School of Molecular and Cell Biology

Faculty of Science

University of the Witwatersrand

Evaluating FOXP2 homo-oligomerisation interfaces and their role in structure and DNA binding

Thesis to be submitted in fulfilment of the requirements

for the degree of Philosophiae Doctor

Naadira Pahad

(684121)

Supervisor: Dr Sylvia Fanucchi

July 2020

Declaration

I, Naadira Pahad (684121), am a student registered for the degree of Philosophiae Doctor (PhD) at the University of the Witwatersrand.

I hereby declare the following;

- I am aware that plagiarism is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- 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 work of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is evidence that this is not my own unaided work or that I have failed to acknowledge the sources of the ideas in my writing.

Ms N Pahad

01/07/2020

Abstract

Forkhead box P2 (FOXP2) is a transcription factor involved in neural development, as well as the development of the gut, lung, and brain. Mutations in FOXP2 are associated with various diseases including developmental verbal dyspraxia, schizophrenia, and both breast cancer and prostate cancer, two of the most prevalent cancers worldwide. FOXP2 is a large multi-domain protein, containing a glutamine-rich region, zinc finger domain, leucine zipper domain, and DNA-binding forkhead domain. Like other FOXP proteins, FOXP2 is able to form homo- and hetero-oligomers, and FOXP subfamily members undergo oligomerisation-dependent DNA binding and transcriptional activity, a phenomenon that is unique to the FOXP subfamily of FOX transcription factors. Previous studies have suggested that only the leucine zipper is necessary for FOXP oligomerisation. However, the FOXP forkhead domain has the unique ability to form domain-swapped dimers, providing an additional interface for protein-protein interactions. The effects of a disruption to the domain-swapped dimer have only been studied on the FOXP2 forkhead domain. In this study of FOXP2, the R401D mutation of the leucine zipper and the A539P mutation of the forkhead domain were used to investigate the contributions of each interface to FOXP2 homo-oligomerisation, and to explore the effects of disruptions to each interface on FOXP2 structure and DNA binding. Constructs used in this study comprise the leucine zipper to the C-terminal end of the protein (leucine zipper - end). The effects of the mutations were established by comparisons of the mutants with wild-type FOXP2 (leucine zipper - end) and the combined effects of the mutations were investigated using the R401D/A539P FOXP2 (leucine zipper - end) construct. *In silico* predictions did not predict any binding sites other than the leucine zipper and forkhead domain of FOXP2 (leucine zipper - end), therefore these two domains formed the only interfaces for protein-protein and protein-DNA interactions. The A539P mutation alone had no apparent effects on the oligomeric state of the protein, since according to size exclusion chromatography both wild-type and A539P FOXP2 (leucine zipper - end) existed as a single large species of approximately hexameric size. R401D FOXP2 (leucine zipper - end) was observed to form an intermediate species of approximately trimeric size. The effects of the A539P mutation were apparent in R401D/A539P FOXP2 (leucine zipper - end), which existed as a mixture of the intermediate species and a small species of approximately dimeric size. Electrophoretic mobility shift assays showed that the same oligomeric states occurred upon DNA binding. *In silico* methods predicted that all FOXP2 (leucine zipper - end) variants shared similar secondary structures, and were predominantly disordered, which was confirmed with far-UV circular dichroism spectropolarimetry. The large proportion of disordered regions increases the propensity of FOXP2 (leucine zipper - end) for protein-protein interactions, and confers conformational freedom on the protein to facilitate any structural rearrangements required for protein-protein and protein-DNA interactions. Fluorescence

anisotropy DNA-binding assays showed that although the R401D mutation decreased the DNA-binding affinity of FOXP2 (leucine zipper - end), the decrease was less significant than that observed for A539P FOXP2 (LeuZip - end). This could be due to the R401D mutation not preventing leucine zipper associations completely, since R401D FOXP2 (leucine zipper - end) was not monomeric. Despite adopting the same quaternary structure as the wild-type, A539P FOXP2 (leucine zipper - end) had a significantly weaker affinity for DNA, emphasising the importance of forkhead domain domain-swapping for DNA binding and revealing that the formation of a hexamer specifically was not significant for DNA binding. R401D/A539P FOXP2 (leucine zipper - end) had the weakest affinity for DNA, due to the combined effects of disruptions at both the leucine zipper and forkhead domain interfaces. Intrinsic tryptophan fluorescence spectroscopy of wild-type FOXP2 (leucine zipper - end) was the most quenched upon DNA binding, under non-reducing conditions. Since wild-type FOXP2 (leucine zipper - end) had the highest affinity for DNA, this could be due to the formation of additional DNA-protein contacts that were not present in the mutants, or due to the binding of more protein molecules to DNA resulting in an enhanced quenching effect. Thermal unfolding revealed that wild-type FOXP2 (leucine zipper - end) was the least stable variant, whereas R401D/A539P FOXP2 (leucine zipper - end) was the most stable, possibly indicating a more transient hexameric association and more stable dimeric association. Stern-Volmer constants showed that hexameric wild-type and A539P FOXP2 (leucine zipper - end) experienced the largest conformational changes upon DNA binding. Although the R401D/A539P mutant experienced a conformational change upon DNA binding, the change was not significant. In addition, although the forkhead domain itself is more structured upon DNA binding, far-UV circular dichroism spectropolarimetry showed an increase in global disorder upon DNA binding, under non-reducing conditions, possibly increasing the propensity for protein-protein interactions. These results suggest that the FOXP2 (leucine zipper - end) hexamer could dissociate – possibly into the more stable dimeric form – in the presence of DNA in order to efficiently locate the appropriate *cis*-regulatory element and then re-associate upon binding DNA. The possible existence of multiple oligomeric states could regulate FOXP2-controlled gene expression.

باسمه تعلی

In the name of the One through whom anything is possible

To my family

“When you reach the end of what you should know,
you will be at the beginning of what you should sense.”
– Khalil Gibran

Acknowledgements

To my supervisor, Dr Sylvia Fanucchi, thank you for being a wonderful source of knowledge, expertise, and advice. You have guided me throughout my postgraduate career, and rekindled my curiosity for understanding life's mysteries.

Thank you to my friends and colleagues at the PSFRU. A special thank you to the FOXy group – Dr Monare Thulo, Ashleigh Blane, Ilan Kirkel, Joni Symon, Heather Donald, and Cardon Perumal – and to my good friends Roland Worth, Mia Færch, Zaahida Sheik Ismail, and Dean Sherry. You made the difficult lab days easier, and I will never forget all our laughs and adventures.

To my parents, thank you for being my home, and for believing in me in everything I do. I appreciate all that you do for me. Thank you to my sisters for being my best friends and personal cheerleaders.

Thank you to Waseem for always being so patient, kind, and supportive, and for taking such good care of me when I was too stressed to take care of myself. Thank you to Waseem's parents for supporting me in my endeavours.

I am grateful to the University of the Witwatersrand and the South African National Research Foundation for their financial support.

Research outputs

OUTPUTS RELATED TO THIS PROJECT

8 - 11 July 2018

Joint Congress of the South African Society of Biochemistry and Molecular Biology (SASBMB) and the Federation of African Societies of Biochemistry and Molecular Biology (FASBMB)

Presentation type: Poster

Title: Structural and DNA-binding effects of FOXP2 oligomer formation

Authors: Naadira Pahad and Sylvia Fanucchi

29 - 30 October 2018

9th Wits Cross Faculty Graduate Symposium

Presentation type: Poster

Title: Dimers and tetramers of the FOXP2 protein

Authors: Naadira Pahad and Sylvia Fanucchi

29 November 2018

Molecular Biosciences Research Thrust (MBRT) Postgraduate Research Day

Presentation type: Poster

Title: Structural and DNA-binding effects of FOXP2 oligomer formation

Authors: Naadira Pahad and Sylvia Fanucchi

3 - 4 September 2019

10th Wits Cross Faculty Graduate Symposium

Presentation type: Gradflash (video)

Title: Discovering more about FOXP2

Creator and producer: Naadira Pahad

28 November 2019

Molecular Biosciences Research Thrust (MBRT) Postgraduate Research Day

Presentation type: Oral

Title: Structural and DNA-binding effects of FOXP2 homoassociations

Presenter: Naadira Pahad

Recognition: Silver award

OTHER OUTPUTS

*Morris, G., *Pahad, N., Dirr, H.W., and Fanucchi, S. (2018). A conserved cation binding site in the DNA binding domain of forkhead box transcription factors regulates DNA binding by FOXP2. Archives of Biochemistry and Biophysics. 657, 56–64.

* Joint first authors

Contents

List of abbreviations

3D	three-dimensional
A	absorbance
ATF2	activating transcription factor 2
bHLH	basic-region helix-loop-helix
BLAST	Basic Local Alignment Search Tool
bZip	basic-region leucine zipper
CaBLAM	Cα Based Low-resolution Annotation Method
CASP	Critical Assessment of Structure Prediction
CD	circular dichroism
DBD	DNA-binding domain
DISOPRED	Disorder Prediction
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
DVD	developmental verbal dyspraxia
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
ERG	ETS-related gene
ETS	erythroblast transformation-specific
ExPASy	Expert Protein Analysis System
F	fluorescence intensity in the presence of quencher
F_0	fluorescence intensity in the absence of quencher
FHD	forkhead domain
FOX	forkhead box
FPLC	Fast Protein Liquid Chromatography
FT	flow-through
GCN4	General Control Nondepressible 4
GST	glutathione-S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	poly-histidine tag
HSF1	heat shock factor 1

HTH	helix-turn-helix
I	insoluble fraction
I-TASSER	Iterative Threading Assembly Refinement
I_h/I_v	intensity of horizontally/vertically polarised light
IMAC	immobilised metal-affinity chromatography
IPEX	immune dysregulation, polyendocrinopathy, enteropathy and X-linked inheritance
IPTG	isopropyl-1-thio-β-D-galactopyranoside
IW	imidazole wash
K_D	dissociation constant
K_{SV}	Stern-Volmer constant
LB	lysogeny broth
LeuZip	leucine zipper
mFoxp3	mouse Foxp3
MRE	mean residue ellipticity
MW	molecular weight marker
MyRF	myelin-gene regulatory factor
n	number of peptide bonds in the protein chain
NLS	nuclear localisation signal
OD	optical density
OPLS	Optimized Potential for Liquid Simulations
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PMSF	phenylmethylsulfonyl fluoride
PSI-BLAST	Position-Specific Iterative-BLAST
PSIPRED	PSI-BLAST based secondary structure prediction
PSSM	position-specific scoring matrix
Q	quencher
Q-rich	glutamine-rich
r	anisotropy
R_f	relative mobility
RMSD	root-mean-square deviation
RNA	ribonucleic acid
S	soluble fraction

SDS	sodium dodecyl sulfate
SDW	salt/detergent wash
SOC	Super Optimal broth with Catabolite repression
TBE	Tris-Cl/borate/EDTA
TF	transcription factor
TFEB	transcription factor EB
Tris-Cl	tris(hydroxymethyl)aminomethane hydrochloride
UV	ultraviolet
V_e/V_o	elution volume/void volume
YT	yeast extract-tryptone
ZnF	zinc finger

The IUPAC abbreviations for the chemical elements and the IUPAC-IUBMB abbreviations for amino acids and DNA nucleotide bases were used.

List of figures

List of tables

Chapter 1: Introduction

Whole genome sequencing has revealed that vertebrates have approximately twice as many genes as invertebrates, but this increase in number is largely due to the duplication of invertebrate genes rather than the evolutionary development of new genes (Adams *et al.,* 2000; Baltimore, 2001; Lander *et al.,* 2001; Ruvkun and Hobert, 1998). Therefore, the complexity of vertebrates is not due to a complexity of the genome. Evidence suggests that this complexity has arisen due to more intricate regulation of gene expression, especially through transcription factors (TFs), since there is a large increase in the number of TFs per gene in vertebrates compared to invertebrates (Aoyagi and Wassarman, 2000; Lander *et al.,* 2001; Levine and Tjian, 2003; Wyrick and Young, 2002). Furthermore, the most complex multicellular organisms express the most diverse and complex range of TFs (de Mendoza *et al.,* 2013).

A recent study identified 1558 human TF-encoding genes, thus TFs comprise a large group of proteins in humans (Wingender *et al.,* 2013). Since some TF-encoding genes give rise to multiple TF isoforms due to alternative splicing, post-translational processing, and epigenetic events, this results in over 2900 classified human TFs (Wingender *et al.,* 2013). TFs are key players in the regulation of gene expression (Spitz and Furlong, 2012). TFs activate or repress gene expression in a context-specific manner by binding distinct deoxyribonucleic acid (DNA) sequences and either activating or repressing the recruitment and binding of transcriptional machinery (Spitz and Furlong, 2012). TFs display the ability to influence the rate of transcription from DNA to ribonucleic acid (RNA) (Latchman, 1993). Upstream of an open reading frame lies the promoter, a stretch of DNA of variable length that is important for regulating the expression of that gene (Figure 1.1A) (Lee and Young, 2000). The promoter consists of numerous *cis*-regulatory elements – *cis*-regulatory elements of the core promoter to which transcription initiation factors and RNA polymerase II bind, and those of the distal regions to which TFs and co-activators bind (Burley and Roeder, 1996; Lee and Young, 2000; Zou *et al.,* 2011). The distal *cis*-regulatory elements to which TFs bind are of different types; enhancers to which transcriptional activators bind, and silencers to which transcriptional repressors bind (Lee and Young, 2000). Regulation of transcription relies upon the interaction of each of the proteins involved in transcription (such as the initiation factors, RNA polymerase II, TFs and co-activators) with specific DNA *cis*-regulatory elements, and the interactions of the proteins with each other, often requiring the bending of DNA to facilitate the formation of a large DNA-protein complex (Figure 1.1B and 1.1C) (Gaston *et al.,* 1992; Lee and Young, 2000). This provides a remarkable mechanism by which gene expression is regulated in a tightly controlled manner. Additional regulatory mechanisms, such as post-translational modifications of TFs, allow for further control of gene expression and

post-translational modifications of protein products allow for increased complexity of multicellular organisms (Benayoun and Veitia, 2009; Smith and Kelleher, 2013; Tootle and Rebay, 2005).

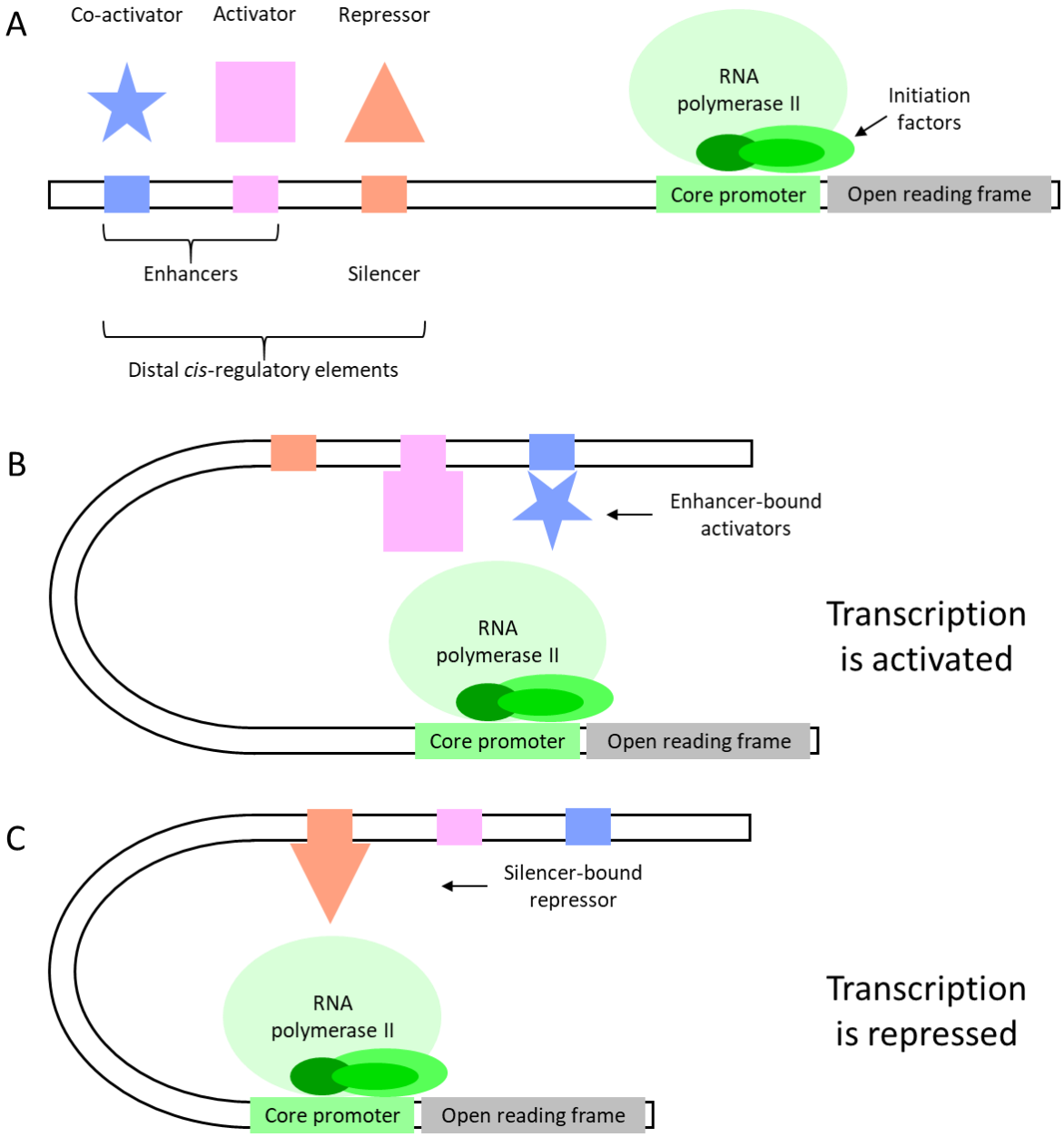


Figure 1.1: The regulation of gene expression by transcription factors. (A) The core promoter, located directly upstream of an open reading frame, contains the *cis*-regulatory elements for the binding of RNA polymerase II and other general transcription initiation factors. Gene expression is controlled by TFs (activators, co-activators, and repressors) which bind to *cis*-regulatory elements in the distal regions. **(B)** Transcription is initiated upon binding of the activator/co-activator to enhancers. Necessary interactions between the activator/co-activator and the proteins bound to the core promoter may require DNA bending. **(C)** Transcription is repressed upon binding of the repressor to a silencer. Necessary interactions between the repressor and the proteins bound to the core promoter may require DNA bending.

1.1 TRANSCRIPTION FACTOR SUPERCLASSES

The binding of TFs to *cis*-regulatory elements is mediated via DNA-binding domains (DBDs). According to a classification system based on DBDs that was first used to classify human TFs and has since been expanded to include non-human mammalian TFs, TFs are classified into ten superclasses (Figure 1.2) which are described briefly (Wingender *et al.,* 2013, 2018). TFs of the basic DBD superclass interact with DNA via basic regions that lack structure when unbound, but become α-helical upon DNA binding (Figure 1.2A) (Wingender *et al.,* 2018). TFs with zinc-coordinating DBDs are those that bind DNA via a Cys2His2, Cys4, Cys6 or other zinc finger motif (Figure 1.2B) (Wingender *et al.,* 2018). A typical helix-turn-helix (HTH) DBD binds to DNA via a recognition helix that inserts into the major groove of DNA, with the second α-helix interacting with the DNA backbone non-specifically (Figure 1.2C) (Wingender *et al.,* 2018). The first three superclasses (basic DBDs, zinc-coordinating DBDs, and HTH DBDs) are the largest of the superclasses and TFs of all three superclasses bind DNA by the interactions of an α-helix with the major groove of DNA (Wingender *et al.,* 2013, 2018).

The other all-α-helical DBD superclass comprises TFs that bind DNA via α-helical DBDs (Figure 1.2D) that are not basic, zinc-coordinating, or HTH DBDs (Wingender *et al.,* 2018). The fifth TF superclass comprises TFs that bind DNA via α-helices exposed by β-structures, and the DNA-binding α-helices of this superclass interact with DNA by packing against the DNA rather than insertion of the α-helices into a DNA groove (Figure 1.2E) (Wingender *et al.,* 2018). The sixth TF superclass comprises TFs with DBDs with immunoglobulin folds (Wingender *et al.,* 2018). The core of the immunoglobulin fold DBD is primarily composed of β-structures and DNA binding is mediated via loops that extend beyond the β-core (Figure 1.2F) (Wingender *et al.,* 2018). TFs with β-hairpins exposed by an α/β-scaffold bind DNA via the β-hairpin that inserts into the major groove of DNA (Figure 1.2G) (Wingender *et al.,* 2018). Members of the eighth superclass of TFs are able to bind DNA via β-stranded DBDs (Figure 1.2H) (Wingender *et al.,* 2018). Any TF that binds DNA via a β-structure, but is unlike the seventh and eighth superclasses, is classified as having a β-barrel DBD (Figure 1.2I) (Wingender *et al.,* 2018). Lastly, TFs that exhibit sequence-specificity when binding to DNA but lack identifiable or currently characterised DBDs comprise the superclass of TFs with undefined DBDs (Wingender *et al.,* 2018).

The first three and largest of the superclasses (basic DBDs, zinc-coordinating DBDs, and HTH DBDs) shall be described in more detail.

