Probing the Protein-Protein Interactions of the FOXP1, FOXP2 and FOXP3 Forkhead Domains

by

Paulina Mhlongo

(1054013)

Dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

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

Supervisor: Dr Sylvia Fanucchi

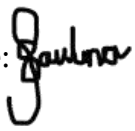
July 2023

Declaration by the Student

I, Paulina Mhlongo (1054013), am a student registered for the degree of Master of Science in the academic year 2021.

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 not submitted this work before for any other degree or examination at this or any other University.
- 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.

Signature: 

Signed on **28** day of **July** **2023**

Abstract

The FOXP proteins are classified amongst the forkhead box superfamily of transcription factors due to their highly conserved forkhead winged-helix domain (FHD). It is through this domain that FOX transcription factors are able to bind DNA in order to perform crucial roles in the regulation of gene transcription from development through adulthood. The FHD of FOXP1, FOXP2 and FOXP3 is remarkably unique in its ability to establish domain-swapped dimerization, postulated to drive interchromosomal interactions to regulate the transcription of distal genetic material. The FOXP1 and FOXP2 proteins are co-expressed and have also been demonstrated to directly interact *in vivo* and *in vitro* via a domain upstream of the FHD. Similarly, the FOXP1 and FOXP3 proteins have been established to form direct heterotypic interactions to function in regulatory T-cells (T_{reg}). Provided that the FOXP FHD has the exceptional capability for dimerization, the study of the heterodimerization of the FHDs of these FOXP transcription factors may provide insight into the role of FHD-dimerization and its importance in the physiological roles of these proteins. Therefore, the aim of this work was to investigate whether a FOXP FHD heterodimerization event can occur between the FOXP1 and FOXP2 FHDs as well as between the FOXP1 and FOXP3 FHDs. The three FHD proteins were expressed and isolated for downstream interaction studies. The activity of the purified FHDs was studied using basic electrophoretic mobility shift assay. Their secondary and tertiary structures were characterized with circular dichroism and intrinsic fluorescence spectroscopies. Concentration-dependent size exclusion chromatography was employed to study their propensity for dimerization and fluorescence anisotropy was used to investigate both the homo- and heterodimerization of the FHDs. It was revealed that FOXP3 FHD showed the highest propensity to homodimerize, whilst FOXP2 FHD showed the weakest propensity. Despite this, the homodimers of FOXP1 FHD appear to be less stable than that of the FOXP2 FHD. The heterodimerization studies suggest that FOXP1 FHD has preference in forming heterotypic associations with FOXP3 FHD rather than FOXP2 FHD.

Acknowledgements

To Dr Fanucchi, my supervisor, I would like to express my sincere gratitude for your unwavering support and belief in me throughout my research project. Your guidance, expertise, and encouragement have been vital in completing my project, and I am grateful for the opportunity to have worked under your guidance.

To my friends from the FOXy group and colleagues from PSFRU, I am immensely grateful for your help and support in the laboratory. Special thanks to Dr Joseph Kabogo, Dr Aasiya Lakhi, Dr Monare Thulo, Dr Blessing Oyiogu, and Dr Heather Donald. Your guidance, patience, and expertise have been crucial in helping me complete my research project.

To my best friends, Malefo Seeletse and Alicia Joshua, I am so grateful for the emotional support throughout my research project. And to my friends, Keiran Mcinnes and Dr Aasiya Lakhi, thank you for all your kindness and willingness to lend a helping hand. All of you brought laughter, calm and shine through the hard moments.

I am extremely grateful to the National Research Foundation (NRF) for the generous support and funding.

Heavenly Father, this submission would have not been possible without You. Thank You for this blessing. Amen.

Table of Contents:

List of Abbreviations 7

List of Figures 8

Introduction 10

1.1) Transcription Factors 10

1.1.1 Transcription Factor Dimerization 12

1.2) Forkhead Box (FOX) Proteins 15

1.2.1 The FOXP Subfamily 15

1.2.2 DNA Binding by FOXP Proteins 20

1.2.3 FOXP Dimerization 21

1.2.4 Domain Swapping 23

1.2.5 Domain Swapping of the FOXP FHDs 25

1.3) Rationale 26

Aims and Objectives 27

2.1) Main Aim: 27

2.2) Objectives: 27

Materials and Methods 28

3.1) DNA Plasmids Encompassing Sequences Encoding the FOXP1, FOXP2 and FOXP3 FHDs and DNA Plasmid Transformation: 28

3.1.1 DNA Plasmids bearing sequences encoding the FOXP1, FOXP2 and FOXP3 FHDs: 28

3.1.2 Transformation of E.coli Bacterial Cells: 29

3.2) Expression of the FOXP1, FOXP2 and FOXP3 FHDs: 30

3.3) Purification of the FOXP1, FOXP2 and FOXP3 FHD proteins 31

3.3.1 Protein Extraction of FOXP1, FOXP2 and FOXP3 FHDs in Preparation for Purification: 31

3.3.2 Purification of the FHDs by immobilised metal-affinity chromatography (IMAC): 32

3.4) Thrombin cleavage of the 6XHis-tag of FOXP1 and FOXP2 FHDs: 33

3.5) Evaluation of Protein Purity and Concentration Determination: 34

3.5.1 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE): 34

3.5.2 Absorbance Spectroscopy 35

3.6) DNA-binding studies of the FHDs with their cognate DNA sequence: 36

3.7) Structural Characterisation of the FOXP1, FOXP2 and FOXP3 FHDs: 37

3.7.1 Circular Dichroism Spectroscopy: 37

3.7.2 Intrinsic Fluorescence Spectroscopy: 38

3.7.3 Analytical Size Exclusion Chromatography: 39

3.8) Fluorescence Anisotropy (FA): 40

3.8.1 Homodimerization of the FOXP1, FOXP2 and FOXP3 FHDs: 41
3.8.2 Heterodimerization of the FOXP1, FOXP2 and FOXP3 FHDs: 41
Results 43
4.1) Transformation: 43
4.2) The FOXP1, FOXP2 and FOXP3 FHDs were purified by IMAC: 45
4.3) Determining protein purity: 48
4.4) FOXP1, FOXP2 and FOXP3 FHDs bind Nelson DNA: 49
4.5) Structural Characterisation of FOXP1, FOXP2 and FOXP3 FHDs: 51
4.5.1 Circular Dichroism Spectroscopy: 51
4.5.2 Intrinsic Fluorescence Spectroscopy: 51
4.6) Analytical Size Exclusion Chromatography: 53
Discussion 59
References 66

List of Abbreviations

APS	Ammonium Persulphate
BSA	Bovine Serum Albumin
DNA	Deoxyribo-Nucleic Acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
FA	Fluorescence Anisotropy
FHD	Forkhead Domain
FOX	Forkhead Box Protein
HTH	Helix-Turn-Helix motif
IPEX	Immune dysregulation, Polyendocrinopathy, Enteropathy, X-linked syndrome
IPTG	Isopropyl β-D-1-thioglalctopyranoside
KCl	Potassium Chloride
PEG	Polyethylene Glycol
SDS-PAGE	Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis
SOC	Super Optimal broth with Catabolite repression
TEMED	Tetramethyl ethylenediamine
TF	Transcription Factor
Tris-HCl	Tris(hydroxymethyl)aminomethane Hydrochloride

List of Figures

Figure 1: Common structural classes of DNA-binding domains 11

Figure 2: Dimerization of transcription factors 14

Figure 3: Schematic diagram showing the general architecture of FOXP transcription factors 16

Figure 4: Comparison of the forkhead domains of FOX proteins folded into the winged helix motif 17

Figure 5: Sequence alignment of the forkhead domain of the FOXP proteins and their selected siblings 19

Figure 6: FOXP3 homodimerization and oligomerization via the leucine zipper motif 21

Figure 7: The domain swapping process and domain-swapped dimers of FOXP3 and FOXP3 forkhead domains 24

Figure 8: Structure of FOXP2 and FOXP3 domain-swapped dimer bound to DNA 25

Figure 9: SDS-PAGE gels (16%) showing the induction trials of FOXP1 and FOXP3 FHDs, respectively. 44

Figure 10: Purification of the forkhead domains of FOXP1, FOXP2 and FOXP3 by immobilised metal affinity chromatography 46

Figure 11: Molecular weight determination of the forkhead domains of FOXP1, FOXP2 and FOXP3 proteins 47

Figure 12: Cleavage of the 6XHis-tags of the FOXP1 and FOXP2 FHDs 48

Figure 13: Absorbance spectra of the FOXP1, FOXP2 and FOXP3 FHDs 49

Figure 14: Electrophoretic Mobility Shift Assay gels of FOXP1, FOXP2 and FOXP3 FHDs 50

Figure 15: Far-UV circular dichroism spectra of FOXP1 (green), FOXP2 (blue) and FOXP3 (red) forkhead domains (FHD) 51

Figure 16: NMR and Crystal structures of the FOXP1, FOXP2 and FOXP3 FHDs, indicating the location of their tryptophan residues 52

Figure 17: Fluorescence spectra of the FOXP1, FOXP2 and FOXP3 FHDs at 4 μM monomer concentration 53

Figure 18: The characterisation of the oligomeric states of the FOXP1, FOXP2 and FOXP3 FHDs 55

Figure 19: Graph of [monomer]^2 against [dimer] constructed to obtain K_D for the homotypic association of the FOXP1, FOXP2 and FOXP3 FHDs 56

Figure 20: Fluorescence anisotropy assays of the homodimerization of the FHDs 57

Figure 21: Fluorescence anisotropy assays of the heterodimerization of the FHDs 58

Chapter 1

Introduction

1.1) Transcription Factors

The up- and down-regulation of genes form the basis of gene regulation and is governed by a large group of diverse proteins known as transcription factors (TFs). Essentially, TFs regulate gene expression through their interactions with specific DNA sequences in the *cis*-regulatory regions and therefore all TFs contain a domain for DNA binding (Jacob and Monod, 1961; Zhu and Huq, 2011). Since these DNA binding domains are conserved in structure, TFs are classified into families based on the structure of their DNA binding domain (Latchman, 1993). These proteins contain additional domains for dimerization and transcriptional activation or repression, making them modular in nature (Yang, 1998). Since their discovery, over 1 600 TFs have been identified in the mammalian genome (Lambert *et al*., 2018). When bound to DNA, these TFs may either promote or prevent the recruitment of the transcriptional machinery to the gene promoter region and thereby either activate or repress transcription (Castellanos *et al.*, 2020). In turn, by regulating gene expression, TFs coordinate numerous critical biological processes including, but not limited to, reproduction, immunity, development, and tissue specificity (Villicaña *et al.*, 2014; Zhou *et al.*, 2017). Examples of common DNA-binding domain classes include the leucine zipper, zinc finger, helix-turn-helix (HTH) and helix-loop-helix (Figure 1) (Pabo and Sauer, 1992). The leucine zipper motif encompasses two alpha helices that form a coiled-coil structure (Ma *et al*., 1994; Yesudhas *et al*., 2017). The zinc finger motif involves the coordination of a zinc ion by cysteine and histidine residues, which stabilizes the protein-DNA interaction (Pabo and Sauer, 1992). The basic helix-loop-helix motif comprises of two alpha helices separated by a loop whilst in the helix-turn-helix motif, a turn separates the two alpha-helices, which fits into the major groove of the DNA to facilitate hydrogen bonding with the base pairs (Ma *et al*., 1994; Yesudhas *et al*., 2017). Since the DNA in every cell of a multicellular organism is similar, tissue specificity is achieved through unique expression patterns of TFs (Sonawane *et al.*, 2017). For instance, specific TFs may be expressed in a particular tissue and repressed in another. This regulation of TF activity through activation and repression is controlled either through specific stimuli, alterations in protein-protein associations, the binding of a ligand or post-translational modifications such as

phosphorylation (Latchman, 1993; Pabo and Sauer, 1992). In this way, the expression of a specific set of genes is controlled.

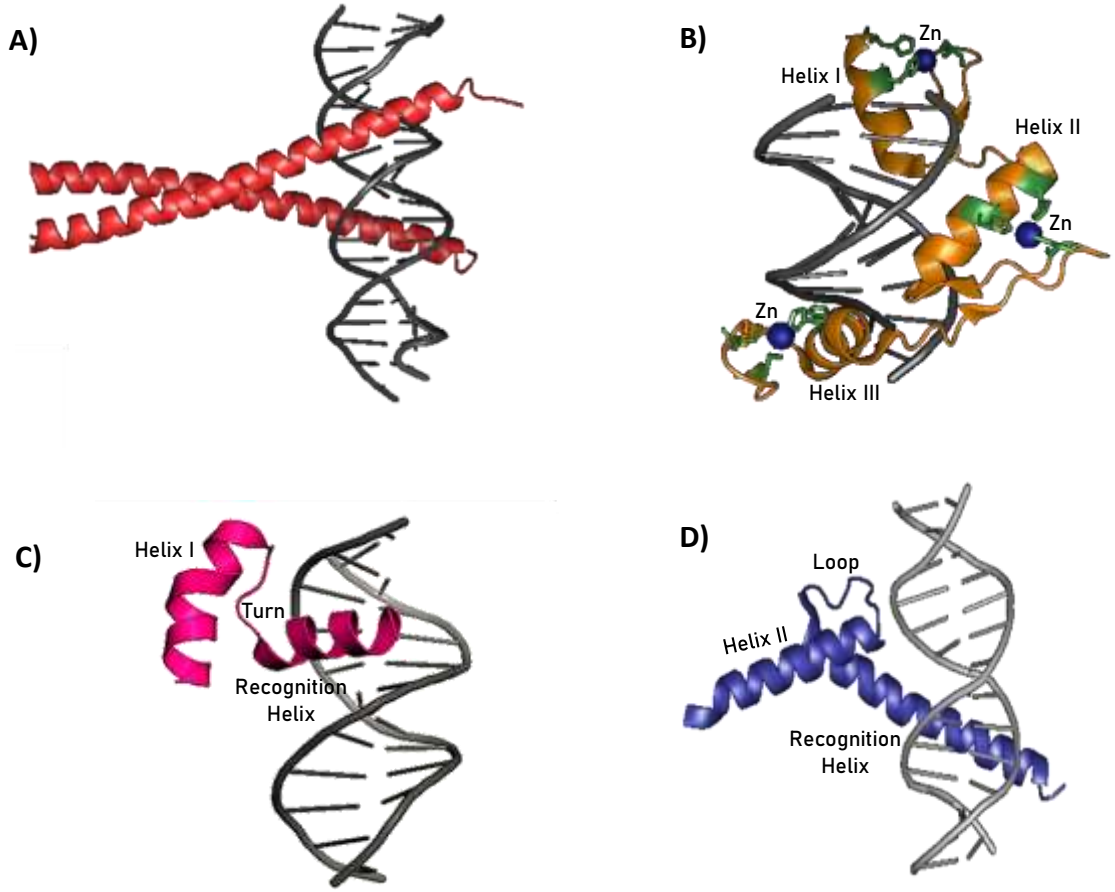


Figure 1: Common structural classes of DNA-binding domains. The DNA-binding domains of all transcription factors exhibit a particular motif used to classify them. A) The leucine zipper motif encompasses two alpha-helices that form a coiled-coil structure. Both alpha-helices interact with DNA via the major groove to regulate transcription. The leucine-zipper structure was rendered using PyMol (PDB: 1ysa). B) The general structure of a basic zinc finger domain, consisting of the coordination of zinc ions by cysteine and histidine residues, which stabilize the protein-DNA interaction. To interact with DNA, helix I, helix II and helix III bind to the major grooves of DNA. PyMol was used to render this structure (PDB: 4m9e). C) The basic structure of a helix-turn-helix domain, composed of a turn separating two alpha-helices. The recognition helix (helix II) interacts with the major groove of DNA to facilitate hydrogen bonding with the base pairs. The structure was rendered by PyMol (PDB: 5d5v). D) Structure representing a basic helix-loop-helix motif protein comprised of two alpha-helices separated by a loop. The recognition helix (helix I) interacts with the minor groove of DNA. This structure was rendered by PyMOL (PDB: 1mdy). Adapted from (Ma et al., 1994; Yesudhas *et al*., 2017)

The regulation of gene expression presents a very complex process. For successful transcription to occur, TFs need to bind either to the promoter or enhancer regions of the target gene. In the case of transcriptional repression, TFs may bind to the silencer regions. In all these regions, the TFs recognize a short sequence of approximately 6-10 base-pairs in length (Badis *et al.*, 2009; Bryne *et al.*, 2008). Both specificity and affinity are important in the binding of TFs to specific DNA sequences (Rohs *et al.*, 2010). When a TF recognizes and binds its specific site on the basis of the shape and charge complementarity between the bases of DNA and the amino acid side chains, this is referred to as specificity (Rohs *et al.*, 2010). The strength and number of interactions occurring between them is linked to the affinity a TF has for a particular sequence such that a low-affinity or high-affinity between the TF and DNA suggests a weaker or greater attraction, respectively (Rohs *et al.*, 2010). However mere binding of TFs to these target sites is insufficient (Latchman, 1993). These proteins are also required to interact with additional factors including other TFs, cofactors and RNA polymerase II itself, using their activation domains to mediate transcription (Latchman, 1993).

Since numerous genes are required to be expressed simultaneously in each eukaryotic cell, a crucial feature required by all TFs is the ability to quickly recognize their target sequence (Bintu *et al.*, 2005). This may be achieved by two distinct mechanisms of diffusion, functioning together to accelerate the process for a TF to locate its target site (Esadze *et al.*, 2014). Firstly, due to the fact that TFs are able to bind with low affinity to any DNA sequence (von Hippel and Berg, 1989), this non-specificity allows the TFs to slide along the DNA in search for their high affinity target site (Berg *et al.*, 1981; Tempestini *et al.*, 2018). The second mechanism involves the transfer of the TFs from one region of DNA to another in transient spatial proximity (Gorman and Greene, 2008). In addition to DNA binding specificity and affinity, oligomerization of TFs also enhances accessibility of the TF target sites (Amoutzias *et al.*, 2008). Indeed, numerous eukaryotic TFs regulate gene expression through the fundamental process of oligomerization (Amoutzias *et al.*, 2008).

1.1.1 Transcription Factor Dimerization

Dimerization is a form of oligomerization process that occurs when two monomers interact to form a functional protein complex (Klemm *et al.*, 1998). Dimers formed between two identical and non-identical protein monomers are referred to as homo- and heterodimers, respectively. A number of transcriptional regulators consist of specialized domains for protein-protein

oligomerization (Hu *et al.*, 1990). The oligomerization process occurs either at the DNA-binding domain or through adjacent domains such as a leucine zipper motif (Funnell and Crossley, 2012). Most TFs require the formation of active dimers to bind DNA and fulfil their functions (Funnell and Crossley, 2012). This is accomplished by specific DNA-binding structural classes that depend on specific homo- and heterodimer formation to generate a functional DNA recognition site (Hu *et al.*, 1990). For instance, several DNA-binding structural classes including the leucine zipper and the helix-loop-helix domains require dimerization before the TFs are able to bind DNA (Figure 2) (Funnell and Crossley, 2012).

The dimerization process may either have a positive (activation) or a negative (repression) effect on gene transcription. Indeed, many of the biological processes, including metabolism, the cell cycle, apoptosis and even development itself, rely on the formation of TF dimers (Amoutzias *et al.*, 2008). Dimerization enhances the DNA-binding specificity of the TFs because it results in an increased surface area, providing greater potential for DNA-protein interactions (Klemm *et al.*, 1998). Furthermore, since dimerization is mediated via highly conserved domains, dimers increase the stability of the TF (Marianayagam *et al.*, 2004). A single TF monomer is able to interact with different protein partners to form dimers, and the different dimers exhibit different properties, thereby mediating differential gene regulation and functions (Amoutzias *et al.*, 2008).

The leucine zipper is an integral example of a dimerization motif that has been studied immensely (Kouzarides and Ziff, 1988; Landschulz *et al*., 1988; O'Shea *et al*., 1989). Structurally, this motif is alpha-helical and is identified by the presence of leucine residues at every seventh position (Hu *et al.*, 1990). During homo- and heterodimerization, two leucine repeats form a coiled-coil conformation by aligning in a parallel fashion as observed in the GCN4 (Figure 2B) and Jun-Fos transcription factors (Kouzarides and Ziff, 1988; Landschulz *et al*., 1988; O'Shea *et al*., 1989). The Jun and Fos proto-oncogenes exhibit a role in signal transduction pathways and have a strong tendency to form heterodimers with each other, consequentially binding DNA with high affinity (Kouzarides and Ziff, 1988). These heterodimers therefore offer increased complexity which in turn offers increased stability as well as regulation of binding site accessibility (Kouzarides and Ziff, 1988).

The helix-loop-helix motif has also been described for a class of DNA-binding TFs with a dimerization domain (Murre *et al*., 1989; Jones, 2004). A number of TFs involved in the regulation of cellular proliferation and differentiation contain the helix-loop-helix motif (Murre

et al., 1989; Jones, 2004). The defining feature of proteins bearing this motif is the remarkable ability of the helix-loop-helix to form heteromeric associations with a variety of proteins similarly bearing the helix-loop-helix motif (Murre *et al*., 1989; Jones, 2004). For instance, the muscle-specific factors MyoD and myogenin, as well as the immunoglobulin gene enhancer binding proteins E12 and E47 are capable of forming heteromeric complexes with each other (Murre *et al*., 1989; Jones, 2004).

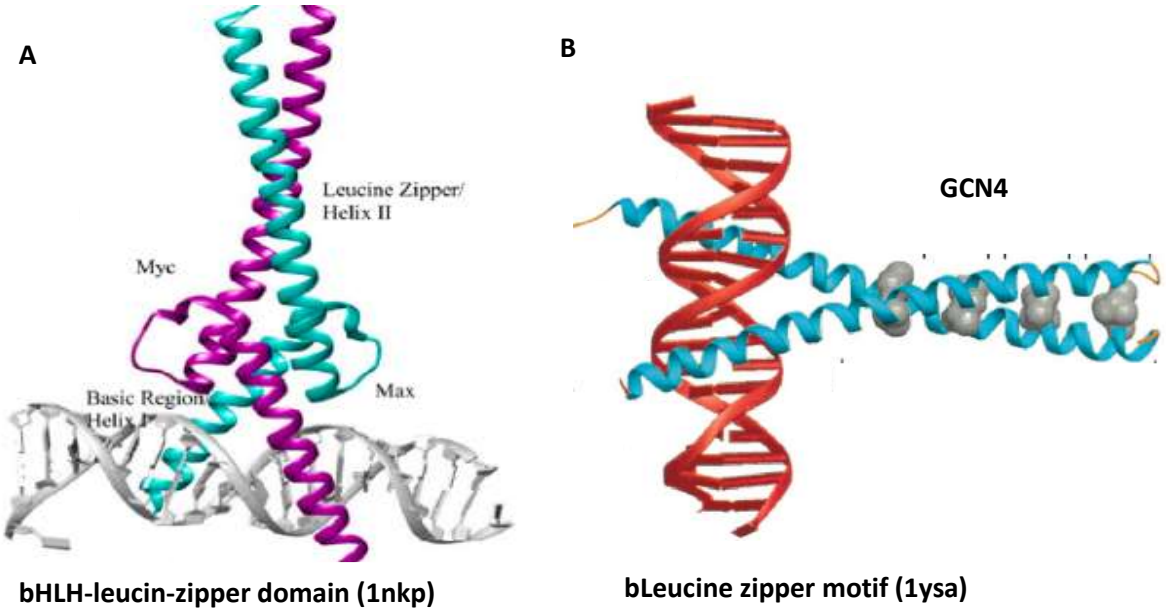


Figure 2: Dimerization of transcription factors. A) Representing the dimerization of the two helix-loop-helix motifs. Both the Myc (purple) and Max (blue) are monomeric TFs each containing two alpha-helices (helix I and helix II) separated by a loop. Since their helix II comprises of heptad repeats of leucine residues, these alpha-helices dimerize, forming a coiled-coil leuzine zipper. This dimerization is an example of a heterotypic interaction and mediates DNA binding via helices I (PDB: 1nkp). B) The structure of the monomeric general control transcription factors. Each alpha-helix comprises of heptad repeats of leucine residues, interacting to form an active leucine zipper dimer that binds to the major groove of DNA. This interaction is an example of homodimerization, whereby two similar monomers interact to form an active dimer that is essential for transcriptional regulation (PDB: 1ysa). Dimerization of transcription factors regulate the transcriptional activity of these proteins. Adapted from (Vinson *et al*., 2002; Yesudhas *et al*., 2017).

