and Jones, D. T. (2000) 'The PSIPRED protein structure prediction server', *Bioinformatics*, 16(4), pp. 404–405. doi: 10.1093/bioinformatics/16.4.404.

Medina, E., Córdova, C., Villalobos, P., Reyes, J., Komives, E. A., Ramírez-Sarmiento, C. A. and Babul, J. (2016) 'Three-Dimensional Domain Swapping Changes the Folding Mechanism of the Forkhead Domain of FoxP1', *Biophysical Journal*, 110(11), pp. 2349–2360. doi: 10.1016/j.bpj.2016.04.043.

Meerschaut, I., Rochefort, D., Revençu, N., Pètre, J., Corsello, C., Rouleau, G. A., Hamdan, F. F., Michaud, J. L., Morton, J., Radley, J., Ragge, N., García-Miñaúr, S., Lapunzina, P., Bralo, M. P., Mori, M. A., Moortgat, S., Benoit, V., Mary, S., Bockaert, N., Oostra, A., Vanakker, O., Velinov, M., De Ravel, T. J. L., Mekahli, D., Sebat, J., Vaux, K. K., DiDonato, N., Hanson-Kahn, A. K., Hudgins, L., Dallapiccola, B., Novelli, A., Tarani, L., Andrieux, J., Parker, M. J., Neas, K., Ceulemans, B., Schoonjans, A. S., Prchalova, D., Havlovicova, M., Hancarova, M., Budisteanu, M., Dheedene, A., Menten, B., Dion, P. A., Lederer, D. and Callewaert, B. (2017) 'FOXP1-related intellectual disability syndrome: A recognisable entity', *Journal of Medical Genetics*, 54(9), pp. 613–623. doi: 10.1136/jmedgenet-2017-104579.

Mercer, F. and Unutmaz, D. (2009) 'The Biology of FoxP3: A Key Player in Immune Suppression during Infections, Autoimmune Disease and Cancer', *Advances in Experimental Medicine and Biology*, 665, pp. 47–59. doi: 10.1016/j.micinf.2011.07.011.Innate.

Miles, E. W., Rhee, S. and Davies, D. R. (1999) 'The Molecular Basis of Substrate Channeling', *Journal of Biological Chemistry*, 274(18), pp. 12193–12197.

Miller, J., McLachlan, A. D. and Klug, A. (1985) 'Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes', *EMBO Journal*, 4(6), pp. 1609–1614. doi: 10.1002/jtra.1022.

Miller, M. C., Le, H. T., Dean, W. L., Holt, P. A., Chaires, J. B. and Trent, J. O. (2011) 'Polymorphism and resolution of oncogene promoter quadruplex-forming sequences', *Organic & Biomolecular Chemistry*, 9(22), p. 7633. doi: 10.1039/c1ob05891f.

Mintseris, J. and Weng, Z. (2003) 'Atomic Contact Vectors in Protein-Protein Recognition', *Proteins: Structure, Function and Genetics*, 53(3), pp. 629–639. doi: 10.1002/prot.10432.

Mittag, T., Kay, L. E. and Forman-Kaya, J. D. (2010) 'Protein dynamics and conformational disorder in molecular recognition', *Journal of Molecular Recognition*, 23(2), pp. 105–116. doi: 10.1002/jmr.961.

Mohd, M. S., Wahid, M. M., Mohammad, E. and Islam, A. (2017) 'Intronic miRNA miR-3666 Modulates its Host Gene FOXP2 Functions in Neurodevelopment and May Contribute to Pathogenesis of Neurological Disorders Schizophrenia and Autism', *Journal of Applied Biotechnology & Bioengineering*, 2(1), pp. 32–45. doi: 10.15406/jabb.2017.02.00022.

Morris, G. and Fanucchi, S. (2016) 'A Key Evolutionary Mutation Enhances DNA Binding of the FOXP2 Forkhead Domain', *Biochemistry*, 55(13), pp. 1959–1967. doi: 10.1021/acs.biochem.5b01271.

Morris, G., Stoychev, S., Naicker, P., Dirr, H. W. and Fanucchi, S. (2018) 'The forkhead domain hinge-loop plays a pivotal role in DNA binding and transcriptional activity of FOXP2', *Biological Chemistry*, pp. 1–13.

Munishkina, L. A., Phelan, C., Uversky, V. N. and Fink, A. L. (2003) 'Conformational behavior and aggregation of α-synuclein in organic solvents: Modeling the effects of membranes', *Biochemistry*, 42(9), pp. 2720–2730. doi: 10.1021/bi027166s.

Myatt, S. S. and Lam, E. W. F. (2007) 'The emerging roles of forkhead box (Fox) proteins in cancer', *Nature Reviews Cancer*, 7, pp. 847–859. doi: 10.1038/nrc2223.

Nagai, Y., Inui, T., Popiel, H. A., Fujikake, N., Hasegawa, K., Urade, Y., Goto, Y., Naiki, H. and Toda, T. (2007) 'A toxic monomeric conformer of the polyglutamine protein', *Nature Structural and Molecular Biology*, 14(4), pp. 332–340. doi: 10.1038/nsmb1215.

Nagradova, N. K. (2001) 'Interdomain interactions in oligomeric enzymes : creation of asymmetry in homo-oligomers and role in metabolite channeling between active centers of hetero-oligomers', *FEBS Letters*, 487, pp. 327–332.

Nakamura, H. (1996) 'Roles of electrostatic interaction in proteins', *Quarterly Reviews of Biophysics*, 29(1), pp. 1–90.

Nelson, C. S., Fuller, C. K., Fordyce, P. M., Greninger, A. L., Li, H. and DeRisi, J. L. (2013) 'Microfluidic affinity and ChIP-seq analyses converge on a conserved FOXP2-binding motif in chimp and human, which enables the detection of evolutionarily novel targets', *Nucleic Acids Research*, 41(12), pp. 5991–6004. doi: 10.1093/nar/gkt259.