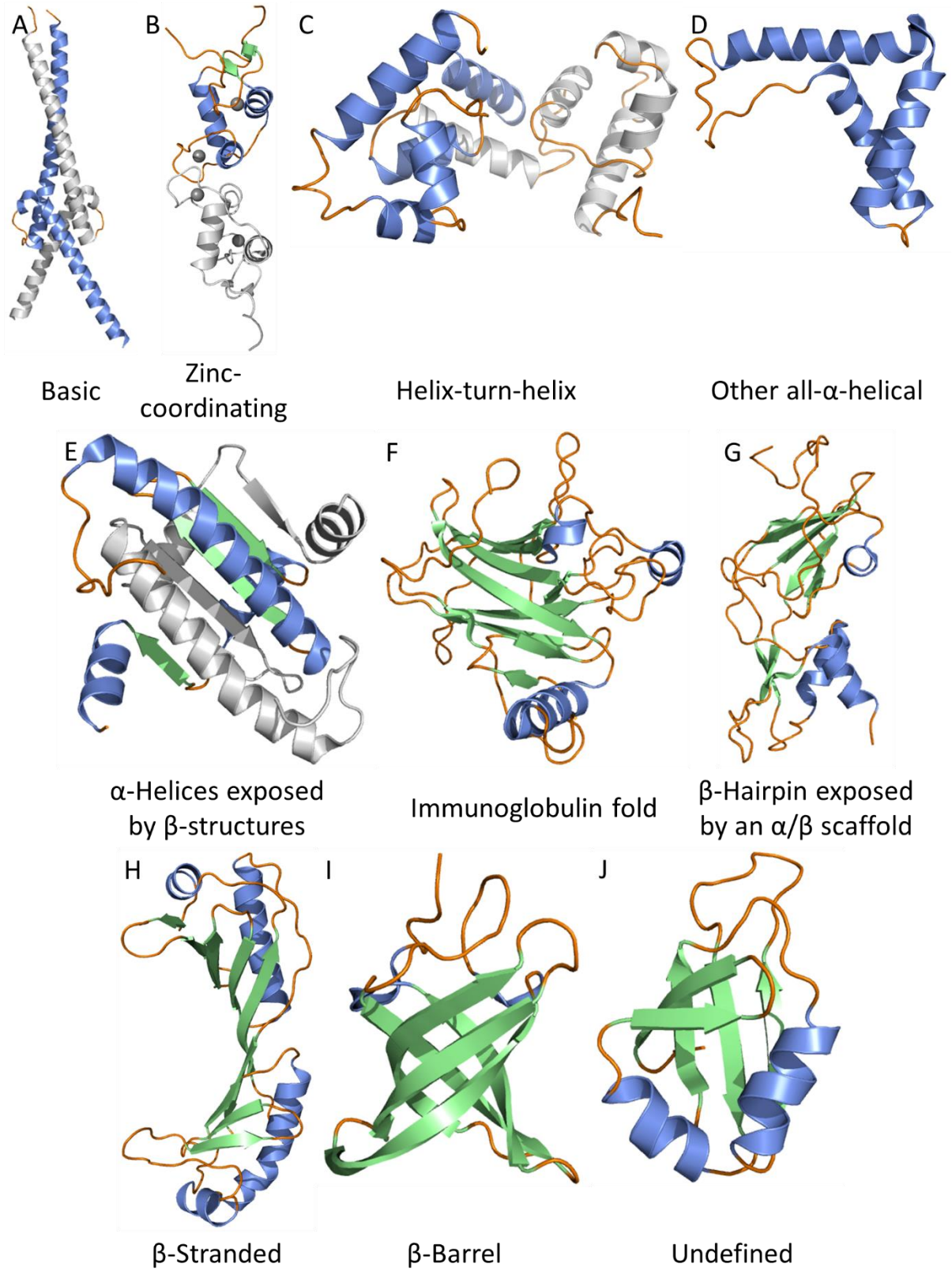


Figure 1.2: Representative DNA-binding domains of each transcription factor superclass. According to a classification system devised by Wingender *et al.* (2018), mammalian TFs belong to one of ten superclasses based on their DBDs. Representatives of each superclass are as follows: **(A)** Basic DBD: an example is the Myc-Max heterodimer (Myc is blue and Max is grey). The DBD (bottom of the image) becomes α-helical upon DNA binding (DNA not shown) (PDB 1NKP) (Nair and Burley, 2003), **(B)** Zinc-coordinating DBD: an example is the glucocorticoid receptor (shown with the zinc ions depicted as

dark grey spheres, and with one monomer in grey) (PDB 1R4O) (Luisi *et al.*, 1991), **(C)** HTH DBD: an example is the λ-repressor homodimer (shown with one monomer in grey) (PDB 1LMB) (Beamer and Pabo, 1992), **(D)** Other all-α-helical DBD: an example is SOX18 (PDB 4Y60) (Klaus *et al.,* 2016), **(E)** α-Helices exposed by β-structures: an example is the MEF2A homodimer (shown with one monomer in grey) (PDB 3KOV) (Wu *et al.,* 2010), **(F)** Immunoglobulin fold DBD: an example is p53 (PDB 2FEJ) (Canadillas *et al.,* 2006), **(G)** β-Hairpin exposed by an α/β-scaffold: an example is GCM1 (PDB 1ODH) (Cohen *et al.,* 2003), **(H)** β-Stranded DBD: an example is TBP (PDB 1CDW) (Nikolov *et al.,* 1996), **(I)** β-Barrel DBD: an example is Lin28 (PDB 3ULJ) (Mayr *et al.,* 2012), **(J)** Undefined DBD: an example is hnRNP D (PDB 1X0F) (Enokizono *et al.,* 2005). α-Helices are blue, β-strands are green, and loops are orange. Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

1.1.1 BASIC DNA-BINDING DOMAINS

The basic DBD superclass of TFs is the third largest of the ten superclasses (Wingender *et al.,* 2018). The most distinguished DBDs of this superclass are the basic-region helix-loop-helix (bHLH) and basic-region leucine zipper (bZip) DBDs (Wingender *et al.,* 2018). DNA-binding of the bHLH and bZip domains is mediated via N-terminal basic regions, which consist of two α-helices that are rich in basic residues (Figure 1.3) (Ellenberger, 1994). In solution, basic DBDs are highly disordered and lack structure due to the large positive charge of the basic amino acid residues (Liu *et al.*, 2006; O'Neil *et al.*, 1990; Patel *et al.*, 1990; Shuman *et al.*, 1990). However, due to the negative charge of DNA, basic DBDs are able to fold into stable α-helices upon binding DNA (Grosberg *et al.*, 2002; Lipfert *et al.*, 2014; Weiss *et al.*, 1990). The N-terminal α-helices interact with the major groove of DNA, forming base-pair and backbone contacts that rely on shape complementarity between amino acid side chains and the major groove (Ellenberger, 1994; Miller *et al.,* 2003).

In addition to facilitating DNA-binding, both bHLH and bZip domains exhibit dimerisation properties (Ellenberger, 1994). In bHLH TFs, dimerisation occurs via the α-helical C-terminal region of each HLH, leading to the formation of a four-helix bundle (Brownlie *et al.,* 1997; Ellenberger, 1994). Dimerisation of two bZip TFs occurs by interactions predominantly of leucine residues at the C-terminal leucine zipper (LeuZip) regions (Glover and Harrison, 1995; Miller *et al.,* 2003). Dimerisation of both bHLH and bZip domains occurs through packing interactions of hydrophobic side chains, though the domains sometimes differ in the spacing and arrangement of conserved hydrophobic residues (Ellenberger, 1994).

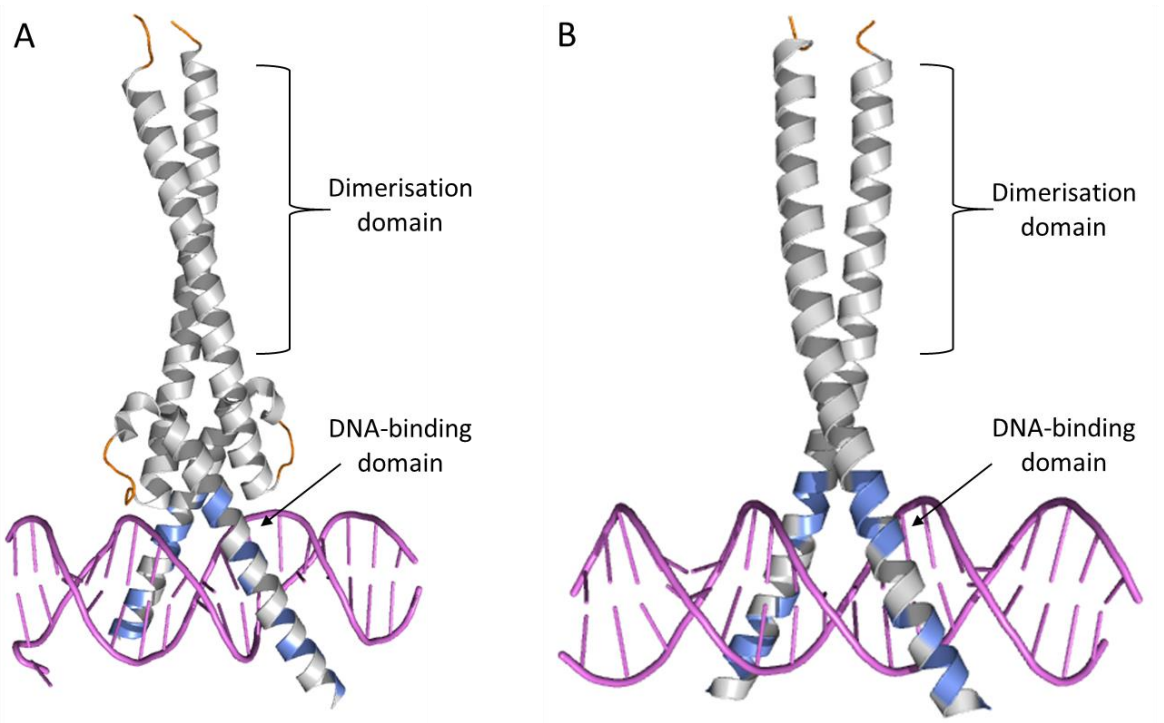


Figure 1.3: Basic-region helix-loop-helix (A) and basic-region leucine zipper (B) DNA-binding domains. (A) The Myc-Max bHLH heterodimer-DNA complex is shown. The C-terminal dimerisation motif mediates Myc-Max heterodimerisation, and the N-terminal basic domain allows for DNA binding at the major groove (PDB 1NKP) (Nair and Burley, 2003). **(B)** The Jun bZip homodimer-DNA complex is shown. The C-terminal leucine-rich region mediates Jun homodimerisation, and the N-terminal basic domain allows for DNA binding at the major groove (PDB 1JNM) (Kim and Podust, unpublished). α-Helices are grey, loops are orange, basic amino acid residues of the basic domains are blue, and DNA is pink. Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

1.1.2 ZINC-COORDINATING DNA-BINDING DOMAINS

The zinc-coordinating DBD superclass of TFs is the largest of the ten superclasses (Wingender *et al.,* 2018). The most distinguished TFs of this superclass are the nuclear receptor TFs, which are common drug targets for a range of diseases (Amoutzias *et al.*, 2008; Wingender *et al.*, 2018). Nuclear receptor TFs contain an N-terminal zinc finger domain for DNA binding (Figure 1.4A). The prototypical zinc finger is the Cys2His2 motif, consisting of approximately 30 amino acid residues with strategically positioned pairs of cysteine and histidine residues coordinating a zinc cation (Wingender *et al.,* 2018; Wolfe *et al.,* 2000). Coordination of the zinc cation positions the cysteine and histidine pairs, thereby stabilising the fold of the zinc finger (Wolfe *et al.,* 2000). A typical zinc finger folds to form a ββα structure and it is the α-helical portion of the zinc finger that inserts into the major groove of DNA (Wolfe *et al.,* 2000). TFs that bind DNA via zinc fingers often contain clusters of zinc fingers, allowing for simultaneous binding of multiple zinc fingers at successive DNA major grooves (Nielsen *et al.,* 2004; Wolfe *et al.,* 2000). Some zinc fingers are variations of the prototypical Cys2His2 motif – for example the Nizp1 TF contains, in addition to four Cys2His2 zinc fingers, a Cys2HisArg zinc finger that is capable

of mediating DNA binding and hence transcriptional repression (Berardi *et al.,* 2016; Nielsen *et al.,* 2004). Zinc fingers can mediate protein-DNA, protein-RNA and protein-protein interactions (Wolfe *et al.,* 2000). In addition to the N-terminal zinc finger domain, nuclear receptor TFs contain a C-terminal ligand-binding domain (Figure 1.4B) (Amoutzias *et al.,* 2008). Apart from its ligand-binding properties, the ligand-binding domain exhibits dimerisation properties, with the dimerisation interface typically composed of α-helices and loops (Pawlak *et al.*, 2012). Thus, dimerisation of nuclear receptor TFs occurs via both the zinc finger and ligand-binding domains (Amoutzias *et al.,* 2008).

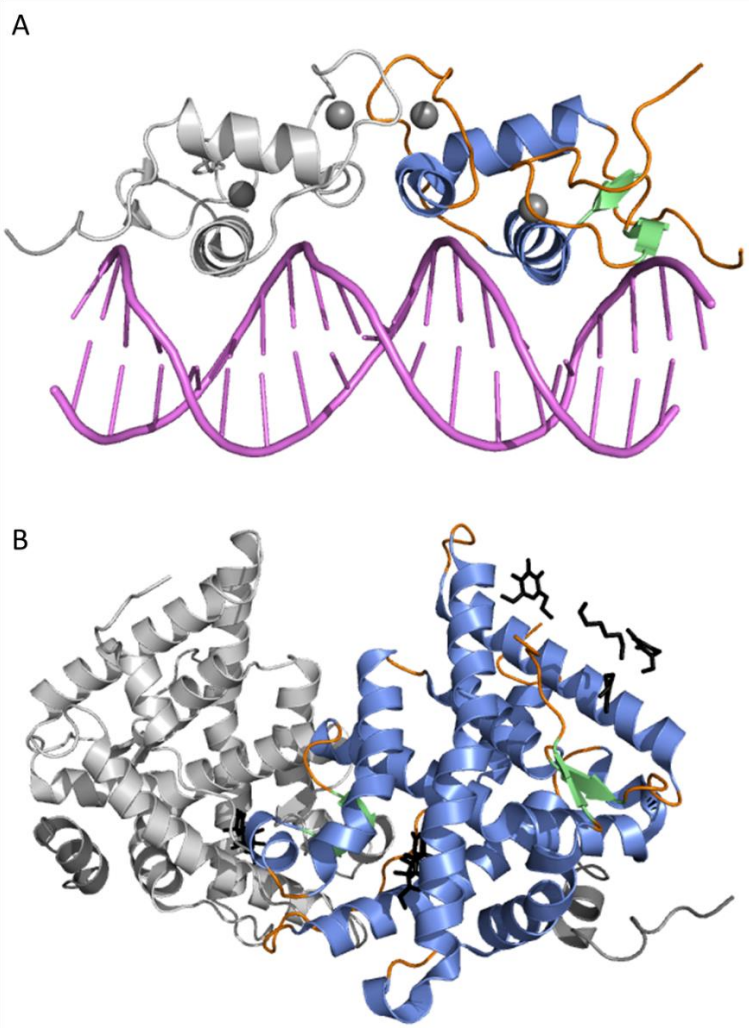


Figure 1.4: DNA-binding domain (A) and ligand-binding domain (B) of representative nuclear receptor transcription factor. **(A)** The N-terminal glucocorticoid receptor DNA-binding domain is shown. This zinc finger is larger than the typical ββα zinc finger and folds to form additional α-helices, enabling the coordination of two zinc cations by each monomer of the homodimer. Each zinc cation is coordinated by a Cys4 motif, a variation of the prototypical Cys2His2 zinc finger. DNA binding of each monomer occurs by insertion of an α-helix into the major groove. **(B)** The C-terminal glucocorticoid receptor ligand-binding domain is shown. This domain contributes to dimerisation. One monomer of each domain is shown in grey (left-hand side). α-Helices are blue, loops are orange, DNA is pink, zinc ions (**A**) and cofactors (**B**) are dark grey spheres and α-helices, respectively, and ligands (**B**) are shown in stick representation (black). The image was rendered on The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC (PDB 1R4O (**A**), 1M2Z (**B**)) (Bledsoe *et al.,* 2002; Luisi *et al.,* 1991).

1.1.3 HELIX-TURN-HELIX DNA-BINDING DOMAINS

HTH TFs comprise the second largest of the ten TF superclasses (Wingender *et al.,* 2018). The HTH DBD is considered one of the most evolutionarily conserved transcription elements (Aravind and Koonin, 1999). It is typically composed of three α-helices in an approximately triangular arrangement when associated with DNA (Figure 1.5A). The HTH DBD forms DNA-binding contacts via the third helix (helix 3) at the major groove (Figure 1.5B) (Aravind and Koonin, 1999). Helix 3 is thus referred to as

the recognition helix. The second helix (helix 2) typically forms non-specific contacts with the DNA backbone, and any additional α-helices serve to stabilise the DNA-protein interaction (Wingender *et al.,* 2018). The rigid 'turn' of the HTH domain occurs between the second and third helices (Aravind *et al.,* 2005). Contrastingly, the loop occurring between the first and second helices is less rigid and modifications in this region lead to variations of the typical HTH DNA-binding motif (Aravind *et al.,* 2005). An example of a variation of the typical HTH DBD is the winged-HTH DBD, usually composed of three α-helices succeeded by two β-strands (the wing), though some winged-HTH DBDs contain up to four interspersed β-strands with the additional β-strands forming in the loop between helix 1 and helix 2 (Aravind *et al.,* 2005). The wings form additional DNA contacts at the minor groove (Aravind *et al.,* 2005). Monomeric HTH DBDs can associate to form dimers (Figure 1.5B) (Amoutzias *et al.*, 2008; Aravind *et al.*, 2005; Beamer and Pabo, 1992). HTH TFs often have a LeuZip located adjacent to the HTH DBD, on the C-terminal side, which contributes to the association of monomers (Amoutzias *et al.,* 2008).

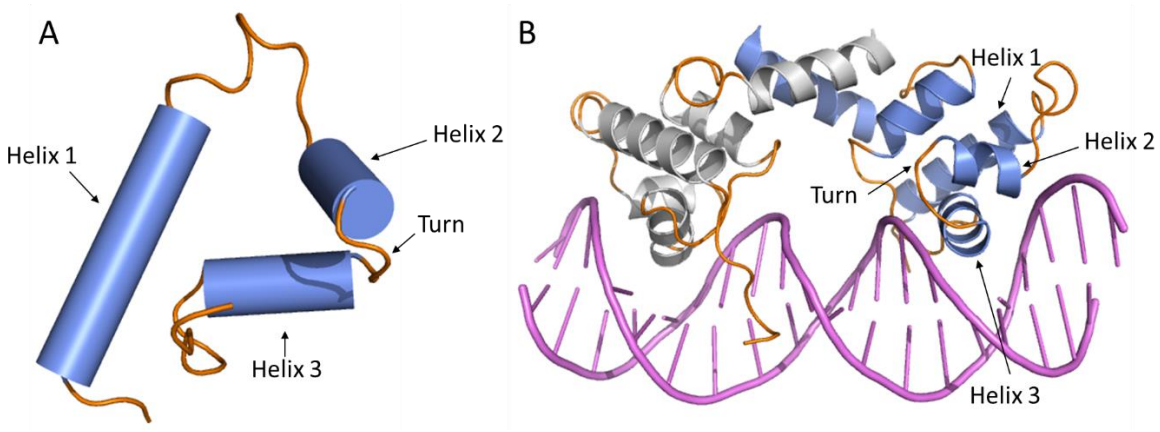


Figure 1.5: Helix-turn-helix DNA-binding domain. (A) The triangular arrangement of the three helices of the bacterial λ-repressor HTH DBD is shown. The rigid 'turn' occurs between helix 2 and helix 3, whereas the loop connecting helix 1 and helix 2 is more flexible. **(B)** The homodimeric HTH DBD of bacterial λ-repressor is shown interacting with DNA. Blue and grey are used to depict each monomer, respectively. Helix 3 (the recognition helix) of each monomer forms DNA-binding contacts at the major groove. The structures labelled for the monomer on the right-hand side (blue) are viewed posteriorly relative to the monomer on the left-hand side (grey). α-Helices are blue/grey, loops are orange, and DNA is pink. Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC (PDB 1LMB) (Beamer and Pabo, 1992).

1.2 Regulation of gene expression

The binding of TFs to *cis*-regulatory elements and resulting control of gene expression is regulated by intricate networks of related mechanisms. The mechanisms influencing gene expression are numerous and include time delays, negative and positive feedback loops, TF oligomerisation, and post-translational modifications of TFs, amongst others (Benayoun and Veitia, 2009; Smolen *et al.,*

2000). The regulation of TF activity is important for ensuring that, upon receiving a stimulus, the appropriate response is elicited by the cell and for ensuring that gene expression patterns are controlled both spatially and temporally in a context-dependent manner (Smolen *et al.,* 2000). Transcriptional dysregulation has been associated with a wide range of diseases such as cancer, cardiovascular diseases, neurological disorders, and autoimmune diseases (Lee and Young, 2014). Transcriptional dysregulation can arise due to mutations in *cis*-regulatory elements or in proteins that interact with these DNA regions, including TFs (Lee and Young, 2014).

1.2.1 POST-TRANSLATIONAL MODIFICATIONS

It is common for a single TF to control a diverse range of processes. For example, the forkhead box O3 TF has been implicated in cell cycle control, DNA repair, erythroid differentiation, the prevention of cardiac hypertrophy, female infertility, cell survival, tumour suppression and human longevity (Castrillon *et al.,* 2003; Flachsbart *et al.,* 2009; Greer and Brunet, 2005, 2008; Salih and Brunet, 2008; Willcox *et al.,* 2008). In order to trigger the appropriate response to a signal and ensure that the TF binds to the appropriate *cis*-regulatory element, a single TF is often subjected to a range of distinct post-translational modifications targeting different amino acid residues, which affect and control the TF DNA-binding specificity, or regulate its affinity for the same *cis*-regulatory element, or alter its protein-protein interactions (Benayoun and Veitia, 2009). The two most commonly occurring post-translational modifications, as observed experimentally, are phosphorylation (typically targeting serine, threonine, tyrosine and histidine residues), and acetylation (typically targeting lysine residues) (Khoury *et al.,* 2011). Lysine acetylation usually targets large protein complexes involved in nuclear processes such as transcription, with competing ubiquitination and sumoylation of TF lysine residues allowing for multi-layered regulation of gene expression (Choudhary *et al.,* 2009; Tootle and Rebay, 2005). Protein mutants are useful mimetics to study the effects of post-translational modifications *in vitro*. For example, the substitution of lysine with glutamine or aspartic acid has been used as a mimic to investigate the effects of lysine acetylation (Gorsky *et al.,* 2016; Song *et al.,* 2012; Wang and Hayes, 2008).

1.3 TRANSCRIPTION FACTOR OLIGOMERS

An astounding number of TF oligomers have been identified and studied, emphasising the importance of TF oligomerisation (Anckar and Sistonen, 2011; Bujalka *et al.*, 2013; Chen and Privalsky, 1995; Chen *et al.*, 2001; Grandori *et al.*, 2000; Hai and Curran, 1991; Hinde *et al.*, 2016; Kim *et al.*, 2017, 1996; Levy and Marié, 2012; Li *et al.*, 2013b; Lüscher, 2001; Rabindran *et al.*, 1993; Sinha *et al.*, 1995, 1996; Vitelli *et al.*, 2000; Zhang and Darnell, 2001; Zhao *et al.*, 2013). Interestingly, TFs from more than one

superclass are able to interact with DNA as oligomers, though oligomerisation usually occurs between members of the same TF class or family (Amoutzias *et al.*, 2008; Wingender *et al.*, 2018). Indeed, oligomerisation is a prerequisite for DNA binding for some TFs (Wingender *et al.,* 2018). The DBDs of these TFs are the most conserved regions, whereas the regions conferring oligomerisation abilities to these TFs are less conserved (Alvarez-Buylla *et al.,* 2000; Amoutzias *et al.,* 2006, 2004; Bailey *et al.,* 2003; Bertrand *et al.,* 2004; De Bodt *et al.,* 2003; Deppmann *et al.,* 2006; Gauthier and Degnan, 2008; Jakoby *et al.,* 2002; Nakamura *et al.,* 2006; Shore and Sharrocks, 1995; Simionato *et al.,* 2007; Wang and Levy, 2006). TF oligomers are composed of different numbers of monomers and are varied with regards to the role of oligomerisation for TF function, and the role of the oligomeric TF in the cell. The effects of TF oligomerisation include the control of nuclear trafficking, altered DNA-binding specificities, inhibitory effects, and regulatory effects (Chen and Privalsky, 1995; Chen *et al.*, 2001; Fisher *et al.*, 1991; Hai and Curran, 1991; Hinde *et al.*, 2016; Vitelli *et al.*, 2000).

TF homo- and hetero-oligomer formation is driven by a number of factors. As with other proteins, the oligomerisation of TFs can be thermodynamically driven, with the resultant oligomer being more stable and energetically favoured than the constituent monomers (Barducci *et al.,* 2013; Beckett, 2001). Under *in vivo* conditions, macromolecular crowding drives oligomerisation due to the resulting oligomer occupying a smaller volume, a phenomenon that has been seen to drive the formation of both homo- and hetero-oligomers (Bookchin *et al.,* 1999; Lindner and Ralston, 1997; Rivas *et al.,* 2001; Snoussi and Halle, 2005; Wilf and Minton, 1981; Zimmerman and Trach, 1988). The accessible surfaces differ between an oligomer and its constituent monomers, which could enable the regulation of gene expression by selective protein-protein and protein-DNA interactions of each oligomeric state, therefore oligomer formation could be driven by the gene expression requirements of the cell (Vinkemeier *et al.,* 1998; Wingelhofer *et al.,* 2018).