1.2) Forkhead Box (FOX) Proteins

The forkhead box (FOX) superfamily refers to one of the largest classes of transcriptional regulators (Clark *et al.*, 1993; Thulo *et al.*, 2021). Since their discovery in a wide variety of species ranging from yeast to human, hundreds of transcription factors have been identified (Benayoun *et al.*, 2011; Golson and Kaestner, 2016; Weigel *et al.*, 1989). In humans, approximately 50 *FOX* members have been reported (Benayoun *et al.*, 2011). These proteins are unified by a highly conserved winged helix motif termed the forkhead domain (FHD), with which these transcription factors bind DNA to regulate transcription of numerous genes (Kaufmann and Knöchel, 1996; Mazet *et al.*, 2003). While the presence of the FHD is distinctive of the FOX superfamily, comparison of the sequence of the TFs has further classified them into 19 subfamilies (FOXA-FOXS) on the basis of sequence similarity both in the FHD and the remaining sequence (Hannenhalli and Kaestner, 2009). The roles that these FOX proteins play extend beyond the development of various organs such as the heart, pancreas, lungs and brain (Hannenhalli and Kaestner, 2009). They have also been established to regulate key biological processes such as metabolism, cellular proliferation and differentiation, ageing, and immunity amongst many other processes (Friedman and Kaestner, 2006; Kaufmann and Knöchel, 1996). Due to the diverse functions in which the FOX proteins partake, they are expressed ubiquitously during early development as well as throughout adult life (Lu *et al.*, 2002), and it is therefore no surprise that these proteins are implicated in a variety of diseases including congenital disorders and cancer (Lam *et al.*, 2013; Stroud *et al.*, 2006).

1.2.1 The FOXP Subfamily

1.2.1.1 Structure

The FOXP subfamily encompasses four transcription factors (FOXP1-FOXP4), all of which contain multiple domains, particularly unique to their subfamily. These proteins share a conserved leucine zipper domain and zinc finger motif in addition to the FHD (Figure 3) (Li, C. and Tucker, 1993). Furthermore, with the exception of FOXP3, the other FOXP members contain a highly acidic region of 70 amino acid residues at the C-terminus while a poly-glutamine tract is evident at the N-terminus consisting of approximately 37 to 40 residues (Li, S. *et al.*, 2004). In the FOXP3 protein, the N-terminal poly-glutamine tract is replaced by a poly-proline stretch (Xie *et al.*, 2015), and the protein is truncated at the C-terminus compared to the other FOXP proteins where it ends with approximately 11 amino acid residues after the

FHD (Schubert *et al.*, 2001). The role of each structural element of the FOXP proteins has been well studied, indicating that while the leucine zipper is required for homo- and heterotypic dimerization, the role of mediating protein-protein associations has been linked to the Cys_2/His_2 zinc finger and poly-glutamine tract, whereby this tract is crucial for the activity of repression by these FOXP proteins (Monaghan *et al.*, 1993). Indeed, disruption of the poly-glutamine stretch inhibits the repression of target gene transcription (Koon *et al.*, 2007). Interestingly, the zinc finger has also been implicated in non-direct dimerization, and this may be linked to the fact that there is a direct link between the α-helices of the zinc finger and the leucine zipper motifs (Song *et al.*, 2012).

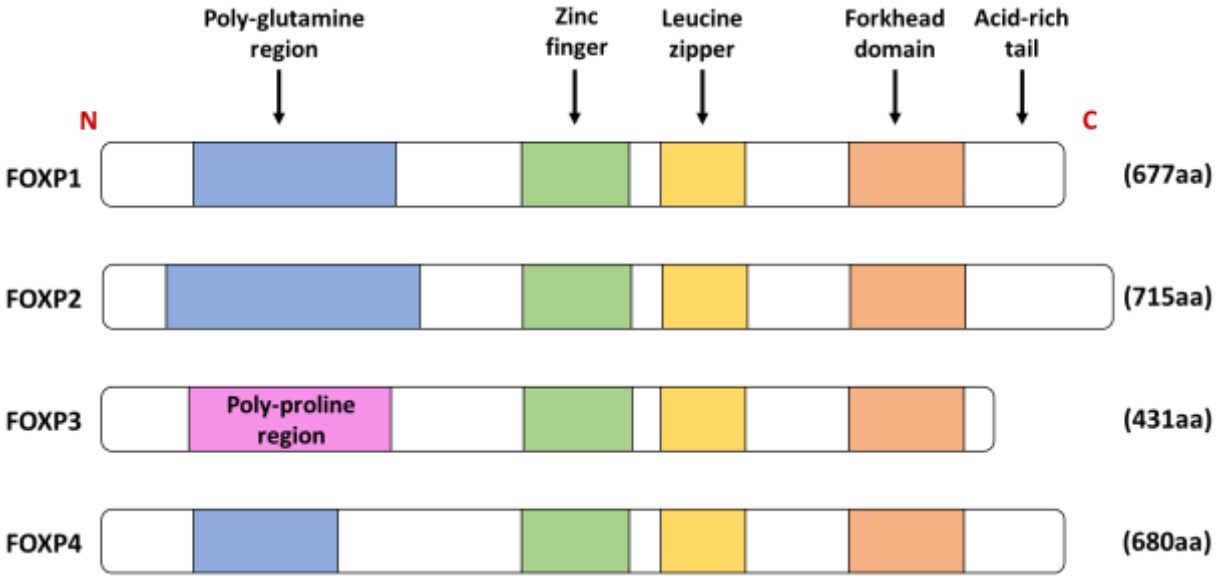


Figure 3: Schematic diagram showing the general architecture of FOXP transcription factors. The FOXP proteins consist of multiple domains including the highly conserved C terminus Forkhead domain which binds DNA, a C_2H_2 zinc finger and leucine zipper motif. While FOXP3 exhibits an N-terminal polyproline region, this is replaced by a poly-glutamine region in the other FOXP transcription factors. FOXP1, FOXP2 and FOXP4 also exhibit a C-terminal acidic tail which is absent in FOXP3. (Image constructed using Microsoft PowerPoint).

In addition to the fact that the abovementioned structural elements are particularly unique to the FOXP proteins, there are also slight differences in FHD structure of the FOXP transcription factors compared to that of the other FOX proteins. Firstly, the FHDs of FOXP proteins consist

of a 3_{10} helix, termed helix 4 (H4), which occurs in the turn between H2 and H3 (Figure 4) (Stroud *et al.*, 2006). Secondly, the wing regions W1 and W2 have been shown to be truncated and shortened, respectively (Figure 4 and 5) (Stroud *et al.*, 2006). Indeed, in conventional FOX proteins, a 5-7 amino acid insert (W1) occurs between strands S2 and S3 whereas these strands are joined closely by a simple type I turn in the FOXP proteins (Stroud *et al.*, 2006). Additionally, unlike the other FOX proteins, which contain an extended loop (W2) that is used to extensively contact the DNA, this region is replaced by a shortened helix 5 (H5) in the FOXP members, resulting in fewer contacts between H5 and DNA (Figure 4) (Stroud *et al.*, 2006). Interestingly, the FHD of most FOX transcription factors is located at the N-terminal side of the protein, but this is not the case with the FOXP proteins which exhibit their FHD at the C-terminus (Clark *et al.*, 1993). Within this canonical FHD of the P subfamily, FOXP1 shares a very high protein identity of 88% and 89% with FOXP2 and FOXP4, respectively, and a protein identity of 76% with FOXP3 (Figure 5) (Chu *et al.*, 2011) .

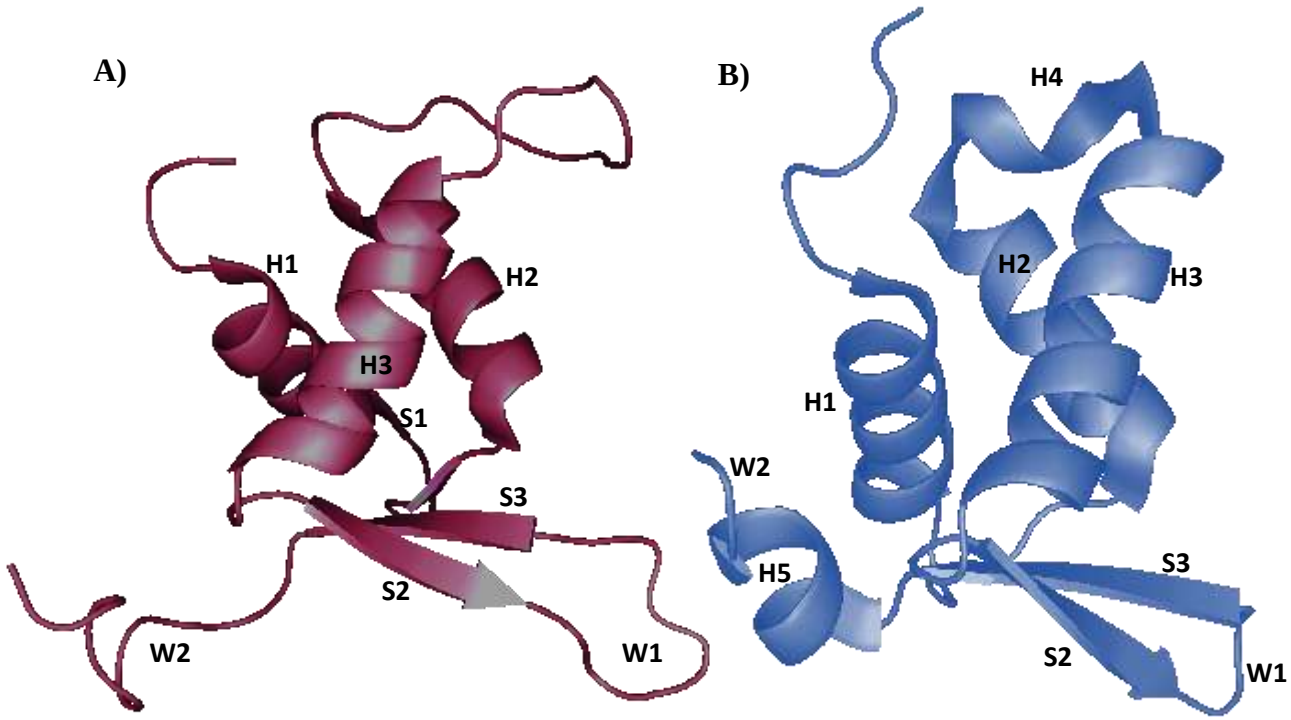


Figure 4: Comparison of the forkhead domains of FOX proteins folded into the winged helix motif. A) Forkhead domain structure of FOXO3 showing the typical architecture of FOX proteins consisting of 3 α-helices (H1, H2, H3), 3 β-strands (S1, S2, S3) and two flexible loops or "wings" (W1, W2). B) FOXP proteins however exhibit slight structural differences such that the stretch occurring between H2 and H3 is replaced with a 3_{10} helix (H4) in FOXP proteins, and W1 and W2 are truncated and shortened, respectively. PDB codes: 2UZK (FOXO3) and 2A07 (FOXP2). Structures were rendered employing PyMOL.

1.2.2.2 Physiological Roles

Approximately 55-65% sequence similarity is observed in the full length FOXP1, FOXP2 and FOXP4 transcription factors while their FOXP3 counterpart shows structural divergence (Teufel *et al.*, 2003). The sequence of a protein determines its final structure and the functions in which it will participate. Indeed, while the FOXP1, FOXP2 and FOXP4 proteins are highly expressed and exhibit overlapping developmental roles in the gut, lung and brain (Shu *et al.*, 2001), the expression of FOXP3 is limited to regulatory T cells where it participates in the immune response (Fontenot *et al.*, 2003). Rather than activating transcription, the members of the FOXP subfamily have been suggested to act as transcriptional repressors, and mutations introduced particularly in the FHD of these proteins result in disease (Li, S. *et al.*, 2004).

The FOXP1 protein was first discovered in B-cells during a transcription factor-detection screening test (Li, C. and Tucker, 1993). Since then, this protein has been established to participate in a plethora of functions including the development and differentiation of B-cells and monocytes, respectively (Hu *et al.*, 2006; Shi *et al.*, 2004). In the heart, FOXP1 has been observed to facilitate the development of the cardiac valves (Wang *et al.*, 2004), and participate in lung development (Shu *et al.*, 2001). This TF is extensively expressed in the brain where it regulates neural connectivity and also plays a major role in cognition (Wang *et al.*, 2003). Dysregulation of FOXP1 has been greatly associated with autism (Co *et al.*, 2020). Furthermore, FOXP1 has been linked to a variety of cancers, where it is considered as a tumour suppressor gene (Huebner, 2001; Zabarovsky *et al.*, 2002). This is because *FOXP1* is situated in the chromosome 3p14.1 region that often facilitates loss of heterozygosity in distinct tumours (Huebner, 2001; Zabarovsky *et al.*, 2002). Therefore the inhibition of transcriptional activities of this TF, due to decreased *FOXP1* mRNA or protein levels, causes the development and progression of cancer (Banham *et al.*, 2001).

Similarly, the FOXP2 protein is extensively expressed in the brain. Particularly, this TF is expressed in regions of the brain that are greatly associated with speech and language function and include the cerebellar Purkinje cells as well as the striatum medium spiny and cortical neurons (Enard *et al.*, 2002; Lang *et al.*, 2019). The *FOXP2* gene is situated on chromosome 7q31 (Lang *et al.*, 2019). Through genome-wide studies, this locus has been established to be susceptible to mental disorders (Lang *et al.*, 2019). Indeed, mutations in FOXP2 lead to schizophrenia (Co *et al.*, 2020; Lang *et al.*, 2019). Additionally, like most FOX proteins, FOXP2 has a diverse repertoire of functions in signal transduction pathways as well as in the

development of distinct tissues such as the kidneys, intestines, skeletal muscles, and the distal epithelia of the lungs (Bowers and Konopka, 2012; Kurt *et al.*, 2012; Shu *et al.*, 2001).

The major difference between FOXP3 and the other FOXP proteins is its ability to regulate transcription by acting both as an activator and a repressor (Li, S. *et al.*, 2004). This protein serves as the marker for the differentiation of regulatory T cells in the immune system (Fontenot *et al.*, 2003), and interruption of the FOXP3 function results in a rare immunological syndrome known as IPEX syndrome (Immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome) (Estruch *et al.*, 2018).

FOXP4, the least studied member, functions within the kidneys, lungs, liver, heart and brain where it is co-expressed with FOXP1 and FOXP2 (Teufel *et al.*, 2003). A recent study has shown that FOXP4 is expressed at high levels in adipocytes where it regulates thermogenic gene expression and function (Perie *et al.*, 2021).

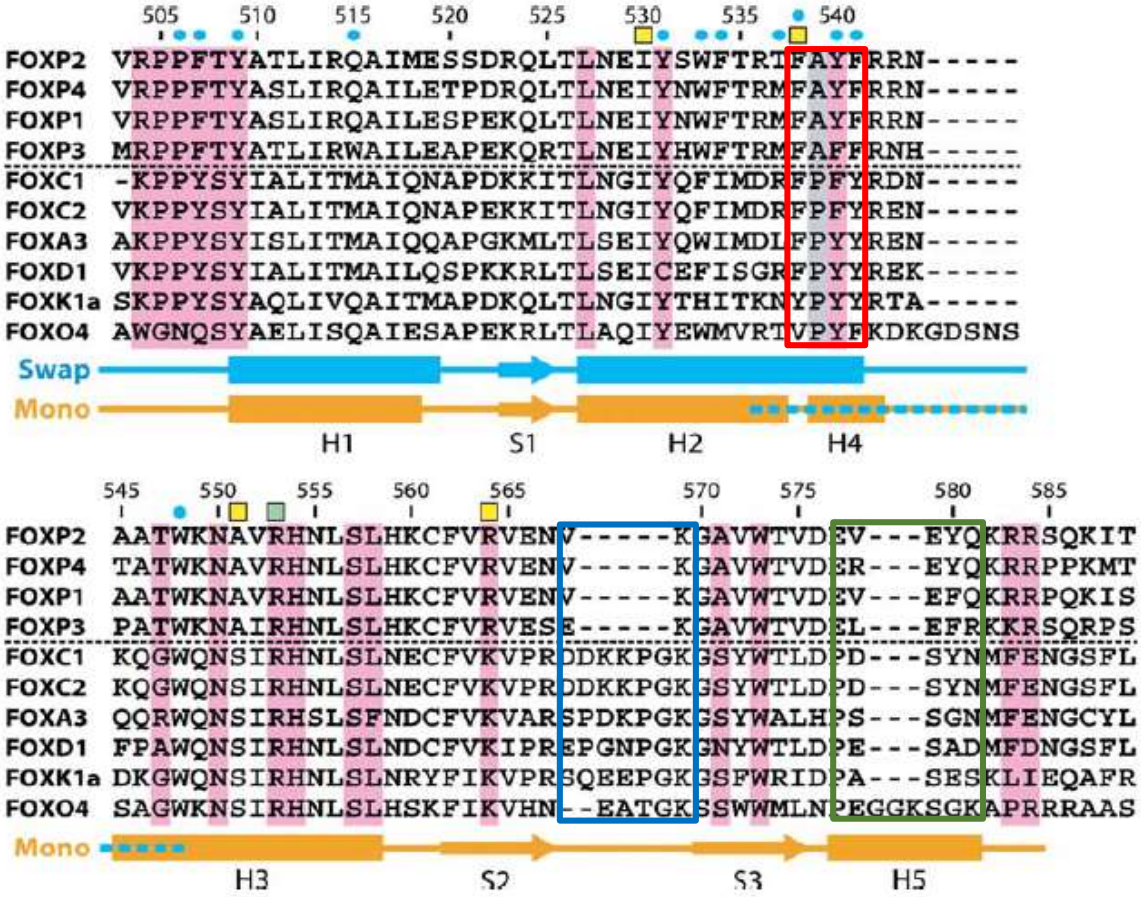


Figure 5: Sequence alignment of the forkhead domain of the FOXP proteins and their selected siblings. The FHD of FOXP transcription factors show slight differences in their sequences compared to the other FOX proteins. FOXP proteins contain a helix 4 (H4) between helices H2 and H3. Furthermore, the wing region (W1) is truncated (blue box) and (W2) is shortened, creating a helix (H5) (green box). Additionally, the sequence in the hinge loop region indicate a base substitution from proline to alanine amino acid residue in FOXP proteins (red box). The figure was adapted from Stroud *et al.*, 2006.

1.2.2 DNA Binding by FOXP Proteins

In order to carry out their transcriptional regulatory roles, transcription factors are required to interact with their target DNA sequences through their DNA-binding domain. For the FOX proteins, this domain is referred to as the canonical FHD (Kaufmann and Knöchel, 1996), and upon binding, the recognition helix (H3), contacts the major groove of the DNA in a distinct sequence-specific manner (Webb *et al.*, 2017). For instance, the FOXP proteins are capable of binding the same DNA sequence, though with varying affinities (Li, S. *et al.*, 2004; Webb *et al.*, 2017). Although, non-specific contacts between the DNA and FHD of the FOXP proteins drive DNA binding, there are also specific interactions including hydrogen bonding and ionic interactions that serve a crucial role in DNA binding (Morris *et al*., 2018). *In vitro* studies suggest that the binding of FOXP proteins, such as FOXP2, to DNA is both greatly enthalpically driven and entropically favourable which has confirmed that FOXP FHD recognises and binds specifically to the major groove of DNA (Morris and Fanucchi, 2016). These studies indicate that upon DNA binding, the FOXP2 FHD changes conformation as a result of the increased entropy (Morris and Fanucchi, 2016). Essentially, these FHD proteins lose their flexibility upon DNA binding and their backbone becomes less solvent exposed (Morris *et al*., 2018). This is supported by the fact that ionic contacts are the major factors contributing largely to DNA-FOXP2 FHD interactions, facilitating high affinity (Morris *et al*., 2018).

The FOX transcription factors are said to interact with DNA in the monomeric form (Li, S. *et al.*, 2004). However, although this holds true for most FOX proteins, those belonging to the FOXP subfamily differ in that these proteins exhibit the ability to form homo- and heterodimers (Li, S. *et al.*, 2004). Studies suggest that in order for the FOXP members to bind to DNA *in vivo*, it is crucial that they dimerize via their leucine zipper motif (Li, S. *et al.*, 2004). This may be supported by the fact that gene regulation is controlled by a crucial mechanism where transcriptional regulators perform in a combinatorial manner to generate complexes specific to their roles such as the development of organs (Silbereis *et al.*, 2016; Smith and Matthews, 2016). Therefore, transcription factors rely on the co-expression of other transcription factors such as the case with FOXP1, FOXP2 and FOXP4 (Estruch *et al.*, 2018). Although, leucine-zipper associations are a requirement for DNA binding by FOXP proteins *in vivo*, this necessity is annulled by the fact that isolated FHDs maintain their ability to bind DNA *in vitro* (Morris and Fanucchi, 2016; Stroud *et al.*, 2006).

1.2.3 FOXP Dimerization

There are two dimerization interfaces described in the FOXP proteins. As stated above, one of these interfaces is the leucine zipper, and the second interface described is the canonical FOXP FHD. The leucine zipper interface has been shown to facilitate both homo- and heterotypic associations of the full-length protein, resulting in dimers and higher order oligomers (Song *et al.*, 2012; Thulo *et al*., 2020). Associations via the leucine zipper domain mediate DNA binding of FOXP proteins and are crucial for transcriptional regulation (Sin *et al*., 2015). Indeed, specific regulation of gene transcription has been shown to be mediated by homo- and heterodimerization of FOXP1, FOXP2 and FOXP4 (Sin *et al*., 2015). Furthermore, studies suggest that the FOXP3 leucine zipper is both pivotal and sufficient to mediate heterodimerization with FOXP1 (Wang *et al.*, 2003), but there has been nothing discovered about FOXP3 heterodimerization with FOXP2 and FOXP4. The leucine zipper of FOXP3 has been solved by crystallography and shows that the motif exhibits an anti-parallel coiled coil structure that is loosely packed (Song *et al.*, 2012). Oligomerization via the FOXP3 leucine zipper is achieved by high order packing between dimers (Figure 6) (Song *et al.*, 2012). Since a high protein similarity between the members of the P subfamily is observed, the mechanisms of oligomerization may be similar, indicating that FOXP1, FOXP2 and FOXP4 may exhibit capabilities of forming oligomers through their leucine zipper.