Nguyen, T. N. and Goodrich, J. A. (2006) 'Protein-protein interaction assays: eliminating false positive interactions Tuan', *Nature Methods*, 3(2164), pp. 135–139. doi: 10.1126/scisignal.2001449.Engineering.

Niepmann, M. and Zheng, J. (2006) 'Discontinuous native protein gel electrophoresis', *Electrophoresis*, 27(20), pp. 3949–3951. doi: 10.1002/elps.200600172.

Noguchi, A., Ikeda, A., Mezaki, M., Fukumori, Y. and Kanemori, M. (2014) 'DnaJ-promoted binding of DnaK to multiple sites on ??32 in the presence of ATP', *Journal of Bacteriology*, 196(9), pp. 1694–1703. doi: 10.1128/JB.01197-13.

Nooren, I. M. A. and Thornton, J. M. (2003a) 'Diversity of protein±protein interactions', *EMBO Journal*, 22(14), pp. 3486–3492.

Nooren, I. M. A. and Thornton, J. M. (2003b) 'Structural characterisation and functional significance of transient protein-protein interactions', *Journal of Molecular Biology*, 325(5), pp. 991–1018. doi: 10.1016/S0022-2836(02)01281-0.

Norel, R., Sheinerman, F., Petrey, D. and Honig, B. (2008) 'Electrostatic contributions to protein-protein interactions: Fast energetic filters for docking and their physical basis', *Protein Science*, 10(11), pp. 2147–2161. doi: 10.1110/ps.12901.

O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S., Karakoc, E., MacKenzie, A. P., Ng, S. B., Baker, C., Rieder, M. J., Nickerson, D. A., Bernier, R., Fisher, S. E., Shendure, J. and Eichler, E. E. (2011) 'Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations', *Nature Genetics*, 43(6), pp. 585–589. doi: 10.1038/ng.835.

O'Shea, E. K., Rutkowski, R. and Kim, P. S. (1989) 'Evidence that the leucine zipper is a coiled coil.', *Science (New York, N.Y.)*, 243(December), pp. 538–542. doi: 10.1126/science.2911757.

Ogihara, N. L., Ghirlanda, G., Bryson, J. W., Gingery, M., Degrado, W. F. and Eisenberg, D. (2001) 'Design of three-dimensional domain-swapped dimers and fibrous oligomers', *Proceedings of the National Academy of Sciences*, 98(4), pp. 1404–1409.

Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. and Wüthrich, K. (1990) 'Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution.', *The EMBO Journal*,

9(10), pp. 3085–92. doi: 10.1002/j.1460-2075.1990.tb07505.x.

Ozbabacan, S. E. A., Engin, H. B., Gursoy, A. and Keskin, O. (2011) 'Transient proteinprotein interactions', *Protein Engineering, Design and Selection*, 24(9), pp. 635–648. doi: 10.1093/protein/gzr025.

Pan, Y., Tsai, C. J., Ma, B. and Nussinov, R. (2010) 'Mechanisms of transcription factor selectivity', *Trends in Genetics*, 26(2), pp. 75–83. doi: 10.1016/j.tig.2009.12.003.

Pantoja-Uceda, D., Santiveri, C. M. and Jiménez, M. A. (2006) 'De novo Design of Monomeric b-Hairpin and b-sheet peptides', in Guerois, R. and Lopez de la Paz, M. (eds) *Methods in Molecular Biology*. Totowa: Humana Press Inc, pp. 27–51.

Pariani, M. J., Spencer, A., Jr, J. M. G. and Rimoin, D. L. (2010) 'A 785 kb deletion of 3p14.1p13, including the FOXP1 gene, associated with speech delay, contractures, hypertonia and blepharophimosis', *European Journal of Medical Genetics*, 52, pp. 123–127. doi: 10.1016/j.ejmg.2009.03.012.A.

Park, S. H., Reyes, J. A., Gilbert, D. R., Kim, J. W. and Kim, S. (2009) 'Prediction of protein-protein interaction types using association rule based classification', *BMC Bioinformatics*, 10. doi: 10.1186/1471-2105-10-36.

Partridge, L. and Brüning, J. C. (2008) 'Forkhead transcription factors and ageing', *Oncogene*, 27(16), pp. 2351–2363. doi: 10.1038/onc.2008.28.

Pavletich, N. P. and Pabo, C. (1991) 'Zinc Structure of a Recognition : Complex Zif268-DNA', *Advancement Of Science*, 252(5007), pp. 809–817. doi: 10.1126/science.2028256.

Payre, F., Buono, P., Vanzo, N. and Vincent, A. (1997) 'Two types of zinc fingers are required for dimerization of the serendipity delta transcriptional activator', *Molecular and Cellular Biology*, 17(6), pp. 3137–3145. doi: 10.1128/MCB.17.6.3137.

Peng, S. L. (2007) 'Immune regulation by Foxo transcription factors', *Autoimmunity*, 40(6), pp. 462–469. doi: 10.1080/08916930701464913.

Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G. and Orengo, C. (2010) 'Transient Protein-Protein Interactions: Structural, Functional, and Network Properties', *Structure*. Elsevier Ltd, 18(10), pp. 1233–1243. doi: 10.1016/j.str.2010.08.007.

Perumal, K., Dirr, H. W. and Fanucchi, S. (2015) 'A Single Amino Acid in the Hinge Loop

Region of the FOXP Forkhead Domain is Significant for Dimerisation', *The Protein Journal*. Springer US, 34(2), pp. 111–121. doi: 10.1007/s10930-015-9603-4.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) 'UCSF Chimera - A visualization system for exploratory research and analysis', *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. doi: 10.1002/jcc.20084.