In order for oligomerisation to occur, it is necessary that the interacting interfaces have the appropriate properties to favour oligomer formation. Analyses of protein-protein interfaces have shown that although diversity exists amongst these interfaces, some overarching properties of the interacting surfaces may influence oligomer formation (Jones and Thornton, 1996; Talavera *et al.,* 2011). Protein-protein interfaces are typically segmented with respect to secondary structural elements, rather than consisting of a single secondary structural element, and the types of secondary structures observed at interfaces are mixed, consisting of α-helices, β-strands, and loops (Jones and Thornton, 1996; Talavera *et al.,* 2011). Interestingly, homo-oligomer interfaces tend to be more intertwined whereas hetero-oligomer interfaces are more planar (Jones and Thornton, 1996). Current

evidence regarding the preferential amino acid composition of protein-protein interfaces is inconclusive. Some evidence suggests that all protein-protein interfaces are characterised by a high frequency of hydrophobic residues, aromatic residues and arginine residues (Bahadur *et al.,* 2003; Chakrabarti and Janin, 2002; Jones and Thornton, 1997; Zhou and Qin, 2007). Other evidence indicates that homo-oligomerisation is preferentially mediated via hydrophobic interfaces which have a high frequency of hydrophobic amino acid residues, whereas hetero-oligomerisation is preferentially mediated via less hydrophobic interfaces which form more hydrogen bonds (Jones and Thornton, 1996). An extension of this model states that despite the preference of hydrophobic interfaces for homo-oligomeric interactions, electrostatic and polar interactions are also important for both homo- and hetero-oligomer formation and the hydrophobic regions of homo-oligomer interfaces are usually segmented and occur as patches (Larsen *et al.,* 1998; Sheinerman *et al.,* 2000). Nevertheless, regardless of the specific amino acid composition of the interfaces, the feasibility and favourability of protein-protein interactions, whether homo- or heterotypic, rely on shape and electrostatic complementarity of the interacting partners (Jones and Thornton, 1996; Li *et al.*, 2013a; Zhang *et al.*, 2011b). The key amino acid residues at a particular protein-protein interface, or the contribution of each interface to proteins that interact via more than one interface, can be determined on a case-by-case basis, for example by using point mutations to identify the contributions of specific amino acid residues or specific interfaces to the overall stability of the oligomer.

1.4 LEUCINE ZIPPERS IN TRANSCRIPTION FACTORS

The leucine zipper (LeuZip) domain, first identified in 1988 by Landschulz and coworkers, is a common structural feature of TFs and mediates TF oligomerisation (Ferre-D' Amare *et al.,* 1994; Fisher *et al.,* 1991; Landschulz *et al.,* 1988; Roman *et al.,* 1992; Song *et al.,* 2012). The LeuZip motif forms a stable α-helix of approximately 30 amino acid residues (Ellenberger, 1994; Landschulz *et al.,* 1988). During the formation of a LeuZip dimer, the interdigitation of leucine residues of each monomer facilitates dimer formation (Figure 1.6) (Ellenberger, 1994; Landschulz *et al.,* 1988). As seen in the mouse Foxp3 (mFoxp3) LeuZip (Figure 1.6), the resulting dimer typically forms a coiled-coil (Landschulz *et al.,* 1988; Song *et al.,* 2012). The characteristic heptad leucine repeat is what distinguishes a LeuZip from other coiled-coil structures (Landschulz *et al.*, 1988). In some instances, strategically positioned hydrophobic or charged residues at the core of the interface enable the formation of additional hydrophobic interactions or hydrogen bonds which further stabilise LeuZip associations (Baxevanis and Vinson, 1993). LeuZips only facilitate oligomerisation and are unable to bind DNA directly. Thus, interactions of TFs with DNA require a DBD that is distinct from the LeuZip. LeuZip motifs are not exclusive to TFs;

for example, a LeuZip motif facilitates homodimerisation of the proline dipeptidase enzyme, and loss of proline dipeptidase LeuZip dimerisation results in a loss of enzymatic activity (Chiravuri *et al.,* 2000).

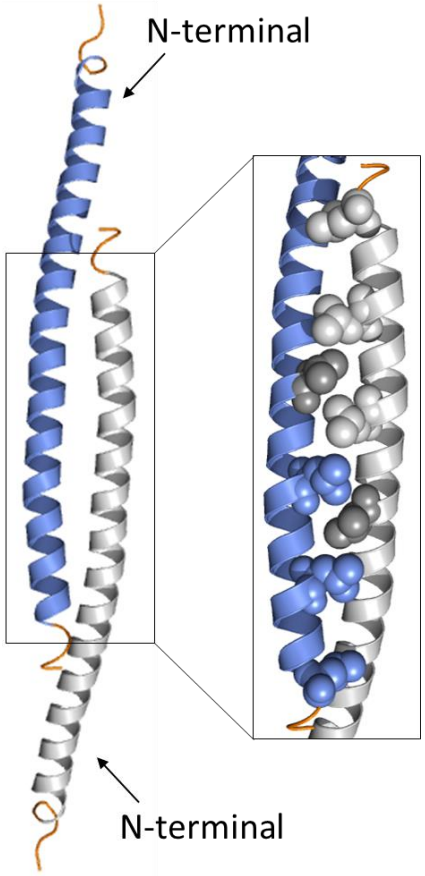


Figure 1.6: Leucine zipper dimerisation. The mFoxp3 LeuZip homodimer is shown with blue and grey used to depict each monomer, respectively. The LeuZip (enlarged on the right-hand side) allows for homotypic associations through interdigitation of leucine residues (spheres) resulting in an anti-parallel coiled-coil. The coiled-coil extends beyond the LeuZip at the N-terminal ends, as indicated. α-Helices are blue/grey, loops are orange, and the conservative valine substitution at the first position is depicted as dark grey spheres. Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC (PDB 4I1L) (Song *et al.,* 2012).

A LeuZip typically spans the region of four or five heptad leucine repeats (Figure 1.7A), with a leucine residue therefore occurring every second turn of the α-helix, accommodating interdigitation of leucine residues from two LeuZips (Landschulz *et al.,* 1988). Amino acid residues at positions *a*, *d*, *e* and *g* (using standard nomenclature for the heptad repeat of a coiled-coil) are important for both stability and specificity of the LeuZip interaction since they occur at the coiled-coil interface (Figure 1.7B) (Ellenberger, 1994; Landschulz *et al.,* 1988; Song *et al.,* 2012). In general, hydrophobic amino acid residues are favoured at positions *a* and *d*, whereas polar and charged amino acid residues are favoured at other positions of the heptad repeat (*b*, *c*, *e*, *f*, and *g*), but variations to this pattern are tolerated (Ellenberger, 1994; Song *et al.,* 2012). The hydrophobic residues (*a* and *d*) form a

hydrophobic core, and hydrogen bonds between residues located at the *e* and *g* positions are frequently observed to further stabilise the interaction (Baxevanis and Vinson, 1993). LeuZips mediate both homo- and hetero-dimerisation, and LeuZip dimers are capable of assuming either parallel or anti-parallel conformations (Deng *et al.,* 2008; Ellenberger, 1994; Landschulz *et al.,* 1988; Li *et al.,* 2004; O'Shea *et al.,* 1991; Song *et al.,* 2012).

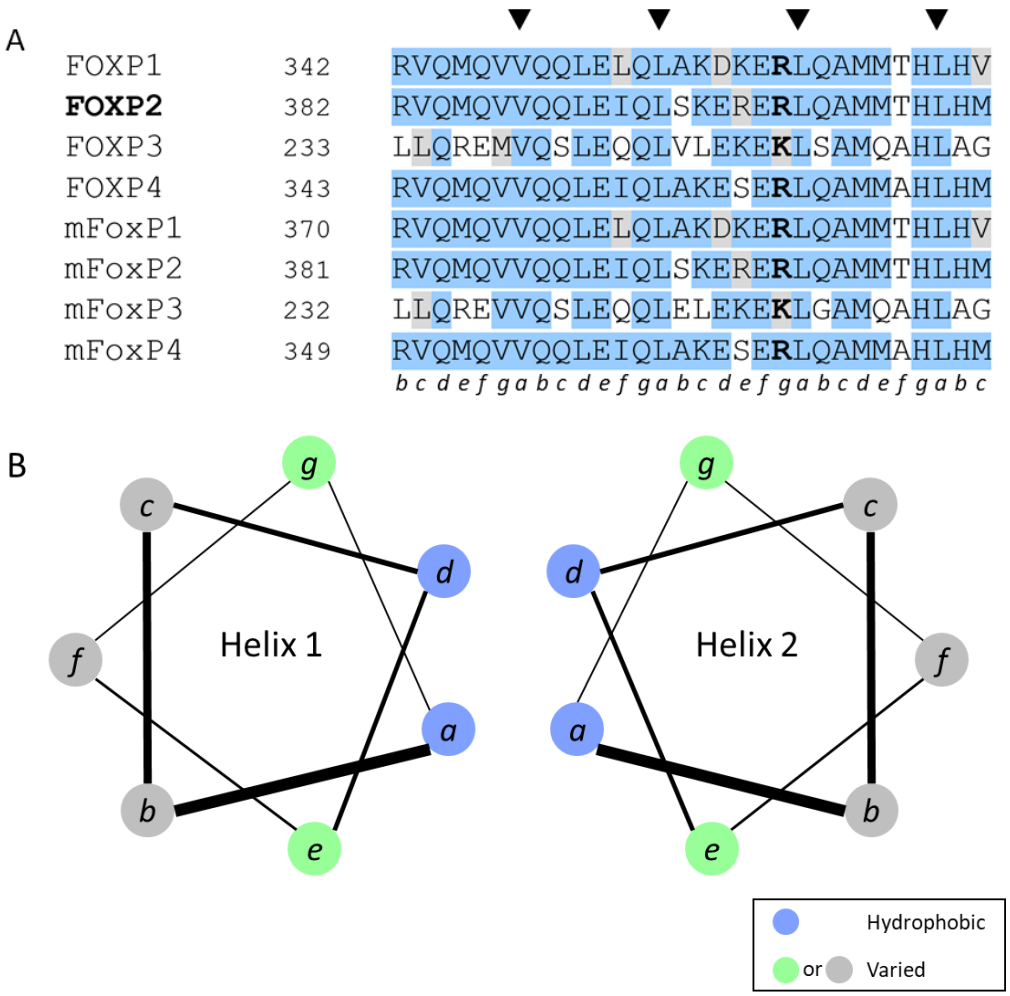


Figure 1.7: (A) Multiple sequence alignment of leucine zippers of human and mouse (m) FOXP subfamily members, and (B) helical wheel diagram of residues forming the heptad repeat of an anti-parallel leucine zipper dimer. (A) An example of a LeuZip found in all members of the FOXP subfamily of TFs is shown. The region is highly conserved between humans and mice, especially amongst FOXP1/2/4. Blue shading denotes identical amino acids, grey shading denotes similar amino acids and no shading denotes different amino acids. Pairwise alignments of the sequences show a mean similarity of 82% and identity of 72%, and the global alignment shows a mean similarity of 71%. Triangles (above) indicate heptad repeats of leucine residues (with a conservative valine substitution at the first position). Letters *a* - *g* (below) are standard nomenclature for the heptad repeat of a coiled-coil. Residues at positions *a* and *d* are predominantly hydrophobic. Residues at positions *e* and *g* are varied. mFoxp3 Lys251 is indicated in boldface and corresponds with FOXP2 Arg401. Amino acid positions are numbered relative to the full-length proteins (left-hand side). The multiple sequence alignment was generated using the online T-Coffee server (Notredame *et al.,* 2000). **(B)** The arrangement of residues are shown for a dimer, the simplest of LeuZip oligomers, adopting an anti-parallel coiled-coil configuration. Positions *a* and *d* (blue) are typically occupied by hydrophobic residues which form a hydrophobic core at the interface. Residues at all other positions are usually polar or charged. Positions *e* and *g* (green) form part of the interface, and hydrogen bonds between residues at these positions contribute to oligomer stability. Sometimes, hydrophobic residues occur at positions *e* and/or *g*, which increases the buried surface area of the hydrophobic core, and contributes to the formation of higher order oligomers (composed of three or more monomers).

In addition to the formation of dimers, LeuZip domains are able to form higher order oligomers (Aoki *et al.,* 1999; Friedman *et al.,* 1995; Kasai *et al.,* 1997; Song *et al.,* 2012). LeuZip tetramers are common in TFs and often occur due to the dimerisation of dimers (Chakerian *et al.,* 1991; Friedman *et al.,* 1995; Song *et al.,* 2012). Although positions *e* and *g* are usually occupied by polar or charged residues, the occurrence of hydrophobic residues at these positions can increase the buried surface area and contribute to higher order LeuZip oligomerisation. Thus, LeuZip tetramers often form on the basis of hydrophobic interactions of side chains at the *a* and *d* positions, and a third hydrophobic residue at either the *e* or *g* position (Deng *et al.,* 2006; Hill *et al.,* 2000; Liu *et al.,* 2007; Solan *et al.,* 2002; Yadav *et al.,* 2006). In the case of the former, tetramer formation is initiated by the linkage of the α-helices in an anti-parallel configuration by hydrophobic side chains at position *a*, resulting in subsequent interactions of residues at positions *d* and *e* which stabilise the interface (Liu *et al.,* 2007). In the case of the latter, tetramer formation is similar but it is the residues at position *d* that initiate tetramer formation, resulting in the subsequent interdigitation of residues at positions *a* and *g* (Deng *et al.,* 2006). Considering the helical arrangement of the side chains (Figure 1.7B), it is thus the central residue that initiates linkage of anti-parallel LeuZips during oligomer formation, and it is clear that residues at positions *a*, *d*, *e* and *g* determine the interactions resulting in oligomerisation and influence the final quaternary structure. LeuZip trimers, hexamers, and octamers are less common but have been reported for a few TFs (Aoki *et al.,* 1999; Harbury *et al.,* 1993; Kasai *et al.,* 1997).

The LeuZip is an important structural feature of many TFs, as evinced by the many examples of TF LeuZip-mediated associations influencing interactions with DNA and transcriptional activity (Anckar and Sistonen, 2011; Aoki *et al.*, 1999; Bujalka *et al.*, 2013; Chakerian *et al.*, 1991; Fisher *et al.*, 1991; Grandori *et al.*, 2000; Hai and Curran, 1991; Hentze *et al.*, 2016; Kim *et al.*, 2017; Li *et al.*, 2013b; Lüscher, 2001; Rabindran *et al.*, 1993). The significance and diversity of TF LeuZip associations will be expounded through the use of some illustrative examples of dimers, trimers, and higher order oligomers.

1.4.1 LEUCINE ZIPPER TRANSCRIPTION FACTOR DIMERS

Dimers are the simplest of oligomers, composed of just two individual subunits or monomers. Dimerisation plays important roles in both gene activation and repression (Amoutzias *et al.,* 2008). Each dimer partner and each cellular context leads to a specified cell fate (Amoutzias *et al.,* 2008). Thus, one TF monomer that is able to bind multiple partners facilitates gene regulation due to the different properties and roles of each dimer (Amoutzias *et al.,* 2008). An example of LeuZip TF heterodimerisation that occurs in the control of gene expression is the Max protein that can

heterodimerise with either Myc or Mad, whereas Myc and Mad are unable to heterodimerise with each other (Grandori *et al.,* 2000; Lüscher, 2001). Myc-Max and Mad-Max dimerise via HLH-LeuZips and bind DNA via basic domains (Nair and Burley, 2003). Figure 1.3A shows the Myc-Max heterodimer interacting with DNA. Heterodimerisation of Max with Myc leads to the recruitment of histone acetyl transferases required for gene expression, whereas heterodimerisation of Max with Mad leads to gene silencing due to the recruitment of histone deacetylases (Grandori *et al.,* 2000; Lüscher, 2001).

Alternatively, differential homo- and heterodimerisation allows for the regulation of gene expression since TF heterodimers often demonstrate properties and binding specificities that are distinct from the homodimers of their constituents (Amoutzias *et al.,* 2008). An example is the Jun-activating transcription factor 2 (ATF2) bZip heterodimer, which differs from both the Jun and the ATF2 homodimers with respect to binding specificities (Hai and Curran, 1991). The Jun homodimer binds well to the activator protein-1 site, which mediates responses to environmental stimuli such as mitogens or viral infections (Hai and Curran, 1991). Conversely, the Jun-ATF2 heterodimer shows a weak affinity for the activator protein-1 site, but binds more strongly to the proenkephalin regulatory region than either homodimer (Hai and Curran, 1991).

1.4.2 LEUCINE ZIPPER TRANSCRIPTION FACTOR TRIMERS

As with dimerisation, TF trimer formation has been shown to regulate gene expression. An example is the myelin-gene regulatory factor (MyRF), a TF essential for the maturation of oligodendrocytes, the maintenance of mature oligodendrocytes, and myelination of the central nervous system (Emery *et al.,* 2009; Koenning *et al.,* 2012). Full-length MyRF forms membrane-bound homo-trimers that undergo auto-proteolysis, releasing stable homo-trimeric N-terminal domains from the endoplasmic reticulum (Bujalka *et al.,* 2013; Kim *et al.,* 2017; Li *et al.,* 2013b). The homo-trimeric N-terminal domains of MyRF function as TFs, and the formation of these homo-trimers is a prerequisite for transcriptional activity (Bujalka *et al.,* 2013; Kim *et al.,* 2017; Li *et al.,* 2013b). MyRF contains a LeuZip adjacent to the DBD, and with the use of point mutations it was shown that the LeuZip is required for both trimer formation and auto-proteolysis (Li *et al.*, 2013b). Furthermore, MyRF homo-trimers bind DNA more tightly than monomers, and exhibit sequence specificity that is not seen in the monomers which bind DNA more promiscuously (Kim *et al.,* 2017).

In some examples, TFs associate via LeuZips upon receiving a stimulus in order to regulate gene expression, thereby eliciting a response to the stimulus. This is seen in the heat shock transcription factor 1 (HSF1) (Anckar and Sistonen, 2011; Rabindran *et al.,* 1993). The heat shock response allows

organisms to survive harsh environmental conditions, not limited to elevated temperatures but also including conditions such as oxidative stress and inflammation (Jolly and Morimoto, 2000). HSF1 is a key regulator of the eukaryotic heat shock response, and exists as a monomer under normal conditions but associates to form a LeuZip-mediated homo-trimer in response to cell stress (Anckar and Sistonen, 2011; Rabindran *et al.,* 1993). Using elevated temperatures as a stimulus, it was observed that the trimeric LeuZip region of HSF1 becomes more tightly packed with increasing levels of stress, emphasising the importance of trimer formation for HSF1 function (Hentze *et al.*, 2016; Rabindran *et al.*, 1993).

1.4.3 LEUCINE ZIPPER TRANSCRIPTION FACTOR TETRAMERS AND HIGHER ORDER OLIGOMERS

TF LeuZips are capable of forming a 'dimer of dimers' with the association of two dimers giving rise to a tetrameric arrangement. The resulting tetramer often has important functional implications. For example, the lactose repressor TF LeuZip facilitates the dimerisation of dimers (Figure 1.8), and disruption of this interface renders the TF unable to bind DNA (Chakerian *et al.,* 1991). In another example, TFEB (transcription factor EB) requires an intact LeuZip for homo- and heterotypic associations (Fisher *et al.,* 1991). TFEB is only able to bind DNA as a homo- or heterodimer, but is observed to associate with another dimer to form a tetramer in its unbound form (Fisher *et al.,* 1991). The transition between dimer-dimer and dimer-DNA complexes could potentially provide a regulatory point for gene expression (Fisher *et al.,* 1991).

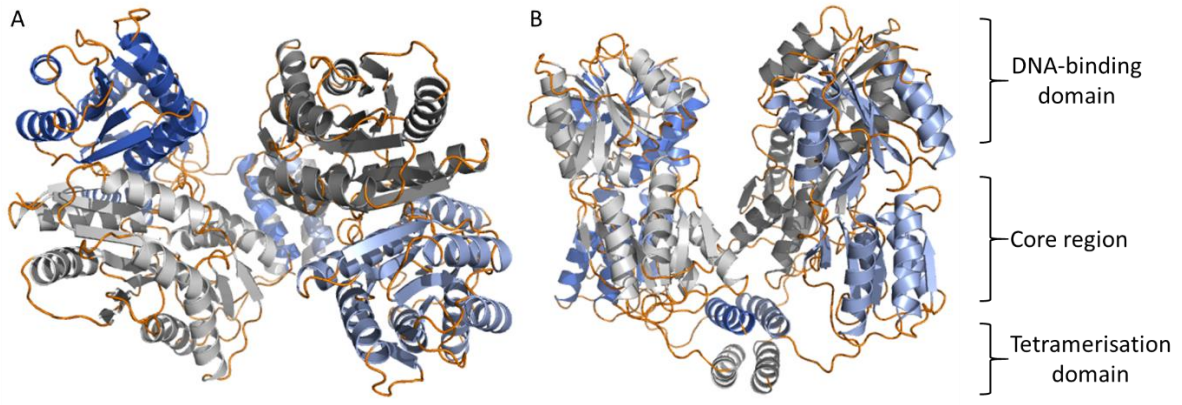


Figure 1.8: Lactose repressor forms a leucine zipper-mediated dimer of dimers. Top view **(A)** and side view **(B)** of lactose repressor tetramer formed by the association of dimers. One monomer of each dimer is depicted in grey. Monomers associate via short C-terminal LeuZips to form an anti-parallel dimer, and dimers then associate to form an anti-parallel tetramer, resulting in a four-helix bundle (tetramerisation domain). DNA binding occurs at the N-terminal HTH DBD (N-terminal DBD is viewed proximally in **A**). The core region stabilises the tetramer and regulates DNA binding. α-Helices/β-strands are blue/grey, and loops are orange. Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC (PDB 1TLF) (Friedman *et al.,* 1995).

LeuZips can facilitate the formation of TF oligomers larger than tetramers, as observed in the translin TF (Aoki *et al.,* 1999; Kasai *et al.,* 1997). Translin binds both DNA and RNA and has been associated with chromosomal translocations, mRNA storage and translocation, telomere maintenance, DNA repair, and other important processes involving nucleic acids (Aoki *et al.,* 1995; Finkenstadt *et al.,* 2002; Han *et al.,* 1995; Jacob *et al.,* 2004). Translin monomers associate via LeuZips to form an octameric ring (Figure 1.9) (Aoki *et al.,* 1999; Kasai *et al.,* 1997). The formation of the octamer, as is frequently observed with the formation of LeuZip tetramers, is suggested to result from the association of dimers (Aoki *et al.,* 1999; Kasai *et al.,* 1997). The open octameric ring forms an internal cavity that serves as an inlet for DNA, enabling the formation of DNA-binding contacts (Eliahoo *et al.,* 2015). Point mutations of the translin DBD result in a loss of DNA-binding activity, without a loss of quaternary structure (Aoki *et al.,* 1999). However, point mutations of the translin LeuZip result in a loss of both quaternary structure and DNA binding, emphasising the importance of LeuZip-mediated oligomerisation for DNA binding and hence function (Aoki *et al.,* 1999).

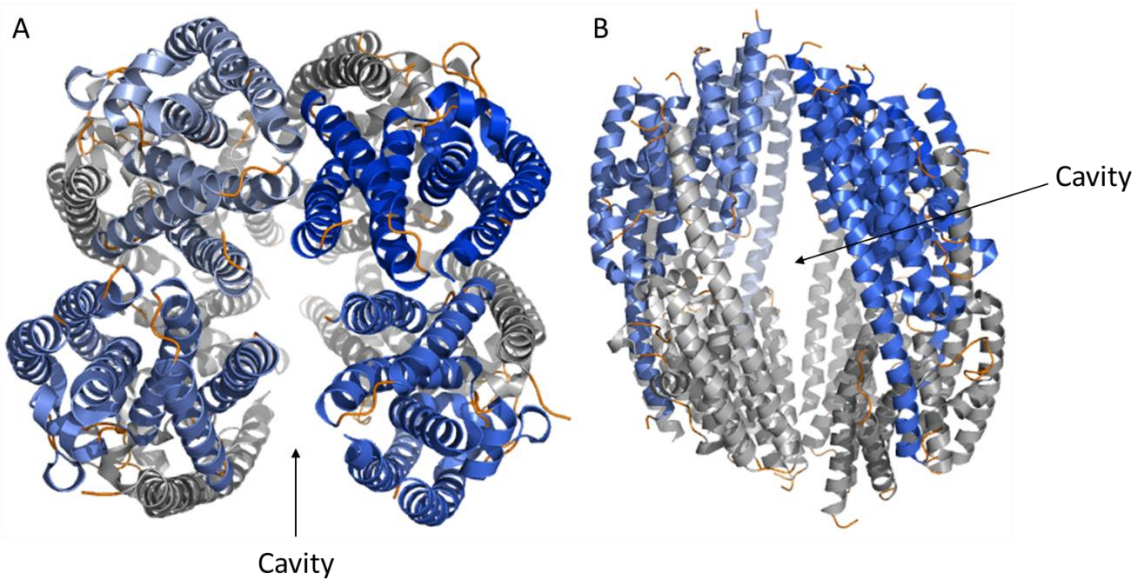


Figure 1.9: Translin forms a leucine zipper-mediated octamer. Top view **(A)** and side view **(B)** of translin octameric ring formed by the association of monomers via C-terminal LeuZips. The functional octameric ring is composed of four anti-parallel dimers (one monomer of each dimer is depicted in grey). Disruption of the LeuZip interface results in a loss of both quaternary structure and function. When closed, the ring shows no apparent entry point for nucleic acids (not shown). However, the open ring conformation forms a cavity that could provide an entry point for nucleic acids (C-terminal LeuZips are viewed proximally in **A** and at the top of **B** for subunits depicted in blue). α-Helices are blue/grey and loops are orange. Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC (PDB 4WYV) (Eliahoo *et al.,* 2015).