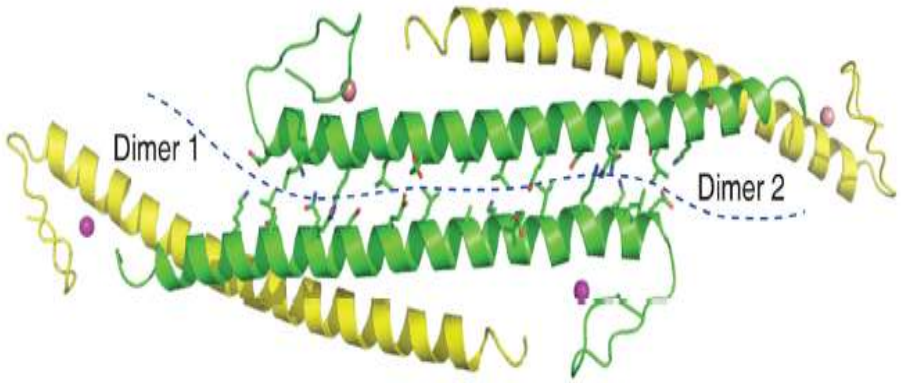


Figure 6: FOXP3 homodimerization and oligomerization via the leucine zipper motif. The FOXP3 protein has the ability to form dimers through the interactions of its leucine zipper. This protein may also form oligomers via the leucine zipper through the interactions of FOXP3 dimers. Adapted from Song *et al*., 2012.

The FOXP FHDs have been suggested to form homodimers in solution (Li, S. *et al.*, 2004). Indeed, crystal structures have shown that the FHD of the FOXP proteins exist in two forms including the monomeric form which is alike the FHDs of the other superfamily members, and the dimeric form which is a characteristic unique to the FOXP members (Stroud *et al.*, 2006).

Despite the fact that the FOXP proteins are able to dimerize via their FHD, it has been shown that, in solution, the FHDs of each FOXP protein show varying dimerization propensities (Medina *et al.*, 2016; Stroud *et al.*, 2006). For instance, under similar conditions, the FHDs of FOXP2 and FOXP1 exist as a combination of monomers and dimers, while FOXP3 tends to exist solely as a homodimer (Medina *et al.*, 2016; Stroud *et al.*, 2006). Indeed, the FOXP3 homodimer is the most stable dimer compared to that of FOXP2 and FOXP1 (Figure 7B) (Stroud *et al*., 2006). The differences in stability between the FOXP3 homodimer and the other FOXP homodimers is suggested to be the result of the orientation of the H1 helices. In the FOXP2 homodimer, the H1 helical elements are turned away from each other whereas in the FOXP3 homodimer, these helices generate a stabilising secondary interface (Bandukwala *et al.*, 2011).

A network of core hydrophobic amino acid residues stabilizes the FHD homodimers of all FOXP proteins. Particularly, in the FOXP3 FHD, this hydrophobic interface is strongest and includes tyrosine (Tyr364), tryptophan (Trp366, Trp381) and phenylalanine (Phe367, Phe371, Phe373, Phe374) residues, which are responsible for the increased stability of the FOXP3 FHD dimer (Stroud *et al*., 2006; Bandukwala *et al.*, 2011). Unlike FOXP3, studies suggest that at physiological concentrations, the FHD of FOXP2 exists predominantly as a monomer (Perumal *et al*., 2015). In accordance with the notion that the dimer is stabilised by a hydrophobic interface, a residue substitution has been suggested to be associated with FOXP3 FHD's increased propensity for dimerization (Perumal *et al*., 2015). In the FHD of FOXP2, the tyrosine amino acid residue substitution to a phenylalanine (Y540F), resembling the residue found in FHD of FOXP3 at a similar position, has been shown to significantly increase the propensity of FOXP2 FHD to dimerize (Perumal *et al*., 2015).

The ability of the FOXP transcription factors to dimerize through their FHDs occurs through a specialized process known as three-dimensional domain swapping (Figure 7A) (Stroud *et al.*, 2006; Bandukwala *et al*., 2011; Medina *et al.*, 2016).

1.2.4 Domain Swapping

Domain swapping is an oligomerization process whereby specific, yet identical, structural elements of associated monomers are exchanged to form intertwined dimers and oligomers (Figure 7) (Rousseau *et al.*, 2003). The structural elements that are exchanged may vary considerably with reference to size. That is, domain swapping can occur simply between secondary structural elements such as α-helices and β-strands or even between entire domains (Rousseau *et al.*, 2003). The exact mechanism of domain swapping is yet to be discovered. However, it appears that domain-swapping proteins do not all follow the same mechanism (Liu, Y. and Eisenberg, 2002). Although approximately 40 proteins that undergo this oligomerization process have been studied (Kundu and Jernigan, 2004), only a few researchers have investigated the domain-swapping mechanisms (Hayes *et al.*, 1999; Kuhlman *et al.*, 2001; Rousseau *et al.*, 2003; Xu *et al.*, 1998). Some of these researchers have particularly focused on the theory underlying the monomer-to-dimer transition mechanism (Gouldson *et al.*, 1998; Xu *et al.*, 1998). There is still so much that remains unknown, however, it is clear that the compact monomers must unfold, breaking the intramolecular interactions between the protein and the region of domain swapping, to achieve an open monomeric state before they are able to intertwine by regaining intermolecular interactions that are similar to those of the monomer (Bennett *et al.*, 1995; Kundu and Jernigan, 2004; Liu, L. *et al.*, 2012; Rousseau *et al.*, 2003). In all cases of domain swapping, it is also clear that the hinge loop is pivotal for the manifestation of this oligomeric state (Liu, Y. and Eisenberg, 2002). Several domain-swapping proteins are functional only when they reach this oligomeric state (Kundu and Jernigan, 2004).

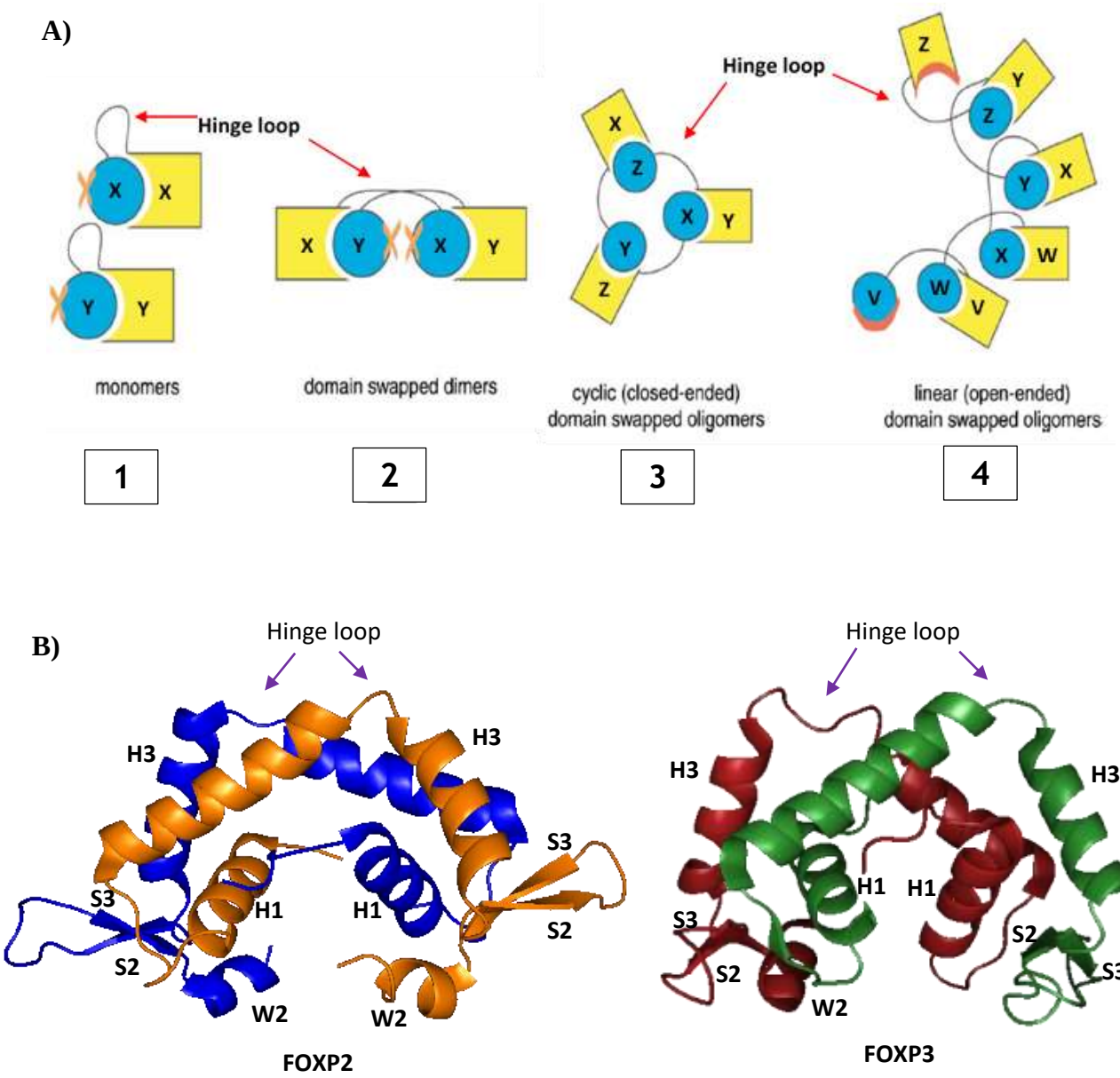


Figure 7: The domain swapping process and domain-swapped dimers of FOXP3 and FOXP3 forkhead domains. A) Schematic diagram showing the basic principles of domain swapping. 1) Representation of monomers consisting of motifs (blue circle and yellow square), separated by the hinge loop 2, 3 and 4) During domain swapping, certain elements are exchanged. 2) Domain-swapped dimer. 3) Higher order oligomers are able to arrange themselves in a cyclic conformation or 4) in an extended linear conformation. (Adapted from Rousseau *et al*., 2003). B) Comparison of the domain-swapped dimers of FOXP2 (each monomer of the domain-swapped dimer is distinguished either by orange or blue) and FOXP3 (each monomer of the domain-swapped dimer is distinguished either by red or green). The elements that are exchanged include a helix (H3) and two adjacent strands (S2 and S3). The FOXP3 FHD homodimer appears more compact and globular than FOXP2 FHD dimer. Compared to that of FOXP2, the hinge loop region in FOXP3 is longer. While the H1 helical elements are turned away from each other in FOXP2, those in FOXP3 present an additional stabilising interface (Bandukwala *et al*., 2011). PDB codes: 3QRF (FOXP3) and 2A07 (FOXP2). Structures were rendered employing PyMOL.

1.2.5 Domain Swapping of the FOXP FHDs

The FOXP proteins exhibit the ability to form domain-swapped dimers via their FHDs. The FHD regions of exchange include a helix (H3), and two adjacent strands (S2 and S3) (Figure 7B) (Stroud *et al.*, 2006). Domain swapping in the FOXP subfamily is mediated predominantly when helix 2 is extended into this hinge loop (Bennett *et al.*, 1995; Stroud *et al.*, 2006). A highly conserved region in helix 2 consisting of a critical sequence, FPYF, is a feature of the FOX family with the exception of the FOXP proteins (Figure 5) (Morris and Fanucchi, 2016). Studies suggest that the members of the FOXP subfamily exhibit an alanine residue instead of the proline which is essentially responsible for extension of helix 2 (Stroud *et al.*, 2006; Morris and Fanucchi, 2016). Indeed, evidence has shown that a residue substitution of alanine (Ala539) in the FHD of FOXP2 to proline (as is highly conserved in other members of the FOX superfamily) does indeed prevent extension into hinge loop and, in turn, prevents domain swapping (Stroud *et al.*, 2006). Recently, it has been established that a conserved histidine residue in the FOXP proteins (His59 in FOXP1) demonstrates pH-dependence of dimer stability. The deprotonation of this specific histidine residue favours the dimer state as a result of increased rates of association (Medina *et al.*, 2019).

It is still unclear what the physiological relevance of domain swapping is. However, this event has been suggested to assist in bridging two DNA molecules in order to facilitate long range chromosomal interactions that are required for the functions of these proteins (Figure 8) (Chen, Yongheng *et al.*, 2015; Webb *et al.*, 2017).

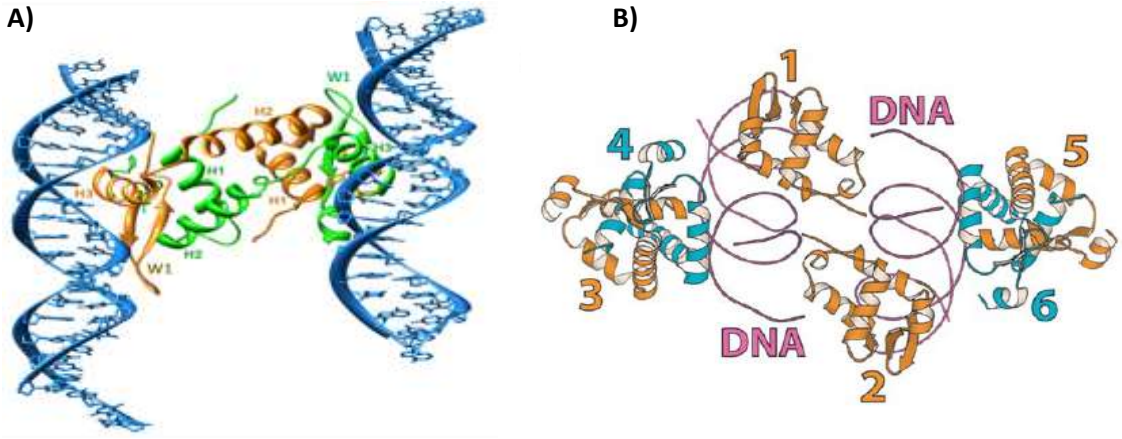


Figure 8: Structure of FOXP2 and FOXP3 domain-swapped dimers bound to DNA. The formation of the FOXP2 and FOXP3 dimers through the mechanism of domain-swapping involves the exchange of helix 3 (H3) and beta strands S2 and S3 through the extension of helix 2 (H2). A) FOXP3 FHDs exists predominantly as domain-swapped dimers. The interaction of two monomers in this manner is crucial for FOXP3 for its transcriptional activities by bridging two segments of DNA in order to facilitate long range chromosomal interactions. B) FOXP2 FHDs exists as a combination of monomers (orange) and dimers (blue and orange). Unlike the FOXP3 FHD, the FOXP2 domain-swapped dimer binds weakly to one DNA strand only. Adapted from (Lalmansingh *et al.*, 2012; Stroud *et al.*, 2006).

1.3) Rationale

The forkhead domains of the FOXP subfamily exhibit the remarkable ability to form domain-swapped homodimers, which is not evident in other FOX transcription factors. This implies that these homodimers may have an impact in the manner in which these proteins bind DNA and regulate transcription. Indeed, the FHD dimers have been suggested to have physiological relevance by binding DNA with both dimer subunits and mediating interchromosomal associations (Chen *et al*., 2015). To date, domain-swapped homodimerization has only been documented for the FOXP1, FOXP2 and FOXP3 FHDs (Chu *et al*., 2011; Stroud *et al*., 2006; Medina *et al*., 2016). The ability of the FOXP4 FHD to homodimerize is yet to be elucidated. For this reason, along with the fact that FOXP4 is the least studied member of the subfamily, FOXP4 was excluded from this study.

While the FHDs of the FOXP1, FOXP2 and FOXP3 proteins are able to undergo homodimerization, each exhibits a unique propensity to dimerize. The FOXP2 FHD is predominantly monomeric in solution (Stroud *et al*., 2006) whilst FOXP3 FHD remains exclusively dimeric (Bandukwala *et al*., 2011). The FOXP1 FHD is suggested to exhibit the best of both worlds, existing as both monomeric and dimeric species at equilibrium (Chu *et al*., 2011; Medina *et al*., 2019). This is remarkable because, *in vivo*, the FOXP1 and FOXP2 transcription factors are co-expressed in several tissues and share overlapping roles in those tissues (Wang *et al.*, 2003). Additionally, FOXP1 shares binding sites with FOXP3 in regulatory T cells and therefore both are likely to share overlapping roles in immunity (Konopacki *et al*., 2019).

The fact that the FHD of these FOXP proteins possess the ability to dimerization, only homodimerization of the FHDs have been considered. To date, there is no work pertaining to probable heterodimerization of these FHDs, which poses as the gap in research. Therefore, this study aims to probe the protein-protein interactions of these FHDs to investigate whether heterodimerization is able to occur between the FOXP1 and FOXP2 FHDs as well as the FOXP1 and FOXP3 FHDs. This is important because if these FHD proteins do heterodimerize, this will provide insight into whether these heterodimers function in the same way as the homodimers and how these heterodimers regulate transcription. Given that FOXP2 does not appear to share physiological roles with FOXP3 and the fact that their propensities of their FHDs to dimerize do not overlap, the heterodimerization studies of their FHDs were excluded.

Chapter 2

Aims and Objectives

2.1) Main Aim:

To investigate whether a homo- and heterotypic interactions can be established between the monomeric forkhead domains of FOXP1 and FOXP2 as well as FOXP1 and FOXP3.

2.2) Objectives:

1. Overexpress the FOXP1, FOXP2 and FOXP3 forkhead domains (FHD) utilising an *E.coli* expression system, and then purify each FHD protein employing immobilised metal affinity chromatography.
2. Assess the purity of the FHDs both qualitatively and quantitatively using sodium dodecyl sulphate-polyacrylamide gel electrophoresis and absorbance spectroscopy, respectively.
3. Confirm the functional integrity of the FHD proteins by investigating the interaction of these FHDs with their cognate DNA sequences using electrophoretic mobility shift assay.
4. Characterize the secondary and tertiary structures of the FHDs using circular dichroism and intrinsic fluorescence spectroscopies, respectively.
5. Determine the oligomeric states of the FHDs using concentration-dependent size exclusion-high performance liquid chromatography and determine the dissociation constant for each FHD protein.
6. Monitor the homotypic associations of the FOXP1, FOXP2 and FOXP3 FHDs using fluorescence anisotropy and determine dissociation constants of the interactions.
7. Determine whether a heterodimerization can be established between FOXP1 FHD and FOXP2 FHD as well as FOXP1 FHD and FOXP3 FHD utilising fluorescence anisotropy and determine dissociation constants of the interactions.

Chapter 3

Materials and Methods

Materials:

The competent cells were purchased from New England Biolabs (Canada). Both the ampicillin salt and isopropyl β-D-1-thiogalactopyranoside (IPTG) were purchased from Melford (USA). DNase and lysozyme were obtained from Merck (Germany) and Sigma Aldrich (USA), respectively. Thrombin was purchased from Sigma Aldrich (USA). From Inqaba Biotechnical Industries (South Africa), dithiothreitol (DTT) was purchased and the Nelson DNA was obtained from IDT, Whitehead Scientific (South Africa). All other reagents and chemicals were standard analytical grade.

Methods:

3.1) DNA Plasmids Encompassing Sequences Encoding the FOXP1, FOXP2 and FOXP3 FHDs and DNA Plasmid Transformation:

3.1.1 DNA Plasmids bearing sequences encoding the FOXP1, FOXP2 and FOXP3 FHDs:

In order to carry out the objectives stated in this study, it was crucial to obtain plasmid vector bearing sequences encoding the FOXP1, FOXP2 and FOXP3 FHD proteins for transformation and expression. Therefore, the coding sequences of these proteins were codon-optimized (Genscript, USA). Codon optimization is a key step in recombinant protein expression and refers to the establishment of synonymous substitutions in the sequence encoding a protein of interest to ensure that there are no restriction sites entrenched in the sequence whilst manipulating its level of expression based on the expression system that will be used (Gustafsson *et al.*, 2004). In the case of this study, the coding sequences of the FHD proteins were codon-optimised specifically for expression in an *Escherichia coli* bacterial expression system and involved looking at codons that are frequently used in bacteria and thereafter utilizing redundancy of the genetic code to mimic those specific *E. coli* codons. For each FHD sequence, a hexahistidine tag (6XHis-tag) followed by a thrombin restriction site were

introduced to the N-terminus of the coding sequences in order to assist with downstream purification of the proteins (Genscript, USA). For DNA plasmid preparation, each codon-optimised sequence was inserted into the multiple cloning site of separate pET-11a plasmids (Novagen, Germany), under the control of a T7 promoter.

3.1.2 Transformation of E.coli Bacterial Cells:

T7 Express pLysS *E.coli* cells were transformed with synthesized FOXP1, FOXP2 and FOXP3 FHD plasmids. These cells were selected as host due to their robustness and dependability in addition to their ability to offer high expression yields of proteins encoded by genes that are under control of the T7 promoter.

Initially, 5 µl DNA plasmid was incorporated with 50 µl thawed *E.coli* cells. The uptake of the plasmids by the cells was facilitated through heat-shock (42°C) over 45 seconds. This is because the high temperature subjected to the cells causes changes in the state of fluidity, and therefore permeability, of their cellular membrane. After heat-shock, the newly-transformed cells were returned to 4°C for 5 minutes. The cells were then subjected to 950 µl pre-warmed SOC (Super Optimal broth with Catabolite repression) media and incubated at 37°C for 1 hour with shaking at 230 rpm. The constituents making up the SOC medium included tryptone (1%, w/v), yeast extract (0.5%, w/v), NaCl (10%, w/v) and 0.02 M glucose which facilitated the recovery of the cells from heat-shock, ensured that transformation was retained, and, lastly, enhanced cell growth by providing nutrients whilst maintaining constant pH and osmolarity.

The cells were then spread on LB (lysogeny broth) agar plates, prepared with 1.5% agar, NaCl (1%), tryptone (1%) and yeast extract (0.5%). The aim was to solely select for pET-11a-positive transformants whilst prohibiting growth of cells that were not transformed with DNA plasmid. Therefore, given that the pET-11a plasmids contain an ampicillin-resistant gene, the broth was supplemented with 100 µg/ml ampicillin for selection. Growth of colonies was enabled over 16 hours at 37°C.

After incubation, single colonies were selected and transferred into flasks containing 100 ml 2X YT media [1% yeast extract, 1.6% tryptone, 0.5% NaCl and 100 µg/ml ampicillin]. These flasks were then incubated overnight at 37°C on the 230 rpm shaker. Following incubation, glycerol stocks of the cultured cells were prepared in eppendorf tubes with 500 µL cultured cells and 500 µL sterile glycerol (80%) and then frozen at -80°C until required for expression.

By creating glycerol stocks of transformed bacterial cells, these cells could be stored for lengthy periods of time. This is because glycerol prevents bacterial cell damage and keeps the cells alive by stabilising the frozen bacteria

3.2) Expression of the FOXP1, FOXP2 and FOXP3 FHDs:

For this study, it was necessary to conduct expression trials on FOXP1 and FOXP3 FHDs to determine optimal conditions for expression particularly in T7 Express pLysS *E.coli* cells. Since these proteins are very similar to the FOXP2 FHD (which has been previously studied extensively in our laboratory), expression trials for FOXP1 and FOXP3 FHDs were initially conducted utilising a 0.5 mM IPTG (isopropyl-1-thio-β-D-galactopyranoside) induction concentration at 20°C for 2 – 22 hours to examine whether these proteins could express at similar conditions as their FOXP2 FHD counterpart. Whilst FOXP1 FHD showed expression under these conditions, there was no expression observed for FOXP3 FHD (*see section 4.1*). Therefore additional expression trials for FOXP3 FHD were conducted utilising 0.2 mM and 0.5 mM IPTG concentrations at 30°C for 2 – 22 hours to determine optimal conditions for overexpression.