Pollard, T. D. (2010) 'MBoC TECHNICAL PERSPECTIVE: A Guide to Simple and Informative Binding Assays', *Molecular Biology of the Cell*, 21, pp. 4061–4067. doi: 10.1091/mbc.E10-08-0683.

Pope, B. and Kent, H. M. (1996) 'High efficiency 5 min transformation of *Escherichia coli*', *Nucleic Acids Research*, 24(3), pp. 536–537.

Poulin, G., Lebel, M., Chamberland, M., Paradis, F. W. and Drouin, J. (2000) 'Specific protein-protein interaction between basic helix-loop-helix transcription factors and homeoproteins of the Pitx family', *Mol. Cell. Biol.*, 20(13), pp. 4826–4837. doi: 10.1128/MCB.20.13.4826-4837.2000.

Poulin, G., Turgeon, B. and Drouin, J. (1997) 'NeuroD1/beta2 contributes to cell-specific transcription of the proopiomelanocortin gene.', *Mol Cell Biol*, 17(11), pp. 6673–6682. doi: 10.1016/j.ijproman.2010.03.003.

Povrozin, Y. and Barbieri, B. (2016) 'Fluorescence Spectroscopy', in Kurtz, M. (ed.) *Handbook of Measurement in Science and Engineering*. John Wiley & Sons, pp. 2474–2499.

Presman, D. M., Ganguly, S., Schiltz, R. L., Johnson, T. A., Karpova, T. S. and Hager, G. L. (2016) 'DNA binding triggers tetramerization of the glucocorticoid receptor in live cells', *Proceedings of the National Academy of Sciences*, 113(29), pp. 8236–8241. doi: 10.1073/pnas.1606774113.

Privalov, P. L., Dragan, A. I., Crane-Robinson, C., Breslauer, K. J., Remeta, D. P. and Minetti, C. A. S. A. (2007) 'What Drives Proteins into the Major or Minor Grooves of DNA?', *Journal of Molecular Biology*, 365(1), pp. 1–9. doi: 10.1016/j.jmb.2006.09.059.

Rambo, R. P. (2017) 'Biological Small Angle Scattering: Techniques, Strategies and Tips', *Biological Small Angle Scattering: Techniques, Strategies and Tips*, 1009, pp. 31–45. doi: 10.1007/978-981-10-6038-0.

Ramsey, J. E., Daugherty, M. A. and Kelm, R. J. (2006) 'Hydrodynamic studies on the quaternary structure of recombinant mouse Purβ', *Journal of Biological Chemistry*, (13). doi: 10.1074/jbc.M609356200.

Raushel, F. M., Thoden, J. B. and Holden, H. M. (2003) 'Enzymes with Molecular Tunnels', *Accounts of Chemical Research*, 36(7), pp. 539–548.

Ravasi, T., Suzuki, H., Cannistraci, C. V, Katayama, S., Bajic, V. B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., Carninci, P., Daub, C. O., Forrest, A. R. R., Gough, J., Grimmond, S., Han, J. H., Hashimoto, T., Hide, W., Hofmann, O., Kamburov, A., Kaur, M., Kawaji, H., Kubosaki, A., Lassmann, T., van Nimwegen, E., MacPherson, C. R., Ogawa, C., Radovanovic, A., Schwartz, A., Teasdale, R. D., Tegner, J., Lenhard, B., Teichmann, S. A., Arakawa, T., Ninomiya, N., Murakami, K., Tagami, M., Fukuda, S., Imamura, K., Kai, C., Ishihara, R., Kitazume, Y., Kawai, J., Hume, D. A., Ideker, T. and Hayashizaki, Y. (2010) 'An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man (vol 140, pg 744, 2010)', *Cell*, 141(2), p. 369. doi: 10.1016/j.cell.2010.04.004.

Reynolds, K. A., Mclaughlin, R. N. and Ranganathan, R. (2011) 'Hot Spots for Allosteric Regulation on Protein Surfaces', *Cell*. Elsevier Inc., 147(7), pp. 1564–1575. doi: 10.1016/j.cell.2011.10.049.

Rial, D. V. and Ceccarelli, E. A. (2002) 'Removal of DnaK contamination during fusion protein purifications', *Protein Expression and Purification*, 25(3), pp. 503–507. doi: 10.1016/S1046-5928(02)00024-4.

Richardson, J. S. and Richardson, D. C. (2002) 'Natural -sheet proteins use negative design to avoid edge-to-edge aggregation', *Proceedings of the National Academy of Sciences*, 99(5), pp. 2754–2759. doi: 10.1073/pnas.052706099.

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B. and Mann, R. (2012) 'Origins of specificity in protein-DNA recognition', *Annual Reviews Biochemistry*, 79, pp. 233–269. doi: 10.1146/annurev-biochem-060408-091030.Origins.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S. J., Vidal, M., *et al.* (2014) 'A proteome-scale map of the human interactome network', *Cell*, 159(5), pp. 1212–1226. doi: 10.1016/j.cell.2014.10.050.

Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in Escherichia

coli : advances and challenges', *Frontiers in Microbiology*, 5, pp. 1–17. doi: 10.3389/fmicb.2014.00172.

Rousseau, F., Schymkowitz, J. W. H., Wilkinson, H. R. and Itzhaki, L. S. (2004) 'Intermediates Control Domain Swapping during Folding of p13 suc1 *', *The Journal of Biological Chemistry*, 279(9), pp. 8368–8377. doi: 10.1074/jbc.M310640200.

Roy, S. and Kumar, V. (2011) 'Estimation of cavern configurations from subsidence data', *Bulletin of Engineering Geology and the Environment*, 70(1), pp. 53–61. doi: 10.1007/s10064-010-0269-z.