1.5 FOX TRANSCRIPTION FACTORS

The forkhead box (FOX) proteins are a family of HTH TFs identified by the distinctive DNA-binding forkhead domain (FHD), first discovered in the *Drosophila melanogaster* embryo (Weigel *et al.,* 1989). Genes encoding FOX TFs have been identified in many species ranging from yeast to rodents (Kaufmann and Knöchel, 1996; Lai *et al.*, 1993). In humans, FOX TFs are ubiquitously expressed from embryonic development to adulthood, and play important roles in diverse processes such as development of the heart, liver, pancreas, and lungs, as well as glucose metabolism, regulation of the immune system, chromatin remodelling, and lowered metabolism associated with aging (Friedman and Kaestner, 2006; Hannenhalli and Kaestner, 2009; Jonsson and Peng, 2005; Ma *et al.,* 2014; Zhu, 2016).

The fifty identified human FOX-encoding genes are classified into nineteen subfamilies (FOXA to FOXS) based on sequence similarity of the FHD (Figure 1.10) (Kaestner *et al.,* 2000). All other regions of FOX proteins, beyond the DNA-binding FHD, exhibit great variation (Kaestner *et al.,* 2000). In addition to having multiple protein domains, FOX TFs interact with a large number of protein binding partners and cofactors, adding to the diversity of genes regulated by FOX TFs, as well as to the diversity of the downstream effects of FOX transcriptional activity (Li *et al.,* 2015).

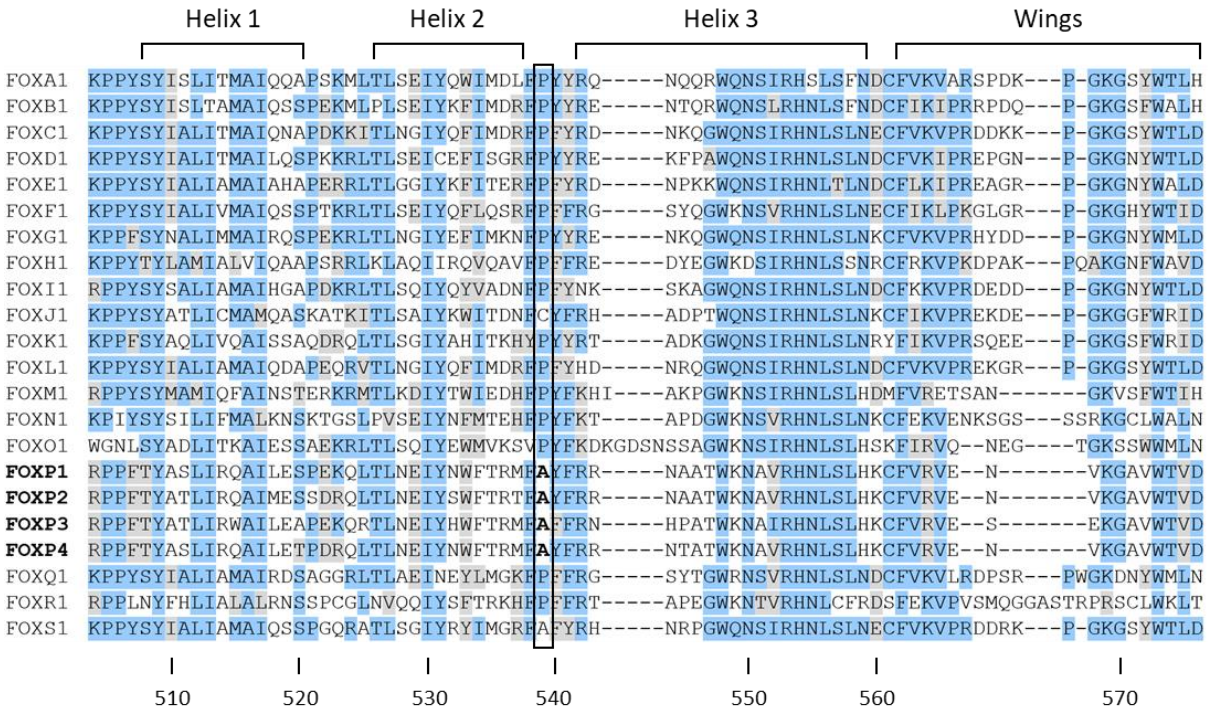


Figure 1.10: Multiple sequence alignment of representative FOX DNA-binding forkhead domains. The FHD sequence of at least one member of each FOX subfamily (A - S) is shown. Blue shading denotes identical amino acids, grey shading denotes similar amino acids, no shading denotes different amino acids, and dashes denote gaps in the alignment. Pairwise alignments of the sequences show a mean similarity of 65% and identity of 51%, and the global alignment shows a mean similarity of 47%. The highly conserved proline residue of the hinge loop, which is replaced by alanine in the FOXP subfamily (boldface), is boxed. Approximate locations of conserved secondary structural elements are indicated above the alignment. Amino acid positions are numbered relative to full-length FOXP2 below the alignment. The multiple sequence alignment was generated using the online T-Coffee server (Notredame *et al.,* 2000).

The FOX FHD is a winged-HTH domain – a derivative of the typical HTH DBD (Clark *et al.,* 1993). The generalised FOX FHD is a monomer of approximately 70 - 100 amino acid residues, folded to form three N-terminal α-helices and two C-terminal loops or wings, with wing 1 often forming two anti-parallel β-strands (Figure 1.11) (Clark *et al.,* 1993). Variations of this prototype arise from differences in the loops between the α-helices, and occasionally in the wing region, with the formation of additional secondary structures in these regions (Aravind *et al.,* 2005; Clark *et al.,* 1993). For example, the FHDs of FOXC2, FOXD3, and FOXO1 form an additional short α-helix between helix 2 and helix 3, the FHDs of FOXD3, FOXK1, and FOXP2 form an α-helix at the C-terminal end, in the wing region, and the loop between helix 2 and helix 3 is extended in the FHDs of the FOXO proteins, compared with other FOX proteins (Brent *et al.*, 2008; Van Dongen *et al.*, 2000; Jin and Liao, 1999; Marsden *et al.*, 1998; Obsil and Obsilova, 2011; Psenakova *et al.*, 2019; Stroud *et al.*, 2006; Tsai *et al.*, 2006). The FOX FHD binds to DNA by the insertion of helix 3, the recognition helix, into the major

groove (Clark *et al.*, 1993; Tsai *et al.*, 2007). Wing 1 and wing 2 form additional major groove, minor groove and DNA backbone contacts, further stabilising the FOX-DNA complex (Clark *et al.*, 1993).

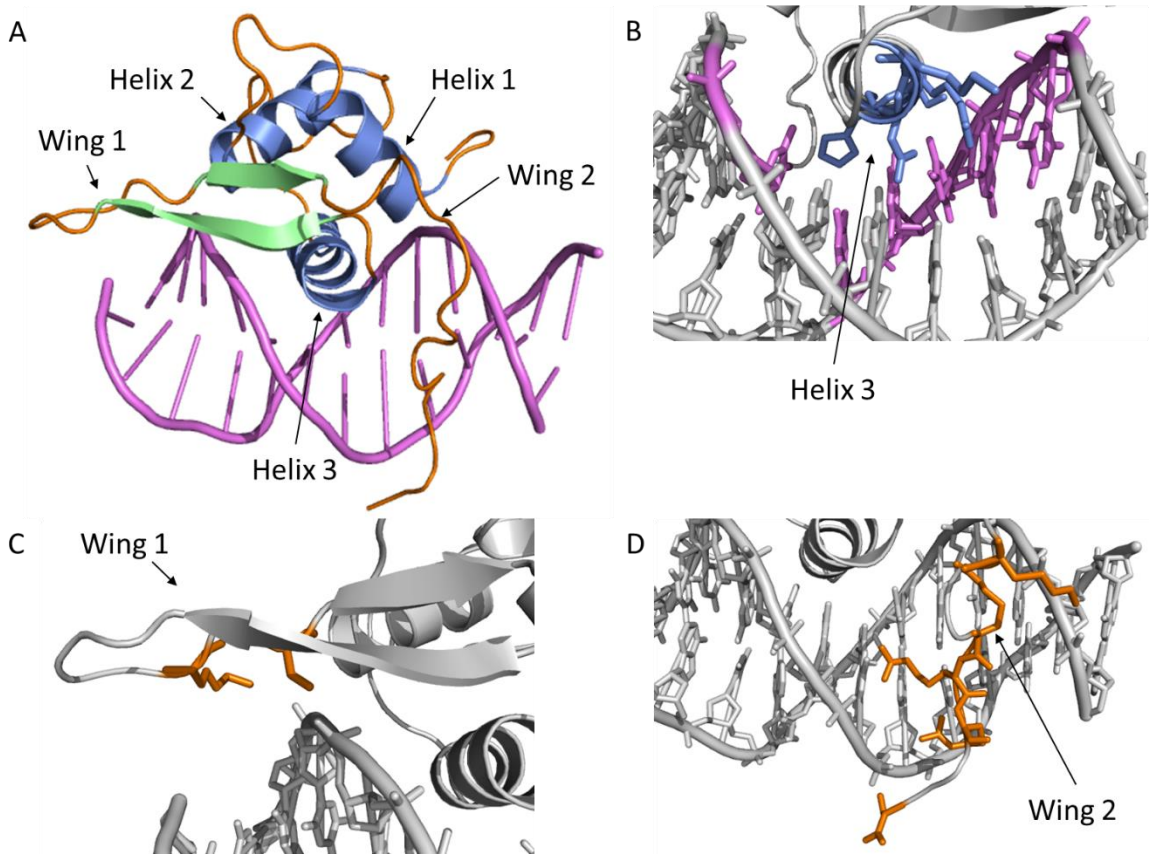


Figure 1.11: FOXO3 forkhead domain-DNA complex. (A) The FOXO3 FHD presents helix 3, the recognition helix, to the DNA major groove. α-Helices are blue, β-strands are green, loops and wings are orange, and DNA is pink. **(B)** Helix 3 forms base-specific DNA-binding contacts at the major groove. Residues and bases involved in the interaction are shown as blue and pink sticks, respectively (viewed posteriorly relative to **A**). In FOXO3, wing 1 **(C)** and wing 2 **(D)** form additional DNA-binding contacts through interactions with the DNA backbone. Residues involved in the interactions are shown as orange sticks. Images were rendered using PyMOL v 0.99 Schrödinger, LLC (PDB 2UZK) (Tsai *et al.,* 2007).

Studies of the DNA-binding specificities of both human and non-human FOX TFS have shown that most FOX TFs exhibit specificity for and bind to three DNA binding sites, or variants thereof (Badis *et al.*, 2009; Kaufmann *et al.*, 1995; Luo *et al.*, 2012; Overdier *et al.*, 1994; Pierrou *et al.*, 1994; Schlake *et al.*, 1997; Zhu *et al.*, 2009, 2012). Based on these DNA binding sites, an analysis of the evolution of FOX DNA-binding specificities was conducted and it was revealed that the DNA-binding specificities evolved independently for each FOX subfamily (Nakagawa *et al.,* 2013). Interestingly, it was observed that several FOX TFs are able to recognise and bind to two or three of the DNA binding sites, and this lead to the theory of multiple FOX DNA-binding "modes" (Nakagawa *et al.*, 2013). According to this theory, a single FHD is able to adopt multiple DNA-binding conformations, allowing for multiple

DNA-binding specificities (Nakagawa *et al.,* 2013). The evolution of FOX DNA-binding specificities has been attributed to this proposed ability of the FOX FHD to adopt multiple DNA-binding conformations (Nakagawa *et al.*, 2013). Importantly, this theory also postulates that by adopting multiple DNA-binding conformations, FOX TFs are able to allosterically expose otherwise inaccessible surfaces and domains, thereby enabling differential protein-protein interactions through the controlled exposure of protein-protein interfaces (Nakagawa *et al.,* 2013). This would allow for an additional regulatory point of FOX TF activity through the DNA binding site-dependent regulation of FOX protein binding partners (Nakagawa *et al.,* 2013).

The detrimental effects of the dysregulation of FOX TFs is evident, since FOX TF dysregulation has been associated with numerous diseases including congenital heart disease, diabetes, and immunological diseases (Hannenhalli and Kaestner, 2009; Jonsson and Peng, 2005; Zhu, 2016). Many FOX TFs have also been associated with the cancers of various tissues and organs (Hannenhalli and Kaestner, 2009; Jonsson and Peng, 2005; Laissue, 2019; Myatt and Lam, 2007). For example, FOXC2 and FOXM1 are associated with several cancers, including cancers of the breast, lung, and colon, amongst others (Balli *et al.*, 2011; Hollier *et al.*, 2013; Kim *et al.*, 2006; Wang *et al.*, 2009; Yoshida *et al.*, 2007). Both FOXC2 and FOXM1 are involved in the epithelial-mesenchymal transition, the induction of cancer stem cell properties, and metastasis (Hollier *et al.*, 2013; Lam *et al.*, 2013). In addition, FOXM1 has been associated with cancer initiation and angiogenesis (Lam *et al.*, 2013). Interestingly, a number of small molecule inhibitors have exhibited the ability to specifically target and inhibit the effects of FOXC2 and FOXM1 in cancer and, in some cases, result in cancer cell apoptosis (Bhat *et al.*, 2009; Castaneda *et al.*, 2018; Hollier *et al.*, 2013; Lam *et al.*, 2013; Liu *et al.*, 2014; Zhu, 2014). Thus, FOX TFs are promising drug targets for the treatment of FOX-linked cancers and possibly other FOX-linked diseases.

1.6 FOXP SUBFAMILY

The FOXP subfamily consists of four members, FOXP1 to FOXP4 (Jackson *et al.,* 2010). Unlike most other FOX TFs that are transcriptional activators, evidence suggests that FOXP subfamily members are transcriptional repressors (Li *et al.,* 2004; Roll *et al.,* 2010; Shu *et al.,* 2001; Zuo *et al.,* 2007a). FOXP TFs regulate genes involved in a variety of processes. FOXP1, FOXP2, and FOXP4 control gene expression in tissues of the lung, brain, spleen, and gut, and play an important role in glucose homeostasis (Lu *et al.,* 2002; Shu *et al.,* 2001; Spaeth *et al.,* 2015). FOXP3 is specifically expressed and involved in the control of autoimmune response by regulatory T cells (Fontenot *et al.,* 2003).

All members of the FOXP subfamily have been linked to disease. Loss of FOXP1, FOXP2, and FOXP4 functioning has been identified in several cancers (Banham *et al.,* 2001; Cuiffo *et al.,* 2014; Teufel *et al.,* 2003). FOXP1 mutations have been identified in patients with speech defects, intellectual disabilities, emotional lability and symptoms of autism disorder (Hamdan *et al.,* 2010; Urreizti *et al.,* 2018). Mutations in FOXP3 cause immune dysregulation, polyendocrinopathy, enteropathy and X-linked inheritance (IPEX), an autoimmune disease (Bennett *et al.,* 2001). IPEX is characterised by diabetes, eczema, anaemia, hypothyroidism and enteropathy (Wildin *et al.,* 2002). A FOXP4 mutation has been associated with neurological disease symptoms such as congenital heart disease, defects of the larynx and developmental delay (Charng *et al.,* 2016).

FOXP2 is a TF involved in the development of specific brain regions (Lai *et al.,* 2001). A FOXP2 missense mutation leading to the substitution of an arginine by a histidine (R553H) was detected in multiple generations of a British family and in an unrelated individual and lead to the finding that FOXP2 is linked to developmental verbal dyspraxia (DVD), a hereditary disorder that modifies speech and language severely in affected individuals through the impairment of orofacial movements (Lai *et al.,* 2001). DVD also leads to impairments in language processing and grammatical skills (Gopnik and Crago, 1991; Vargha-Khadem *et al.,* 1995). Thus, FOXP2 is a key player in the development of speech and language and the *Foxp2* gene was the first gene to be linked to the development of neural structures required for speech and language (Lai *et al.,* 2001). FOXP2 has subsequently been associated with several other neurological diseases, including autism disorder and schizophrenia (Gong *et al.,* 2004; Sanjuán *et al.,* 2006).

The dysregulation of FOXP2 has been associated with numerous cancers, including breast cancer, prostate cancer, gastric cancer, and colorectal cancer (Cuiffo and Karnoub, 2016; Cuiffo *et al.,* 2014; Jia *et al.,* 2016; Stumm *et al.,* 2013; Tang *et al.,* 2010). However, the mechanism of FOXP2 dysregulation differs between cancer types, suggesting that FOXP2 has both tumour suppressive and oncogenic functions, and alternations between the two could depend on the signalling pathway. For example, malignant clinical breast cancer has been shown to display decreased FOXP2 expression levels (Cuiffo and Karnoub, 2016). An investigation of the effects of decreased FOXP2 expression levels in breast cancer cells has shown that FOXP2 knockdown is sufficient for cancer stem cell propagation, secondary tumour initiation and metastasis (Cuiffo and Karnoub, 2016). Thus, FOXP2 appears to have a tumour suppressive function in breast cancer. In 50% of prostate cancers, an androgen responsive gene is fused to a member of the erythroblast transformation-specific (ETS) family of TFs, known as the ETS-related gene (ERG) (Stumm *et al.,* 2013). This leads to a high ERG expression, with the

occurrence of downstream molecular changes in fusion-positive prostate cancers (Stumm *et al.,* 2013). Fusion-negative cancers, comprising the remaining 50% of prostate cancers, have a poorer prognosis than fusion-positive cancers, and have shown adverse effects linked to increased FOXP2 expression levels (Hermans *et al.,* 2009; Stumm *et al.,* 2013). Thus, FOXP2 appears to have an oncogenic function in prostate cancer, which could be counteracted by ERG in fusion-positive cancers (Stumm *et al.,* 2013).

1.6.1 FOXP STRUCTURE

FOXP subfamily members contain a glutamine-rich region, a zinc finger domain, a LeuZip domain and a FHD from N- to C-terminal (Shu *et al.,* 2001). The domain organisation of FOXP2, which typically consists of 715 amino acid residues, is outlined in Figure 1.12 (Lai *et al.*, 2001; Shu *et al.*, 2001).

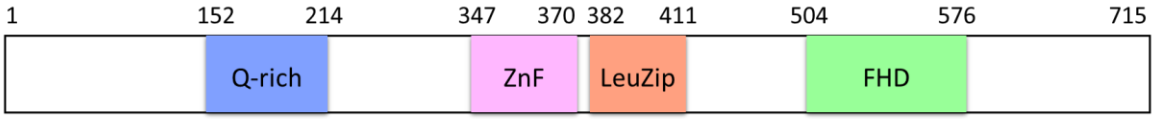


Figure 1.12: Domain organisation of FOXP2. FOXP2 contains a glutamine-rich region (Q-rich), zinc finger domain (ZnF), LeuZip domain and DNA-binding FHD, from N- to C-terminal. Numbers indicate amino acid positions of each domain relative to the full-length protein.

The FOXP glutamine-rich region may facilitate regulatory protein-protein interactions, and is the longest poly-glutamine tract in the human genome, in healthy individuals (Butland *et al.,* 2007; Chamberlain *et al.,* 1994; Lai *et al.,* 2001; Wang *et al.,* 2003). Zinc fingers have been found to play roles in stabilising protein fold and in binding to DNA, RNA, protein, and lipid (Brown, 2005; Gamsjaeger *et al.,* 2007; Hall, 2005; Klug, 1999; Low *et al.,* 2002; Matthews and Sunde, 2002; Miller *et al.,* 2010). Recent evidence suggests that, in the presence of the LeuZip, the FOXP2 zinc finger plays a role in mediating homotypic protein-protein interactions (Häußermann *et al.*, 2019). However, this has not been confirmed in other FOXP subfamily members, and the possibility of additional roles for the FOXP zinc finger have not been fully investigated. FOXP LeuZips are required for *in vivo* homo-oligomerisation and hetero-oligomerisation between FOXP subfamily members (Chae *et al.,* 2006; Li *et al.,* 2004; Wang *et al.,* 2003). In addition, FOXP LeuZips could be involved in non-specific interactions with DNA due to the requirement of an intact LeuZip domain for FOXP DNA binding (Chae *et al.*, 2006; Li *et al.*, 2004). Evidence suggests that the zinc finger and LeuZip form an uninterrupted α-helix which, during oligomerisation, assembles into an extended zinc finger-LeuZip coiled-coil (Song *et al.,* 2012). The FHD is the DBD and is linked to binding specificity (Koh *et al.,* 2009; Stroud *et al.,* 2006; Wang *et al.,* 2003; Webb *et al.,* 2017). FOXP3 exhibits the most structural variation, as it is

shorter than other FOXP subfamily members and the glutamine-rich region is replaced by a proline-rich region in FOXP3. Deletion of the FOXP3 proline-rich region results in the loss of FOXP3-mediated gene regulation, most likely due to the role of this region in the recruitment of FOXP3 binding partners such as histone deacetylases (Xie *et al.,* 2015).

1.6.2 FOXP DNA BINDING

As with other FOX TFs, FOXP proteins bind DNA by interactions of the FHD with DNA (Bandukwala *et al.*, 2011; Chen *et al.*, 2015; Stroud *et al.*, 2006; Wu *et al.*, 2006). The FOXP FHD binds specific DNA sites via the recognition helix (helix 3), which inserts into the major groove and forms base-specific contacts (Bandukwala *et al.*, 2011; Chen *et al.*, 2015; Stroud *et al.*, 2006; Wu *et al.*, 2006). Helix 1 and the wing regions form additional DNA contacts, predominantly with the DNA backbone, which stabilise the interaction (Bandukwala *et al.*, 2011; Chen *et al.*, 2015; Stroud *et al.*, 2006; Wu *et al.*, 2006). DNA binding of the FOXP2 FHD has been studied in detail, and it was shown that electrostatic and polar interactions of regions beyond H3, especially the wing regions, with the DNA backbone contribute to DNA-binding specificity (Morris *et al.*, 2018a). Furthermore, the flexibility of the loop between helix 2 and helix 3 is an important regulator of both DNA-binding specificity and affinity, through its effects on electrostatic interactions with the DNA backbone and on the depth of insertion of helix 3 into the major groove, respectively (Morris *et al.*, 2018a). The formation of the DNA-protein complex reduces the flexibility of the FOXP2 FHD through the formation of DNA-protein contacts which restrict the motion of the FHD (Morris *et al.*, 2018a).