Once optimal conditions were determined for these FHDs, 1 ml glycerol stocks (prepared in section 3.1.2) of FOXP1, FOXP2 and FOXP3 FHDs were thawed and inoculated into separate flasks containing 100 ml 2X YT media, supplemented with 100 µg/ml ampicillin. These flasks were incubated for 16 hours at 37°C while shaking at 230 rpm to promote the growth of the cells.

The aim was to overexpress sufficient amounts of protein for downstream applications. Therefore FOXP1, FOXP2 and FOXP3 FHDs were expressed in 3 L batches as follows: the overnight cultures were transferred into fresh 3 L 2X YT media, supplemented with 100 µg/ml ampicillin, in a 1:50 dilution. These bacterial cells were allowed to grow at 37°C with 220 rpm shaking until an optical density (OD_{600}) between 0.6 and 1.0 was reached. Thereafter, expression of FOXP1 and FOXP2 FHDs was induced with 0.5 mM IPTG and 0.2 mM IPTG was used for FOXP3 FHD. Cells expressing FOXP1 and FOXP2 FHDs were incubated at 20°C for 22 hours stirring at 220 rpm whilst FOXP3 FHD was allowed to overexpress at a shaking speed of 150 rpm for 4 hours at 30°C .

When the incubation period was complete, the bacterial cells with overexpressed proteins were harvested at 4°C, where centrifugation was performed at 5000 xg for 15 minutes, resulting in the production of a pellet which, in turn, was re-suspended in HisTrap^{TM} Equilibration Buffer [20 mM Tris-HCl, pH 7.5, 500 mM NaCl, and 30 mM imidazole] and stored at -20 °C until required for purification.

3.3) Purification of the FOXP1, FOXP2 and FOXP3 FHD proteins

3.3.1 Protein Extraction of FOXP1, FOXP2 and FOXP3 FHDs in Preparation for Purification:

Cell lysis is a key step in enabling the isolation of proteins from bacterial and mammalian cells. The major difference between these cell types is that the bacterial cells have a sturdy cell wall and thus isolation of the expressed proteins from these cells required both enzymatic and mechanical means of cell lysis.

Initially, the cells that were re-suspended in equilibration buffer were thawed to room temperature. Thereafter, 10 µg/ml lysozyme, 10 µg/ml PMSF (phenylmethylsulfonyl fluoride) and 1 µg/ml DNase were added to the re-suspensions which were left rotating at room temperature for 20 minutes. The use of the lysozyme was to assist with bacterial cell lysis by destabilizing the cell wall, whilst the use of PMSF and DNase was to maintain the integrity of the protein by inhibiting protease activity and disintegrating the bacterial DNA, respectively, thereby significantly reducing DNA contamination. This was followed by the sonication of the cells which was performed at 4°C to completely lyse the cells whilst preventing protein denaturation caused by heat. The sonication process involved 5 rounds of 30 second pulses whereby a sonicator probe was inserted into the solution of re-suspended cells and sound waves of high frequencies were generated, breaking the cell via shearing forces. Upon successful cell lysis, the cell lysates were centrifuged (18 000 xg) for 20 minutes at 4°C to separate cellular debris from the soluble proteins in the supernatant.

3.3.2 Purification of the FHDs by immobilised metal-affinity chromatography (IMAC):

Recombinant protein purification following expression is a crucial procedure in biochemical studies. The fusion of a peptide affinity tag, such as His-tag, to a protein of interest facilitates protein purification with the use of affinity chromatography (Bornhorst and Falke, 2000). As mentioned earlier, the FOXP1, FOXP2 and FOXP3 FHDs were fused with an N-terminal 6X His-tag, essentially allowing these proteins to be purified using immobilized metal-affinity chromatography (IMAC).

IMAC, a type of affinity chromatography, is a versatile technique widely-used to purify polyhistidine affinity-tagged proteins. In IMAC, transition metal ions such as cobalt (II), zinc (II), copper (II) and nickel (II) are immobilized on a stationary matrix in order to bind the side chains of specific amino acids (Bornhorst and Falke, 2000). Histidine residues have an extremely high affinity for transition metals, particularly nickel (II), due to the aromatic imidazole ring of the amino acid whose electron donor groups readily form coordination bonds with the metals (Bornhorst and Falke, 2000). In this way, only the protein of interest should be retained in the column, thereby resulting in purification of these proteins. The protein of interest is then eluted either by altering the pH of the column buffer or introducing a metal ion-binding competitor.

In this study, the FOXP1, FOXP2 and FOXP3 FHDs were purified separately by IMAC utilising an ÄKTA Fast Protein Liquid Chromatography (FPLC) system. Pre-charged 5 ml HisTrap^{TM} Ni^{2+} columns (GE Healthcare Bio-Sciences, Sweden) were used for the IMAC procedure. Initially, these columns were equilibrated with HisTrap Equilibration Buffer consisting of 20 mM Tris-Cl, pH 7.5, 30 mM imidazole and 500 mM NaCl. Following equilibration of the column, the soluble proteins that were collected from the supernatant after centrifugation of the cell lysate were filtered with 0.45 µm filters and then loaded onto the Ni^{2+}-charged columns, reinforcing the binding of the 6xHis-tag to the nickel ions. Subsequently, a high salt wash of 1.5 M NaCl was performed on the columns in order to remove any cellular DNA fragments that may be bound to the FOXP1, FOXP2 or FOXP3 FHDs. Lastly, the FHD proteins were eluted from their respective columns with a HisTrap Elution Buffer [300 mM imidazole, 20 mM Tris-Cl, pH 7.5 and 500 mM NaCl] and were subsequently collected as fractions of 2 ml. The importance of containing a high concentration of imidazole in the elution buffer was to assist eluting the histidine-tag fused proteins as the imidazole competes with the 6XHis-tag for the nickel ions.

3.4) Thrombin cleavage of the 6XHis-tag of FOXP1 and FOXP2 FHDs:

Thrombin cleavage is a procedure that involves the removal of purification protein tags that were added to a recombinant protein (Liu *et al*., 2008). Thrombin, a serine protease, is often used to cleave the tags and specifically recognises the amino acid sequence Leucine-Valine-Proline-Arginine-Glycine-Serine and cleaves between the arginine and serine residues (Waugh, 2011). Initially, the sample containing the protein tag is incubated with thrombin (Waugh, 2011). Thereafter, both the cleaved tag and thrombin are removed from the protein sample. A variety of methods may be used to achieve this and include IMAC and benzamidine chromatography.

Benzamidine chromatography is an affinity chromatography that is specifically useful for purifying serine proteases such as thrombin (Owens *et al*., 1992; Zhang *et al.*, 2013). A column is charged with benzamidine molecules and the proteins containing benzamidine-binding sites are loaded (Owens *et al*., 1992; Zhang *et al.*, 2013). These proteins are retained in the column. In this way, thrombin may be removed from recombinant proteins after thrombin cleavage.

The 6XHis-tags of the FOXP1 and FOXP2 FHDs were cleaved using thrombin. Initially, the FHDs were dialysed against thrombin cleavage buffer consisting of Tris-HCl (20 mM) of pH 8.0, NaCl (100 mM) and KCl (2 mM). Following the overnight dialysis, the FHD proteins were subjected to centrifugation at 12 000 xg for 12 minutes to remove aggregation. Thereafter, these proteins were incubated at 20°C with 1 unit of thrombin per mg of protein for 4 hours to mediate the cleavage of their N-terminal 6XHis-tags. Following incubation with thrombin, these proteins were centrifuged at 12 000 xg for 12 minutes, again removing aggregation. The FHD proteins were then purified by IMAC and benzamidine chromatography.

Initially, the benzamidine column was set up above the Ni^{2+}-charged IMAC column, and both columns were equilibrated with the HisTrap™ equilibration buffer. The FHDs were loaded onto their respective columns to separate digested FHDs from undigested proteins, simultaneously removing thrombin from the proteins. After collecting the digested FHDs, the IMAC columns were subjected to HisTrap Elution Buffer and the undigested FHDs were collected. Fractions of the cleaved FHDs that were collected from the purifications were resolved on SDS-PAGE alongside the undigested fractions for comparison and to assess the purity of the cleaved FHD proteins.

3.5) Evaluation of Protein Purity and Concentration Determination:

3.5.1 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE):

The fractions that were collected from IMAC were resolved on SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) gels. These were conducted so as to qualitatively assess the purity of the isolated FOXP1, FOXP2 and FOXP3 FHD monomers as well as to ensure that the desired protein was expressed by sizing them according to the molecular weight marker. This is a crucial technique which confirms that any information that is obtained in downstream studies is reliable.

First established in 1970, discontinuous SDS-PAGE refers to an electrophoretic technique that separates proteins solely on the basis of their size. That is, proteins of higher molecular weights migrate slower in the gel than proteins of smaller molecular weights. Initially, the protein samples to be resolved are reduced of cystines with either DTT (dithiothreitol) or β-mercaptoethanol. The samples are further denatured by heat in the presence of SDS. This anionic detergent functions by adhering uniformly to the proteins, causing the proteins to linearize and exhibit an overall negative charge. By doing so, proteins migrate from the cathode to the anode under the influence of an electric field, and are separated solely based on their size whilst excluding separation caused by other factors such as the charge and shape of proteins (Laemmli, 1970).

The denatured samples are then resolved through a discontinuous gel consisting of a stacking gel polymerized atop a separating gel. Both polyacrylamide gels are prepared differently, consisting of varying properties including the acrylamide concentration, pore size and pH. This allows electrophoresis to occur in two stages. The first stage occurs in the stacking gel whereby samples migrate and gradually form a tightly compacted protein layer at the interface between the stacking and separating gels. At the second stage of electrophoresis, protein samples enter the separating gel simultaneously and are resolved until the dye front approaches the bottom of the gel. A molecular weight marker which is run alongside the samples is used to identify the protein of interest based on its molecular weight.

Samples obtained from the purification of the FOXP1, FOXP2 and FOXP3 FHDs were prepared for electrophoresis. SDS reducing buffer was added to the samples achieving a 1:2 dilution of samples. The SDS reducing buffer used in the samples was prepared with 150 mM Tris-HCl, pH 7.0, 12% (w/v) SDS, 30% (v/v) glycerol, 6% (v/v) β-mercaptoethanol, and 0.05%

(w/v) Coomassie Brilliant Blue G-250. The samples were thereafter incubated on a heating block at 95°C for 5 minutes before they were loaded into the wells of a 4% stacking gel and resolved through a 16% tricine separating gel. Electrophoresis was conducted at 160 V for 1.5 hours. The resolved proteins were then detected by staining the gel for an hour with Coomassie stain solution prepared with 0.1 g Coomassie Brilliant Blue R-250 in 100 ml of acetic acid-methanol-water (1:5:4) solution and then left in de-stain solution overnight. The molecular weight of the FHDs were determined relative to a 3-colour pre-stained protein molecular weight marker which confirmed whether the desired protein had been isolated.

3.5.2 Absorbance Spectroscopy

The JASCO V-630 UV/Vis absorbance spectrophotometer was employed to obtain absorbance spectra for each FHD protein within the wavelength range of 240 nm and 340 nm. Subsequently, the values at 260 nm and 280 nm were substituted into the A280/A260 ratio to verify that the proteins were free of DNA contamination. Absorbance spectrophotometry was also performed to determine the concentrations of the FHD proteins whereby five serial dilutions were prepared and their measurements were recorded at 280 nm and 340 nm. Thereafter, the measurements at 340 nm were subtracted from the 280 nm values to account for aggregation. After taking aggregation into account, the concentrations were calculated as per the Beer Lambert Law:

$$c = \frac{A}{\varepsilon \,.\, l}$$

where c defines the concentration of protein (M), A presents the absorbance at 280 nm wavelength (A.U), ε is the extinction coefficient ($M^{-1}.cm^{-1}$) and l is the path length (cm). The Expert Protein Analysis System (ExPASy) ProtParam tool (http://web.expasy.org/protparam/) was used to obtain the extinction coefficient for each FHD and were determined as 22460 $M^{-1}.cm^{-1}$, 22460 $M^{-1}.cm^{-1}$ and 24980 $M^{-1}.cm^{-1}$ for FOXP1, FOXP2 and FOXP3, respectively.

3.6) DNA-binding studies of the FHDs with their cognate DNA sequence:

The electrophoretic mobility shift assay (EMSA) is a rapid technique used in studies of DNA-protein interactions. The method was first described approximately two decades ago, and is at present the widely-used method to study transcription factors (Laniel *et al.*, 2001; Hellman and Fried, 2007). Although the EMSA is primarily qualitative, under appropriate conditions it may additionally provide quantitative data that may be exploited to determine binding affinities (Hellman and Fried, 2007).

Important components involved in this procedure include a non-denaturing gel matrix, buffers, DNA probes and the proteins of interest (Laniel *et al.*, 2001). The principle of EMSA is based on the fact that free DNA possesses a smaller molecular weight than protein-bound DNA, and is therefore separated from the complex during electrophoresis due to the greater electrophoretic mobility it exhibits through the gel matrix (Hellman and Fried, 2007). The resulting shift in DNA may then be detected. In classical EMSA techniques, the shifts in DNA are detected using DNA end-labelling radioisotopes such as ^{32}P (Hellman and Fried, 2007). However, due to the dangers involved in the use of radioisotopes and the need for expensive laboratory equipment, the use of fluorophores or biotin as detection agents has been used over the years (Hellman and Fried, 2007).

In this study, EMSAs were conducted to assess the structural integrity of the FOXP1, FOXP2 and FOXP3 FHDs, ensuring that their correct folded, native three-dimensional structure was maintained following purification. Since the structure and function of proteins are closely linked, the detection of DNA-protein complex bands in EMSA is an indication of native, functional proteins. The EMSAs were conducted utilising the Nelson DNA duplex sequence: 5'-TTAGG**TGTTTAC**TTTCATAG-3' (Nelson *et al.*, 2013). This sequence is important in the study of FHDs of the FOXP subfamily because studies have shown the DNA sequence to bind effectively to the FOXP2 FHD with a fairly high affinity (Webb *et al*., 2017).

Initially, FOXP1, FOXP2 and FOXP3 FHDs were dialysed separately against 10 mM HEPES buffer, pH 7.5, containing 100 mM NaCl. Thereafter, binding reactions of each protein were prepared by incubating increasing concentrations (0 µM – 2.5 µM) of protein with 500 nM cognate DNA in binding buffer [10 mM HEPES (pH 7.5), 100 mM KCl, 1 mM $MgCl_2$, glycerol (10%, w/v), and 0.1 mg/ml BSA]. The addition of the binding buffer to the samples ensured that optimal and physiological conditions of salt concentration and pH were achieved. Additionally, the glycerol in the buffer served the purpose of stabilizing labile proteins and

enhancing the interactions between the monomers and cognate sequence, whilst the BSA (bovine serum albumin) protein minimised the non-specific losses of the binding FHD monomers throughout sample handling. Once samples were prepared, samples were incubated for a period of 30 minutes at 4 °C to ensure DNA-protein complexes were formed prior to gel loading.

The DNA-protein complexes were resolved through continuous polyacrylamide gels. The FOXP1 and FOXP2 FHD samples were resolved on a 10% polyacrylamide gel and the FOXP3 FHD samples were resolved on an 8% gel. These gels were initially subjected to an electrophoresis pre-run (100 V for 1 hour) in order to prevent smearing in the gels caused by the inevitable dissociation of DNA-protein complexes during electrophoresis. Thereafter, 10 µl samples were loaded into the wells and the electrophoresis was performed at 160 V for 1.5 hours at 4°C. Following, electrophoresis, gels were stained using GelRed® Nucleic Acid Gel Stain and the bands were detected under UV-light using a ChemiDoc™ XRS+ Imaging System (Bio-Rad, USA).

3.7) Structural Characterisation of the FOXP1, FOXP2 and FOXP3 FHDs:

3.7.1 Circular Dichroism Spectroscopy:

Circular Dichroism (CD) is a form of spectrophotometric technique utilised for various applications including the determination of protein secondary structure. The backbone of a protein has optical activity due to the fact that amino acid residues are chiral molecules (Wei *et al.*, 2014). This suggests that the polypeptide backbone can potentially interact with polarised light. CD involves the differential absorption of left- and right-handed circularly polarised light by a chiral molecule (Wei *et al.*, 2014). In the far-UV region (170 - 250 nm), peptide bonds exhibit two electronic transitions including the π to π^* transition occurring at wavelengths between 185 – 200 nm, and the n to π^* transition which occurs at wavelengths ranging between 215 – 230 nm (Wei *et al.*, 2014). Therefore, depending on the secondary structural elements exhibited by proteins, these proteins will show a specific characteristic behaviour which can be observed on a CD spectrum (Wei *et al.*, 2014). Typically, the CD spectrum of a protein that is predominantly α-helical will show two negative troughs (bands) at about 208 nm and 222 nm with a positive peak (band) near a wavelength of 190 nm (Wei *et al.*, 2014). The β-sheet

CD spectrum however will show one negative band at approximately 218 nm and a positive band at a wavelength that is near 195 nm (Wei *et al.*, 2014).

For the FHDs of FOXP1, FOXP2 and FOXP3, far-UV CD was performed at 20 °C to assess their secondary structures and confirm that the FHDs were folded in their characteristic secondary structure consisting of a predominantly α-helical conformation. These experiments were conducted on 10 µM protein [10 mM sodium phosphate, 50 mM Na_2SO_4, 2 mM DTT]. The measurements were obtained by making use of the JASCO J-810 Spectrophotometer with a 1.00 nm band width, 200 nm/min scanning speed and 0.2 nm data pitch, and were then converted from millidegrees (mdeg) to mean residue ellipticity ($deg.cm^2.dmol^{-1}$) using the following equation (where n is the number of amino acids in the protein):

$$\theta_{MRE} = \frac{100 \times \theta}{[protein](mmol) \times pathlength(cm) \times n}$$

3.7.2 Intrinsic Fluorescence Spectroscopy:

Fluorescence is based on excitation and emission of electromagnetic radiation (EMR) whereby electrons of a fluorophore are excited to higher energy levels by EMR of specific wavelength and then as they return to their ground state emit EMR of longer wavelength. In intrinsic fluorescence spectroscopy of proteins, amino acids such as tryptophan and tyrosine act as fluorophores as they produce good fluorescence signals due to their aromatic groups, allowing them to be excited at 280 nm wavelength. Tryptophan has the additional ability to be excited at 295 nm (Chen, Yu and Barkley, 1998), and these properties allow the tertiary structure of proteins to be assessed as these fluorophores are sensitive to the polarity of their environment. That is, fluorophores that are buried within the core of a protein in the nonpolar environment cause a blue (hyposochromic) shift in the fluorescence spectrum and those that contact the polar environment, residing on the surface of the protein, cause a red (bathochromic) shift in the fluorescence spectrum (Chen, Yu and Barkley, 1998).

Intrinsic fluorescence of all FHDs was performed using the JASCO FP-6300 spectrofluorometer with a 0.5 nm data pitch, excitation and emission bandwidths of 5 nm, and 1 cm path length quartz cuvettes. The FHDs were diluted to a final concentration of 4 µM and were excited at a wavelength of 295 nm. Spectra were recorded from 280 nm to 500 nm.

3.7.3 Analytical Size Exclusion Chromatography:

Size exclusion chromatography is a rapid and reproducible technique providing a means of quantitatively analysing the quaternary structure of active proteins (Hong *et al.*, 2012). SEC separates proteins of a heterogeneous solution according to their hydrodynamic volume (Hong *et al.*, 2012). The column which is used to characteristically separate these proteins is fixed with pores of particular size (Mojsiewicz-Pieńkowska, 2014). Essentially, separation of the protein molecules is established due to the differences in flow delays of the proteins of varying hydrodynamic volumes (Mojsiewicz-Pieńkowska, 2014). The flow delays are caused by the ability of molecules to diffuse into the pores. Therefore proteins that are smaller than the size of the pores will diffuse into these pores and thereby experience a greater flow delay and consequently elute at a slower rate (Mojsiewicz-Pieńkowska, 2014). Conversely, proteins larger than the size of the pore will elute rapidly from the column because they are unable to enter the pore but rather flow in the spaces between them (Mojsiewicz-Pieńkowska, 2014). Therefore SEC can be performed to analyse the quaternary structure of proteins. The oligomerization of the proteins results in larger hydrodynamic volumes which distinguishes oligomers from monomers. While the analysis of hydrodynamic volume of a protein may provide information about the oligomeric state of a protein, analysis may be used to further measure the dissociation constant (K_D) of the oligomers at equilibrium to evaluate the stability of the quaternary structures under varying conditions such as protein concentration.

The oligomeric states of the FOXP1, FOXP2 and FOXP3 FHDs were investigated, qualitatively and quantitatively, by the Size Exclusion-High Performance Liquid Chromatography procedure. Studies have shown that FOXP2 and FOXP1 FHDs exist as both monomers and homodimers at physiological conditions while FOXP3 exists solely as homodimers (Medina *et al.*, 2016; Stroud *et al.*, 2006). At concentrations up to 300 µM, the FHD of FOXP2 has been established to exist almost entirely as a monomer (Blane and Fanucchi, 2015; Morris and Fanucchi, 2016; Perumal *et* al., 2015). Therefore the proportions of native dimeric and monomeric species were analysed by separating increasing concentrations of the FHD proteins (5 – 100 µM). First, the Yarra^{TM} 3U SEC-3000 LC column was fitted into the SHIMADZU SPD20A HPLC machine and then pre-equilibrated for an hour with HPLC Equilibration Buffer composed of sodium phosphate (10 mM), pH 6.5, DTT (2 mM) and NaCl (500 mM). The use of DTT in the buffer was to ensure that intra- and intermolecular disulphide bonds between the protein cysteine residues were not established. Thereafter, calibration of the column was performed employing 6.5 kDa – 75 kDa Bio-Rad

Standards. Varying concentrations of each FHD protein (20 µl) was loaded onto the column and the experiments were conducted at a flow rate of 1 ml/minute at 20°C. The dissociation constant (K_D) for each FHD was then measured from the elution profile. This was achieved by utilizing the areas of each peak and subsequently constructing a graph of monomer concentration ([monomer]^2) against dimer concentration ([dimer]) (Perumal *et al* ., 2015). Thereafter the gradient of the graph was used to determine K_D using the K_D = [monomer]^2/[dimer] equation (Perumal *et al*., 2015).

3.8) Fluorescence Anisotropy (FA):

Fluorescence anisotropy measures the changes in orientation of a fluorescently-labelled molecule in a solution with reference to the time between the absorption and emissions events by using the phenomenon of fluorescence (Zhang *et al.*, 2015). The fluorescence phenomenon involves the emission of light by a fluorescence molecule after it has been excited by electromagnetic radiation. In terms of FA, the fluorescence molecule, which is referred to as a fluorophore, is excited with vertically polarized light, in turn resulting in the emission of polarised light, by the insertion of a polarisation filter in front of the detector, oriented either perpendicular or parallel to the polarised light (Zhang *et al.*, 2015). Therefore, the degree of polarisation of the emission, including both the vertical and horizontal emission components, may be described in terms of the anisotropy (Zhang *et al.*, 2015). The emission of polarised light is affected by a number of factors, including rotational diffusion (Bashardanesh *et al.*, 2019).