Sackett, S. D., Li, Z., Hurtt, R., Gao, Y., Wells, R. G., Brondell, K., Kaestner, K. H. and Greenbaum, L. E. (2009) 'Foxl1 is a marker of bipotential hepatic progenitor cells in mice', *Hepatology*, 49(3), pp. 920–929. doi: 10.1002/hep.22705.

Sagardoy, A., Martinez-Ferrandis, J. I., Roa, S., Bunting, K. L., Aznar, M. A., Elemento, O., Shaknovich, R., Fontán, L., Fresquet, V., Perez-Roger, I., Robles, E. F., De Smedt, L., Sagaert, X., Melnick, A. and Martinez-Climent, J. A. (2013) 'Downregulation of FOXP1 is required during germinal center B-cell function', *Blood*, 121(21), pp. 4311–4320. doi: 10.1182/blood-2012-10-462846.

Sakai, Y., Shaw, C. A., Dawson, B. C., Dugas, D. V., Al-Mohtaseb, Z., Hill, D. E. and Zoghbi, H. Y. (2011) 'Protein interactome reveals converging molecular pathways among autism disorders', *Science Translational Medicine*, 3(86). doi: 10.1126/scitranslmed.3002166.

Schägger, H. and von Jagow, G. (1991) 'Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form', *Analytical Biochemistry*, 199(2), pp. 223–231. doi: 10.1016/0003-2697(91)90094-A.

Schechtman, D., Mochly-rosen, D. and Ron, D. (no date) 'Glutathione S-Transferase Pull-Down Assay', in Newton, A. C. (ed.) *Methods in Molecular Biology: Protein Kinase C Protocols*. Totowa: Humana Press Inc, pp. 345–350.

Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lehrach, H. and Wanker, E. E. (1997) 'Huntingtin-Encoded Polyglutamine Expansions Form Amyloid-like Protein Aggregates *In Vitro* and *In Vivo*', *Cell*, 90, pp. 549–558.

Schleif, R. F. (2013) 'Modulation of DNA binding by gene-specific transcription factors',

Biochemistry, 52(39), pp. 6755–6765. doi: 10.1021/bi400968e.

Schlunegger, M. P., Bennett, M. J. and Eisenberg, D. (1997) 'Oligomer formation by 3d domain swapping: a model for protein assembly and misassembly', *Advances in Protein Chemistry*, 50, pp. 61–121.

Schmitz, K. S. and Phillies, G. D. J. (1991) '*An Introduction to Dynamic Light Scattering by Macromolecules*', *Physics Today*, 44(5), pp. 66–66. doi: 10.1063/1.2810116.

Schreiber, G. and Fersht, A. R. (1996) 'Rapid, electrostatically assisted association of proteins', *Nature Structural Biology*, 3(5), pp. 427–431. doi: 10.1038/nsb0596-427.

Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F. and Ziegler, S. F. (2001) 'Scurfin (FOXP3) Acts as a Repressor of Transcription and Regulates T Cell Activation', *Journal of Biological Chemistry*, 276(40), pp. 37672–37679. doi: 10.1074/jbc.M104521200.

Sezonov, G., Joseleau-Petit, D. and D'Ari, R. (2007) '*Escherichia coli* Physiology in Luria-Bertani Broth', *Journal of Bacteriology*, 189(23), pp. 8746–8749. doi: 10.1128/JB.01368-07.

Shapiro, A. L., Viñuela, E. and V. Maizel Jr., J. (1967) 'Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels', *Biochemical and Biophysical Research Communications*, 28(5), pp. 815–820. doi: 10.1016/0006-291X(67)90391-9.

Shea, E. K. O., Rurkowski, R., Iii, W. F. S. and Kim, P. S. (1989) 'Preferential Heterodimer Formation by Isolated Leucine Zippers from Fos and Jun Author (s): Erin K . O ' Shea , Rheba Rutkowski , Walter F . Stafford III and Peter S . Kim Published by : American Association for the Advancement of Science Stable URL : h', *Science*, 245(4918), pp. 646–648.

Sheinerman, F. B. and Honig, B. (2002) 'On the role of electrostatic interactions in the design of protein-protein interfaces', *Journal of Molecular Biology*, 318(1), pp. 161–177. doi: 10.1016/S0022-2836(02)00030-X.

Sheinerman, F. B., Norel, R. and Honig, B. (2000) 'Electrostatic aspects of protein–protein interactions', *Current Opinion in Structural Biology*, 10, pp. 153–159. doi: 10.1016/S0959-440X(00)00065-8.

Shu, W., Lu, M. M., Zhang, Y., Tucker, P. W., Zhou, D. and Morrisey, E. E. (2007) 'Foxp2 and Foxp1 cooperatively regulate lung and esophagus development', *Development*, 134(10), pp. 1991–2000. doi: 10.1242/dev.02846.

Shu, W., Yang, H., Zhang, L., Lu, M. M. and Morrisey, E. E. (2001) 'Characterization of a New Subfamily of Winged-helix/Forkhead (Fox) Genes That Are Expressed in the Lung and Act as Transcriptional Repressors', *Journal of Biological Chemistry*, 276(29), pp. 27488–27497. doi: 10.1074/jbc.M100636200.

Sin, C., Li, H. and Crawford, D. A. (2014) 'Transcriptional Regulation by FOXP1, FOXP2, and FOXP4 Dimerization', *Journal of Molecular Neuroscience*, 55(2), pp. 437–448. doi: 10.1007/s12031-014-0359-7.

Singh, S. M., Cabello-Villegas, J., Hutchings, R. L. and Mallela, K. M. G. (2010) 'Role of partial protein unfolding in alcohol-induced protein aggregation', *Proteins: Structure, Function and Bioinformatics*, 78(12), pp. 2625–2637. doi: 10.1002/prot.22778.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordân, R. and Rohs, R. (2014) 'Absence of a simple code: How transcription factors read the genome', *Trends in Biochemical Sciences*, 39(9), pp. 381–399. doi: 10.1016/j.tibs.2014.07.002.