1.7 OLIGOMERISATION IN THE FOXP SUBFAMILY

It is necessary that TF activity is tightly controlled. For FOXP TFs, this is achieved through a number of mechanisms; for example, FOXP2 DNA binding is regulated by environmental factors such as pH, the binding of cations to a conserved cation-binding site, and post-translational modifications such as phosphorylation (Blane and Fanucchi, 2015; Blane *et al.*, 2018; Morris *et al.*, 2018b). The FHD cation-binding site is conserved amongst FOX TFs across different subfamilies, and could serve as a universal regulator of FOX activity (Morris *et al.*, 2018b). A regulatory mechanism that appears to be unique to the FOXP subfamily, however, is the formation of homo-oligomers (Häußermann *et al.*, 2019; Li *et al.*, 2007; Song *et al.*, 2012). In addition, although other FOX TFs have shown the ability to form heterotypic oligomers, the formation of hetero-oligomers amongst members of the same subfamily appears to be unique to the FOXP subfamily, and serves to regulate FOXP transcriptional activity (Blount *et al.*, 2009; Li *et al.*, 2007, 2004; Lozano *et al.*, 2017; Madureira *et al.*, 2006; Ren *et al.*, 2019; Rudra *et al.*, 2012; Seoane *et al.*, 2004; Sin *et al.*, 2014; van der Vos and Coffer, 2008).

LeuZip motifs have been shown to facilitate the formation of TF dimers and higher order oligomers (see 1.4 Leucine zippers in transcription factors). FOXP subfamily members all contain a conserved LeuZip motif (Figure 1.12), and studies have shown that the LeuZip facilitates FOXP homo-oligomerisation (Chae *et al.*, 2006; Häußermann *et al.*, 2019; Li *et al.*, 2007, 2004; Lopes *et al.*, 2006; Song *et al.*, 2012; Wang *et al.*, 2003). In particular, the isolated mFoxp3 LeuZip homodimer forms a transient dimer, resulting in a dimer of dimers and giving rise to a tetrameric or higher order arrangement (Song *et al.,* 2012). In full-length FOXP3, the LeuZip mediates the formation of a number of homotypic quaternary structures, ranging from dimers to oligomers larger than tetramers (Li *et al.,* 2007). Indeed, there is evidence of the formation of homotypic tetramers amongst all FOXP TFs (Häußermann *et al.*, 2019; Li *et al.*, 2007; Mendoza, 2011; Song *et al.*, 2012). However, further work is required to fully clarify the native oligomeric state of each FOXP protein, as well as the role of higher order oligomers for FOXP DNA binding and hence function.

Hetero-oligomerisation between FOX proteins of different subfamilies, and of FOX TFs with other TFs, is a common means of regulating transcriptional activity (Blount *et al.*, 2009; Madureira *et al.*, 2006; Seoane *et al.*, 2004; van der Vos and Coffer, 2008). For example, the associations of FOXO3 with both FOXG1 and FOXM1 have been reported (Madureira *et al.*, 2006; Seoane *et al.*, 2004). However, in addition to enabling homo-oligomerisation, the FOXP LeuZip enables the association of FOXP subfamily members with each other, a phenomenon that is unique to the FOXP subfamily (Lam *et al.*, 2013; Li *et al.*, 2004; Rousso *et al.*, 2012; Shu *et al.*, 2007). There is currently limited information regarding the oligomeric state and configuration of heterotypic FOXP oligomers, but a recent study has shown that FOXP1 and FOXP2 constructs associate to form a number of quaternary structures *in vitro* (Thulo, 2019). In addition, associations of FOXP1 and FOXP2 with each other are relatively weak compared with the homotypic associations of these proteins, suggesting a more transient FOXP1-FOXP2 hetero-oligomeric association (Thulo, 2019).

Tissue-specific expression patterns of FOXP1, FOXP2, and FOXP4 share significant overlap, particularly in the developing brain (Konopka *et al.,* 2009; Takahashi *et al.,* 2008a, 2008b, 2009; Teramitsu *et al.,* 2004). Interestingly, these members of the FOXP subfamily have been shown to co-localise subcellularly (Sin *et al.,* 2014). An investigation was conducted on 10 downstream FOXP2 gene targets including genes involved in the Wnt signalling pathway, Notch signalling pathway, and nervous system development (Sin *et al.,* 2014). Wnt signalling is important for embryonic development and adult homeostasis whilst Notch signalling plays roles in differentiation, proliferation, and apoptosis (Artavanis-Tsakonas *et al.,* 1999; Sin *et al.,* 2014). It was discovered that FOXP1/2/4 homo- and

hetero-oligomerisation have distinct effects on gene expression, with each homo- and hetero-oligomer differentially activating and repressing the transcription of specific gene targets (Sin *et al.,* 2014). For example, FOXP1 and FOXP4 homo-oligomers both lead to the activation of the nuclear receptor co-repressor 2 gene involved in the Notch signalling pathway, whereas the FOXP1-FOXP4 hetero-oligomer causes transcriptional repression of this gene (Sin *et al.,* 2014).

A study of FOXP3 revealed that the inhibition of FOXP3 homo-oligomerisation alone results in a loss of function, in particular with respect to the control of regulatory T cells (Lozano *et al.,* 2017). Studies have also shown that FOXP1 and FOXP3 co-localise at many sites in the nuclei of regulatory T cells (Li *et al.,* 2007; Song *et al.,* 2012). FOXP3-FOXP1 associations are crucial to the function of regulatory T cells; loss of FOXP1 results in a decrease in FOXP3 transcriptional activity and can lead to autoimmune diseases due to a loss of suppressive immune activity (Li *et al.,* 2007; Ren *et al.,* 2019). Interestingly, evidence has shown that FOXP3 controls FOXP1 expression levels in regulatory T cells, suggestive of a regulatory FOXP1-FOXP3 feedback loop (Rudra *et al.,* 2012). Thus, the formation of homo- and heterotypic FOXP oligomers has been observed for all FOXP subfamily members, and is a form of transcriptional regulation that could allow for the control of tissue-specific gene expression (Li *et al.*, 2007, 2004; Lozano *et al.*, 2017; Ren *et al.*, 2019; Rudra *et al.*, 2012).

1.7.1 FOXP OLIGOMER INTERFACES

Most evidence suggests that the LeuZip is the key FOXP oligomer interface (Chae *et al.,* 2006; Li *et al.,* 2004; Lopes *et al.,* 2006; Wang *et al.,* 2003). However, the FOXP DNA-binding FHD also exhibits unique dimerisation abilities, which have not been observed in any other FOX protein (Bandukwala *et al.,* 2011; Chu *et al.,* 2011; Stroud *et al.,* 2006). Furthermore, a recent study comparing mutant and truncated FOXP2 variants suggests that the FHD could contribute to the overall affinity of FOXP2 homo-oligomerisation interactions (Häußermann *et al.,* 2019). The relative importance and contribution of each interface to FOXP homo- and hetero-oligomerisation has not been fully characterised.

In vitro studies have shown that the isolated FOXP1 and FOXP2 FHDs exist as a mixture of monomer and dimer, and the isolated FOXP3 FHD is exclusively dimeric (Bandukwala *et al.,* 2011; Chu *et al.,* 2011; Stroud *et al.,* 2006). Crystal structures have revealed that FOXP FHD dimerisation occurs by domain swapping (Bandukwala *et al.,* 2011; Stroud *et al.,* 2006). Domain swapping is a process involving the exchange of equivalent parts between two identical proteins in the formation of an oligomer (Bennett *et al.,* 1994). In the FOXP FHD, domain swapping occurs by the exchange of helix 1

and helix 2 between two monomers, resulting in an intertwined dimer as shown in Figure 1.13 (Bandukwala *et al.,* 2011; Stroud *et al.,* 2006). Several disease-linked mutations of the FOXP FHD, including mutations linked to IPEX and autism disorder, have been shown to impair domain swapping, emphasising the significance of FHD domain swapping for FOXP function (Bandukwala *et al.*, 2011; Siper *et al.*, 2017). Although FOXP proteins are capable of forming hetero-oligomers with each other, there is currently limited evidence of heterotypic interactions of the FHD (Thulo, 2019). Furthermore, the domain-swapped FHD (Figure 1.13) is more intertwined than the coiled-coil LeuZip (Figure 1.6), suggesting that heterotypic FOXP interactions occur via the LeuZip whereas FHD domain-swapping occurs predominantly between monomers of the same FOXP protein (Bandukwala *et al.*, 2011; Jones and Thornton, 1996; Li *et al.*, 2004; Stroud *et al.*, 2006).

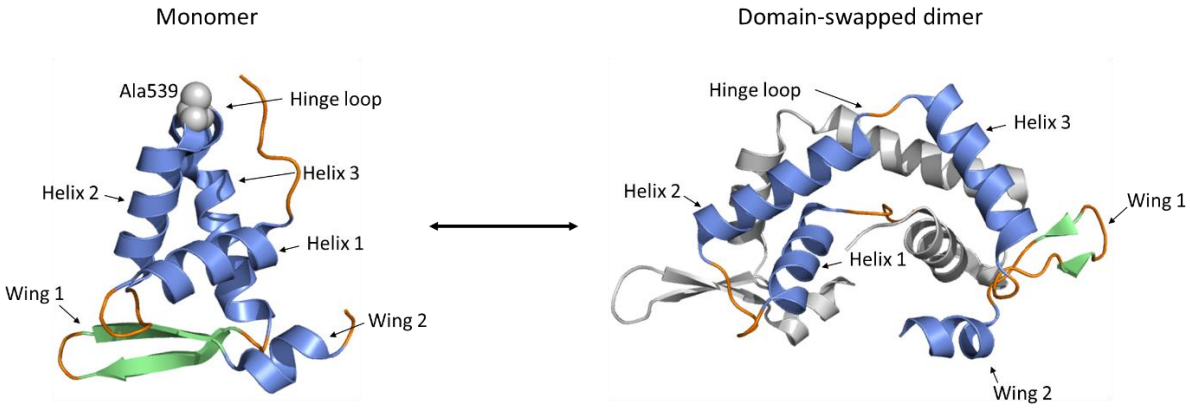


Figure 1.13: The FOXP2 forkhead domain forms a domain-swapped dimer. FOXP2 Ala539 confers flexibility to the FHD hinge loop, thereby allowing for dimerisation by domain swapping. Exchange of helix 1 and helix 2 between two monomers allows for the formation of an intertwined dimer. A single monomer is shown on the left-hand side and a domain-swapped dimer is shown on the right-hand side, with one monomer of the domain-swapped dimer in grey (structures are labelled for the monomer shown in blue). In the FOXP2 FHD, wing 1 forms two anti-parallel β-strands and wing 2 forms an α-helix. α-Helices are blue, β-strands are green, loops are orange, and Ala539 is grey and represented as spheres (left-hand side). Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC (PDB 2A07) (Stroud *et al.,* 2006).

Formation of the FHD domain-swapped dimer requires flexibility of the turn or hinge loop connecting helix 2 and helix 3 (Bandukwala *et al.,* 2011; Stroud *et al.,* 2006). As seen in Figure 1.10, a key proline residue situated in the hinge loop is conserved amongst all other FOX proteins and replacement by alanine in FOXP subfamily members (Ala539 in FOXP2) confers flexibility to this region (Bandukwala *et al.,* 2011; Stroud *et al.,* 2006). Therefore, it is not surprising that mutation of Ala539 to a proline reduces hinge loop flexibility and prevents formation of the FOXP2 FHD domain-swapped dimer (Stroud *et al.,* 2006). It is worth noting that the conserved hinge loop proline is replaced by cysteine in FOXJ1 and alanine in FOXS1 (Figure 1.10), but no structural characterisation has been conducted on the FHDs of these proteins.

Specific interactions and residues involved in FOXP1/2/4 LeuZip oligomerisation have not been studied. However, the M406T missense mutation of the FOXP2 LeuZip was detected in a patient and associated with cognitive and language impairments, and epilepsy (Roll *et al.,* 2010). *In vitro* analyses showed that M406T FOXP2 retained partial transcriptional activity (Roll *et al.,* 2010). Furthermore, a parent and both siblings of the patient were carriers of the mutation but showed no phenotypical effects or symptoms, so it is plausible that the effects of the M406T mutation, as seen in the patient, are dependent on environmental or other factors, and on its own, the mutation may have little or no effect on FOXP2 LeuZip oligomerisation, although it should be highlighted that no additional FOXP2 mutations were detected in the patient (Roll *et al.,* 2010).

Mutations of the FOXP3 LeuZip eliminate homo- and hetero-oligomerisation of full-length FOXP3 (Li *et al.,* 2007). An extensive study of the mFoxp3 zinc finger-LeuZip coiled-coil conducted by Song *et al.* (2012) found that this region facilitates the formation of anti-parallel dimers, and the resulting dimers associate in an anti-parallel configuration to form a tetramer. Key residues for mFoxp3 oligomerisation and oligomer stability were identified. Leu241 and Leu245 are located at core positions *d* and *a*, respectively (corresponding with Leu242 and Leu246 in human FOXP3, refer to Figure 1.7A) and contribute to the formation of a hydrophobic core at the dimer interface, as depicted in Figure 1.14A (Song *et al.,* 2012). Leu241 and Leu245 lie at the symmetric centre of the coiled-coil (Song *et al.,* 2012). The L242P mutation in human FOXP3 is an IPEX mutation and introducing the equivalent mutation (L241P) leads to the disruption of mFoxp3 LeuZip homodimeric associations (Song *et al.,* 2012). Lys249 is at position *e* (corresponding with Lys250 in human FOXP3) and forms an intersubunit hydrogen bond with Glu242 at position *e* of the second monomer, which stabilises the mFoxp3 LeuZip dimer (Figure 1.14B) (Song *et al.,* 2012). The deletion of Lys250 in human FOXP3 is another IPEX mutation and introducing the equivalent mutation (deletion of Lys249) was also found to disrupt mFoxp3 LeuZip homodimeric associations (Song *et al.,* 2012).

Interestingly, the mFoxp3 zinc finger-LeuZip coiled-coil has a higher proportion of polar amino acid residues at core positions (*a* and *d*) than is usual for LeuZips and other coiled-coil structures (Song *et al.,* 2012). The side chain edges of these polar amino acid residues are twisted out of the buried hydrophobic core, a unique phenomenon that promotes higher order oligomerisation (Song *et al.,* 2012). It is worth noting that, in humans and mice, the core polar amino acid residues are conserved amongst the FOXP subfamily (Figure 1.7A), which, like the unique alanine of the FHD hinge loop, could promote oligomerisation in FOXP subfamily members. One such amino acid is Gln234 (core position *d*) that forms an intersubunit hydrogen bond with Glu248 (position *f*). Lys251 (position *g*) is a key

residue for this interaction, and initiates and stabilises the interaction by positioning the Glu248 side chain optimally for bond formation (Figure 1.14C) (Song *et al.,* 2012). Other FOXP subfamily members have an arginine residue in the position equivalent to that of Lys251 (Figure 1.7A), and mutation of mFoxp3 Lys251 to arginine had little effect on LeuZip oligomerisation (Song *et al.,* 2012). Acetylation is an important post-translational modification for the regulation of FOXP3 DNA binding and function (Samanta *et al.,* 2008; Tao *et al.,* 2007). There is much evidence in support of the significance of FOXP3 acetylation for its function; for example, acetylation promotes regulatory T cell function and protects FOXP3 from polyubiquitination and proteasomal degradation (van Loosdregt *et al.,* 2010, 2011; Tao *et al.,* 2007). Through the use of mass spectrometry, it was found that human FOXP3 undergoes regulatory acetylation at Lys250 and Lys252 (corresponding with Lys249 and Lys251 in mFoxp3), resulting in the alteration of the charges at these positions (Song *et al.,* 2012). Interestingly, introducing the K251D mutation which alters the charge at this position was found to disrupt homotypic mFoxp3 LeuZip associations, as well as FOXP3-FOXP1 interactions (Song *et al.,* 2012).

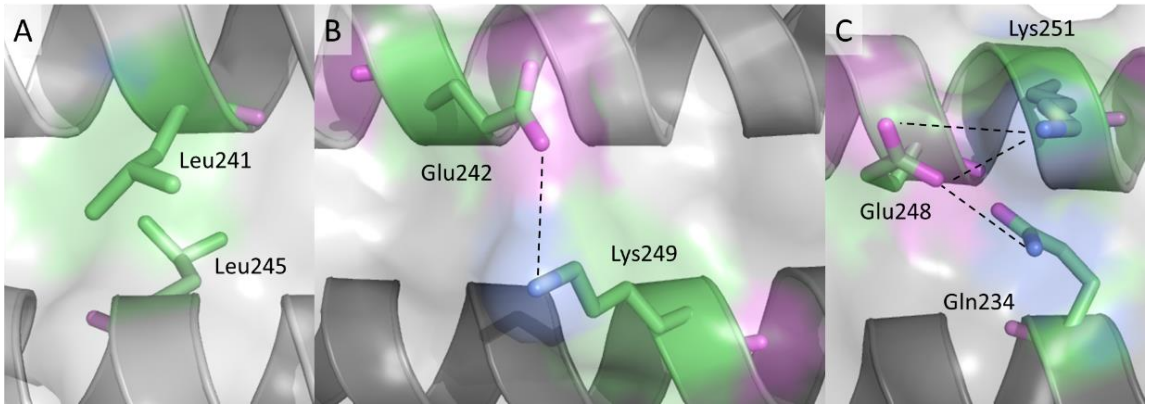


Figure 1.14: Key residues at the mFoxp3 leucine zipper dimer interface. (A) The hydrophobic leucine residues at positions 241 and 245 are located at the symmetric centre of the coiled-coil. **(B)** An intersubunit hydrogen bond between Lys249 and Glu242 stabilises the LeuZip dimer. **(C)** Lys251 positions the Glu248 side chain correctly for formation of an intersubunit hydrogen bond between Glu248 and Gln234, which stabilises the mFoxp3 LeuZip dimer. Carbon atoms are green, oxygen atoms are pink, nitrogen atoms are blue, and hydrogen bonds are dashed lines (black). Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC (PDB 4I1L) (Song *et al.,* 2012).

The FOXP2 LeuZip is situated N-terminal to the FHD (Lai *et al.,* 2001; Shu *et al.,* 2001). The N-terminal ends of the FOXP2 FHD monomers are in close proximity in the isolated FHD domain-swapped dimer (Stroud *et al.,* 2006). Thus, the FOXP2 FHD, upon dimerisation, adopts a configuration that would enable simultaneous dimerisation of the LeuZip and FHD (Li *et al.,* 2004). However, the possibility of higher order oligomerisation should be considered since there is substantial evidence for the formation of FOXP2 oligomers that are larger than dimers (Häußermann *et al.*, 2019; Mendoza, 2011; Thulo, 2019). Based on the mFoxp3 LeuZip, it is likely that the FOXP2 LeuZip forms an anti-parallel

dimer of dimers. Figure 1.15 shows a plausible tetrameric arrangement for the LeuZip and FHD regions of FOXP2, with the LeuZips forming an anti-parallel dimer of dimers. The FHDs of one copy of each dimer could be in close proximity, depending on the flexibility of the regions connecting the LeuZips to the FHDs. Thus, the proposed arrangement would allow for simultaneous dimerisation of the FHDs, and could hold true for all FOXP proteins.

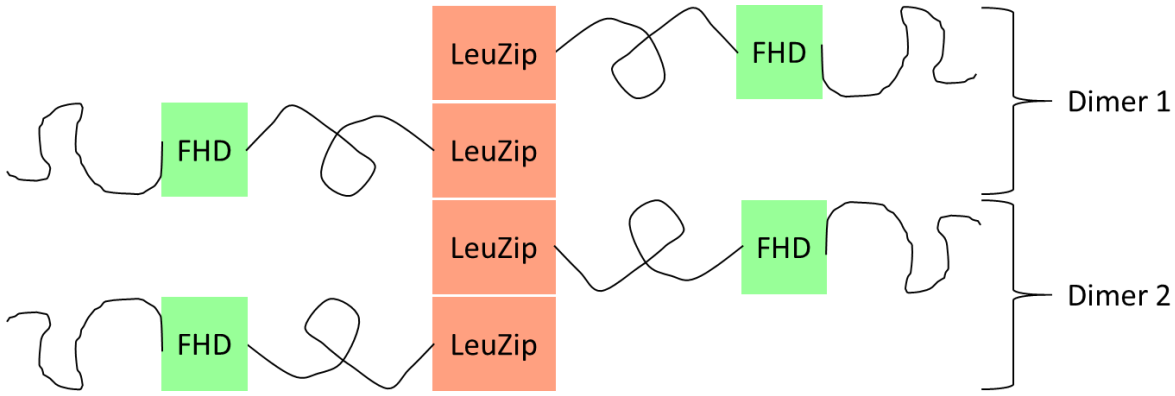


Figure 1.15: Proposed tetrameric arrangement of the LeuZip and forkhead domain regions of FOXP2. In the proposed model, two FOXP2 monomers dimerise in an anti-parallel configuration, via the LeuZip domains (dimer 1). Two dimers then associate, again in an anti-parallel configuration via the LeuZip domains, resulting in a dimer of dimers or tetramer. The flexibility of the regions connecting the LeuZip and FHD could allow for simultaneous dimerisation of the FHDs.

1.7.2 FOXP OLIGOMERS AND DNA BINDING

Crystal structures of the FOXP2 and FOXP3 DNA-binding FHDs reveal unique DNA-bound FHD dimers, whereas the FHDs of other FOX proteins bind DNA exclusively as monomers (Bandukwala *et al.*, 2011; Chen *et al.*, 2015; Clark *et al.*, 1993; Li *et al.*, 2004; Obsil and Obsilova, 2008; Stroud *et al.*, 2006; Tsai *et al.*, 2007). As previously mentioned, this has been attributed to the increased flexibility of the FOXP FHD hinge loop due to the replacement of a conserved proline residue by alanine in FOXP subfamily members (Bandukwala *et al.*, 2011; Morris and Fanucchi, 2016; Stroud *et al.*, 2006). This confers flexibility to the FOXP FHD in general, thereby predisposing the FOXP FHD to the formation of a larger number of non-specific DNA interactions compared to the FHDs of other FOX TFs, and possibly resulting in an increased DNA-binding affinity (Morris and Fanucchi, 2016; Morris *et al.*, 2018a). Interestingly, in the FOXP2 and FOXP3 FHD dimers, the recognition helices (helix 3) of each monomer are separated and positioned in such a manner that would allow each monomer to simultaneously bind a distinct DNA site, and the dimers display DNA interactions differing to those of the monomers (Bandukwala *et al.,* 2011; Chen *et al.,* 2015; Stroud *et al.,* 2006). Hence, FOXP2 and FOXP3 could play roles in the assembly of large transcription complexes, organising genes, DNA looping or establishing interchromosomal links (Bandukwala *et al.,* 2011; Chen *et al.,* 2015; Stroud *et al.,* 2006). A regulatory

equilibrium between monomeric and oligomeric forms could have significant *in vivo* functional implications (Chen *et al.,* 2015; Stroud *et al.,* 2006).

In vivo studies have shown that deletion of the LeuZip regions of FOXP1, FOXP2, and FOXP4 disrupts homo- and heterotypic associations amongst this group of proteins, and abolishes DNA binding and hence function (Li *et al.,* 2004). The requirement of oligomerisation for DNA binding in FOXP1/2/4 is further supported by evidence suggesting that DNA binding can be restored by the addition of a glutathione S-transferase (GST) tag, which is an obligate dimer in solution (Li *et al.,* 2004). Similarly, the LeuZip of FOXP3 is required for homo-oligomerisation and hetero-oligomerisation with FOXP1, and mutations in this region disrupt both homo- and heterotypic associations (Chae *et al.,* 2006; Song *et al.,* 2012; Wang *et al.,* 2003). FOXP3 LeuZip mutations, and the resulting disruption of homo- and hetero-oligomerisation, decreases FOXP3 DNA binding and function significantly *in vivo* (Chae *et al.,* 2006; Li *et al.,* 2007; Lopes *et al.,* 2006). Taken together, these discoveries indicate that oligomerisation is a prerequisite for FOXP DNA binding.