In solution, molecules constantly experience rotational diffusion due to molecule-molecule interactions and interactions that occur between the molecules and their environment, thereby causing changes in the orientation of the molecules (Bashardanesh *et al.*, 2019). Rotational diffusion ensures that an overall equilibrium of the orientation of all the molecules in solution is maintained (Bashardanesh *et al.*, 2019). Typically, a smaller molecule will orientate more quickly than a larger molecule and thus results in the emission of light that is significantly depolarised (Jameson and Ross, 2010). Therefore FA is a widely used tool for studies related to protein-protein and DNA-protein interactions, providing quantitative analysis. Accordingly, free fluorescently-labelled DNA or protein will orientate faster. However, DNA-protein or protein-protein complexes reduce the rate of orientation, resulting in reduced depolarisation and in turn resulting in an increased anisotropy.

In this study, FA assays were performed to obtain dissociation constants (K_D) for (1) the homodimerization of FOXP1 FHD, FOXP2 FHD and FOXP3 FHD, and (2) the heterodimerization of FOXP1 FHD with FOXP2 FHD, and FOXP1 FHD with FOXP3 FHD, depending on whether these heterotypic associations occur.

3.8.1 Homodimerization of the FOXP1, FOXP2 and FOXP3 FHDs:

For the homotypic association studies, 5 µM of each FHD protein was labelled with Alexa Fluor 555 dye. Alexa Fluor 555 is a fluorescent dye with a propensity to bind cysteine amino acid residues. The FHD proteins were incubated with dye 3X the protein concentration for 1 hour at 20°C to allow the dye to bind to the proteins accordingly. Thereafter, these samples were incubated overnight at 4°C. Upon incubation, excess dye was removed utilising dye removal columns.

Increasing concentrations of unlabelled FOXP1, FOXP2 and FOXP3 FHD (0 µM – 40 µM) were incubated with 0.6 µM of their corresponding labelled counterpart at 20°C. The fluorescence anisotropy experiments were conducted employing the PerkinElmer LS-50B fluorescence spectrophotometer at 20°C in buffer consisting of 10 mM sodium phosphate, pH 7.5, 100 mM NaCl and 2 mM TCEP (Tris(-carboxyethyl)phosphine hydrochloride). These samples were excited at a wavelength of 556 nm and the measurements of fluorescence anisotropy were recorded at 572 nm wavelength. These sets of data were used as controls for the heterotypic association experiments. Therefore, the K_D values for these experiments were determined and used to compare with the K_D values of the heterodimerization studies.

3.8.2 Heterodimerization of the FOXP1, FOXP2 and FOXP3 FHDs:

In order to determine whether the FOXP1 FHD can form domain-swapped dimers with FOXP2 FHD as well as FOXP3 FHDs, fluorescence anisotropy of each heterotypic association was conducted. FOXP1 and FOXP3 FHDs were labelled with Alexa Fluor 555. These proteins were left at 20°C for 1 hour to allow labelling to occur. Thereafter, an overnight incubation of the FHD proteins was implemented at 4°C. Excess dye was removed using dye removal columns.

The heterotypic association fluorescence anisotropies were performed in buffer consisting of 10 mM sodium phosphate, pH 7.5, 100 NaCl and 2 mM TCEP at 20°C. For the FOXP1-FOXP2

FHD experiment, increasing concentrations of unlabelled FOXP2 FHD were titrated into 0.6 μM labelled FOXP1 FHD. Similarly, for the FOXP1-FOXP3 FHD experiment, increasing concentrations of unlabelled FOXP1 FHD were titrated into 0.6 μM labelled FOXP3 FHD. The samples were excited at a wavelength of 556 nm and the measurements of fluorescence anisotropy were recorded at 572 nm wavelength. These assays were used to determine the K_D for the heterodimerization between these FHDs.

Chapter 4

Results

Many studies pertaining to the forkhead domain of the FOXP transcription factors have focused particularly on their homodimerization. There is no research thus far pertaining to the possible heterodimerization of these FHDs despite the fact that it has been established that the homodimers are important for their function. Therefore, in this research study, the possible heterotypic interactions between the FOXP1 FHD with FOXP2 FHD and FOXP1 FHD with FOXP3 FHD were studied. The FOXP1 and FOXP2 proteins are co-expressed in certain regions of the brain, and FOXP1 and FOXP3 TFs have been shown to interact via an upstream domain. Analytical size exclusion chromatography and fluorescence anisotropy were conducted to gain insight and essentially obtain the dissociation constants for their propensities for homodimerization as well as binding affinities for their homo- and heterotypic associations.

4.1) Transformation:

Initially, the coding sequences of the FOXP1, FOXP2 and FOXP3 FHDs were incorporated into pET-11A plasmids and thereafter, these plasmids were used to transform T7 PLysS *E.Coli* cells. It was crucial to perform induction trials to determine optimal conditions for the expression of FOXP1 and FOXP3 FHDs. Although these proteins have a high similarity to FOXP2 FHD, only the FOXP1 FHD showed expression (Figure 9A) under similar expression conditions as FOXP2 FHD (0.5 mM IPTG at 20°C for 22 hours). In regards to FOXP3 FHD, expression failed to occur under these conditions (Figure 9B). Therefore, this prompted additional trials to be conducted. Since FOXP3 FHD has been previously shown to express in Rosetta Cells in our laboratory, the conditions investigated for the trials in T7 Express pLysS *E.coli* cells included 0.2 mM and 0.5 mM IPTG concentrations under 30°C. Essentially, optimal conditions for FOXP3 FHD expression were achieved with 0.2 mM IPTG for 4 hours at 30°C (Figure 9C), and although expression also occurred with induction of 0.5 mM IPTG concentration (Figure 9D), expression was reduced in the supernatant and mostly occurred in the pellet.

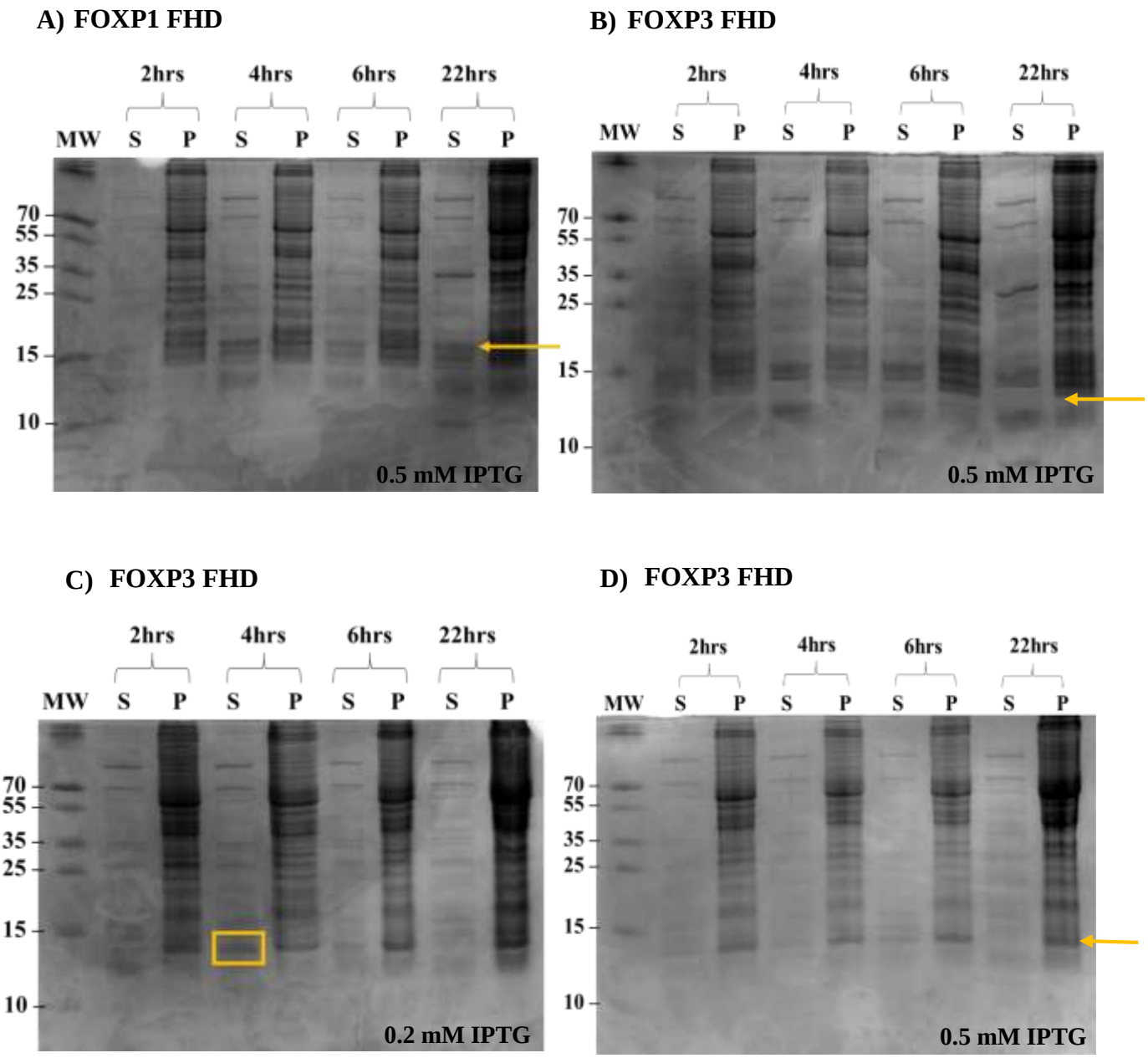


Figure 9: SDS-PAGE gels (16%) showing the induction trials of FOXP1 and FOXP3 FHDs. In each case, 1 ml of cells were collected at four time points and were immediately centrifuged. The pellets were re-suspended in HisTrap^{TM} Equilibration buffer and subjected to sonication. Following sonication, cells were centrifuged and both the supernatant (S) and pellet (P) were resolved through SDS-PAGE gels. A) Induction trials were conducted for FOXP1 FHD (expected size 15.6 kDa) with 0.5 mM IPTG at 20°C between 2 – 22 hours. Optimal soluble expression occurred at 22 hours as indicated by the arrow. B) Induction trials for FOXP3 FHD (expected size 12.9 kDa) conducted with 0.5 mM IPTG at 20°C between 2 – 22 hours. No expression occurred under these conditions which is indicated by an arrow. C) FOXP3 FHD induction trials employing 0.2 mM IPTG at 30°C between 2 – 22 hours. Optimal soluble expression occurred at 4 hours as indicated by the box. D) Induction trials conducted for FOXP3 FHD 0.5 mM IPTG at 30°C between 2 – 22 hours. Very small amount of expression observed with most of the expression occurring in the pellet as indicated by the arrow.

4.2) The FOXP1, FOXP2 and FOXP3 FHDs were purified by IMAC:

FOXP1, FOXP2 and FOXP3 FHDs were expressed under their respective optimal conditions. Following expression, the supernatants of each FHD, collected from cell lysates, were loaded separately onto their corresponding Ni^{2+}-charged IMAC columns. Binding of these proteins to the nickel ions via their 6XHis-tags ensured that the FHD proteins would separate from the bulk contamination. Furthermore, a 1.5 M NaCl wash step was performed to ensure that DNA contamination was eliminated. Eluents from each step of the purification were collected as fractions and resolved through tricine SDS-PAGE gels to assess the degree of purity of each of the FHDs (Figure 10). It was observed that the FHD proteins effectively bound to the IMAC columns as these proteins did not appear in the flow through (FT) eluent. However, both FOXP1 and FOXP2 FHD gels showed that some of the protein eluent fractions contained a considerable amount of contaminants. Therefore, the contaminated fractions for each FHD were pooled and the 6XHis-tag was cleaved so as to further purify the proteins separately. Successful purification of each FHD was achieved (Figure 12). This is because the dialysis buffer at pH 8 aggregates the contaminants from each FHD sample. Additionally, any contaminants that did not separate from the FHDs as aggregates bind to the IMAC column during purification, thereby separating from the cleaved proteins.

A molecular weight marker of known protein sizes was resolved alongside the samples in order to determine the molecular weights of the FHD proteins and in turn confirm whether they correspond to the theoretical molecular weights obtained from ExPASy Molecular weight determination was accomplished by constructing standard curves (Figure 11). The molecular weights of the FOXP1, FOXP2 and FOXP3 FHDs were calculated to be 15.6 kDa, 13.2 kDa and 12.2 kDa, respectively (Figure 11D). Comparing these values with their theoretical molecular weights, which are 15.6 kDa, 13.0 kDa and 12.9 kDa, respectively, confirms that the desired FHD proteins were successfully expressed and collected from purification.

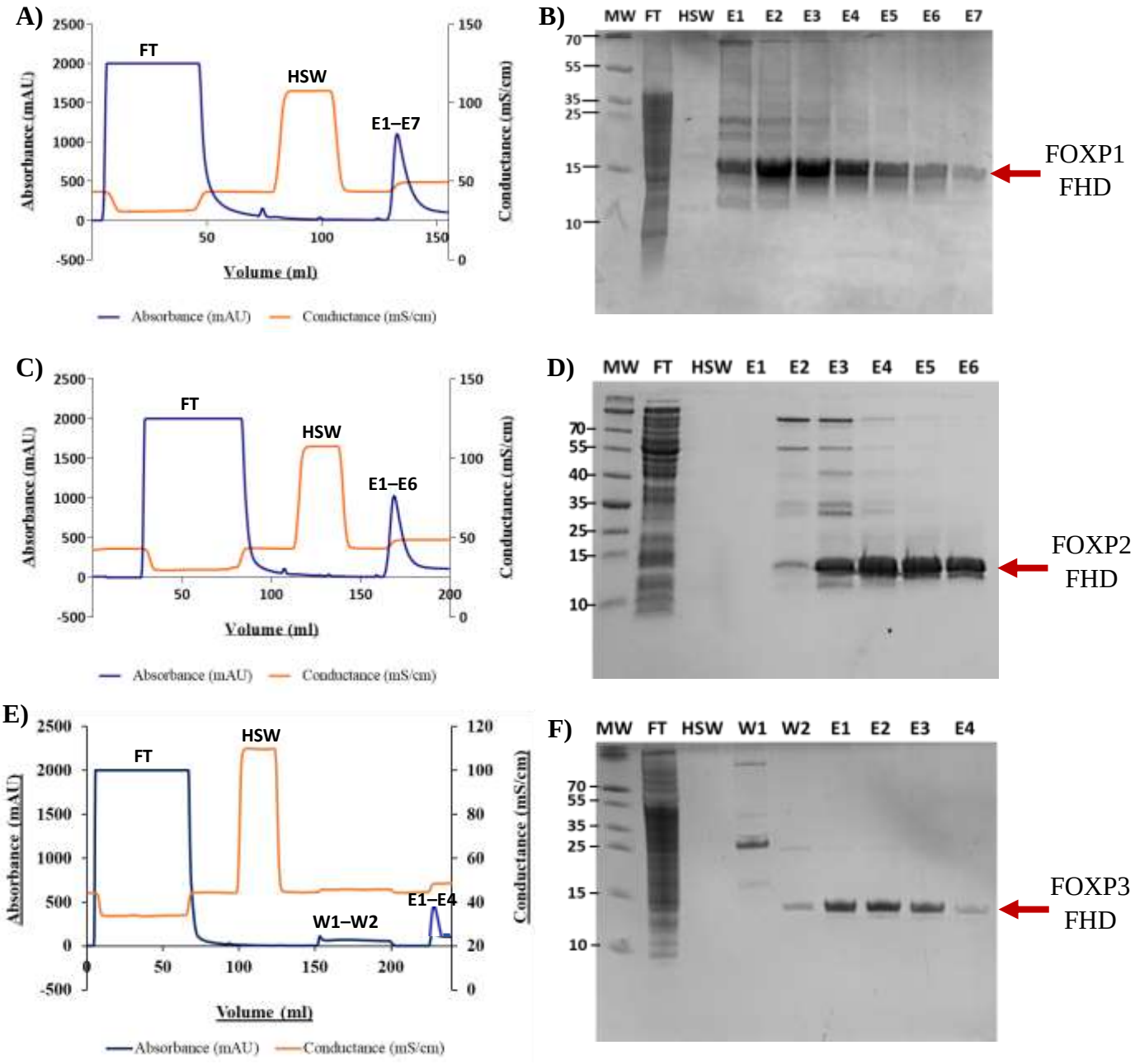
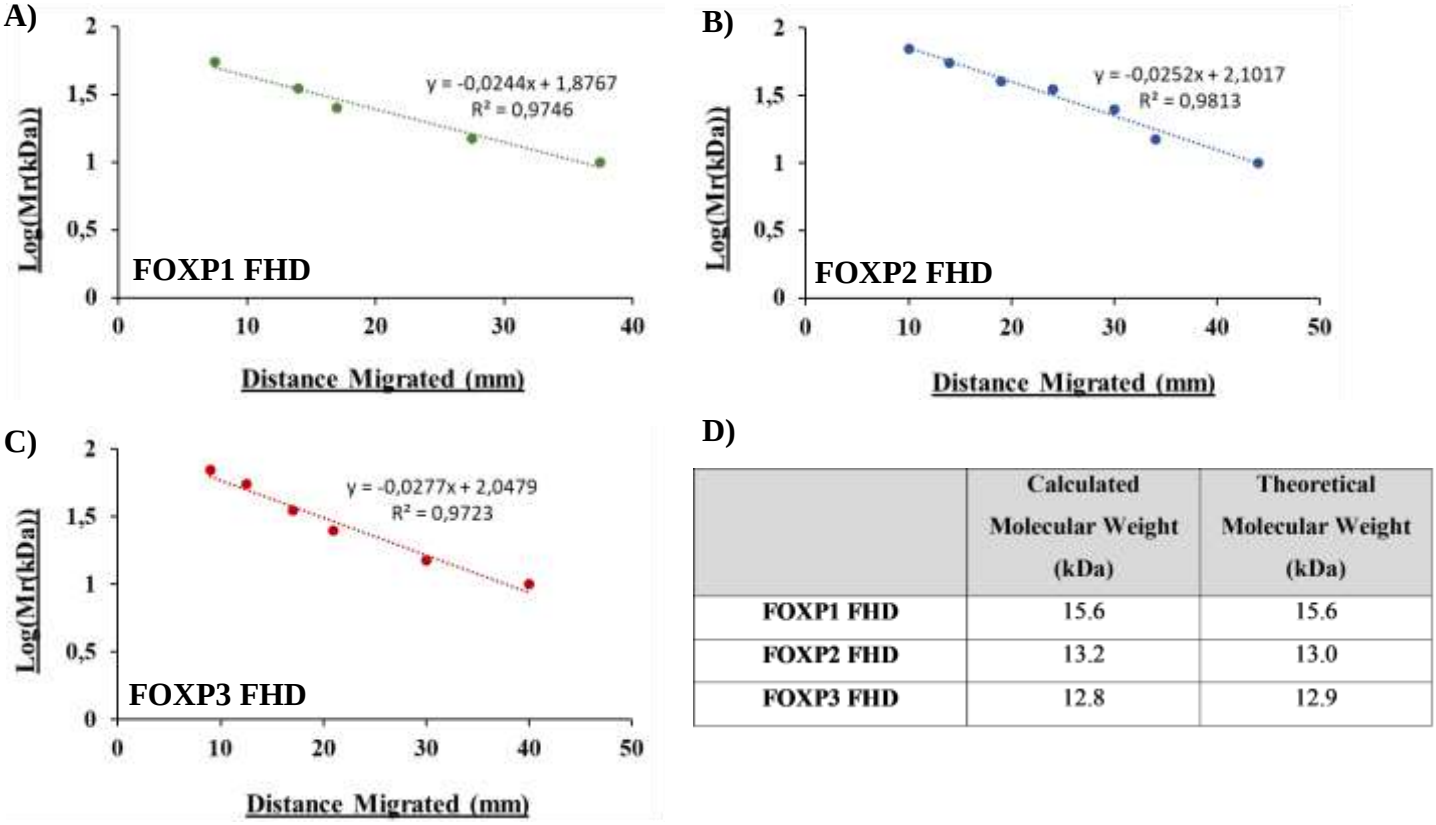


Figure 10: Purification of the forkhead domains of FOXP1, FOXP2 and FOXP3 by immobilised metal-affinity chromatography. Soluble cell lysate of each protein was loaded onto their respective nickel (II) ion-charged column and the course of purification was recorded in the form of elution profiles. (A, C and E) Elution profiles of FOXP1, FOXP2 and FOXP3 FHDs, respectively. In each elution profile, the blue line presents the absorbance recorded at a wavelength of 280 nm, and the orange line presents the conductance, which is affected by the concentration of NaCl in the buffers. (B, D and F) 16% SDS-PAGE gels of FOXP1, FOXP2 and FOXP3 FHDs, respectively, indicate that the proteins were each successfully purified. MW represents the molecular weight marker, FT is the flow through, HSW is the high salt wash, W is the wash and E presents the eluted protein fractions.



D)

	Calculated Molecular Weight (kDa)	Theoretical Molecular Weight (kDa)
FOXP1 FHD	15.6	15.6
FOXP2 FHD	13.2	13.0
FOXP3 FHD	12.8	12.9

Figure 11: Molecular weight determination of the forkhead domains of FOXP1, FOXP2 and FOXP3 proteins. Standard curves were constructed from the SDS-PAGE gels of the resolved proteins. The molecular weights of FOXP1, FOXP2 and FOXP3 FHDs were calculated as 15.6 kDa, 13.2 kDa and 12.8 kDa, respectively. A) FOXP1 FHD standard curve with R-squared value calculated as 0.9746. B) Standard curve of FOXP2 FHD with an R-squared correlation of 0.9813. C) FOXP3 FHD standard curve with R-squared correlation of 0.9723. D) A table comparing the molecular weights of the FHD proteins determined from the standard curve to their theoretical molecular sizes.

The 6XHis-Tags of the FOXP1 and FOXP2 FHDs were cleaved so as to further purify these proteins from the contaminants that were observed in Figure 10. Additionally, since interaction studies were going to be conducted, cleavage of the 6XHis-tags was important in ensuring that the tags were not going to interfere with domain swapping. Initially, each FHD was incubated separately with thrombin for 4 hours. Following incubation with thrombin, the FOXP1 and FOXP2 FHDs were purified separately employing IMAC and benzamidine chromatography (Figure 12A and C). Fractions from the purification procedure were collected and resolved on 16% polyacrylamide gels (Figure 12B and D). It was observed that the cleavage was successful. Furthermore, successful purification was achieved for both proteins.