Slutsky, M. and Mirny, L. A. (2004) 'Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential', *Biophysical Journal*. Elsevier, 87(6), pp. 4021–4035. doi: 10.1529/biophysj.104.050765.

De Smedt, L., Palmans, S., Govaere, O., Moisse, M., Boeckx, B., De Hertogh, G., Prenen, H., Van Cutsem, E., Tejpar, S., Tousseyn, T. and Sagaert, X. (2015) 'Expression of FOXP1 and Colorectal Cancer Prognosis', *Laboratory Medicine*, 46(4), pp. 299–311. doi: 10.1309/LM7IHV2NJI1PHMXC.

Smock, R. G. and Gierasch, L. M. (2009) 'Sending Signals Dynamically', *Science*, 324(5924), pp. 198–203. doi: 10.1126/science.1169377.Sending.

Soghoian, D. (2004) *Nickel vs Cobalt columns*.Caltech

Sollis, E., Graham, S. A., Vino, A., Froehlich, H., Vreeburg, M., Dimitropoulou, D., Gilissen, C., Pfundt, R., Rappold, G. A., Brunner, H. G., Deriziotis, P. and Fisher, S. E. (2016) 'Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder', *Human Molecular Genetics*, 25(3), pp. 546–557. doi: 10.1093/hmg/ddv495.

Song, X., Li, B., Xiao, Y., Chen, C., Wang, Q., Liu, Y., Berezov, A., Xu, C., Gao, Y., Li, Z., Wu, S. L., Cai, Z., Zhang, H., Karger, B. L., Hancock, W. W., Wells, A. D., Zhou, Z. and

Greene, M. I. (2012) 'Structural and Biological Features of FOXP3 Dimerization Relevant to Regulatory T Cell Function', *Cell Reports*. The Authors, 1(6), pp. 665–675. doi: 10.1016/j.celrep.2012.04.012.

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., Vernes, S. C., Fisher, S. E., Ren, B. and Geschwind, D. H. (2007) 'Identification of the Transcriptional Targets of FOXP2, a Gene Linked to Speech and Language, in Developing Human Brain', *The American Journal of Human Genetics*, 81(6), pp. 1144–1157. doi: 10.1086/522237.

Spitz, F. and Furlong, E. E. M. (2012) 'Transcription factors: From enhancer binding to developmental control', *Nature Reviews Genetics*. Nature Publishing Group, 13(9), pp. 613–626. doi: 10.1038/nrg3207.

Stein, A., Pache, R. A., Bernadó, P., Pons, M. and Aloy, P. (2009) 'Dynamic interactions of proteins in complex networks: A more structured view', *FEBS Journal*, 276(19), pp. 5390–5405. doi: 10.1111/j.1742-4658.2009.07251.x.

Stites, W. E. (1996) 'Protein−Protein Interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis', *Chemical Reviews*, 97, pp. 1233–1250.

Stroud, J. C., Wu, Y., Bates, D. L., Han, A., Nowick, K., Paabo, S., Tong, H. and Chen, L. (2006) 'Structure of the forkhead domain of FOXP2 bound to DNA', *Structure*, 14(1), pp. 159–166. doi: 10.1016/j.str.2005.10.005.

Struhl, K. (1989) 'Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins', *Trends in Biochemical Sciences*, 14(4), pp. 137–140. doi: 10.1016/0968-0004(89)90145-X.

Stryer, L. (1968) 'Fluorescence spectroscopy of proteins', *Science*, 162(3853), pp. 526–533. doi: 10.1126/science.162.3853.526.

Suh, W.-C., Burkholder, W. F., Lu, C. Z., Zhao, X., Gottesman, M. E. and Gross, C. A. (1998) 'Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ', *Proceedings of the National Academy of Sciences*, 95(26), pp. 15223–15228. doi: 10.1073/pnas.95.26.15223.

Sussel, L., Marin, O., Kimura, S. and Rubenstein, J. L. (1999) 'Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum', *Development*, 126(15), pp.

3359–3370. doi: 10393115.

Suzuki, H., Ogawa, C., Usui, K. and Hayashizaki, Y. (2004) ‘In vitro pull-down assay without expression constructs’, *BioTechniques*, 37(6), pp. 918–919.

Szklarczyk, D., Santos, A., Von Mering, C., Jensen, L. J., Bork, P. and Kuhn, M. (2016) ‘STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data’, *Nucleic Acids Research*, 44(D1), pp. D380–D384. doi: 10.1093/nar/gkv1277.

Takahashi, H., Takahashi, K. and Liu, F.-C. (2009) ‘FOXP Genes, Neural Development, Speech and Language Disorders’, in *Madame Curie Bioscience Database*, pp. 117–129. doi: 10.1007/978-1-4419-1599-3_9.

Takeuchi, K., Nakatani, Y. and Hisatomi, O. (2014) ‘Accuracy of Protein Size Estimates Based on Light Scattering Measurements’, *Open Journal of Biophysics*, (April), pp. 83–91. doi: 10.4236/ojbiphy.2014.42009.

Takeuchi, R., Ohtsuka, K., Ochi, N. and Inaguma, Y. (2010) ‘Post-translational modification of FOXP2 , a transcriptional fact verbal dyspraxia or , involved in developmental Rika Takeuchi’, *Annual Report of Research Institute for Biological Function*, 10(2001), pp. 81–91.

Tanabe, Y., Fujita, E. and Momoi, T. (2011) ‘FOXP2 promotes the nuclear translocation of POT1, but FOXP2(R553H), mutation related to speech-language disorder, partially prevents it’, *Biochemical and Biophysical Research Communications*. Elsevier Inc., 410(3), pp. 593–596. doi: 10.1016/j.bbrc.2011.06.032.