It has been observed that the monomeric FOXP2 FHD is capable of binding DNA, which appears to contradict studies showing that FOXP2 undergoes oligomerisation-dependent DNA binding (Li *et al.*, 2004; Morris and Fanucchi, 2016; Stroud *et al.*, 2006). However, it is only in the absence of the LeuZip that a FOXP2 monomer has exhibited DNA binding (Morris and Fanucchi, 2016; Stroud *et al.*, 2006). Therefore, it is likely that only the isolated FHD is capable of binding DNA as a monomer, whereas DNA binding of full-length FOXP2 is oligomerisation-dependent. Nevertheless, even in the isolated FOXP2 FHD, the disruption of domain swapping decreases DNA-binding affinity and alters the DNA binding mechanism significantly (Morris *et al.*, 2018a).

Current findings are clear in emphasising the importance of homo- and hetero-oligomerisation as a regulatory mechanism in the FOXP subfamily, with FOXP protein binding partners dictating the downstream effects of FOXP transcriptional activity. Dysregulation of FOXP homo- and hetero-oligomerisation could be the basis for many FOXP-linked diseases (Lozano *et al.,* 2017; Ren *et al.,* 2019; Sin *et al.,* 2014). Further investigation is required to fully understand the role of oligomerisation in FOXP structure and function.

1.8 PROBLEM STATEMENT

Homo-oligomerisation is a common phenomenon amongst TFs. For some TFs, homo-oligomerisation is a prerequisite for DNA binding and hence function, whereas for other TFs the existence of multiple oligomeric states enables regulation of TF activity through alternate DNA-protein, and sometimes protein-protein, interactions for each oligomeric state (Vinkemeier *et al.,* 1998; Wingelhofer *et al.,* 2018; Wingender *et al.,* 2018). A number of motifs have been identified at TF oligomer interfaces. The most common TF protein-protein interfaces are the HLH motif of bHLH TFs, the LeuZip motif of bZip TFs, and both the zinc finger and ligand-binding domains of nuclear receptor TFs (Amoutzias *et al.,* 2008; Brownlie *et al.,* 1997; Ellenberger, 1994; Glover and Harrison, 1995; Miller *et al.,* 2003). The LeuZip interface, particularly, is able to mediate TF oligomer formation for oligomers as large as octamers (Aoki *et al.,* 1999; Kasai *et al.,* 1997).

FOX TFs are ubiquitously expressed throughout eukaryotes (Kaestner *et al.,* 2000). In humans, FOX TFs regulate a large number of diverse processes from embryonic development through to adulthood, and the dysregulation of FOX TFs has been linked with an equally large number of diseases, emphasising the important role of this TF family in gene regulation (Friedman and Kaestner, 2006; Hannenhalli and Kaestner, 2009; Jonsson and Peng, 2005; Koon *et al.,* 2007; Laissue, 2019; Lee *et al.,* 2005; Ma *et al.,* 2014; Myatt and Lam, 2007; Wan *et al.,* 2005; Zhu, 2016). Members of the FOXP subfamily exhibit the unique ability to form dimers and higher order oligomers, a trait that has not been observed in any other FOX subfamily (Bandukwala *et al.,* 2011; Chen *et al.,* 2015; Li *et al.,* 2004; Sin *et al.,* 2014; Stroud *et al.,* 2006). Homo- and heterotypic interactions amongst FOXP subfamily members regulate gene expression, with different homo- and hetero-oligomer combinations each having distinct effects on the activation or repression of transcriptional activity and gene expression (Li *et al.,* 2007; Ren *et al.,* 2019; Sin *et al.,* 2014). Both the FOXP LeuZip and the DNA-binding FHD exhibit dimerisation/oligomerisation properties, though most evidence suggests that the LeuZip is the major FOXP oligomerisation interface (Bandukwala *et al.,* 2011; Chae *et al.,* 2006; Chu *et al.,* 2011; Li *et al.,* 2004; Lopes *et al.,* 2006; Song *et al.,* 2012; Stroud *et al.,* 2006; Wang *et al.,* 2003).

It is evident that oligomerisation is an important regulator of FOXP DNA binding and function. However, there is currently limited information regarding the relative contributions of the LeuZip and FHD to FOXP oligomerisation. The isolated LeuZip and FHD interfaces are well characterised, but the combined effects of both interfaces on structure, homo-oligomerisation, and DNA binding are still somewhat unclear. An understanding of the FOXP LeuZip and FHD could reveal important information regarding FOXP protein-protein and protein-DNA interactions, as well as the mechanism of FOXP

oligomerisation and DNA binding. This information could shed light on the essential roles of FOXP TFs in regulating transcriptional activity. Importantly, a thorough understanding of FOXP proteins is required in order to develop novel therapeutics, and to explore the possibility of diagnostics, targeting FOXP proteins in FOXP-linked diseases such as cancer.

In this study of homotypic FOXP2 associations, a FOXP2 construct extending from the LeuZip to the C-terminal end of the protein was used (LeuZip - end). The LeuZip and FHD interfaces were investigated using point mutations of each interface. This was used to clarify the effects of disruptions to each interface, as well as the combined effects of disrupting both interfaces, on the homo-oligomerisation, structure, and DNA binding of FOXP2. The novel R401D mutation of the FOXP2 LeuZip was designed based on the equivalent K251D mutation of mFoxp3 (Song *et al.*, 2012). The FOXP2 R401D mutation was designed to disrupt homotypic FOXP2 LeuZip interactions, in order to probe the role of the FOXP2 LeuZip interface in FOXP2 structure and DNA binding. The A539P mutation is well-characterised in the isolated FHD and results in an exclusively monomeric FHD (Morris and Fanucchi, 2016; Morris *et al.*, 2018a; Stroud *et al.*, 2006), but this mutation has not been investigated when combined with the LeuZip. Therefore, the A539P mutation was used to create a novel FOXP2 construct for investigating the effects of an exclusively monomeric FHD on the overall quaternary structure of the protein, in the presence of the LeuZip oligomer interface. The A539P mutant also allowed for evaluating the importance of the FHD interface in FOXP2 structure and DNA binding. In addition, the combined effects of both mutations were studied through the R401D/A539P mutant, aimed to disrupt homotypic FOXP2 interactions at both the LeuZip and FHD interfaces. A thorough comparison of the R401D, A539P and R401D/A539P mutants with each other, as well as with a wild-type FOXP2 construct, was conducted. The findings of this study could improve our understanding of FOXP oligomerisation through understanding the LeuZip and FHD interfaces, as well as the importance of key residues (Arg401 and Ala539 in FOXP2) at each interface.

1.9 AIM AND OBJECTIVES

1.9.1 AIM

In order to fully understand a TF or any other protein, it is necessary that both the structure and function of the protein is fully characterised. The unique ability of FOXP TFs to form higher order oligomers could play a physiological role in enhancing DNA binding, and hence FOXP-mediated transcriptional repression, and could therefore provide a critical means of regulating gene expression. The aim of this research was to understand the importance of the LeuZip and FHD oligomer interfaces for FOXP oligomerisation, by studying the effects of point mutations of the LeuZip and FHD on FOXP2. A combination of *in silico* and *in vitro* methods was used for a thorough characterisation of wild-type and mutant FOXP2 (LeuZip - end) structure, homo-oligomerisation, and DNA binding.

1.9.2 OBJECTIVES

The aim of this study was achieved through specific objectives, which were to:

1. Use site-directed mutagenesis to produce the R401D, A539P and R401D/A539P mutants of FOXP2 (LeuZip - end).

2. Overexpress and purify all FOXP2 (LeuZip - end) variants.

3. Confirm DNA binding and compare unbound and DNA-bound oligomeric states by electrophoretic mobility shift assays.

4. Elucidate the roles of LeuZip and FHD associations in FOXP2 DNA binding by analysing DNA-binding affinities of FOXP2 (LeuZip - end) variants using fluorescence anisotropy DNA-binding assays.

5. Investigate the secondary and tertiary structure of the FOXP2 (LeuZip - end) construct, and the effects of the LeuZip and FHD mutations on protein structure, using a combination of *in silico* and spectroscopic techniques.

6. Evaluate the effects of the mutations on FOXP2 (LeuZip - end) oligomeric state using size exclusion chromatography to determine the importance of the LeuZip and FHD for FOXP2 homo-oligomerisation.

7. Compare the effects of the LeuZip and FHD mutations, and hence the effects of LeuZip- and FHD-mediated associations of FOXP2, on the flexibility and fold of the unbound and DNA-bound FOXP2 (LeuZip - end) variants by means of Stern-Volmer constants obtained from acrylamide fluorescence quenching.

8. Examine the importance of LeuZip and FHD associations for protein stability, and compare the stabilities of the various FOXP2 oligomeric states, through thermal unfolding of the FOXP2 (LeuZip - end) variants.

Chapter 2: Experimental procedures

2.1 *IN SILICO* PROTEIN STRUCTURE ANALYSES

2.1.1 PROTEIN SECONDARY STRUCTURE AND DISORDER PREDICTIONS

Computational methods of protein structure predictions are becoming increasingly common, especially since the development of new prediction methods and algorithms allows for more accurate predictions (Deng *et al.*, 2018; Nielsen and Mulder, 2019; Yang *et al.*, 2018). Knowledge of protein secondary structures in particular is important, since secondary structures largely affect how proteins fold and thus the resulting protein tertiary structures (Dai and Zhou, 2011; Murzin *et al.*, 1995; Ozkan *et al.*, 2007; Zhou and Karplus, 1999). Furthermore, due to a protein structure being a key determinant of its function, knowledge of protein secondary structures is useful for predicting the functions of proteins (Godzik *et al.*, 2007; Yang *et al.*, 2018). Related to the prediction of protein secondary structures is the prediction of unstructured or disordered regions (Nielsen and Mulder, 2019; Yang *et al.*, 2018). Disordered protein regions are significant since the lack of structure confers conformational freedom to these regions, which increases the range of biological functions that disordered protein regions are capable of assuming (Nielsen and Mulder, 2019). Indeed, disordered proteins or protein regions are involved in a wide variety of biological processes and diseases (Buchan *et al.*, 2010; Dyson and Wright, 2005; Midic *et al.*, 2009; Wright and Dyson, 2015).

The PSIPRED (PSI-BLAST based secondary structure prediction, where PSI-BLAST stands for Position-Specific Iterative Basic Local Alignment Search Tool (Altschul *et al.,* 1997)) server hosts a range of online tools for the prediction and identification of different protein features based on protein primary structures (Buchan and Jones, 2019). One such tool is the PSIPRED secondary structure prediction tool, which predicts protein secondary structures in three stages (Jones, 1999). In the first stage, PSI-BLAST is used to produce a multiple sequence alignment (Jones, 1999). PSI-BLAST produces a position-specific scoring matrix (PSSM) based on the multiple sequence alignment, with the elements of the matrix (scores) representing the likelihood of that amino acid substitution occurring in the alignment (Jones, 1999). The PSSM is used to search for new sequences with each iteration, and is updated after each iteration, allowing PSI-BLAST to detect distant evolutionary relationships, which improves the accuracy of PSIPRED over that of alternative tools (Jones, 1999). PSIPRED uses three iterations of PSI-BLAST (Jones, 1999). Since the iterative nature of PSI-BLAST can lead to the inclusion of random sequences, PSIPRED employs its own unique sequence database

containing non-identical sequences, and with regions of sequences lacking information removed (Jones, 1999). In the second stage, the secondary structure is predicted using neural networking, as a three-state prediction (α-helix, β-strand, or loop) (Jones, 1999). The third stage produces the final three-state prediction by filtering the outputs from the second stage by an additional neural networking step (Jones, 1999). Although the latest version of PSIPRED (PSIPRED 4.0) has not yet been compared with other secondary structure prediction tools, previous versions (including PSIPRED 3.0) have proven to have a high level of accuracy in comparison with most alternative available tools, making PSIPRED the tool of choice for protein secondary structure prediction (Buchan *et al.*, 2010; Koswatta *et al.*, 2011; Yang *et al.*, 2018).

The DISOPRED (Disorder Prediction) method, also available on the PSIPRED server, predicts intrinsically disordered regions within proteins (Jones and Cozzetto, 2015). Critical Assessment of Structure Prediction (CASP) is a biennial experiment, during which independent assessors critically evaluate and compare computational methods for protein structure prediction (Moult *et al.*, 1995). DISOPRED3 has been recognised as one of the most accurate tools for the prediction of disordered protein regions, in the most recent CASP evaluation of disorder prediction tools, since it ranked as the best or one of the best tools in all comparisons (Monastyrskyy *et al.*, 2011, 2014). DISOPRED3 uses a prediction method that is a combination of three modules; a support vector machine, neural networking, and a nearest neighbour classifier. The support vector machine has a high accuracy for predictions of short disordered regions, whereas the neural networking has a high accuracy for predictions of long disordered regions (Jones and Cozzetto, 2015). Since both the support vector machine and neural networking require training, which is time- and labour-intensive, the nearest neighbour classifier was included to allow for DISOPRED to be easily updated, as this module does not require training (Jones and Cozzetto, 2015).

Similarly to PSIPRED, DISOPRED3 predictions rely on a PSSM resulting from three iterations of PSI-BLAST (Jones and Cozzetto, 2015; Jones and Ward, 2003). DISOPRED3 employs its own unique sequence database containing non-identical sequences to generate the PSI-BLAST multiple sequence alignment (Jones and Cozzetto, 2015; Jones and Ward, 2003). Disordered regions are identified based on residues or regions that are missing, or have an occupancy of zero, from high resolution protein structures solved through X-ray crystallography, as these regions are presumably missing due to disorder (Jones and Cozzetto, 2015). Based on this, the support vector machine was trained to classify a residue as either ordered or disordered, using the PSSM (Jones and Cozzetto, 2015). The neural networking is similar to that of PSIPRED, but the prediction is a two-state prediction (either ordered

or disordered), based on the PSSM (Jones and Ward, 2003). The nearest neighbour classifier compares amino acid residues in the target sequence with those of the sequences in the multiple sequence alignment generated by PSI-BLAST, on the basis of PSSM elements (scores) (Jones and Cozzetto, 2015). Amino acid sequences from the alignment with the highest PSSM elements are used to classify the amino acid residues of the target sequence as either ordered or disordered (Jones and Cozzetto, 2015). The results of all three modules (support vector machine, neural networking, and nearest neighbour classifier) are used as the input for a final neural networking stage. DISOPRED3 annotates each amino acid of the target sequence as either ordered or disordered, based on a confidence score that indicates the difference between the confidence of the two possible outputs (disordered - ordered) (Jones and Cozzetto, 2015).

Disordered protein regions are often involved in binding to either proteins or nucleic acids, and the disorder allows these regions to adopt different conformations depending on the interactions required for binding (Jones and Cozzetto, 2015; Ward *et al.*, 2004). The DISOPRED3 method is able to identify protein binding regions amongst the predicted disordered protein regions (Jones and Cozzetto, 2015). Specifically, DISOPRED3 predicts protein binding regions based on proteins that are disordered in solution, and fold upon binding (Jones and Cozzetto, 2015). This is achieved by a support vector machine that has been trained to differentiate between peptides that are bound to either proteins, nucleic acids, ligands, or metal ions, from peptides that are unbound (Jones and Cozzetto, 2015). As with the previously described prediction methods, the protein binding region predictions are based on the PSSM elements generated from three iterations of PSI-BLAST (Jones and Cozzetto, 2015). DISOPRED3 annotates each disordered residue as either forming part of a protein binding region or not, based on a confidence score that indicates the probability of that residue forming part of a protein binding region (Jones and Cozzetto, 2015).

Due to the dependence of PSIPRED 4.0 and DISOPRED3 on multiple sequence alignments, a limitation of these methods when studying the effects of mutations on a protein, is that the mutants could result in the same multiple sequence alignment as that of the wild-type. This would reduce the sensitivity of the predictions to the effects of the mutations. Nevertheless, predictions are useful in obtaining an overview of the secondary structures, disorder, and any predicted binding regions of the proteins. Furthermore, predictions could reveal any local changes in secondary structure or disorder, occurring in close proximity to the sites of mutation.

The amino acid sequence of each FOXP2 (LeuZip - end) protein was analysed online by PSIPRED 4.0 and DISOPRED3 (Jones, 1999; Jones and Cozzetto, 2015). PSIPRED 4.0 was used to predict the protein secondary structures and DISOPRED3 was used to predict intrinsically disordered regions of each protein. Each sequence was returned with each amino acid annotated by PSIPRED 4.0 according to the secondary structural element that it was predicted to form part of (α-helix, β-strand, or loop). In addition, each amino acid was annotated by DISOPRED3 with a confidence score indicating the disorder (Jones and Cozzetto, 2015). Amino acid residues with confidence scores of 0.5 or higher were considered disordered and, from this subset, residues that were predicted to form part of a binding site were identified (Jones and Cozzetto, 2015).

2.1.2 THREE-DIMENSIONAL PROTEIN STRUCTURE PREDICTIONS

In order to fully understand the way a protein functions, as well as its protein-protein and protein-DNA interactions, it is useful to determine the structure of the protein (Slabinski *et al.*, 2007). There are several methods for the *in vitro* determination of protein structure, the most common being X-ray crystallography and nuclear magnetic resonance spectroscopy (Slabinski *et al.*, 2007; Srivastava *et al.*, 2018). However, these methods pose challenges and have limitations. X-ray crystallography requires successful crystallisation of the protein, which is affected by the length of the protein amino acid sequence, its flexibility in solution, and its hydrophobicity (Slabinski *et al.*, 2007). In addition, since protein crystallisation hinders the dynamic motions of protein molecules, X-ray crystallography provides only a single representation of protein structures (Slabinski *et al.*, 2007). Although this limitation is overcome in nuclear magnetic resonance spectroscopy, which allows for the examination of dynamic protein molecules, nuclear magnetic resonance spectroscopy requires large quantities of highly concentrated protein samples, which is not always feasible or possible (Slabinski *et al.*, 2007). Furthermore, the size limit of nuclear magnetic resonance spectroscopy does not usually allow for determining the structures of proteins larger than 25 kDa (Frueh *et al.*, 2013). Thus, *in silico* methods of three-dimensional (3D) protein structure prediction are useful for estimating protein structures, and the improvement in structure prediction tools over the years has allowed for predicting accurate models of protein structures (Deng *et al.*, 2018; Srivastava *et al.*, 2018).

I-TASSER (Iterative Threading Assembly Refinement) is an online platform for predictions of global 3D protein structure (Roy *et al.,* 2010; Yang and Zhang, 2015; Yang *et al.,* 2015). Since 2006, I-TASSER has been recognised as the most accurate tool for 3D protein structure prediction in all CASP experiments (Croll *et al.*, 2019; Kryshtafovych *et al.*, 2018; Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). I-TASSER predicts protein structure using a three-step method. First, template proteins with similar

super-secondary structures or homologous structure templates are identified from the Protein Data Bank (PDB), based on the amino acid sequence submitted to I-TASSER (Yang and Zhang, 2015). The identification is achieved using a meta-threading-server algorithm that combines multiple individual threading servers (Wu and Zhang, 2007). In the second step, the relevant fragments from the templates are used to create full-length models of the protein through the assembly of the fragments based on alignments with segments of the amino acid sequence (Yang and Zhang, 2015). Any missing segments are modelled *ab initio* by Monte Carlo simulations (Yang and Zhang, 2015; Zhang *et al.*, 2003). Theoretically, a protein adopts the fold corresponding with the lowest Gibbs free energy in its native state (Zhang and Skolnick, 2004). However, during 3D protein structure prediction, the lowest Gibbs free energy state is not necessarily the most similar to the native protein fold due to limitations of the force field (Zhang and Skolnick, 2004). Nevertheless, the Gibbs free energy states most similar to the native protein fold can be identified as these are usually the most populated (Zhang and Skolnick, 2004). Based on this, in the third step, the optimal Gibbs free energy states are identified and each of the highly populated Gibbs free energy states is clustered (Yang and Zhang, 2015). The second step is then repeated on the clusters to refine the models (Yang and Zhang, 2015).

The resulting structures then undergo further refinement by two algorithms. The first refinement method is the Fragment Guided Molecular Dynamics algorithm which improves local structural distortions such as steric clashes or poor backbone torsion angles arising due to the assembly of multiple template fragments (Zhang *et al.*, 2011a). The second refinement method is the ModRefiner algorithm which improves both local and global protein structure by first improving the backbone topology and then improving side-chain conformations (Xu and Zhang, 2011). The five best clusters with the lowest final Gibbs free energy states are selected and averaged, resulting in five predicted 3D protein structures (Yang and Zhang, 2015).

Similarly to PSIPRED 4.0 and DISOPRED3, I-TASSER depends on multiple sequence alignments in order to predict 3D protein structures. Therefore, a limitation of this method when studying the effects of mutations on a protein, is that the mutants could result in the same multiple sequence alignment as that of the wild-type. This would reduce the sensitivity of the predictions to the effects of the mutations. Additionally, I-TASSER can only predict the 3D structures of monomeric proteins, therefore I-TASSER predictions cannot be used to assess the oligomeric state of a protein, or the effects of a mutation on oligomerisation. Nevertheless, predictions are useful in obtaining an overview of the 3D structures of the proteins. Furthermore, predictions could reveal any local changes in structure, occurring in close proximity to the sites of mutation.

The online I-TASSER server was used to predict the 3D structures of the FOXP2 (LeuZip - end) proteins based on the amino acid sequence of each protein (Roy *et al.,* 2010; Yang and Zhang, 2015; Yang *et al.,* 2015). No restraints, templates, or secondary structures were specified. I-TASSER returned five models for each of the FOXP2 (LeuZip - end) proteins. Each model was ranked with a C-score indicating the confidence or quality of the model, based on the threading template alignments and assembly of the structure (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015).

Although the I-TASSER prediction of 3D protein structures includes refinement steps, since I-TASSER uses multiple templates rather than a single template, it is common for the final models to have unfavourable backbone torsion angles or other local structural distortions, especially when the protein does not have very similar homologous templates available (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). The local structural distortions can increase the overall Gibbs free energy of the structures, which may be improved through energy minimisation. All five models of wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) were energy minimised using Maestro 12.0 (Maestro, Schrödinger, LLC, New York, NY, 2019). The Protein Preparation Wizard was used to prepare each model prior to energy minimisation. Each model was preprocessed with all hydrogens displayed and all other parameters set to default (Epik, Schrödinger, LLC, New York, NY, 2016; Madhavi Sastry *et al.,* 2013). The Protein Preparation Wizard was then used to optimise hydrogen bonds with hydrogen minimisation of altered species (Madhavi Sastry *et al.,* 2013). A preliminary restrained minimisation was conducted using the Protein Preparation Wizard and the OPLS3e (Optimized Potential for Liquid Simulations 3e) force field using default parameters (Harder *et al.,* 2016; Jorgensen and Tirado-Rives, 1988; Jorgensen *et al.,* 1996; Madhavi Sastry *et al.,* 2013; Shivakumar *et al.,* 2010). Thereafter, energy minimisation was conducted using the Minimization MacroModel and the OPLS3e force field (Harder *et al.,* 2016; Jorgensen and Tirado-Rives, 1988; Jorgensen *et al.,* 1996; MacroModel, Schrödinger, LLC, New York, NY, 2019; Shivakumar *et al.,* 2010). All default parameters were used.

MolProbity is a useful online server for validating protein and other macromolecular structures (Williams *et al.*, 2018). MolProbity evaluates both local and global protein structure with respect to favourable rotamers, backbone torsion angles, the geometry of the amino acid side chains relative to the protein backbone and other covalent geometries, and the positions of all atoms in the structure (Williams *et al.*, 2018). The energy minimised models were analysed and validated using the MolProbity server (Williams *et al.,* 2018). Structure files were uploaded and analyses were conducted on the trimmed files with hydrogens added, using nuclear bond lengths (Williams *et al.*, 2018). Any suggested Asn/Gln/His flips were accepted and programs to analyse local and global protein structure

were run (Williams *et al.,* 2018). Taking the I-TASSER C-scores into consideration, the results of the MolProbity analyses were used to select the best of the five models for each FOXP2 (LeuZip - end) protein. The best model of each FOXP2 (LeuZip - end) variant (wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end)) was used for comparisons of the FOXP2 (LeuZip - end) variants to each other. These models were also used for comparisons to data from *in vitro* experiments.