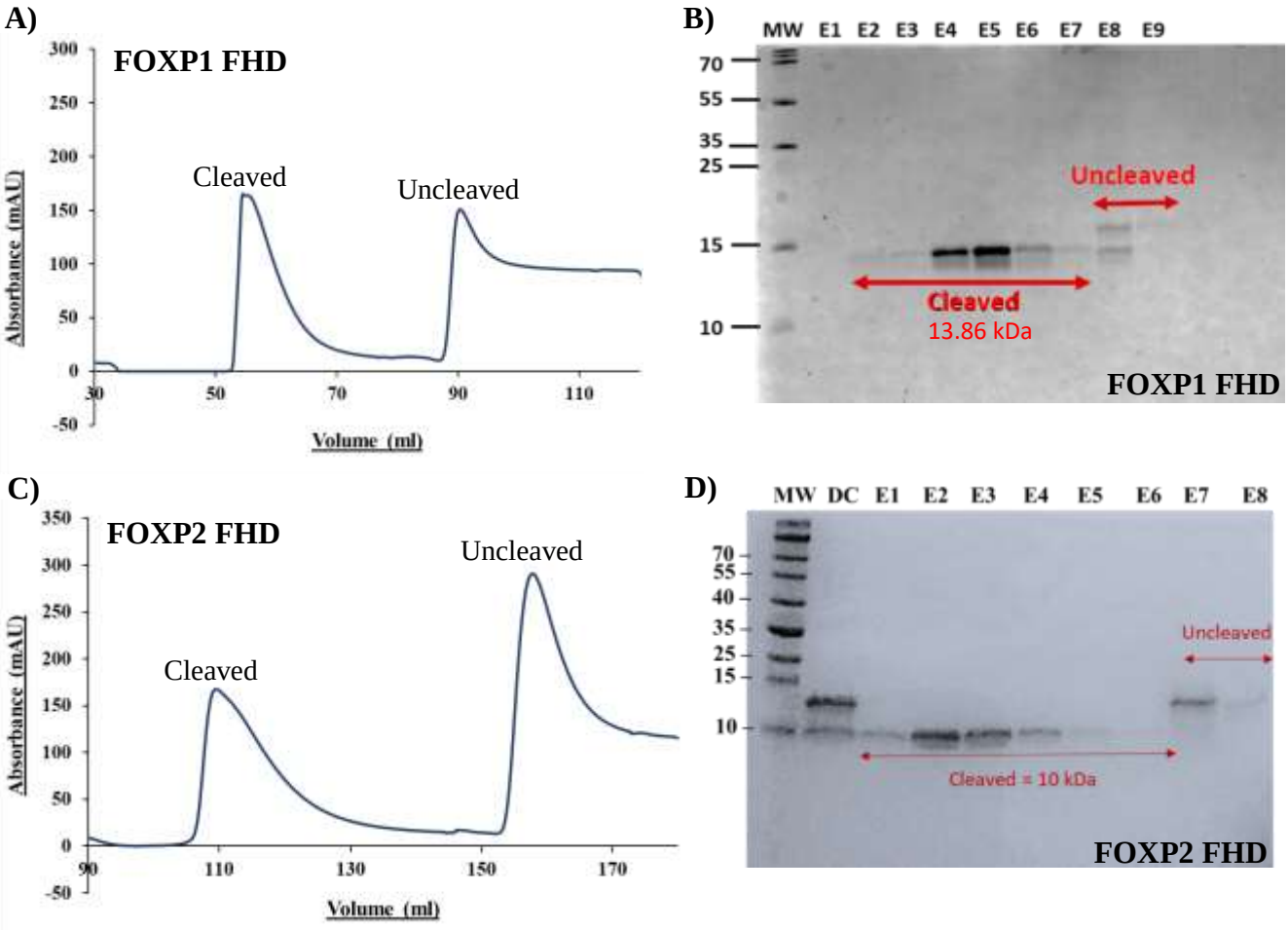


Figure 12: Cleavage of the 6XHis-tags of the FOXP1 and FOXP2 FHDs. The 6XHis-tag of each FHD protein was cleaved employing thrombin and the proteins were subsequently purified by benzamidine and immobilised metal affinity chromatography. A and C) Elution profiles of the FOXP1 and FOXP2 FHDs, respectively. Cleaved FHDs were eluted with equilibration buffer and the uncleaved counterparts were eluted using elution buffer. B and D) SDS-PAGE gels of FOXP1 and FOXP2 FHDs, respectively displaying both the cleaved and uncleaved fractions. Cleaved fractions of both the FOXP1 and FOXP2 FHDs were successfully purified as no contaminations were observed.

4.3) Determining protein purity:

Absorbance spectra of the FOXP1, FOXP2 and FOXP3 FHDs were captured between 240 nm and 340 nm at 20°C. The spectra were utilised to assess the quality and purity of the proteins. At wavelengths of 280 nm and 295 nm, two distinct peaks were observed, representing the absorbance of the tryptophan and tyrosine residues (Figure 13). There was no peak observed at 340 nm wavelength, indicating absence of aggregation and in turn stable proteins. The

A280/A260 ratio determined for each FHD protein was approximately 1.7, confirming that the protein samples were free from DNA contamination.

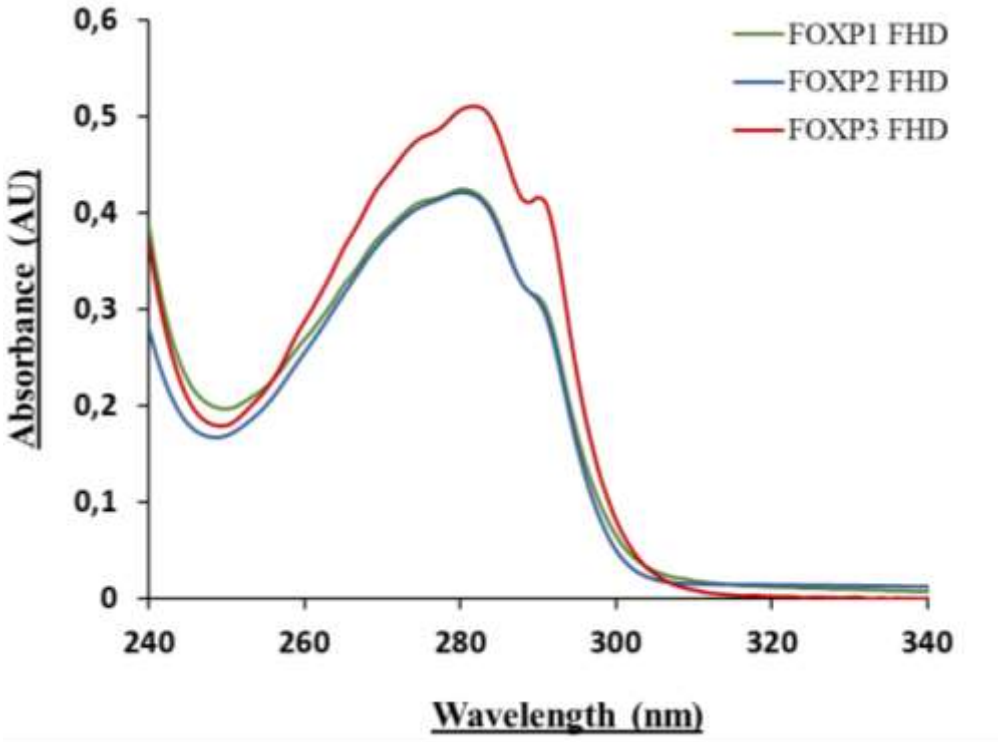
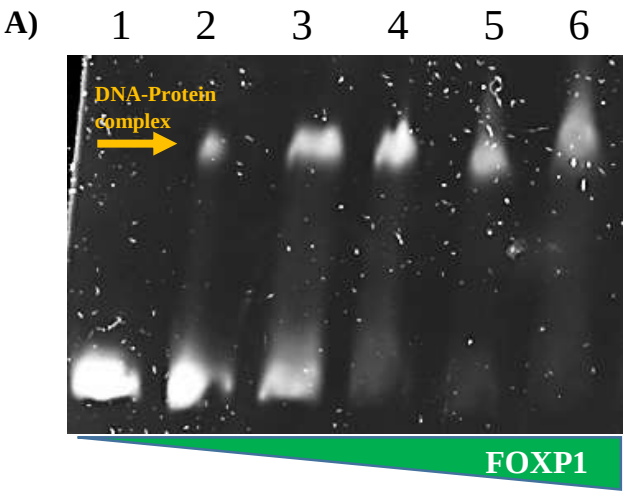


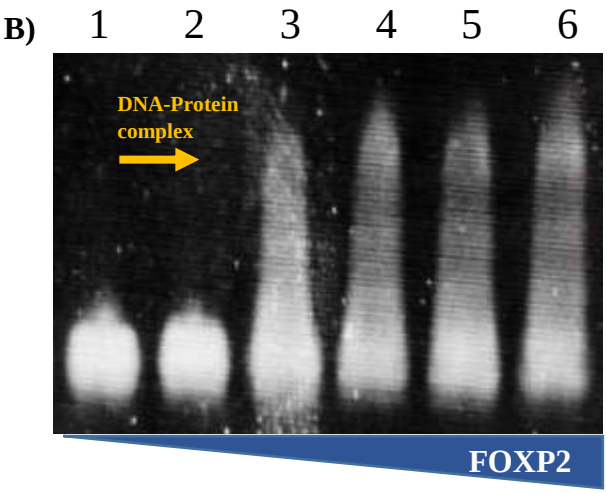
Figure 13: Absorbance spectra of the FOXP1, FOXP2 and FOXP3 FHDs. The spectra were obtained from 240 nm to 340 nm and showed peaks at 280 nm and 295 nm. Furthermore, analysis of the A280/A260 ratio showed a value of 1.7, indicating samples that were free of DNA contamination.

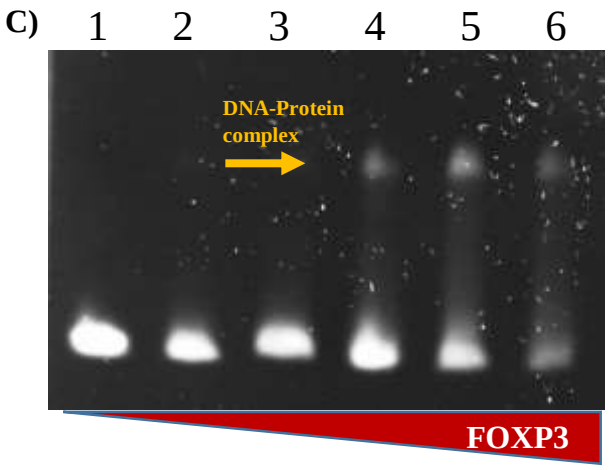
4.4) FOXP1, FOXP2 and FOXP3 FHDs bind Nelson DNA:

The functional integrities of the FOXP1, FOXP2 and FOXP3 FHDs were investigated. Since the FHD is the DNA-binding domain of these transcription factors, DNA-binding studies, in the form of EMSAs, were conducted. As previously mentioned in section 3.6, the Nelson DNA sequence was employed in these assays as this duplex sequence has been reported to effectively interact with the FHD of FOXP proteins (Webb *et al*., 2017). As anticipated, all three FHD proteins showed interaction with the cognate sequence (Figure 14). However, it was observed that FOXP1 FHD established a DNA-protein complex from a 2:1 DNA:protein ratio (lane 2),

whilst FOXP2 and FOXP3 FHDs formed DNA-protein complexes from DNA:protein ratios of 1:1 (lane 3) and 1:2 (lane 4), respectively







D)

Lane	1	2	3	4	5	6
Ratio (DNA:Protein)	1:0	2:1	1:1	1:2	1:3	1:5
DNA concentration (µM)	0.5	0.5	0.5	0.5	0.5	0.5
Protein concentration (µM)	0	0.25	0.5	1	2	3

Figure 14: Electrophoretic Mobility Shift Assay gels of FOXP1, FOXP2 and FOXP3 FHDs. (A) FOXP1 FHD EMSA gel indicating DNA-protein complex formation from a DNA:Protein ratio of 2:1 (lane 2). (B) FOXP2 FHD EMSA gel indicating DNA-protein complex formation from a DNA:Protein ratio of 1:1 (lane 3). (C) FOXP3 FHD EMSA gel indicating DNA-protein complex formation from a DNA:Protein ratio of 1:3 (lane 4) (D) Table presenting the lanes of the EMSA gels and the corresponding DNA and protein concentrations.

4.5) Structural Characterisation of FOXP1, FOXP2 and FOXP3 FHDs:

4.5.1 Circular Dichroism Spectroscopy:

CD spectra of 10 µM monomeric FOXP1, FOXP2 and FOXP3 FHDs were obtained to assess and compare the secondary structure of the proteins. As anticipated, the FHDs assumed a predominantly alpha-helical CD spectrum with distinct troughs at 208 nm and 222 nm (Figure 15), corresponding to their crystal structures.

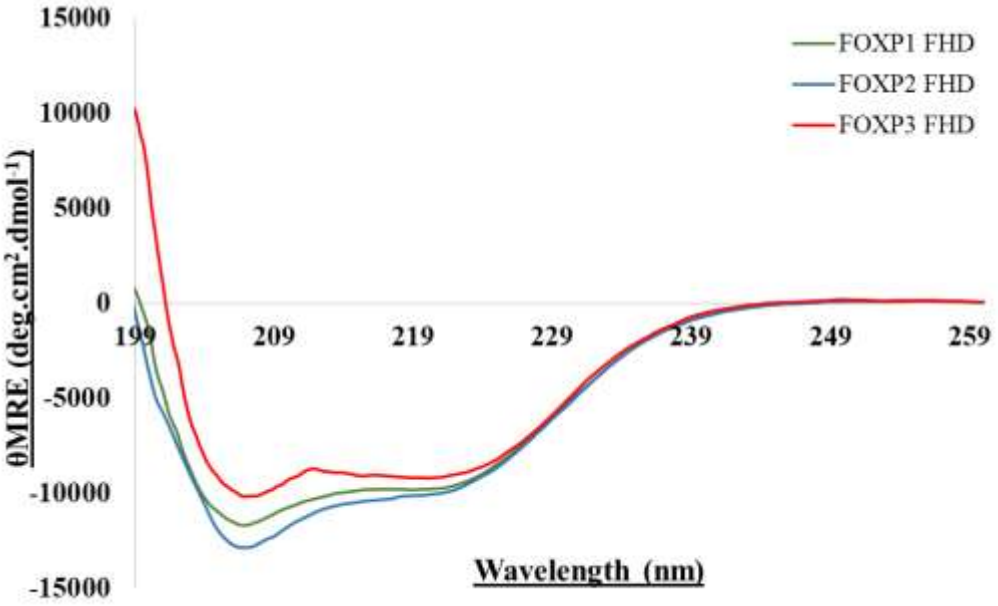


Figure 15: Far-UV circular dichroism spectra of FOXP1 (green), FOXP2 (blue) and FOXP3 (red) forkhead domains (FHD). The spectra were recorded for 10 µM monomeric FHD at 20 °C. All spectra contain troughs at 208 nm and 222 nm, indicating a predominantly alpha-helical characteristic as expected for the FHDs.

4.5.2 Intrinsic Fluorescence Spectroscopy:

The tertiary structure of each FHD protein was monitored using intrinsic fluorescence spectroscopy as a result of the fluorescence signal emitted by tryptophan residues. This was done so as to confirm that the proteins were structurally intact. The signal emitted by these amino acid residues depends on the location of the tryptophan residues in the protein. Both the FOXP1 and FOXP2 FHDs comprise of three tryptophan residues and based on their crystal structures, two of these residues are exposed to the solvent polar environment with the third

one buried in the hydrophobic pocket of the FHDs (Figure 16A and 16B). Conversely, FOXP3 FHD comprises of four tryptophan residues. The crystal structure of the FOXP3 FHD shows two tryptophan residues buried in the hydrophobic environment of the protein whilst the third and fourth tryptophan residue experiences partial and complete exposure to the solvent polar environment, respectively (Figure 16C).

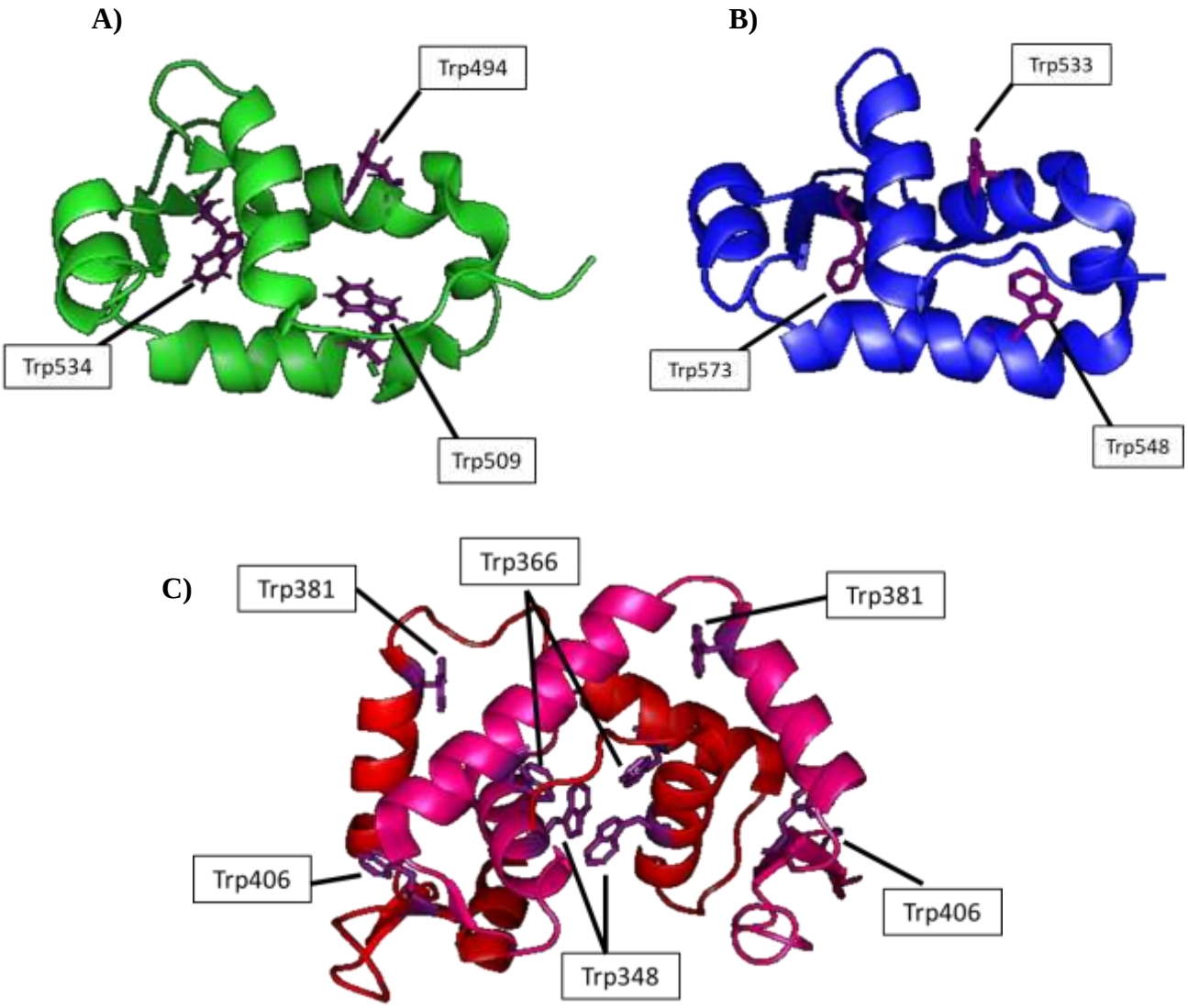


Figure 16: NMR and Crystal structures of the FOXP1, FOXP2 and FOXP3 FHDs, indicating the location of their tryptophan residues. A) NMR structure of FOXP1 FHD. The FOXP1 FHD comprises of three tryptophan residues. Trp494 and Trp534 are located such that they are exposed to the solvent, and Trp509 is buried in the hydrophobic pocket of the protein (PDB: 2KIU). B) FOXP2 FHD crystal structure comprises of three tryptophan residues. Trp533 and Trp573 are located such that they are exposed to the solvent, and Trp548 is buried in the hydrophobic core of the protein (PDB: 2A07). C) FOXP3 FHD crystal structure exhibiting four tryptophan residues. Trp406 is fully exposed to the solvent, Trp381 exhibits partial exposure to the solvent, and both Trp348 and Trp366 are buried in the hydrophobic core of the protein (3QRF). All structures were rendered by PyMol.

As mentioned in section 3.7.2, the tryptophan amino acid residues produce a strong fluorescence signal when excited at 295 nm. Therefore, the FHDs were excited at 295 nm, and emission spectra were captured over a wavelength range of 300 nm and 500 nm (Figure 17). Two emission maxima of 330 nm and 336 nm were observed for FOXP1 and FOXP2 FHDs. FOXP3 FHD showed an emission maximum of 342 nm with a shoulder at 330 nm. Observation of two peaks in a fluorescence spectrum indicates that the tryptophan residues are positioned in different polar environments within the native structure of these proteins. This data is consistent with the NMR structure of FOXP1 FHD and the crystal structures of FOXP2 and FOXP3 FHDs.

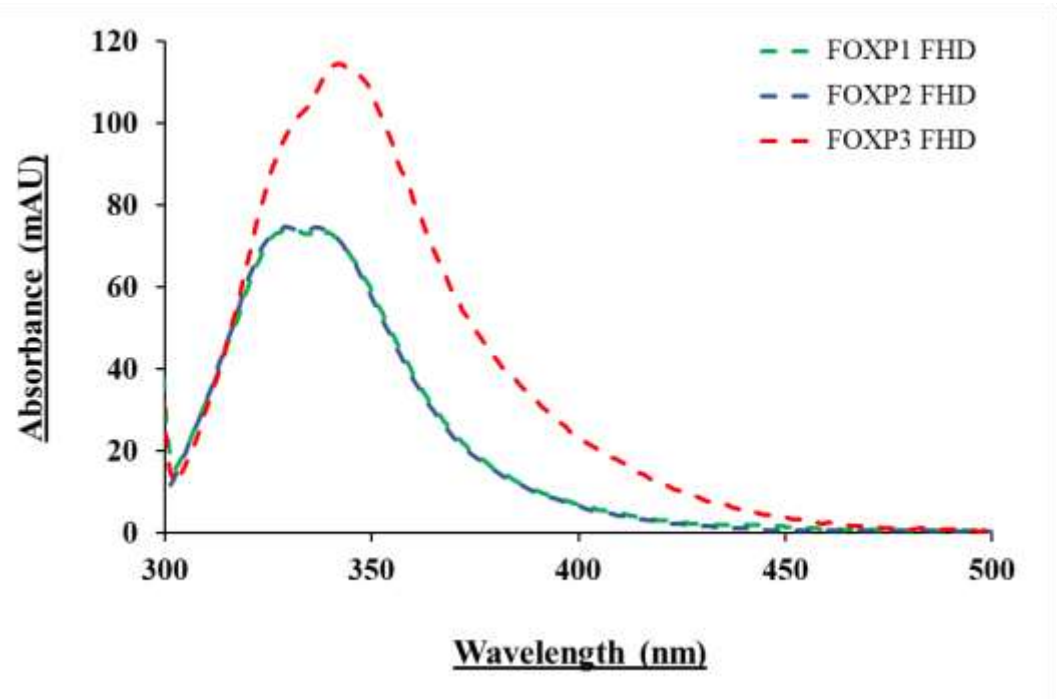


Figure 17: Fluorescence spectra of the FOXP1, FOXP2 and FOXP3 FHDs at 4 µM monomer concentration. For the emission spectra of the FHDs following excitation at 295 nm, both FOXP1 and FOXP2 FHDs exhibit emission maxima of 330 nm and 336 nm, and FOXP3 FHD exhibits emission maxima at 330 nm and 342 nm. The observation of two peaks indicates that they have tryptophan residues that are buried within the core of the protein and exposed to the solvent.