Tarunina, M. and Jenkins, J. R. (1993) ‘Human p53 binds DNA as a protein homodimer but monomeric variants retain full transcription transactivation activity’, *Oncogene*, 8(11), p. 3165—3173. Available at: http://europepmc.org/abstract/MED/8414520.

Taylor, C. M. and Keating, A. E. (2008) ‘Orientation and Oligomerization Specificity of the Bcr Coiled-Coil Oligomerization Domain’, *Biochemistry*, 44(49), pp. 16246–16256. doi: 10.1021/bi051493t.Orientation.

Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H. and White, S. A. (2004) ‘Parallel FoxP1 and FoxP2 Expression in Songbird and Human Brain Predicts Functional Interaction’, *Journal of Neuroscience*, 24(13), pp. 3152–3163. doi: 10.1523/JNEUROSCI.5589-03.2004.

Tetz, G. V., Artemenko, N. K. and Tetz, V. V. (2009) ‘Effect of DNase and antibiotics on biofilm characteristics’, *Antimicrobial Agents and Chemotherapy*, 53(3), pp. 1204–1209. doi:

10.1128/AAC.00471-08.

Tetz, V. V. and Tetz, G. V. (2010) 'Effect of Extracellular DNA Destruction by DNase I on Characteristics of Forming Biofilms', *DNA and Cell Biology*, 29(8), pp. 399–405. doi: 10.1089/dna.2009.1011.

Teufel, A., Wong, E. A., Mukhopadhyay, M., Malik, N. and Westphal, H. (2003) 'FoxP4, a novel forkhead transcription factor', *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1627(2–3), pp. 147–152. doi: 10.1016/S0167-4781(03)00074-5.

Tomblin, J. B., Brien, M. O., Shriberg, L., Williams, C., Murray, J., Bjork, J. and Anderson, S. (2010) 'Language features in a mother and daughter of a chromosome 7;13 translocation involving FOXP2', *Journal of Speech, Language and Hearing Research*, 52(5), pp. 1157–1174. doi: 10.1044/1092-4388(2009/07-0162).Language.

Tonkin, M. L., Halavaty, A. S., Ramaswamy, R., Ruan, J., Igarashi, M., Ngo, H. M. and Boulanger, M. J. (2015) 'Structural and functional divergence of the aldolase fold in Toxoplasma godii', *Journal of Molecular Biology*, 427(4), pp. 840–852. doi: 10.1021/acschemneuro.5b00094.Serotonin.

Torgerson, T. R. and Ochs, H. D. (2007) 'Immune dysregulation, polyendocrinopathy, enteropathy , X-linked : Forkhead box protein 3 mutations and lack of regulatory T cells', *Journal of Allergy and Clinical Immunology*, 120(4), pp. 744–750. doi: 10.1016/j.jaci.2007.08.044.

Tóth-Petróczy, Á., Simon, I., Fuxreiter, M. and Levy, Y. (2009) 'Disordered tails of homeodomains facilitate DNA recognition by providing a trade-off between folding and specific binding', *Journal of the American Chemical Society*, 131(42), pp. 15084–15085. doi: 10.1021/ja9052784.

Tsai, C.-J., Lin, S. L., Wolfson, H. J. and Nussinov, R. (1997) 'Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect', *Protein Science*, 6, pp. 53–64. doi: 10.1002/pro.5560060106.

Tsai, C. J., Xu, D. and Nussinov, R. (1997) 'Structural motifs at protein-protein interfaces: Protein cores versus two-state and three-state model complexes', *Protein Science*, 6(9), pp. 1793–1805. doi: 10.1002/pro.5560060901.

Tuteja, G., Kaestner, K. H. and Gene, H. (2007a) 'SnapShot : Forkhead Transcription Factors

I SnapShot : Forkhead Transcription Factors I', *Cell*, 130, pp. 1160–1161. doi: 10.1016/j.cell.2007.09.005.

Tuteja, G., Kaestner, K. H. and Gene, H. (2007b) 'SnapShot : Forkhead Transcription Factors II SnapShot : Forkhead Transcription Factors II', *Cell*, 131, pp. 192–193. doi: 10.1016/j.cell.2007.09.016.

Uversky, V. N. (2002) 'What does it mean to be natively unfolded? - Uversky - 2003 - European Journal of Biochemistry - Wiley Online Library', *European Journal of Biochemistry*, 12(May 2001), pp. 2–12. doi: 10.1110/ps.4210102.matic.

Uversky, V. N. (2009) 'Intrinsically disordered proteins and their environment: Effects of strong denaturants, temperature, pH, Counter ions, membranes, binding partners, osmolytes, and macromolecular crowding', *Protein Journal*, 28(7–8), pp. 305–325. doi: 10.1007/s10930-009-9201-4.

Uversky, V. N. (2017) 'Paradoxes and wonders of intrinsic disorder: Stability of instability', *Intrinsically Disordered Proteins*. Taylor & Francis, 5(1), p. e1327757. doi: 10.1080/21690707.2017.1327757.

Uversky, V. N., Narizhneva, N. V., Kirschstein, S. O., Winter, S. and Löber, G. (1997) 'Conformational transitions provoked by organic solvents in β-lactoglobulin: Can a molten globule like intermediate be induced by the decrease in dielectric constant?', *Folding and Design*, 2(3), pp. 163–172. doi: 10.1016/S1359-0278(97)00023-0.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. and Luscombe, N. M. (2009) 'A census of human transcription factors: function, expression and evolution', *Nature Reviews Genetics*, 10, pp. 252–263. doi: 10.1038/nrg2538.

van der Vegt, N. F. A. and Nayar, D. (2017) 'The Hydrophobic E ff ect and the Role of Cosolvents', *The Journal of Physical Chemistry B*, 121, pp. 9986–9998. doi: 10.1021/acs.jpcb.7b06453.