2.2 PROTEIN PREPARATION

2.2.1 PLASMID

The FOXP2 (LeuZip - end) proteins used in this study comprise position 372 to position 715 of the full-length FOXP2 amino acid sequence, and includes the LeuZip and FHD (Figure 2.1). The glutamine-rich region and zinc finger domain were excluded since glutamine-rich regions cause protein aggregation, and zinc finger proteins require the addition of zinc to the growth medium during overexpression, resulting in lowered protein yields due to the toxic effect of zinc on *Escherichia coli* (*E. coli*) cells (Barton *et al.*, 2007; Perutz *et al.*, 1994; Yao *et al.*, 2005). A poly-histidine tag (His-tag) was incorporated at the C-terminal end of the FOXP2 (LeuZip - end) amino acid sequence, to facilitate downstream protein purification.

```
                                                    HMASMTGGQQ
372         ALDDRSTAQCRVQMQVVQQLEIQLSKERERLQAMMTHLHMRPS
415         EPKPSPKPLNLVSSVTMSKNMLETSPQSLPQTPTTPTAPVTPI
458         TQGPSVITPASVPNVGAIRRRHSDKYNIPMSSEIAPNYEFYKN
501         ADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRR
544         NAATWKNAVRHNLSLHKCFVRVENVKGAVWTVDEVEYQKRRSQ
587         KITGSPTLVKNIPTSLGYGAALNASLQAALAESSLPLLSNPGL
630         INNASSGLLQAVHEDLNGSLDHIDSNGNSSPGCSPQPHIHSIH
673         VKEEPVIAEDEDCPMSLVTTANHSPELEDDREIEEEPLSEDLE
            SGLVPR|GSHHHHHH
```

Figure 2.1: Wild-type FOXP2 (LeuZip - end) amino acid sequence. The FOXP2 (LeuZip - end) sequence comprises position 372 to position 715 of the full-length FOXP2 amino acid sequence. Amino acid positions are numbered relative to the full-length protein (left-hand side). Sites of mutation for this study (Arg401 and Ala539) are indicated in boldface. Regions are indicated as highlighted: residues in grey are additional residues for improved protein expression, residues in rose comprise the LeuZip domain, residues in green comprise the FHD, residues in yellow are the thrombin cleavage site for removal of the His-tag if necessary, with the vertical line indicating the point of cleavage, and residues in blue are the His-tag.

The gene sequence encoding FOXP2 (LeuZip - end) was cloned into a pET-11a plasmid, with sequence codons optimised for expression in *E. coli* cells (GenScript Biotech Corporation, USA). The pET-11a plasmid multiple cloning site was under the control of a T7 promoter. An ampicillin resistance gene allowed for selection of positive transformants.

2.2.2 MUTAGENESIS

Appropriate primers were designed for the R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) mutations (Table 2.1), with the lowest number of base changes possible, and synthesised by Inqaba Biotechnical Industries (ZA). The R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) mutants were generated from wild-type FOXP2 (LeuZip - end) using the QuikChange Lightning Site-Directed Mutagenesis Kit, according to the manual (Agilent Technologies, USA). Briefly, 50 ng of template DNA was combined with 100 ng each of the forward and reverse primers and a small amount of a deoxyribonucleotide triphosphate (dNTP) mix. QuikChange Lightning Enzyme, QuikChange Lightning Buffer and QuikSolution reagent were added according to the manual. After the polymerase chain reaction (PCR) was complete, the PCR product was incubated with *Dpn* I restriction enzyme for 45 minutes at 37 °C, for digestion of remaining parental template DNA. Sanger DNA sequencing confirmed successful mutagenesis (Inqaba Biotechnical Industries, ZA).

Table 2.1: Primers for R401D and A539P mutations of FOXP2 (LeuZip - end). Incorporation of the R401D mutation required a double base change, which was achieved by mutation of one base at a time. The A539P mutation was introduced to R401D FOXP2 (LeuZip - end) to create the R401D/A539P mutant.

Mutation	**Primer sequence (mismatched base underlined)**	**Original base at site of mutation**
R401D (first mutation)	5’-GTAAAGAACGCGAA<u>G</u>GCCTGCAGGCAATGATG-3’	C
R401D (second mutation)	5’-GTAAAGAACGCGAAG<u>A</u>CCTGCAGGCAATGATG-3’	G
A539P	5’-GGTTTACCCGTACGTTT<u>C</u>CGTACTTCCGCCGTAATG-3’	G

2.2.3 TRANSFORMATION

The pET-11a plasmids encoding all FOXP2 (LeuZip - end) variants were used to transform competent *E. coli* cells for heterologous protein expression. Plasmids encoding wild-type and R401D FOXP2 (LeuZip - end) were used to transform T7 Express pLysS Competent *E. coli* cells, and plasmids encoding A539P and R401D/A539P FOXP2 (LeuZip - end) were used to transform BL21 (DE3) Competent *E. coli* cells. The *E. coli* cell strains were chosen based on trials that produced the highest yields of soluble proteins in these respective strains.

Transformation was initiated by incubating 300 ng of each plasmid with 100 µl of competent *E. coli* cells on ice for 30 minutes. Thereafter, the cells were heat shocked for 42 seconds at 42 °C and immediately incubated on ice for 5 minutes. Sterile Super Optimal broth with Catabolite repression (SOC) (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, and 20 mM glucose) was added to the cells, which were then incubated for 1 hour at 37 °C with shaking. The transformation reactions were spread onto sterile LB (lysogeny broth) agar spread plates consisting of LB (1% tryptone, 0.5% yeast extract, and 1% NaCl) with 2% bacteriological agar and 100 µg/ml ampicillin for selection of positive transformants. Spread plates lacking ampicillin, of SOC without cells, and of cells lacking plasmids were prepared as controls. All plates including control plates were incubated for 16 hours at 37 °C. Single colonies were selected and incubated in fresh LB with 100 µg/ml ampicillin at 37 °C, and grown with shaking to an optical density at 600 nm (OD_{600}) of 0.6. The cultures were used to prepare stocks by adding 50% sterile glycerol, and stocks were stored at - 80 °C.

2.2.4 PROTEIN EXPRESSION

One of the most popular methods of recombinant protein expression is by means of the *E. coli* bacterium. This Gram-negative bacterium can be easily transformed with plasmid DNA and cultured to reproduce in a short period of time, on inexpensive media with good protein yields, making it ideal for heterologous protein expression (Baneyx, 1999). In this manner, any protein can be produced specifically for a study in the form required, and isolation from organic sources is not necessary.

Following transformation of competent *E. coli* cells (see 2.2.3 Transformation) the FOXP2 (LeuZip - end) proteins were produced by heterologous protein expression. Stocks containing transformed cells in 50% sterile glycerol were used to inoculate 50 ml sterile 2 x yeast extract-tryptone (YT) medium (1.6% tryptone, 1% yeast extract, and 0.5% NaCl) with 100 µg/ml ampicillin. The cells were grown for 16 hours at 37 °C with shaking. This was used to inoculate 1.5 - 2 L fresh sterile 2 x YT

medium at a 1:50 dilution. The cells were grown in batches of 100 ml each at 37 °C with shaking to an OD_{600} of 0.6.

Protein overexpression was induced using isopropyl-1-thio-β-D-galactopyranoside (IPTG). Overexpression of wild-type FOXP2 (LeuZip - end) was conducted as previously optimised (Thulo, 2019). Trials were conducted to determine the optimal IPTG concentrations, temperatures and durations of overexpression of R401D, A539P and R401D/A539P FOXP2 (LeuZip - end), to produce the highest yields of soluble proteins (Table 2.2).

Table 2.2: Optimal overexpression conditions of R401D, A539P and R401D/A539P FOXP2 (LeuZip - end). Trials were conducted to determine the optimal IPTG concentration, temperature and duration of overexpression of each FOXP2 (LeuZip - end) mutant, to produce the highest yield of soluble protein.

	[IPTG] (mM)	Temperature (°C)	Duration (hours)
R401D FOXP2 (LeuZip - end)	0.2	20	5
A539P FOXP2 (LeuZip - end)	0.2	20	22
R401D/A539P FOXP2 (LeuZip - end)	0.2	20	22

Cells were harvested by centrifugation for 15 minutes at 5000 x g and 4 °C, and pellets were resuspended in immobilised metal-affinity chromatography (IMAC) equilibration buffer (20 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-Cl), pH 7.5, 500 mM NaCl, and 50 mM imidazole) containing 1 mg/ml chicken egg white lysozyme for lysis and 1 mM phenylmethylsulfonyl fluoride (PMSF) for the prevention of proteolytic degradation. Cell lysates were stored at - 20 °C.

2.2.5 PROTEIN PURIFICATION

Protein chromatographic techniques employ the use of a column that is packed with an insoluble packing material or resin, to separate components of a protein mixture based on a characteristic property (Leonard, 1997). The protein sample mixture is dissolved in a liquid solvent and migrates through the column (Leonard, 1997). A detector measures a quantitative signal upon elution (Leonard, 1997). The properties of the packing material or resin determine the basis of separation; for example, protein mixtures can be separated based on net charge, size, or the affinity for a specific ligand or compound (Leonard, 1997). Protein chromatographic techniques are thus useful analytical techniques for the separation and quantification of proteins of a mixture (Leonard, 1997).

Affinity chromatography is a high-precision isolation procedure that relies on the specific binding of a ligand to a protein. The ligand is fixed to the stationary phase matrix (Cuatrecasas *et al.,* 1968). Therefore, the protein binds to the ligand and is adsorbed to the stationary phase, while all other components of the mixture elute through the column (Cuatrecasas *et al.,* 1968). The column, with the target protein bound to the resin, is then subjected to a change (such as a change in pH or ionic strength) that decreases the affinity of the protein for the ligand, facilitating elution and collection of the isolated protein (Cuatrecasas *et al.,* 1968).

Histidine residues are able to form coordination complexes with divalent cations, particularly nickel (II) and cobalt (II), due to the aromatic imidazole histidine side chain which has a strong affinity for transition metal ions (Sundberg and Martin, 1974). The incorporation of a His-tag on a protein of interest allows for the divalent-cation affinity of histidine to be manipulated for protein isolation and purification. The divalent cation is immobilised on the column stationary phase, thus protein molecules bind to the divalent cation via His-tags (Porath *et al.,* 1975). This is referred to as IMAC. The protein of interest is then eluted by a change in environment or the addition of a competitive metal ion-binder (Porath *et al.,* 1975).

In order to isolate protein following heterologous expression it is necessary to ensure that complete cell lysis has been achieved. Freezing and subsequent thawing of cell lysates as well as the addition of chicken egg white lysozyme all contributed to cell lysis. Once the cell lysates were thawed completely, a sonication step was included to further lyse the cells. Sonication was conducted using 5 - 10 rounds of 20 second pulses, with a 20 second incubation on ice following each pulse to protect proteins from denaturation due to heat build-up. Prior to purification, 1 mM DNase was added to the lysates to remove unwanted DNA contamination and improve lysate viscosities. Insoluble cell debris was separated from soluble fractions by centrifugation for 20 minutes at 18 000 x g and 4 °C.

The FOXP2 (LeuZip - end) proteins were isolated from the soluble fractions by IMAC using an ÄKTA Fast Protein Liquid Chromatography (FPLC) system (GE Healthcare Life Sciences, UK) which was possible due to the incorporation of C-terminal protein His-tags (Figure 2.1). IMAC was conducted using 5 ml HisTrap™ HP columns (GE Healthcare Life Sciences, UK). Columns used for purification of wild-type and R401D/A539P FOXP2 (LeuZip - end) were pre-charged with cobalt (II) and columns used for purification of R401D and A539P FOXP2 (LeuZip - end) were pre-charged with nickel (II). The metal ions were chosen based on trials that optimised for the largest protein yields and highest levels of protein purity.

All steps were conducted at a flow rate of 5 ml/minute. Soluble lysate fractions were loaded onto the columns that were pre-equilibrated with IMAC equilibration buffer (previously described). Purification of all FOXP2 (LeuZip - end) variants required a wash step with a high concentration of salt to remove DNA contamination, and a wash step with detergent to remove the high levels of unwanted protein contaminants. This was performed as a combined salt/detergent wash step. Therefore, once unbound contaminant proteins had eluted completely, 10 column volumes of a salt/detergent wash (20 mM Tris-Cl, pH 7.5, 1.5 M NaCl, 50 mM imidazole, 1% Triton™ X-100, and 0.5% Tween® 20) were used. The column was washed extensively with IMAC equilibration buffer to remove all traces of detergent, then washed with 10 column volumes of an imidazole wash (20 mM Tris-Cl, pH 7.5, 500 mM NaCl, and 100 mM imidazole) to further remove protein contaminants. During the imidazole wash step, the concentration of imidazole was maintained below the required concentration for elution, to avoid decreases in protein yields. The imidazole wash step was not included for purification of R401D/A539P FOXP2 (LeuZip - end) to avoid lowered protein yields, since this mutant bound to the column more weakly than the other FOXP2 (LeuZip - end) variants. Finally, the FOXP2 (LeuZip - end) proteins were eluted using elution buffer containing a high concentration of imidazole (20 mM Tris-Cl, pH 7.5, 500 mM NaCl, and 500 mM imidazole). Pure protein solutions were stored at 4 °C.

2.2.6 Assessment of protein purity and concentration

In order to obtain reliable information and ensure that all conclusions pertain to the protein of interest, it is important to ensure that a protein sample is pure prior to conducting downstream studies. One of the most widely used methods for assessing protein purity is sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE separates individual proteins of a mixture under the influence of an electric field on a polyacrylamide gel (Laemmli, 1970). The polyacrylamide gel matrix has a sieving effect which enables the separation of the protein molecules (Laemmli, 1970). The anionic detergent, SDS, is added to the polyacrylamide gel and protein samples, to denature and linearise each protein molecule, and confer a negative charge to the linearised protein molecule (Laemmli, 1970; Roy and Kumar, 2014). The negative charge is proportional to the length of the protein molecule, therefore the speed of migration through the gel is influenced by the size of the molecule (Roy and Kumar, 2014). Once the samples have been resolved, protein bands are visualised typically by staining with Coomassie dyes (Laemmli, 1970). If a pure protein sample was applied to the gel, it is expected to appear as a single band once stained.

The purity of each FOXP2 (LeuZip - end) protein was assessed by glycine SDS-PAGE (Laemmli, 1970). Protein samples were resolved under reducing conditions, using discontinuous electrophoresis with a 4% stacking gel and a 12.5% resolving gel. Protein molecular weights were determined relative to the migration of the proteins in the 3-Color Prestained Protein Marker (BLIRT, Poland). Gels were stained with Coomassie Brilliant Blue R-250. A Molecular Imager® Gel Doc™ XR System (Bio-Rad Laboratories, USA) was used to conduct densitometric evaluations of the gels for a quantitative assessment of protein purity.

In solution, proteins absorb ultraviolet (UV) light at 280 nm (Schmid, 2001). Protein absorbance at 280 nm (A_{280}) is due to tryptophan, tyrosine and, to a lesser extent, cystines (Schmid, 2001). DNA and RNA absorb UV light at 260 nm (A_{260}), due to the pyrimidine and purine rings of the nucleic acid bases (Schmid, 2001). Protonation of DNA and RNA bases affects the absorbance of the nucleic acids, therefore the absorbance is pH-sensitive (Schmid, 2001). Disulfide bonds between cysteine residues are also able to absorb light at 260 nm, which should be taken into consideration when interpreting protein absorbance spectra (Schmid, 2001).

Protein purity with respect to DNA and RNA contamination was assessed using a NanoDrop™ One (Thermo Fisher Scientific, USA) for absorbance spectroscopy, with the aim of obtaining the highest A_{280}/A_{260} ratio. This was used to ensure that the samples were sufficiently free of nucleic acid contamination, which was especially important for DNA-binding studies.

Protein sample concentrations were assessed by absorbance spectroscopy, specifically at a wavelength of 280 nm. Absorbance data obtained using a Jasco V-630 Spectrophotometer (JASCO, USA) were compared to those obtained using a NanoDrop™ One (Thermo Fisher Scientific, USA) for each FOXP2 (LeuZip - end) variant and showed good agreement. Protein yields were low, especially for R401D/A539P FOXP2 (LeuZip - end). Therefore, subsequent absorbance measurements were conducted using the NanoDrop™ One (Thermo Fisher Scientific, USA) due to the smaller quantity of sample required to conduct measurements on this instrument. All measurements were conducted in triplicate and averages were used to determine the final molar concentrations based on the Beer-Lambert law. The molar absorptivity coefficients used were 28 670 $M^{-1}{\cdot}cm^{-1}$ under non-reducing conditions or 28 420 $M^{-1}{\cdot}cm^{-1}$ under reducing conditions, for all FOXP2 (LeuZip - end) variants, as predicted by the Expert Protein Analysis System (ExPASy) ProtParam tool (Gasteiger *et al.,* 2005). Absorbance spectroscopy at a wavelength of 340 nm was used to ensure that no protein aggregates were present in any samples, and to correct for the baseline of absorbance spectra.

2.3 DNA-BINDING STUDIES

2.3.1 COGNATE DNA SEQUENCE

The DNA sequence to which the FOXP2 DNA-binding FHD binds most favourably is 5'-TGTTTAC-3' (Nelson *et al.*, 2013; Webb *et al.*, 2017). All experiments involving DNA binding were conducted using 12.23 kDa double-stranded oligonucleotides containing this binding sequence flanked by random sequences. The exact DNA sequence used was 5'-TTAGGTGTTTACTTTCATAG-3' (with the region to which the FOXP2 FHD binds most favourably underlined). Binding of the isolated FOXP2 FHD to this DNA sequence has been well characterised (Morris and Fanucchi, 2016). DNA oligonucleotides were synthesised by Integrated DNA Technologies (USA). Lyophilised DNA was solubilised in ultrapure deionised distilled water and stored at - 20 °C. DNA stock concentrations were assessed by absorbance spectroscopy at a wavelength of 260 nm, using a NanoDrop™ One (Thermo Fisher Scientific, USA). Measurements were conducted in triplicate and final molar concentrations were determined from the average. Absorbance spectroscopy at a wavelength of 280 nm was used to ensure that no protein contaminants were present in any DNA stocks.

2.3.2 ELECTROPHORETIC MOBILITY SHIFT ASSAYS

An electrophoretic mobility shift assay (EMSA) is the separation of a DNA-protein mixture under the influence of an electric field on a polyacrylamide or agarose gel (Fried, 1989). The speed of migration through the gel is predominantly influenced by the size of the molecule, so DNA-protein complexes migrate more slowly than either free DNA or free protein (Fried, 1989). Thus, bands representing DNA-protein complexes appear 'shifted' compared to bands representing unbound DNA or unbound protein (Fried, 1989). The charge of the molecule affects its migration, so all molecules are typically given an even charge-to-mass distribution. Due to hydrodynamic effects, the shape of the molecule can also influence its migration through the gel, albeit to a lesser extent (Fried, 1989). Compact structures are expected to migrate through the gel more quickly than relaxed structures.

EMSAs conducted using the Bio-Rad Mini-PROTEAN® Tetra Vertical Cell (Bio-Rad Laboratories, USA) were used routinely to confirm that wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) were all folded correctly, and hence able to bind DNA. DNA binding was observed from these EMSAs but the EMSAs were not large enough to resolve more than one DNA-protein complex. Therefore, the larger Bio-Rad PROTEAN® II xi Cell (Bio-Rad Laboratories, USA) was used to conduct EMSAs to compare the number of DNA-protein complexes formed for each FOXP2 (LeuZip - end) variant. This, combined with size exclusion chromatography data (see 2.4.3 Size Exclusion Chromatography), gave an

indication of the number of oligomeric species present for each of the FOXP2 (LeuZip - end) proteins and whether the number of species differed between the DNA-bound and unbound states.

Tris-Cl/borate/ethylenediaminetetraacetic acid (EDTA) buffer, or TBE buffer, was prepared (89 mM Tris-Cl, pH 8, 89 mM boric acid, and 2 mM $Na_2EDTA{\cdot}H_2O$). Samples containing DNA and protein were resolved on a 20% triethylene glycol/8% acrylamide/1 x TBE gel. Samples were prepared in EMSA buffer (10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.5, 100 mM KCl, 1 mM $MgCl_2$, 10% glycerol, and 0.1 mg/ml bovine serum albumin). Samples included free DNA and mixtures of DNA and protein at varying concentrations (Table 2.3). Experiments were conducted in TBE buffer at 160 V and 4 °C using gels, samples, and TBE buffer that were all pre-equilibrated to 4 °C. Durations of the EMSAs were optimised for each protein, and were varied (Table 2.3). Gels were stained with SYBR® Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific, USA) and visualised immediately under a UV lamp using a Molecular Imager® Gel Doc™ XR System (Bio-Rad Laboratories, USA).

EMSAs comparing non-reducing and reducing conditions were only conducted for those FOXP2 (LeuZip - end) variants that showed differences in the numbers or sizes of oligomers present between non-reducing and reducing conditions in size exclusion chromatography. EMSAs conducted under reducing conditions followed the same protocol outlined above but samples were prepared at 2 mM dithiothreitol (DTT).

Table 2.3: Comparison of sample compositions and durations of EMSAs conducted using the Bio-Rad Mini-PROTEAN® Tetra Vertical Cell and the Bio-Rad PROTEAN® II xi Cell. Due to the difference in size between the two cells, the Bio-Rad Mini-PROTEAN® Tetra Vertical Cell (Bio-Rad Laboratories, USA) was used to conduct routine EMSAs confirming that the proteins were able to bind to DNA, whereas the Bio-Rad PROTEAN® II xi Cell (Bio-Rad Laboratories, USA) was used to conduct EMSAs assessing the number of DNA-protein complexes formed for each protein.

		Bio-Rad Mini-PROTEAN® Tetra Vertical Cell	**Bio-Rad PROTEAN® II xi Cell**
Concentration of DNA (µM)		0.2	0.5
Range of DNA:protein ratios		3:1 to 1:30	3:1 to 1:35
Duration (hours:minutes)	Wild-type FOXP2 (LeuZip - end)	03:00	08:05
	R401D FOXP2 (LeuZip - end)	02:45	10:00 (non-reducing conditions) 07:37 (reducing conditions)
	A539P FOXP2 (LeuZip - end)	02:00	06:34
	R401D/A539P FOXP2 (LeuZip - end)	02:00	07:24 (non-reducing conditions) 08:13 (reducing conditions)

2.3.3 FLUORESCENCE ANISOTROPY DNA-BINDING ASSAYS

Fluorescence is the phenomenon of the emission of light by a molecule (referred to as the fluorochrome) after excitation by electromagnetic radiation (Lakowicz, 2006a). The fluorochrome absorbs light or electromagnetic radiation, causing excitation of electrons (Lakowicz, 2006a). Excited electrons then return to the ground state, releasing electromagnetic radiation of a longer wavelength (or lower energy) than that of the absorbed radiation (Lakowicz, 2006a). The emission is influenced by both the fluorochrome itself and the environment of the fluorochrome (Lakowicz, 2006a). Additionally, the emission is influenced by the polarisation of the excitation light (Favicchio *et al.*, 2009). Thus, if a fluorochrome is excited with plane polarised light, it typically emits light of a longer wavelength that is also plane polarised (Favicchio *et al.*, 2009).

Molecules or particles are constantly in motion, due to interactions with each other and with the environment (Jovin *et al.*, 1981). One such motion is rotational diffusion, in which the orientation of molecules constantly fluctuates, in order to maintain an overall equilibrium of the orientation of all molecules in the system (Bashardanesh *et al.*, 2019; Favicchio *et al.*, 2009; Jovin *et al.*, 1981). The smaller the molecule, the more quickly it reorients itself and therefore the more quickly it rotates (Favicchio *et al.*, 2009; Lakowicz, 2006a; Mukherjee and Bagchi, 2006).