4.6) Analytical Size Exclusion Chromatography:

The proportion of monomeric and dimeric species was investigated for FOXP1, FOXP2 and FOXP3 FHDs using concentration-dependent SE-HPLC. For each FHD, varying

concentrations (5 µM – 100 µM) of the FHD protein were loaded onto the column in order to observe the effect of concentration on oligomeric state. Moreover, K_D values were determined and compared.

It was observed that at all concentrations, the FOXP1 FHD eluted as two peaks (Figure 18A), corresponding to the monomeric and dimeric species. The FOXP2 FHD was observed to be predominantly monomeric at all concentrations (Figure 18B). However, a FOXP2 FHD dimer peak was observed at the 100 µM. Although, studies have shown that FOXP3 FHD exists solely as a homodimer, two peaks were observed for the FOXP3 FHD protein (Figure 18C). However, unlike FOXP1 FHD, the FOXP3 FHD is primarily homodimeric at all concentrations and shows smaller proportions of monomeric species.

From the chromatograms, graphs of squared monomer concentration ($[monomer]^2$) as a function of dimer concentration ([dimer]) (Perumal *et al*., 2015) were constructed (Figure 19). The K_D was determined as the gradient of the graph. The K_D values of the FOXP1, FOXP2 and FOXP3 FHDs were determined as 619 µM, 2.8 mM and 11 µM, respectively, suggesting that the FOXP3 FHD indeed exhibits the greatest propensity for dimerization compared to the other FHDs whereas the FOXP2 FHD exhibits the lowest propensity for dimerization.

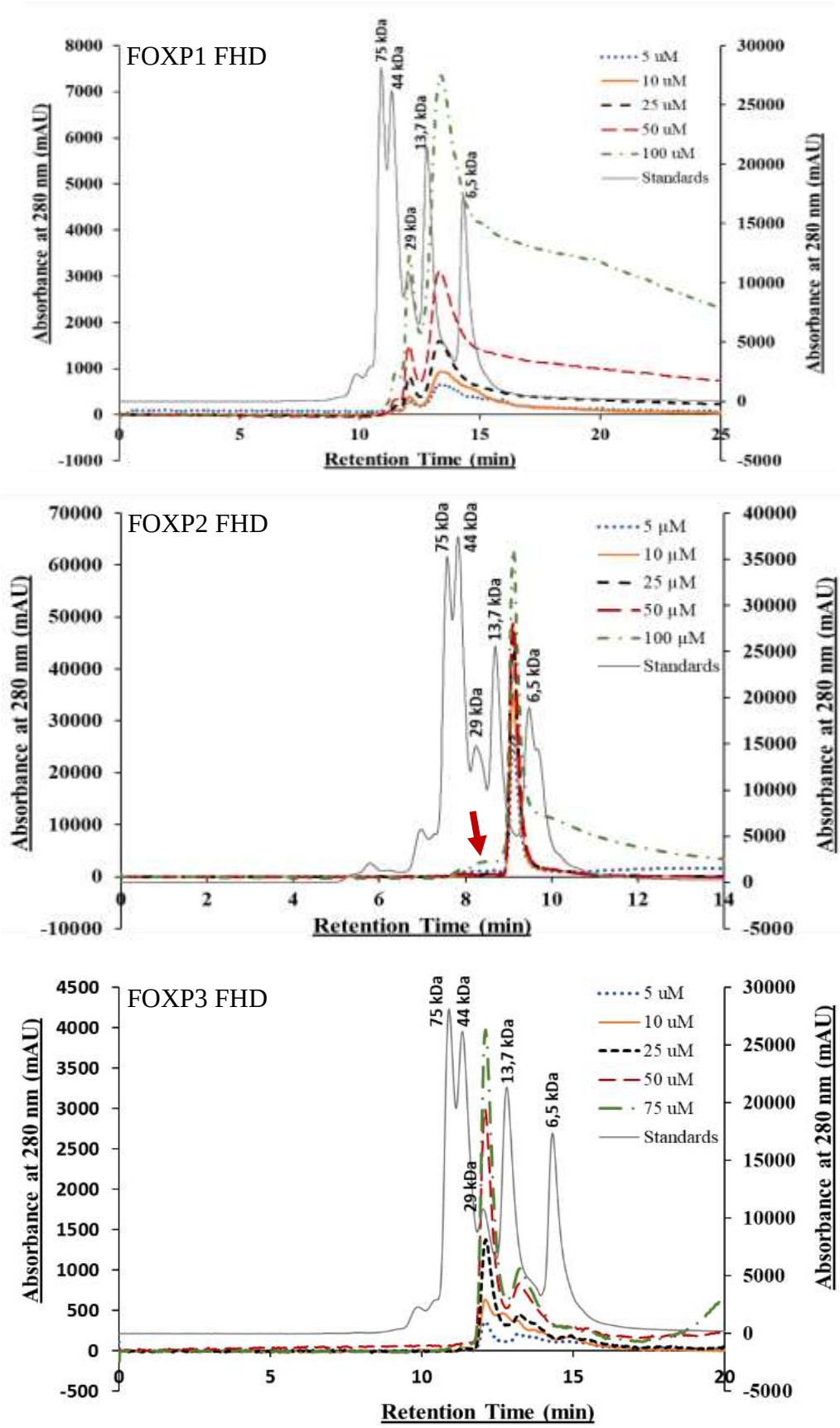


Figure 18: The characterisation of the oligomeric states of the FOXP1, FOXP2 and FOXP3 FHDs. The FHDs of FOXP1, FOXP2 and FOXP3 were run through the SE-HPLC at various concentrations in order to determine their dissociation constants. A) Analytical SEC profile of FOXP1 FHD at 5 – 100 µM concentrations. B) FOXP2 FHD SEC profile at 5 – 100 µM concentrations. The red arrow observed in the graph indicates the dimer peak observed for FOXP2 FHD at 100 µM. C) SEC profile of FOXP3 FHD at 5 – 75 µM concentrations.

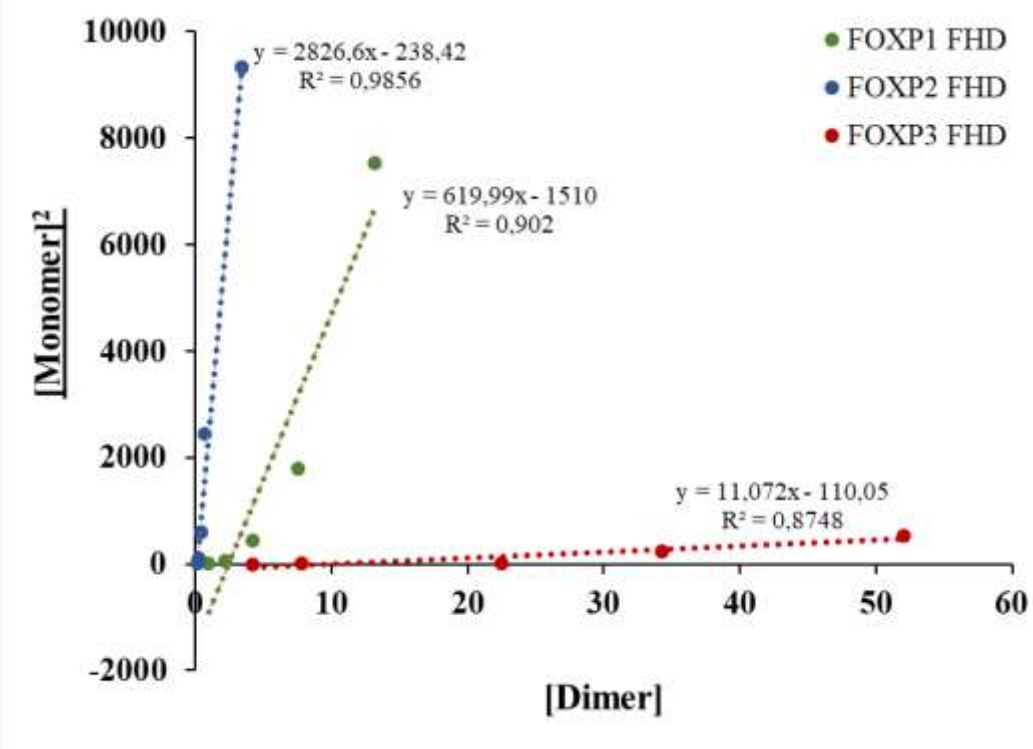


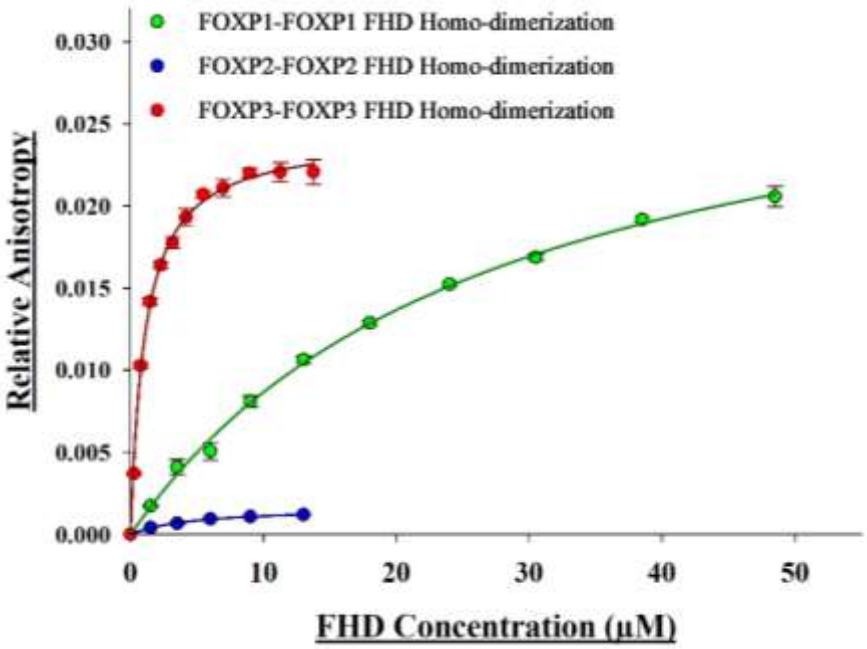
Figure 19: Graph of [monomer]^2 against [dimer] constructed to obtain K_D for the homotypic association of the FOXP1, FOXP2 and FOXP3 FHDs. From the analytical size exclusion chromatography experiments, the areas of the monomer and dimer peaks were employed to obtain the graph of [monomer]^2 against [dimer]. In turn, the values of K_D for FOXP1, FOXP2 and FOXP3 FHDs were determined to be 619 µM, 2 826 µM and 11 µM, respectively.

4.7) Fluorescence Anisotropy:

Since it was established from analytical SEC that all the FHDs are able to form homodimers, fluorescence anisotropies were performed to establish whether these FHDs are capable of heterodimerization. Firstly, fluorescence anisotropies were conducted in triplicate to obtain binding affinities for the homotypic interactions of FOXP1, FOXP2 and FOXP3 FHDs. Thereafter, binding affinities for the heterotypic interactions between FOXP1 FHD and FOXP2 FHD as well as FOXP1 FHD and FOXP3 FHD were investigated. SigmaPlot 15 was used to fit the curves to a single site ligand binding model. Both the homo- and heterotypic interaction data were analysed and compared. For the homodimerization of the FOXP1, FOXP2 and FOXP3 FHDs, the K_D values were determined to be 27.4 ± 1.25 µM, 4.9 ± 0.4 µM and 1.1 ± 0.04 µM, respectively (Figure 20). The K_D for the heterodimerization studies was determined

as 34.0 ± 3.5 µM for the interaction between FOXP1 FHD and FOXP2 FHD and 13.0 ± 1.3 µM for the interaction between FOXP1 FHD and FOXP3 FHD (Figure 21).

A)

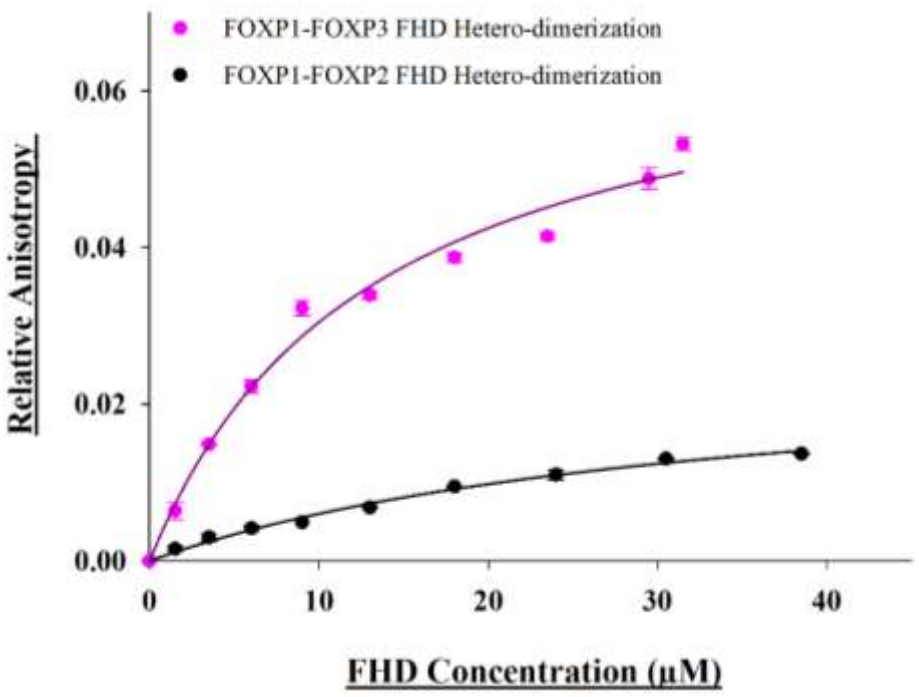


B)

Homo-dimerization	K_D (µM)	R^2
FOXP1-FOXP1 FHD	27.4 ± 1.25	0.9965
FOXP2-FOXP2 FHD	4.9 ± 0.4	0.9919
FOXP3-FOXP3 FHD	1.1 ± 0.04	0.9956

Figure 20: Fluorescence anisotropy assays of the homodimerization of the FHDs. A) Fluorescence anisotropy was performed to determine the binding affinities of the homodimerization of FOXP1, FOXP2 and FOXP3 FHDs. Each protein was labelled with Alexa Fluor 555. Increasing concentration (0 µM – 50 µM) of the unlabelled FHD protein was titrated into 0.6 µM labelled protein. These experiments were performed in triplicates. B) Table showing the K_D values for the homodimerization of the FHDs.

A)



B)

Hetero-dimerization	K_D (µM)	R^2
FOXP1-FOXP2 FHD	34.0 ± 3.5	0.9947
FOXP1-FOXP3 FHD	13.0 ± 1.3	0.9819

Figure 21: Fluorescence anisotropy assays of the heterodimerization of the FHDs. A) Fluorescence anisotropy was performed to determine the binding affinities of the heterodimerization between FOXP1 FHD and FOXP2 FHD as well as between FOXP1 and FOXP3 FHDs. FOXP1 and FOXP3 FHD protein were labelled with Alexa Fluor 555. Increasing concentration (0 µM – 40 µM) of the unlabelled FHD protein was titrated into 0.6 µM labelled protein. These experiments were performed in triplicates. B) Table showing K_D values obtained from the heterodimerization studies where 34 µM was determined for the FOXP1-FOXP2 FHD interaction, and 13 µM was obtained for the FOXP1-FOXP3 FHD interaction.

Chapter 5

Discussion

A unique feature of the forkhead domain of the FOXP transcription factors is their potential to establish domain-swapped dimers through the process of domain swapping. This is evident in the FOXP2 and FOXP3 FHD crystal structures. Research suggests that this feature may have functional relevance due to the fact that there are various mutations that are associated with the collapse of domain swapping, leading to deregulations of transcription (Bandukwala *et al*., 2011). In fact, homodimerization of the FOXP FHD has been suggested to enable regulation of transcription of distal genetic material by promoting interchromosomal associations (Chen *et al*., 2015). Although, only homodimerization of the FOXP FHD has been studied, research shows that these proteins are able to form hetero-oligomers via an upstream domain. Since the FOXP1 and FOXP2 proteins are co-expressed in various regions of the brain and have been suggested to play overlapping roles, and FOXP1 and FOXP3 share genomic binding sites in regulatory T-cells, this study therefore pursued the investigation of probable heterotypic association of their FHDs. The FOXP1 FHD shares both 89% and 76% similarities with the FHDs of FOXP2 and FOXP3, respectively (Chu *et al*., 2011). Despite this, these FHDs contrast in the way they interact with DNA, the functions in which they partake and the genes that they regulate. Additionally, these FHDs display varying propensities and binding affinities of homodimerization.

FOXP1, FOXP2 and FOXP3 FHD Structural Characterization and DNA binding:

To investigate the possibility of the FOXP1-FOXP2 and FOXP1-FOXP3 FHDs to heterodimerize, these proteins were expressed and purified. Since, the function of a protein is closely linked to its structure, it was crucial to ensure that the purified FHDs were folded into their native structures and that they bound DNA. This would ensure that the conclusions derived from the results obtained from downstream experiments pertaining to the homo- and heterodimerization of the FHDs would provide a fairly accurate understanding of what occurs physiologically. Therefore, circular dichroism and intrinsic fluorescence spectroscopies as well

as EMSAs were conducted to ensure that the FHDs were folded correctly and were functional, respectively.

Structurally, the FHD is predominantly alpha-helical, and unlike the FHDs of FOXP1 and FOXP2 which contain three tryptophan residues, the FOXP3 FHD contains four tryptophan residues. Analysis of the data obtained from both the circular dichroism and intrinsic fluorescence spectroscopies indicated that all the FHDs were characteristically folded into their native structures. The IFS further revealed that the fluorescence spectra of the FOXP1 and FOXP2 FHDs are superimposable. The tertiary structure of the proteins is probed by identifying the locations of tryptophan residues around the proteins. This suggests that the tryptophan residues in each FHD of FOXP1 and FOXP2 are in the same degree of solvent exposure. Analysis of the NMR and crystal structures of the FOXP1 and FOXP2, respectively, indicates that it is highly likely that their tryptophan residues are indeed located in similar positions in the proteins. Unlike both these FHDs, the spectrum of the FOXP3 FHD showed an increased fluorescence intensity. This is no surprise since the FHD of FOXP3 comprises of four tryptophan residues.

In addition to being folded correctly, the FHDs of FOXP1, FOXP2 and FOXP3 were observed to be functional as shown in the EMSA studies, indicating that the isolated FHD is indeed able to function without the full-length protein. The Nelson DNA sequence was employed in this study (Nelson *et al*., 2013). This sequence is crucial in binding studies with FOXP FHDs as it has been suggested to have a fairly high affinity for the FOXP2 FHD (Webb *et al*., 2017). Since the FOXP1 and FOXP3 FHDs are closely similar to the FOXP2 FHD, this implies that these proteins may also bind the Nelson DNA with a fairly high affinity as well. As conjectured, all the FHDs bound to the Nelson sequence. However, rather than binding with similar strengths, it was observed that the proteins experienced varying binding affinities. That is, the FHD of FOXP1 showed the greatest binding affinity as it bound DNA already at a 2:1 (DNA:protein) ratio. This is different from both the FOXP2 and FOXP3 FHDs which bound the DNA at ratios of 1:2 and 1:3 (DNA:protein), respectively. The difference in binding strength may be attributed to the fact that although these FHDs are highly conserved, they exhibit subtle variations in their amino acid sequences that may be causing these differences in affinity. As mentioned previously, DNA-binding by TFs is governed by both specificity and affinity (Rohs *et al.*, 2010). Affinity refers to the strength and number of interactions occurring between the protein side chains and the DNA bases (Rohs *et al.*, 2010). It can therefore be suggested that the FOXP1 FHD is able to form large number of strong interactions with the DNA than FOXP2

FHD, thereby binding with higher affinity than FOXP2 FHD. Additionally, when comparing the EMSAs of the FOXP1 and FOXP2, it is observed that the bands of FOXP2 FHD are profoundly smeared. Such smearing is indicative of a DNA-protein interaction that dissociates due to instability (Hellman and Fried, 2007), indicating that the FHD of FOXP2 does indeed have lower affinity than FOXP1 FHD to bind the Nelson DNA. Specificity refers the shape and charge complementarity between the DNA and protein (Rohs *et al.*, 2010). Since the FOXP3 FHD exists exclusively as homodimers in solution, these dimers are compact and rigid, which may cause particular amino acids of FOXP3 FHD to interact weakly, thereby binding with low affinity.

The propensities of FOXP1, FOXP2 and FOXP3 FHD to homodimerize:

Although the FOXP1, FOXP2 and FOXP3 FHDs share high similarity scores, each FHD is unique in that each demonstrates a propensity for dimerization that differs from the next. Indeed, previous studies have revealed that the FOXP1 FHD exists as both monomeric and dimeric species in solution (Chu *et al*., 2011; Medina *et al*., 2016) and that the FOXP2 FHD is predominantly monomeric (Perumal *et al*., 2015; Morris and Fanucchi, 2016) while FOXP3 FHD is exclusively dimeric (Medina *et al.*, 2016; Stroud *et al.*, 2006).

As anticipated, analytical SEC data from this present study showed FOXP1 FHD to occur as a mixture of monomeric and dimeric species in solution. The K_D that was obtained from these concentration-dependent SEC experiments was 619 µM, suggesting that FOXP1 FHD exhibits a low propensity for dimerization. However, this is not concurrent with previous studies, which have suggested FOXP1 FHD to only exhibit high propensities for dimerization, with K_D values ranging between 1 and 30 µM from 17°C to 37°C (Chu *et* al., 2011; Medina *et al*., 2016). It is evident that the dissociation constants for FOXP1 FHD are particularly temperature dependent (Medina *et al*., 2016). These K_D values are also pH dependent (Medina *et al*., 2019). That is, an increase in temperature decreases the propensity of the FHD to dimerize (increases K_D), and an increase in pH increases its propensity to dimerize (decreases K_D). Accordingly, this suggests that, since the analytical SEC experiments in this study were conducted at pH 6.5 at a temperature of 20°C, the expected dissociation constant ranges between 1.30 µM and 3.10 µM (Medina *et* al., 2016; Medina *et al*., 2019). The K_D value obtained in this study is large and falls out of this range. It is therefore evident that the divergence in the dissociation constant is attributable to experimental inaccuracies. That is, the monomeric peaks of the FOXP1 FHD in

the SEC experiment encountered tailing as well as significant drifting, therefore failing to re-establish baseline and in turn yielding incorrect peak area values. Since the graph of [monomer]^2 with respect to [dimer] was constructed from such values, it is no surprise that the R^2-value obtained for the graph was 0.902 which is moderately low.

In the case of FOXP2 FHD, the analytical SEC data in this study were also concurrent with previous studies which indicated that the protein is predominantly monomeric, however exhibiting small proportions of dimeric species. The dimeric species are particularly observed at a FOXP2 FHD protein concentration of 100 µM (Figure 18B). Therefore it is expected that the FHD displays an extremely low propensity for dimerization. Indeed, the K_D that was obtained was 2.8 mM. This K_D is relatively similar to a study by Perumal *et al*., 2015 that showed a K_D of 2.4 mM for FOXP2 FHD. Additionally, the R^2-value from the [monomer]^2 against [dimer] graph shows tight correlation (0.9856), which is indicative of accuracy in the data.