Velazquez-campoy, A., Leavitt, S. A. and Freire, E. (2015) 'Characterization of Protein-Protein Interactions by Isothermal Titration Calorimetry', in Meyerkord, C. L. and Fu, H. (eds) *Methods in Molecular Biology*. New York: Springer Science+Business Media, pp. 183–204. doi: 10.1007/978-1-4939-2425-7.

Vernes, S. C., Nicod, J., Elahi, F. M., Coventry, J. A., Kenny, N., Coupe, A. M., Bird, L. E.,

Davies, K. E. and Fisher, S. E. (2006) 'Functional genetic analysis of mutations implicated in a human speech and language disorder', *Human Molecular Genetics*, 15(21), pp. 3154–3167. doi: 10.1093/hmg/ddl392.

Vihma, H., Luhakooder, M., Pruunsild, P. and Timmusk, T. (2016) 'Regulation of different human NFAT isoforms by neuronal activity', *Journal of Neurochemistry*, 137(3), pp. 394–408. doi: 10.1111/jnc.13568.

Vinson, C., Acharya, A. and Taparowsky, E. J. (2006) 'Deciphering B-ZIP transcription factor interactions in vitro and in vivo', *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1759(1–2), pp. 4–12. doi: 10.1016/j.bbaexp.2005.12.005.

Vivian, J. T. and Callis, P. R. (2001) 'Mechanisms of Tryptophan Fluorescence Shifts in Proteins', *Biophysical Journal*, 80(5), pp. 2093–2109. doi: 10.1016/S0006-3495(01)76183-8.

Vogl, T., Jatzke, C., Hinz, H. J., Benz, J. and Huber, R. (1997) 'Thermodynamic stability of annexin V E17G: Equilibrium parameters from an irreversible unfolding reaction', *Biochemistry*, 36(7), pp. 1657–1668. doi: 10.1021/bi962163z.

Vuzman, D. and Levy, Y. (2012) 'Intrinsically disordered regions as affinity tuners in protein-DNA interactions', *Molecular BioSystems*, 8(1), pp. 47–57. doi: 10.1039/c1mb05273j.

Vuzman, D., Polonsky, M. and Levy, Y. (2010) 'Facilitated DNA search by multidomain transcription factors: Cross talk via a flexible linker', *Biophysical Journal*. Biophysical Society, 99(4), pp. 1202–1211. doi: 10.1016/j.bpj.2010.06.007.

Wang, B., Lin, D., Li, C. and Tucker, P. (2003) 'Multiple Domains Define the Expression and Regulatory Properties of Foxp1 Forkhead Transcriptional Repressors', *Journal of Biological Chemistry*, 278(27), pp. 24259–24268. doi: 10.1074/jbc.M207174200.

Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. and Jones, D. T. (2004) 'The DISOPRED server for the prediction of protein disorder', *Bioinformatics*, 20(13), pp. 2138–2139. doi: 10.1093/bioinformatics/bth195.

Webb, H., Steeb, O., Blane, A., Rotherham, L., Aron, S., Machanick, P., Dirr, H. and Fanucchi, S. (2017) 'The FOXP2 forkhead domain binds to a variety of DNA sequences with different rates and affinities', *Journal of Biochemistry*, 162(1), pp. 45–54. doi: 10.1093/jb/mvx003.

Weigel, D. and Jäckle, H. (1990) 'The fork head domain: a novel DNA binding motif of eukaryotic transcription factors?', *Cell*, 63, pp. 455–456. doi: 10.1016/0092-8674(90)90439-

L.

Weinreis, S. A., Ellis, J. P. and Cavagnero, S. (2011) 'Dynamic Fluorescence Depolarization: A Powerful Tool to Explore Protein Folding on the Ribosome', *Methods*, 52(1), pp. 57–73. doi: 10.1016/j.ymeth.2010.06.001.Dynamic.

Weinstein, D. C., Ruiz i Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M. and Darnell, J. E. (1994) 'The winged-helix transcription factor HNF-3β is required for notochord development in the mouse embryo', *Cell*, 78(4), pp. 575–588. doi: 10.1016/0092-8674(94)90523-1.

Weiss, M. A. (1990) 'Thermal Unfolding Studies of a Leucine Zipper Domain and Its Specific DNA Complex: Implications for Scissor's Grip Reeognition', *Biochemistry*, 29(35), pp. 8020–8024. doi: 10.1021/bi00487a004.

West, M. J., Webb, H. M., Sinclair, A. J. and Woolfson, D. N. (2004) 'Biophysical and Mutational Analysis of the Putative bZIP Domain of Epstein-Barr Virus EBNA 3C', *Journal of Virology*, 78(17), pp. 9431–9445. doi: 10.1128/JVI.78.17.9431.

Winter, C., Henschel, A., Kim, W. K. and Schroeder, M. (2006) 'SCOPPI: a structural classification of protein-protein interfaces', *Nucleic Acids Research*, 34(90001), pp. D310–D314. doi: 10.1093/nar/gkj099.

Wittig, I., Braun, H. P. and Schägger, H. (2006) 'Blue native PAGE', *Nature Protocols*, 1(1), pp. 418–428. doi: 10.1038/nprot.2006.62.

Woody, R. W. (1995) 'Circular Dichroism', *Methods in Enzymology*, 246, pp. 34–70.

Wu, M., Pilling, P., Chandry, P. S. and Mckinstry, W. J. (2013) 'Comparison of different IMAC resins for their ability to purify recombinant His-tagged proteins expressed in Escherichia coli and insect cells .', in *38th Lorne Conference on Protein Strucutre and Function*. Lorne, Victoria.

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A. D., Stroud, J. C., Bates, D. L., Guo, L., Han, A., Ziegler, S. F., Mathis, D., Benoist, C., Chen, L. and Rao, A. (2006) 'FOXP3 Controls Regulatory T Cell Function through Cooperation with NFAT', *Cell*, 126(2), pp. 375–387. doi: 10.1016/j.cell.2006.05.042.

Xie, X., Stubbington, M. J. T., Nissen, J. K., Andersen, K. G., Hebenstreit, D., Teichmann, S. A. and Betz, A. G. (2015) 'The Regulatory T Cell Lineage Factor Foxp3 Regulates Gene

Expression through Several Distinct Mechanisms Mostly Independent of Direct DNA Binding', *PLOS Genetics*, 11(6), p. e1005251. doi: 10.1371/journal.pgen.1005251.

Xu, D., Lin, S. L. and Nussinov, R. (1997) 'Protein binding versus protein folding: The role of hydrophilic bridges in protein associations', *Journal of Molecular Biology*, 265(1), pp. 68–84. doi: 10.1006/jmbi.1996.0712.

Yan, X., Zhou, H., Zhang, Tingting, Xu, P., Zhang, S., Huang, W., Yang, L., Gu, X., Ni, R. and Zhang, Tianyi (2015) 'Downregulation of FOXP2 promoter human hepatocellular carcinoma cell invasion', *Tumor Biology*, 36(12), pp. 9611–9619. doi: 10.1007/s13277-015-3701-y.

Yao, J., Liu, Y., Liang, H. G., Zhang, C., Zhu, J. Z., Qin, X., Sun, M., Qu, S. S. and Yu, Z. N. (2005) 'The effect of zinc(ii) on the growth of E. coli studied by microcalorimetry', *Journal of Thermal Analysis and Calorimetry*, 79(1), pp. 39–43. doi: 10.1007/s10973-004-0559-4.

Yoshikawa, H., Hirano, A., Arakawa, T. and Shiraki, K. (2012a) 'Effects of alcohol on the solubility and structure of native and disulfide-modified bovine serum albumin', *International Journal of Biological Macromolecules*. Elsevier B.V., 50(5), pp. 1286–1291. doi: 10.1016/j.ijbiomac.2012.03.014.

Yoshikawa, H., Hirano, A., Arakawa, T. and Shiraki, K. (2012b) 'Mechanistic insights into protein precipitation by alcohol', *International Journal of Biological Macromolecules*. Elsevier B.V., 50(3), pp. 865–871. doi: 10.1016/j.ijbiomac.2011.11.005.

Young, L., Jernigan, R. L. and Covell, D. G. (1994) 'A role for surface hydrophobicity', *Protein Science*, 3, pp. 717–729.

Zaret, K. S. and Carroll, J. S. (2011) 'Pioneer transcription factors : establishing competence for gene expression', pp. 2227–2241. doi: 10.1101/gad.176826.111.GENES.

Zeng, X., Herndon, A. M. and Hu, J. C. (1997) 'Buried asparagines determine the dimerization specificities of leucine zipper mutants', *Proceedings of the National Academy of Sciences of the United States of America*, 94, pp. 3673–3678.

Zhang, D., Lu, M. and Wang, H. (2011) 'Fluorescence anisotropy analysis for mapping aptamer-protein interaction at the single nucleotide level', *Journal of the American Chemical Society*, 133(24), pp. 9188–9191. doi: 10.1021/ja202141y.

Zhang, H., Wu, Q. and Berezin, M. Y. (2015) 'Fluorescence anisotropy (polarization): from

drug screening to precision medicine', *Expert Opinion on Drug Discovery*, 10(11), pp. 1145–1161. doi: 10.1517/17460441.2015.1075001.Fluorescence.

Zhao, C., Hellman, L. M., Zhan, X., Bowman, W. S., Whiteheart, S. W. and Fried, M. G. (2010) 'Hexahistidine-Tag-Specific Optical Probes for Analyses of Proteins and Their Interactions', *Analytical Biochemistry*, 399(2), pp. 237–245. doi: 10.1016/j.ab.2009.12.028.Hexahistidine-Tag-Specific.

Zhou, B., Zhong, Q., Minoo, P., Li, C., Ann, D. K., Frenkel, B., Morrisey, E. E., Crandall, E. D. and Borok, Z. (2008) 'Foxp2 inhibits Nkx2.1-mediated transcription of SP-C via interactions with the Nkx2.1 homeodomain', *American Journal of Respiratory Cell and Molecular Biology*, 38(6), pp. 750–758. doi: 10.1165/rcmb.2007-0350OC.

Zhu, H. (2016) 'Forkhead box transcription factors in embryonic heart development and congenital heart disease', *Life Sciences*. Elsevier Inc., 144, pp. 194–201. doi: 10.1016/j.lfs.2015.12.001.

Zhu, H., Domingues, F. S., Sommer, I. and Lengauer, T. (2006) 'NOXclass: prediction of protein-protein interaction types', *BMC Bioinformatics*, 7(27), pp. 27–34. doi: 10.1186/1471-2105-7-27.

Zmijewski, M. A., Kwiatkowska, J. M. and Lipińska, B. (2004) 'Complementation studies of the DnaK-DnaJ-GrpE chaperone machineries from Vibrio harveyi and Escherichia coli, both in vivo and in vitro', *Archives of Microbiology*, 182(6), pp. 436–449. doi: 10.1007/s00203-004-0727-8.

Zylicz, M., Ang, D. and Georgopoulos, C. (1987) 'The grpE protein of Escherichia coli. Purification and properties', *Journal of Biological Chemistry*, 262(36), pp. 17437–17442.

Zyliczs, M., Yamamotoo, T., Mckittrick, N., Sell, S. and Georgopoulos, C. (1985) 'Purification and Properties of the dnaJ Replication Protein of *Escherichia coli*', *Journal of Biological Chemistry*, 260, pp. 7591–7598.