Unlike FOXP1 and FOXP2 FHDs, the FHD of FOXP3 exhibits a strong propensity to form homodimers. This is shown in the data obtained from the analytical SEC experiments with K_D of 11 µM. Studies have suggested that homodimerization of the isolated FHD of FOXP3 appears to be exclusive (Bandukwala *et* al., 2011; Medina *et al.*, 2016; Stroud *et al.*, 2006). However, monomeric species of the FOXP3 FHD were detected by SEC (Figure 18C). It is unlikely that the monomer peak is contamination as the purity of this protein was assessed by SDS-PAGE. Therefore, it is likely that the cause of the monomeric peak is a result of denaturation of the FOXP3 FHD. Although the isolated FOXP3 FHD is able to function without the full-length protein, it appears that the FHD becomes unstable when exposed to 20°C for periods as long as thirty minutes.

The fact that strong propensities for dimerization are assumed by the FOXP3 FHD complements previous studies that have suggested these homodimers to have physiological relevance, which involve the regulation of homeostasis of the immune system (Bandukwala *et al*., 2011). Indeed, mutations that disrupt domain swapping result in the IPEX syndrome (Bandukwala *et al*., 2011). In regards to the FOXP2 FHD, since a very low propensity for dimerization was observed, it is postulated that homodimerization of the FHD may be a transient event which assists in the binding of the protein to DNA before the dimer dissociates. This postulation is supported by a previous study that suggested that the FHD homodimer binds DNA with a higher affinity than its monomeric form (Morris and Fanucchi, 2016), alongside

the fact that the FOXP2 FHD crystal structure shows the dimer bound to only one strand of DNA and not two strands as observed in the crystal structure of FOXP3 FHD. Given that the monomeric form of the FHD is also able to bind DNA, its homodimeric form may function in cases in which demanding regulation of gene transcription is required.

There is no crystal structure of the FHD of FOXP1. However, this protein exhibits characteristics of both the FOXP2 and FOXP3 FHDs. That is, the FOXP1 FHD is both monomeric and dimeric. Furthermore, physiologically, this protein is co-expressed with both FOXP2 and FOXP3, sharing similar functions with both proteins. Therefore, it can be expected that its interaction with DNA would be similar to that of both FOXP2 and FOXP3 FHDs. That is, FOXP1 FHD binds DNA in its monomeric and dimeric forms. Furthermore, it is likely that each monomer of the dimer binds DNA simultaneously.

FOXP1, FOXP2 and FOXP3 FHD homotypic associations:

The main aim of this study was to identify whether FOXP1-FOXP2 FHD and FOXP1-FOXP3 FHD heterodimerization could occur. If so, it may be possible that these heterotypic associations are involved in transcriptional regulation by these proteins. This would also imply that the heterotypic associations of the full-length protein that have already been established may contribute to the associations between the FHDs (Thulo *et al*., 2020). Homodimerization studies were also conducted and compared to the heterodimerization studies. The binding affinity of protein-protein interactions are classified according to whether they are low affinity ($K_D > 10^{-6}$ M), high affinity ($K_D < 10^{-10}$ M) or average affinity ($K_D = 10^{-6}$ to 10^{-10} M) which is also attributed to the functions or roles they play physiologically (Kastritis *et al*., 2011).

Fluorescence anisotropy revealed that there is a homotypic interaction of FOXP3 FHD, even though FOXP3 FHD is exclusively dimeric (Bandukwala *et al*., 2011; Medina *et al.*, 2016; Stroud *et al*., 2006). The K_D obtained from these experiments was 1.1 µM, indicating that the binding affinity is moderate. However, since the FOXP3 FHD homodimer has been postulated to be important for its physiological function, the K_D value is expected to show high affinity of binding instead of moderate (Bandukwala *et al*., 2011). Therefore these results could suggest that FOXP3 FHD is likely forming higher order complexes which would not be uncommon for FOXP3. That is, the FOXP3 protein has been suggested to have the capability of forming higher order oligomers via a domain that adjacent to the FHD (Song *et al*., 2012), and therefore it would be no surprise if this behaviour was also assumed by the FOXP3 FHD.

Unexpectedly, the FOXP1 FHD homotypic association exhibits a weaker binding affinity than that of FOXP2 FHD, despite the fact that FOXP1 FHD shows a greater propensity to dimerize. It was observed that the FOXP1 FHD exhibits a low affinity to bind whilst a moderate binding affinity was observed for FOXP2 FHD. This implies that although the FHD of FOXP1 is prone to form homodimers, these dimers are less stable than those of the FOXP2 FHD. However, it may be argued that the lack of sufficient data points in the FOXP2 FHD fluorescence anisotropy experiments may have contributed to the lower K_D and therefore additional data points are required for these experiments.

FOXP1, FOXP2 and FOXP3 FHD heterotypic associations:

To date, research has only been attentive to homotypic associations of the FOXP FHD (Bandukwala *et al*., 2011; Chu *et al.*, 2011; Stroud *et al.*, 2006). Little to no work has been done on heterodimerization of the FHDs even though the FOXP proteins have been observed to form heterotypic associations with each other. Work on possible heterotypic association of the FHDs has only implied indirect interaction of the FHDs, specifically, in the case of FOXP1 and FOXP3 FHDs (Bandukwala *et al*., 2011). Therefore, this study aimed to investigate the direct heterotypic interactions between these FHDs.

In this study, the heterotypic association between the FOXP1 and FOXP2 FHDs were studied. Similarly, the heterotypic association between the FOXP1 and FOXP3 FHDs were investigated. In both cases, a low affinity interaction ($K_D > 10^{-6}$ M) was observed with the FOXP1-FOXP3 FHD heterodimer showing the stronger binding affinity. A low binding affinity is a typical characteristic of transcription factors that establish transient complexes (Erijman *et al*., 2014). Therefore, this indicates that these FOXP1-FOXP2 and FOXP1-FOXP3 FHD heterodimers are likely able to bind other proteins to form various multimolecular complexes that allow them to regulate transcription differentially. This feature is observed with FOXP3 whereby, depending on whether it establishes a protein complex with YY1 and RZH2 or with KAT5, RELA and IKF2, it acts either as an activator or repressor of transcription (Kwon *et al*., 2017).

Conclusion

Dimerization is a property of the FOXP1, FOXP2 and FOXP3 FHDs. In this study, both the homo- and heterotypic associations of these FHDs were investigated. Since this work has never been established the main aim was to determine whether heterodimerization can occur between FOXP1 FHD with FOXP2 FHD and FOXP1 FHD and FOXP3 FHD. From this study, it was established that FOXP3 FHD exhibits the highest propensity for homodimerization, followed by FOXP1 FHD. FOXP2 FHD showed an extremely weak propensity for homodimerization but despite this, its homodimers are more stable than FOXP1 FHD which could be due to the strength of interaction of its hydrophobic amino acids residues at the dimer interface. Although the heterodimerization of these proteins occurs, these heterotypic interactions exhibit low binding affinities, suggesting that these heterodimers allow the flexibility to form various multimolecular complexes in order to differentially regulate transcription. For future work, these established heterodimers will be investigated in the presence of DNA. This would assist in determining how these associations affect DNA-binding and therefore assist in depicting how transcription is regulated by these heterodimers.

Chapter 6

References

Amoutzias, G. D., Robertson, D. L., Van de Peer, Y., and Oliver, S. G. (2008) Choose your partners: dimerization in eukaryotic transcription factors. *Trends in Biochemical Sciences* 33(5): 220–229.

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C.-F., Coburn, D., Newburger, D. E., Morris, Q., Hughes, T. R., and Bulyk, M. L. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science (New York, N.Y.)* 324(5935): 1720–1723.

Bandukwala, H. S., Wu, Y., Feurer, M., Chen, Y., Barbosa, 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): 479–491.

Banham, A. H., Beasley, N., Campo, E., Fernandez, P. L., Fidler, C., Gatter, K., Jones, M., Mason, D. Y., Prime, J. E., Trougouboff, P., Wood, K., and Cordell, J. L. (2001) The FOXP1 Winged Helix Transcription Factor Is a Novel Candidate Tumor Suppressor Gene on Chromosome 3p. *Cancer Research* 61(24): 8820–8829.

Bashardanesh, Z., Elf, J., Zhang, H., and van der Spoel, D. (2019) Rotational and Translational Diffusion of Proteins as a Function of Concentration. *ACS Omega* 4(24): 20654–20664.

Benayoun, B. A., Caburet, S., and Veitia, R. A. (2011) Forkhead transcription factors: key players in health and disease. *Trends in Genetics* 27(6): 224–232.

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

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): 6929–6948.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005) Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development* 15(2): 116–124.

Bornhorst, J. A., and Falke, J. J. (2000) [16] Purification of Proteins Using Polyhistidine Affinity Tags. *Methods in enzymology* 326: 245.

Bowers, J. M., and Konopka, G. (2012) The Role of the FOXP Family of Transcription Factors in ASD. *Disease Markers* 33(5): 251–260.

Bryne, J. C., Valen, E., Tang, M.-H. E., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., and Sandelin, A. (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Research* 36(Database issue): D102-106.

Castellanos, M., Mothi, N., and Muñoz, V. (2020) Eukaryotic transcription factors can track and control their target genes using DNA antennas. *Nature Communications* 11(1): 540.

Chen, Yongheng, 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): 1268–1282.

Chen, Yu, and Barkley, M. D. (1998) Toward Understanding Tryptophan Fluorescence in Proteins. *Biochemistry* 37(28): 9976–9982.

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): 908–924.

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* 364(6436): 412–420.

Co, M., Anderson, A. G., and Konopka, G. (2020) FOXP transcription factors in vertebrate brain development, function, and disorders. *Wiley Interdisciplinary Reviews. Developmental Biology* 9(5): e375.

Dessau, M. A., and Modis, Y. (2011) Protein Crystallization for X-ray Crystallography. *Journal of Visualized Experiments* (47): 2285. doi:10.3791/2285.

Ehsani, R., Bahrami, S., and Drabløs, F. (2016) Feature-based classification of human transcription factors into hypothetical sub-classes related to regulatory function. *BMC Bioinformatics* 17(1): 459.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P., and Pääbo, S. (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418(6900): 869–872.

Esadze, A., Kemme, C. A., Kolomeisky, A. B., and Iwahara, J. (2014) Positive and negative impacts of nonspecific sites during target location by a sequence-specific DNA-binding protein: origin of the optimal search at physiological ionic strength. *Nucleic Acids Research* 42(11): 7039–7046.

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): 1212–1227.

Fontenot, J. D., Gavin, M. A., and Rudensky, A. Y. (2003) Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nature Immunology* 4(4): 330–336.

Friedman, J. R., and Kaestner, K. H. (2006) The Foxa family of transcription factors in development and metabolism. *Cellular and molecular life sciences: CMLS* 63(19–20): 2317–2328.

Funnell, A. P. W., and Crossley, M. (2012) Homo- and Heterodimerization in Transcriptional Regulation. In Matthews, J. M. (Ed.), *Protein Dimerization and Oligomerization in Biology*. New York, NY: Springer doi:10.1007/978-1-4614-3229-6_7.

Golson, M. L., and Kaestner, K. H. (2016) Fox transcription factors: from development to disease. *Development* 143(24): 4558–4570.

Gorman, J., and Greene, E. C. (2008) Visualizing one-dimensional diffusion of proteins along DNA. *Nature Structural & Molecular Biology* 15(8): 768–774.

Gouldson, P. R., Snell, C. R., Bywater, R. P., Higgs, C., and Reynolds, C. A. (1998) Domain swapping in G-protein coupled receptor dimers. *Protein Engineering Design and Selection* 11(12): 1181–1193.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004) Codon bias and heterologous protein expression. *Trends in Biotechnology* 22(7): 346–353.

Hannenhalli, S., and Kaestner, K. H. (2009) The evolution of Fox genes and their role in development and disease. *Nature Reviews Genetics* 10(4): 233–240.

Hayes, M. V., Sessions, R. B., Brady, R. L., and Clarke, A. R. (1999) Engineered assembly of intertwined oligomers of an immunoglobulin chain. *Journal of Molecular Biology* 285(4): 1857–1867.

Hellman, L. M., and Fried, M. G. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature Protocols* 2(8): 1849–1861.

Hong, P., Koza, S., and Bouvier, E. S. P. (2012) A REVIEW SIZE-EXCLUSION CHROMATOGRAPHY FOR THE ANALYSIS OF PROTEIN BIOTHERAPEUTICS AND THEIR AGGREGATES. *Journal of Liquid Chromatography & Related Technologies* 35(20): 2923–2950.

Hu, Y. F., Luscher, B., Admon, A., Mermod, N., and Tjian, R. (1990) Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. *Genes & Development* 4(10): 1741–1752.

Hu, H., Wang, B., Borde, M., Nardone, J., Maika, S., Allred, L., Tucker, P. W., and Rao, A. (2006) Foxp1 is an essential transcriptional regulator of B cell development. *Nature Immunology* 7(8): 819–826.

Huebner, K. (2001) Tumor suppressors on 3p: A neoclassic quartet. *Proceedings of the National Academy of Sciences of the United States of America* 98(26): 14763–14765.

Jacob, F., and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3: 318–356.

Jameson, D. M., and Ross, J. A. (2010) Fluorescence Polarization/Anisotropy in Diagnostics and Imaging. *Chemical reviews* 110(5): 2685–2708.

Jones, S. (2004) An overview of the basic helix-loop-helix proteins. *Genome Biology* 5(6): 226.

Kaestner, K. H., Knöchel, W., and Martínez, D. E. (2000) Unified nomenclature for the winged helix/forkhead transcription factors. *Genes & Development* 14(2): 142–146.

Kaufmann, E., and Knöchel, W. (1996) Five years on the wings of fork head. *Mechanisms of Development* 57(1): 3–20.

Klemm, J. D., Schreiber, S. L., and Crabtree, G. R. (1998) Dimerization as a regulatory mechanism in signal transduction. *Annual Review of Immunology* 16: 569–592.

Koon, H. B., Ippolito, G. C., Banham, A. H., and Tucker, P. W. (2007) FOXP1: a potential therapeutic target in cancer. *Expert opinion on therapeutic targets* 11(7): 955–965.

Kouzarides, T., and Ziff, E. (1988) The role of the leucine zipper in the fos–jun interaction. *Nature* 336(6200): 646–651.

Kuhlman, B., O'Neill, J. W., Kim, D. E., Zhang, K. Y. J., and Baker, D. (2001) Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proceedings of the National Academy of Sciences of the United States of America* 98(19): 10687–10691.

Kundu, S., and Jernigan, R. L. (2004) Molecular Mechanism of Domain Swapping in Proteins: An Analysis of Slower Motions. *Biophysical Journal* 86(6): 3846–3854.

Kurt, S., Fisher, S. E., and Ehret, G. (2012) Foxp2 Mutations Impair Auditory-Motor Association Learning. *PLOS ONE* 7(3): e33130.

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 severe speech and language disorder. *Nature* 413(6855): 519–523.

Lalmansingh, A. S., Karmakar, S., Jin, Y., and Nagaich, A. K. (2012) Multiple modes of chromatin remodeling by Forkhead box proteins. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1819(7): 707–715.

Lam, E. W.-F., Brosens, J. J., Gomes, A. R., and Koo, C.-Y. (2013) Forkhead box proteins: tuning forks for transcriptional harmony. *Nature Reviews. Cancer* 13(7): 482–495.

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 (New York, N.Y.)* 240(4860): 1759–1764.

Lang, X., Zhang, W., Song, X., Zhang, G., Du, X., Zhou, Y., Li, Z., and Zhang, X. Y. (2019) FOXP2 contributes to the cognitive impairment in chronic patients with schizophrenia. *Aging (Albany NY)* 11(16): 6440–6448.

Laniel, M.-A., Béliveau, A., and Guérin, S. L. (2001) Electrophoretic Mobility Shift Assays for the Analysis of DNA-Protein Interactions. In Moss, T. (Ed.), *DNA-Protein Interactions: Principles and Protocols*. Totowa, NJ: Humana Press doi:10.1385/1-59259-208-2:013.

Latchman, D. S. (1993) Transcription factors: an overview. *International Journal of Experimental Pathology* 74(5): 417–422.

Lee, T. I., and Young, R. A. (2013) Transcriptional regulation and its misregulation in disease. *Cell* 152(6): 1237–1251.

Li, C., and Tucker, P. W. (1993) DNA-binding properties and secondary structural model of the hepatocyte nuclear factor 3/fork head domain. *Proceedings of the National Academy of Sciences of the United States of America* 90(24): 11583–11587.

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): 809–822.

Liu, L., Byeon, I.-J. L., Bahar, I., and Gronenborn, A. M. (2012) Domain Swapping Proceeds via Complete Unfolding: A 19F- and 1H-NMR Study of the Cyanovirin-N Protein. *Journal of the American Chemical Society* 134(9): 4229–4235.

Liu, Y., and Eisenberg, D. (2002) 3D domain swapping: as domains continue to swap. *Protein Science: A Publication of the Protein Society* 11(6): 1285–1299.

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. *Mechanisms of Development* 119 Suppl 1: S197-202.

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

Ma, P. C. M., Rould, M. A., Weintraub, H., and Pabo, C. O. (1994) Crystal structure of MyoD bHLH domain-DNA complex: Perspectives on DNA recognition and implications for transcriptional activation. *Cell* 77(3): 451–459.

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

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

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): 2349–2360.

Mojsiewicz-Pieńkowska, K. (2014) Size Exclusion Chromatography a Useful Technique For Speciation Analysis of Polydimethylsiloxanes. In Inamuddin, Dr. and Mohammad, A. (Eds.), *Green Chromatographic Techniques*. Dordrecht: Springer Netherlands doi:10.1007/978-94-007-7735-4_9.

Monaghan, A. P., Kaestner, K. H., Grau, E., and Schutz, G. (1993) Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development* 119(3): 567–578.

Morris, G., and Fanucchi, S. (2016) A Key Evolutionary Mutation Enhances DNA Binding of the FOXP2 Forkhead Domain. *Biochemistry* 55(13): 1959–1967.

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* 399(8): 881–893.

Murre, C., McCaw, P. S., and Baltimore, D. (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56(5): 777–783.

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): 5991–6004.

O'Shea, E. K., Rutkowski, R., Stafford, W. F., and Kim, P. S. (1989) Preferential heterodimer formation by isolated leucine zippers from fos and jun. *Science (New York, N.Y.)* 245(4918): 646–648.

Pabo, C. O., and Sauer, R. T. (1992) TRANSCRIPTION FACTORS: Structural Families and Principles of DNA Recognition. *Annual Review of Biochemistry* 61(1): 1053–1095.

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* 34(2): 111–121.

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., and Mann, R. S. (2010) Origins of specificity in protein-DNA recognition. *Annual Review of Biochemistry* 79: 233–269.

Rousseau, F., Schymkowitz, J. W. H., and Itzhaki, L. S. (2003) The Unfolding Story of Three-Dimensional Domain Swapping. *Structure* 11(3): 243–251.

Rowan, S., Siggers, T., Lachke, S. A., Yue, Y., Bulyk, M. L., and Maas, R. L. (2010) Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes & Development* 24(10): 980–985.

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): 37672–37679.

Shi, C., Zhang, X., Chen, Z., Sulaiman, K., Feinberg, M. W., Ballantyne, C. M., Jain, M. K., and Simon, D. I. (2004) Integrin engagement regulates monocyte differentiation through the forkhead transcription factor Foxp1. *Journal of Clinical Investigation* 114(3): 408–418.

Shimeld, S. M., Degnan, B., and Luke, G. N. (2010) Evolutionary genomics of the Fox genes: Origin of gene families and the ancestry of gene clusters. *Genomics* 95(5): 256–260.

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. *The Journal of Biological Chemistry* 276(29): 27488–27497.

Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M., and Sestan, N. (2016) The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* 89(2): 248–268.

Sin, C., Li, H., and Crawford, D. A. (2015) Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. *Journal of molecular neuroscience* 55(2): 437–448.

Smith, N. C., and Matthews, J. M. (2016) Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Current Opinion in Structural Biology* 38: 68–74.

Sonawane, A. R., Platig, J., Fagny, M., Chen, C.-Y., Paulson, J. N., Lopes-Ramos, C. M., DeMeo, D. L., Quackenbush, J., Glass, K., and Kuijjer, M. L. (2017) Understanding Tissue-Specific Gene Regulation. *Cell reports* 21(4): 1077–1088.

Song, X., Li, B., Xiao, Y., Chen, C., Wang, Q., Liu, Y., Berezov, A., Xu, C., Gao, Y., Wu, S.-L., 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* 1(6): 665–675.

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): 159–166.

Teufel, A., Wong, E. A., Mukhopadhyay, M., Malik, N., and Westphal, H. (2003) FoxP4, a novel forkhead transcription factor. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1627(2–3): 147–152.

Thulo, M., Rabie, M. A., Pahad, N., Donald, H. L., Blane, A. A., Perumal, C. M., Penedo, J. C., and Fanucchi, S. (2021) The influence of various regions of the FOXP2 sequence on its structure and DNA-binding function. *Bioscience Reports* 41(1): BSR20202128.

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(4): 252–263.

Villicaña, C., Cruz, G., and Zurita, M. (2014) The basal transcription machinery as a target for cancer therapy. *Cancer Cell International* 14(1): 18.

Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., and Bonovich, M. (2002) Classification of Human B-ZIP Proteins Based on Dimerization Properties. *Molecular and Cellular Biology* 22(18): 6321–6335.

von Hippel, P. H., and Berg, O. G. (1989) Facilitated target location in biological systems. *The Journal of Biological Chemistry* 264(2): 675–678.

Wang, B., Lin, D., Li, C., and Tucker, P. (2003) Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. *The Journal of Biological Chemistry* 278(27): 24259–24268.

Wang, B., Weidenfeld, J., Lu, M. M., Maika, S., Kuziel, W. A., Morrisey, E. E., and Tucker, P. W. (2004) Foxp1 regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation. *Development (Cambridge, England)* 131(18): 4477–4487.

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* : mvx003. doi:10.1093/jb/mvx003.

Wei, Y., Thyparambil, A. A., and Latour, R. A. (2014) Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1844(12): 2331–2337.

Weigel, D., Jürgens, G., Küttner, F., Seifert, E., and Jäckle, H. (1989) The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. *Cell* 57(4): 645–658.

White, M. A., Parker, D. S., Barolo, S., and Cohen, B. A. (2012) A model of spatially restricted transcription in opposing gradients of activators and repressors. *Molecular Systems Biology* 8: 614.

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): e1005251.

Xu, D., Tsai, C. J., and Nussinov, R. (1998) Mechanism and evolution of protein dimerization. *Protein Science : A Publication of the Protein Society* 7(3): 533–544.

Yang, V. W. (1998) Eukaryotic Transcription Factors: Identification, Characterization and Functions. *The Journal of Nutrition* 128(11): 2045–2051.

Yesudhas, D., Batool, M., Anwar, M. A., Panneerselvam, S., and Choi, S. (2017) Proteins Recognizing DNA: Structural Uniqueness and Versatility of DNA-Binding Domains in Stem Cell Transcription Factors. *Genes* 8(8).

Zabarovsky, E. R., Lerman, M. I., and Minna, J. D. (2002) Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. *Oncogene* 21(45): 6915–6935.

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): 1145.

Zhou, Q., Liu, M., Xia, X., Gong, T., Feng, J., Liu, W., Liu, Y., Zhen, B., Wang, Y., Ding, C., and Qin, J. (2017) A mouse tissue transcription factor atlas. *Nature Communications* 8: 15089.

Zhu, L., and Huq, E. (2011) Mapping functional domains of transcription factors. *Methods Molecular Biology* 754: 167-84.