Due to rotational diffusion, fluorochromes have the inherent ability to emit plane polarised light in a different plane from the plane of absorbed light (Favicchio *et al.*, 2009). This is a phenomenon known as fluorescence anisotropy (Favicchio *et al.,* 2009). Fluorescence anisotropy can be detected by the use of a polarised filter placed before the detector (Lakowicz, 2006a). If a fluorochrome is excited with vertically polarised light the extent or degree of depolarisation may be determined relative to the incident light according to the equation;

$$r = \frac{I_{\mathrm{v}} - I_{\mathrm{h}}}{I_{\mathrm{v}} - 2\,I_{\mathrm{h}}}$$

where r is the anisotropy, and I_{v} and I_{h} are the intensities of vertically and horizontally polarised light, respectively (Chen and Van Der Meer, 1993; Favicchio *et al.,* 2009). As an excited fluorochrome undergoes rotational diffusion, the emission is depolarised (Favicchio *et al.*, 2009). The faster the rotational diffusion, the greater the depolarisation and hence the lower the anisotropy (Favicchio *et al.*, 2009). It follows that a smaller fluorochrome causes a lower anisotropy, whereas a larger fluorochrome causes a higher anisotropy (Favicchio *et al.,* 2009). This is useful when studying DNA-protein interactions since the DNA-protein complex is larger than either the unbound DNA or protein, resulting in an increased anisotropy upon DNA binding (Favicchio *et al.,* 2009). Usually, when conducting DNA-binding assays using fluorescence anisotropy, fluorescently labelled DNA is used to enhance the signal between bound and unbound states since the DNA is smaller than the protein (Favicchio *et al.,* 2009). The fluorescence anisotropy of the DNA is then monitored at increasing protein concentrations, and as the concentration of protein increases, the fluorescence anisotropy increases (Favicchio *et al.*, 2009).

Fluorescence anisotropy assays were used to obtain a dissociation constant (K_{D}) for the interaction of each FOXP2 (LeuZip - end) protein with DNA, to study the effects of the R401D, A539P and R401D/A539P mutations on DNA-binding affinity. This was especially relevant since previous studies have shown that homo-oligomerisation is a regulator of FOXP2 DNA binding, therefore the individual and combined contributions of the LeuZip and FHD interfaces to DNA binding were assessed (Li *et al.,*

2007, 2004). The cognate DNA sequence (see 2.3.1 Cognate DNA sequence) was labelled with 5-carboxy-X-rhodamine at the 5' end. Fluorescence anisotropy was conducted using a Perkin Elmer LS-50B Luminescence Spectrometer (Perkin Elmer, USA). Samples were prepared in DNA-binding buffer (10 mM HEPES, pH 7.5, and 100 mM NaCl). Samples included free DNA, and mixtures of DNA and protein at DNA:protein ratios ranging from 2:1 to 1:30, at a constant DNA concentration of 200 nM. Experiments were conducted at 20 °C with an excitation wavelength of 580 nm, an emission wavelength of 605 nm, an integration time of 3 seconds, and 5 nm slits. All samples were prepared in triplicate and data points were averaged for 10 accumulations. Final averaged data points were corrected relative to the final averaged anisotropy value for free DNA, and used to plot DNA-binding isotherms. Errors were calculated as the standard deviations of the averages. DNA-binding isotherms were fit using the equation for a single site saturation, and dissociation constants were determined based on the equations of the fits. Dissociation constants were analysed using two-sample t-tests to determine the significance of observed differences.

Dissociation constants from the DNA-binding assays were compared with those obtained previously for the isolated FOXP2 DNA-binding FHD (Morris and Fanucchi, 2016). Therefore, the assays were conducted under the same conditions for an accurate comparison. The assays were then repeated under reducing conditions to compare to those conducted under non-reducing conditions, since other experimental work was conducted under reducing conditions. This was used to ensure that the affinities of R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) for DNA relative to the affinity of wild-type FOXP2 (LeuZip - end) for DNA were comparable under non-reducing and reducing conditions. Fluorescence anisotropy DNA-binding assays conducted under reducing conditions followed the same protocol outlined above, but samples were prepared at 2 mM DTT. Data points were averaged for 10 accumulations. Final averaged data points were corrected relative to the final averaged anisotropy value for free DNA, and used to plot DNA-binding isotherms. DNA-binding isotherms were fit using the equation for a single site saturation, and dissociation constants were determined based on the equations of the fits.

2.4 CHARACTERISATION OF PROTEIN STRUCTURE AND STABILITY

2.4.1 CIRCULAR DICHROISM SPECTROPOLARIMETRY

Circular dichroism (CD) is the difference in absorption of left- and right-handed circularly polarised light by a molecule containing a chiral centre. CD spectropolarimetry measures the CD of molecules over a range of wavelengths. This is particularly useful when studying protein structure, since the twenty standard amino acids are chiral molecules (all except glycine) and confer chirality to the peptide bonds of protein backbones. In the far-UV range of wavelengths, there are two electronic transitions of peptide bonds when interacting with circularly polarised light; the *n* to π^* transition which occurs at longer wavelengths (215 - 230 nm) and the π to π^* transition which occurs at shorter wavelengths (185 - 200 nm) (Woody and Koslowski, 2002). The alignment of peptide bonds in regular arrangements produces distinct bands and, in some cases, the bands are split due to exciton interactions (Greenfield, 2006a; Woody and Koslowski, 2002). As a result, CD spectra in the far-UV range of wavelengths produce patterns specific to protein secondary structures, such as α-helices (Hashizume *et al.,* 1967). The typical far-UV CD spectrum of a predominantly α-helical protein displays negative bands at approximately 208 nm and 222 nm, and a positive band near 190 nm (Greenfield, 2006a; Woody and Koslowski, 2002). The typical far-UV CD spectrum of a predominantly β-stranded protein displays a negative band at around 218 nm and a positive band close to 195 nm (Greenfield, 2006a; Woody and Koslowski, 2002). Disordered proteins usually exhibit negative bands at approximately 195 nm, though there is variation in the far-UV CD spectra of disordered proteins (Greenfield, 2006a; Uversky, 2002a). A combination of these secondary structural elements results in a CD spectrum that is a cross between these characteristic spectra. The measured signal is recorded in millidegrees and normalised to mean residue ellipticity (MRE) in deg·cm^2·dmol^{-1} using the equation;

$$\theta_{MRE} = \frac{\theta\ (\mathrm{mdeg}) \times 10^6}{\mathrm{pathlength}\ (\mathrm{mm}) \times [\mathrm{protein}]\ (\mu\mathrm{M}) \times n}$$

where *n* is the number of peptide bonds in the protein chain.

Far-UV CD spectropolarimetry was used to evaluate the secondary structures of the FOXP2 (LeuZip - end) proteins, in the absence and presence of DNA. This was used to determine if DNA binding or if the R401D, A539P or R401D/A539P mutations caused any changes to the secondary structure of FOXP2 (LeuZip - end). Spectra were recorded on a Jasco J-810 Spectropolarimeter (JASCO, USA). The FOXP2 (LeuZip - end) variants were dialysed against spectroscopy buffer (10 mM HEPES, pH 7.5, 50 mM Na_2SO_4). Samples consisted of 5 µM protein, with 2 mM DTT for experiments conducted under reducing conditions. Samples with DNA were of the same composition with the

addition of 2.5 µM DNA. Samples were prepared at a DNA:protein concentration ratio that was determined based on the dissociation constants (see 2.3.3 Fluorescence anisotropy DNA-binding assays), with higher concentrations of DNA-protein complexes than the concentrations of free proteins (or free DNA) at this ratio. Samples were prepared immediately before recording spectra by diluting high concentration protein and DNA stocks with ultrapure deionised distilled water at a ratio of 1:10, to reduce noise due to buffer components. Spectra were recorded from 180 nm to 250 nm using bandwidths of 1 nm, a speed of 200 nm/minute, and a pathlength of 2 mm, at 20 °C. All samples were prepared in triplicate and spectra were averaged for 5 accumulations. Final averaged spectra were corrected using blanks consisting of buffer only or buffer with 2.5 µM DNA for samples containing DNA, and with 2 mM DTT for samples prepared under reducing conditions. Noisy data points with a high tension (voltage) of over 600 V were removed.

The CD spectra were analysed using the online DichroWeb server (Whitmore and Wallace, 2004, 2008). The CONTINLL analysis algorithm was used to provide quantitative information regarding the protein secondary structures with respect to the content of specific structural elements (Provencher and Glockner, 1981; van Stokkum *et al.,* 1990). Reference set 7 was used, which is optimised for the 190 nm to 240 nm range of wavelengths (Sreerama and Woody, 2000). The secondary structural elements of the FOXP2 (LeuZip - end) variants were analysed using two-sample t-tests to determine the significance of differences observed between the variants, and using paired t-tests to determine the significance of differences observed upon DNA binding of each variant.

Data from CD spectropolarimetry was compared with that obtained from protein secondary structure predictions (see 2.1.1 Protein secondary structure and disorder predictions) and from global protein structure predictions (see 2.1.2 Three-dimensional protein structure predictions).

2.4.2 INTRINSIC TRYPTOPHAN FLUORESCENCE SPECTROSCOPY

Protein molecules exhibit intrinsic fluorescence, predominantly due to tryptophan residues (see 2.3.3 Fluorescence anisotropy DNA-binding assays for a description of fluorescence) (Ostashevskii, 1972). Tryptophan absorbs electromagnetic radiation at a wavelength of 295 nm, and tryptophan emission is solvatochromic (Ostashevskii, 1972). Thus, as the polarity of the tryptophan environment or solvent increases, the emission shifts to a longer wavelength, and as the polarity of the solvent decreases, the emission shifts to a shorter wavelength (Lakowicz, 2006b). In addition, interactions of fluorochromes with each other and with molecules in the environment can affect the intensity of the emitted fluorescence (Lakowicz, 2006a). Tryptophan emission, in particular, can be

quenched by protonated amino acid residues, disulfide groups, amide groups, energy transfer with other tryptophan residues, and even the peptide bonds of the protein backbone (Lakowicz, 2006b). The interactions resulting in the quenching of tryptophan fluorescence are largely affected by the distance between the tryptophan residue and the quencher (Lakowicz, 2006b). As a result, tryptophan fluorescence spectroscopy is a good probe of the environment of tryptophan residues, and hence of protein fold or tertiary structure. In the case of there being too few or no tryptophan residues in a protein, fluorescence spectroscopy at an excitation wavelength of 280 nm detects tryptophan, tyrosine and, to a lesser extent, cysteine residues (Ostashevskii, 1972). However, tryptophan residues make the largest contribution to the emission spectra (Ostashevskii, 1972).

The FOXP2 (LeuZip - end) proteins each contain three tryptophan residues, all located within the DNA-binding FHD. Intrinsic tryptophan fluorescence spectroscopy was conducted to compare the folds of the FOXP2 (LeuZip - end) variants, and assess the effects of the R401D, A539P and R401D/A539P mutations on the protein fold, using the tryptophan residues as probes. Intrinsic tryptophan fluorescence spectra were recorded for each FOXP2 (LeuZip - end) protein in the absence and presence of DNA, to study the effects of DNA binding on the environments of the tryptophan residues, and hence on the folds of the proteins. An excitation wavelength of 295 nm was used, and spectra were recorded over an emission range of 280 nm to 500 nm. Experiments were conducted using a Jasco FP-6300 Spectrofluorometer (JASCO, USA) with bandwidths of 5 nm and a speed of 500 nm/minute, at 20 °C. Samples consisted of 5 µM protein, prepared in spectroscopy buffer (previously described) with 2 mM DTT for spectra measured under reducing conditions. Samples with DNA were of the same composition with the addition of 2.5 µM DNA (for an explanation of the DNA:protein concentration ratio, see 2.4.1 Circular dichroism spectropolarimetry). All samples were prepared in triplicate. Final averaged spectra were corrected using blanks consisting of buffer only or buffer with 2.5 µM DNA for samples containing DNA, and with 2 mM DTT for samples prepared under reducing conditions. Normalised emission maxima at 339 nm measured in the presence of DNA were analysed using two-sample t-tests, to determine the significance of differences observed between the DNA-bound FOXP2 (LeuZip - end) proteins.

2.4.3 SIZE EXCLUSION CHROMATOGRAPHY

There are various protein chromatographic techniques for the separation and quantification of proteins of a mixture (see 2.2.5 Protein purification). Size exclusion chromatography is one such technique in which proteins are distinguished and separated based on size (Barth *et al.*, 1994). The column resin consists of pores of a specified size; large molecules pass freely without entering pores whereas smaller molecules travel a longer distance through the column pores, hence larger molecules elute more quickly (Barth *et al.*, 1994).

Size exclusion chromatography is useful for analysing protein quaternary structure since oligomerisation leads to an increase in size. Large protein oligomers elute quickly whereas monomeric or small protein complexes elute more slowly (Barth *et al.,* 1994). Inaccuracies may arise in size exclusion chromatography data, especially with non-globular proteins, since the hydrodynamic volume may lead to a larger apparent size. However, data obtained from non-globular proteins is still useful for comparative purposes, especially with respect to oligomeric state. An alternative technique for the evaluation of protein quaternary structures is native polyacrylamide gel electrophoresis (Niepmann and Zheng, 2006). However, the hydrodynamic volume of a protein influences its migration, and hence its apparent size, during gel electrophoresis (Tanase *et al.*, 2015). Furthermore, size exclusion chromatography is of a higher resolution, and allows for a quicker and simpler recovery of protein samples following elution (Niepmann and Zheng, 2006; Tacal and Ozer, 2002; Tanase *et al.*, 2015). Although the effects of hydrodynamic volume are overcome in mass spectrometry, which can be used to determine protein size, mass spectrometry usually depends on successful cross-linking of monomers when studying homo-oligomers, which requires extensive optimisation (Titeca *et al.*, 2019). Furthermore, the choice of cross-linker can result in either an under- or overestimation of the size of the oligomer, since cross-linkers differ with respect to their reactivities and specificities (Titeca *et al.*, 2019). Thus, size exclusion chromatography was the preferred method for analysing the quaternary structures of the FOXP2 (LeuZip - end) proteins.

The quaternary structures of the FOXP2 (LeuZip - end) proteins were investigated using size exclusion chromatography. This was used to shed light on the relative oligomeric status of each FOXP2 (LeuZip - end) protein, and the effects of the R401D, A539P and R401D/A539P mutations on FOXP2 (LeuZip - end) oligomerisation. Therefore, through size exclusion chromatography, the relative contributions of the LeuZip and the FHD to FOXP2 (LeuZip - end) oligomerisation were investigated. Size exclusion chromatography was conducted using a HiLoad® 16/600 Superdex® 200 pg column and an ÄKTA FPLC system (GE Healthcare Life Sciences, UK). All analyses were conducted at a flow rate of

1 ml/minute at 20 °C. The column was pre-equilibrated with IMAC equilibration buffer (previously described) and calibrated using the Gel Filtration Markers Kit for Protein (Sigma-Aldrich, USA) prepared in IMAC equilibration buffer. Samples consisted of between 25 and 35 µM protein in IMAC equilibration buffer. The numbers and sizes of FOXP2 (LeuZip - end) oligomers present under non-reducing and reducing conditions were compared by preparing samples at 0 mM or 2 mM DTT. All samples were prepared in duplicate.

2.4.4 ACRYLAMIDE FLUORESCENCE QUENCHING

Protein molecules are able to fluoresce, predominantly due to tryptophan residues (see 2.4.2 Intrinsic tryptophan fluorescence spectroscopy). Various small molecules such as acrylamide, potassium iodide, caesium chloride, and succinimide, are able to quench tryptophan fluorescence (Eftink and Ghiron, 1981). Acrylamide, in particular, quenches tryptophan fluorescence by photoinduced electron transfer (Eftink *et al.*, 1987; Evans *et al.*, 1978; Zelent *et al.*, 1998). During the transfer, acrylamide is able to accept electrons from tryptophan, since acrylamide is a good electron scavenger (Eftink *et al.*, 1987). Quenching of tryptophan fluorescence by a quencher (*Q*) depends on the accessibility of the tryptophan residues to the quencher, which depends on the fold and flexibility/dynamics of the protein (Eftink and Ghiron, 1981). Measuring the intrinsic tryptophan fluorescence at increasing concentrations of *Q* can be used to calculate the Stern-Volmer constant (K_{SV}) from the relationship;

$$\frac{F_0}{F} = 1 + K_{SV} \cdot [Q]$$

where F_0 and F are the fluorescence intensities in the absence and presence of *Q*, respectively (Eftink and Ghiron, 1981). Comparisons of Stern-Volmer constants between different proteins, or for the same protein in different environments, could provide insight into the relative flexibilities/dynamics of each protein or for each environment.

Acrylamide fluorescence quenching of the FOXP2 (LeuZip - end) proteins was used to determine the Stern-Volmer constant for each protein. This was used to provide insight into the effects, if any, of the R401D, A539P and R401D/A539P mutations on the fold and flexibility of FOXP2 (LeuZip - end). Stern-Volmer constants were also determined in the presence of DNA to observe the effects of DNA binding on the flexibilities and dynamics of the FOXP2 (LeuZip - end) proteins, and to probe any conformational changes occurring upon DNA binding. Emission maxima at 339 nm were recorded on a Jasco FP-6300 Spectrofluorometer (JASCO, USA) at an excitation wavelength of 295 nm with bandwidths of 5 nm at 20 °C. Samples consisted of 5 µM protein prepared in spectroscopy buffer (previously described) at acrylamide concentrations of 0 mM, 50 mM, 100 mM, 150 mM and 250 mM.

Samples with DNA were of the same composition with the addition of 2.5 µM DNA (for an explanation of the DNA:protein concentration ratio, see 2.4.1 Circular dichroism spectropolarimetry). All samples and blanks were prepared under reducing conditions at 2 mM DTT and all samples were prepared in triplicate. Final averaged measurements were corrected using blanks consisting of buffer only, or buffer with 2.5 µM DNA for samples containing DNA, at each acrylamide concentration. Errors were calculated as the standard deviations of the averages. Stern-Volmer constants were analysed using two-sample t-tests to determine the significance of differences observed between the FOXP2 (LeuZip - end) proteins, and using paired t-tests to determine the significance of differences observed upon DNA binding of each protein.

2.4.5 PROTEIN THERMAL UNFOLDING

Protein unfolding can be achieved by different methods including chemical methods, changes in temperature, or the use of pressure or force (for a review of protein unfolding methods see Lapidus, 2017). Proteins unfold at both high and low temperatures, but the low temperature denaturation midpoint is typically below 0 °C, resulting in difficulties when trying to study protein unfolding due to the freezing of aqueous solvents at temperatures above the denaturation midpoint (Lapidus, 2017; Pastore *et al.*, 2007; Privalov *et al.*, 1986; Yoshidome and Kinoshita, 2009). Elevated temperatures, however, are useful for studying protein unfolding, and data obtained from protein thermal unfolding can be interpreted relatively easily since the effects of temperature on proteins are well-understood (Lapidus, 2017).

A protein produces a distinct spectrum when studied by CD spectropolarimetry in the far-UV range of wavelengths, depending on the secondary structural content of the protein (see 2.4.1 Circular dichroism spectropolarimetry). As a protein unfolds, ordered secondary structures disintegrate and become disordered (Greenfield, 2006b). The loss of secondary structure is accompanied by a distortion to the CD spectrum of the protein, therefore protein unfolding can be easily monitored using CD spectropolarimetry, resulting in an unfolding curve (Greenfield, 2006b). Unfolding curves can be used to identify any thermodynamically stable intermediates in the pathway between native and unfolded states, and therefore provide important clues as to the folding pathway of the protein (Seelig and Schönfeld, 2016). The effects of mutations on protein folding/unfolding and stability can be determined by comparing the unfolding curves of mutant proteins with that of the wild-type (Greenfield, 1999). Furthermore, protein unfolding curves obtained under different conditions can reveal the effects of environmental factors such as pH or ionic strength on the stability of a protein (Kishore *et al.*, 2012; Mao *et al.*, 2007; Metrick *et al.*, 2013).

Thermal unfolding of the FOXP2 (LeuZip - end) variants was monitored using far-UV CD spectropolarimetry. This was used to study the effects of the R401D, A539P and R401D/A539P mutations on the thermal stability of FOXP2 (LeuZip - end). Based on the oligomeric states observed through size exclusion chromatography, thermal unfolding was used to compare the thermal stabilities of the FOXP2 (LeuZip - end) oligomers. The signal at 222 nm was monitored using a Jasco J-810 Spectropolarimeter (JASCO, USA). Samples consisted of protein prepared in spectroscopy buffer (previously described). Measurements were recorded at 5 µM and 15 µM protein over a range of 20 - 80 °C with a temperature slope of 1 °C/minute, using bandwidths of 5 nm. The high tension (voltage) was monitored and kept below 600 V to ensure that no noisy data was included.

Chapter 3: Results

Evidence suggests that oligomerisation is a prerequisite for FOXP2 functioning, and that both the LeuZip and the DNA-binding FHD of FOXP2 are able to interact to facilitate the formation of dimers or higher order oligomers (Li *et al.*, 2004; Stroud *et al.*, 2006). What is currently unclear, however, is whether both interfaces are required simultaneously for FOXP2 oligomerisation, whether the LeuZip and FHD function independently with different strengths, and the effects that LeuZip and FHD associations have on the protein structure and DNA binding. In this study of FOXP2 (LeuZip - end), the contribution of each interface to FOXP2 oligomerisation, and any effects of disrupting each interface on FOXP2 structure and DNA binding were investigated. Each of the R401D and A539P mutants were used to disrupt one of the two interfaces, while the R401D/A539P mutant was used to disrupt both interfaces. A comparative study of wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) was conducted. *In silico* and *in vitro* methods were combined to predict the 3D protein structures, to structurally characterise the FOXP2 (LeuZip - end) variants, to evaluate changes to the protein quaternary structure as a result of each mutation, and to assess DNA-binding affinities.

3.1 *IN SILICO* PROTEIN STRUCTURE ANALYSES

Knowledge of the structure of a protein is integral to understanding its function, and hence forms an essential component of the study of any protein. In order to gain a thorough understanding of the structures of all FOXP2 (LeuZip - end) variants, *in silico* methods were used to predict the protein secondary and 3D structures, as well as the disorder, of each FOXP2 (LeuZip - end) variant. This provided insight into the structure of the wild-type FOXP2 (LeuZip - end) protein. Furthermore, comparisons between the wild-type and mutants allowed for investigating the effects of the R401D, A539P and R401D/A539P mutations on protein structure, and hence the importance of an intact LeuZip and a flexible FHD (due to a flexible hinge loop) for the folding of the protein into its native structure.

3.1.1 PROTEIN SECONDARY STRUCTURE AND DISORDER PREDICTIONS

The secondary structure of a protein represents the first level of protein folding. Protein secondary structures often dictate the global protein fold, and hence protein function (Dai and Zhou, 2011; Godzik *et al.*, 2007; Murzin *et al.*, 1995; Ozkan *et al.*, 2007; Yang *et al.*, 2018; Zhou and Karplus, 1999). Thus, PSIPRED 4.0 was used to predict the secondary structures of the FOXP2 (LeuZip - end) proteins based on the primary structures (Jones, 1999). In addition, this was used to compare the secondary

structures of the mutants with the wild-type FOXP2 (LeuZip - end) protein, to observe the effects of the mutations on the secondary structure of the protein. PSIPRED 4.0 annotated each amino acid residue according to the secondary structural element that it was predicted to form part of. The results of the prediction are summarised in Figure 3.1. All FOXP2 (LeuZip - end) variants exhibited a large proportion of predicted loops, with α-helices and β-strands mostly occurring within the LeuZip which spans the region between Arg382 and Met411, and the FHD which spans the region between Arg504 and Asp576. Only a few interspersed α-helices were predicted in the regions extending from the FHDs to the C-terminal ends of the proteins. There were no differences in the arrangement of the secondary structural elements between the FOXP2 (LeuZip - end) proteins, with only minor differences in the predicted lengths of a few α-helices. The sensitivity of PSIPRED 4.0 to the effects of mutations could be reduced, due to the likelihood of PSIPRED 4.0 generating the same multiple sequence alignment for both wild-type and mutant proteins. Therefore, the similarity between the proteins could be as a result of this limitation. However, data obtained provided a useful overview of the secondary structures of the FOXP2 (LeuZip - end) proteins.

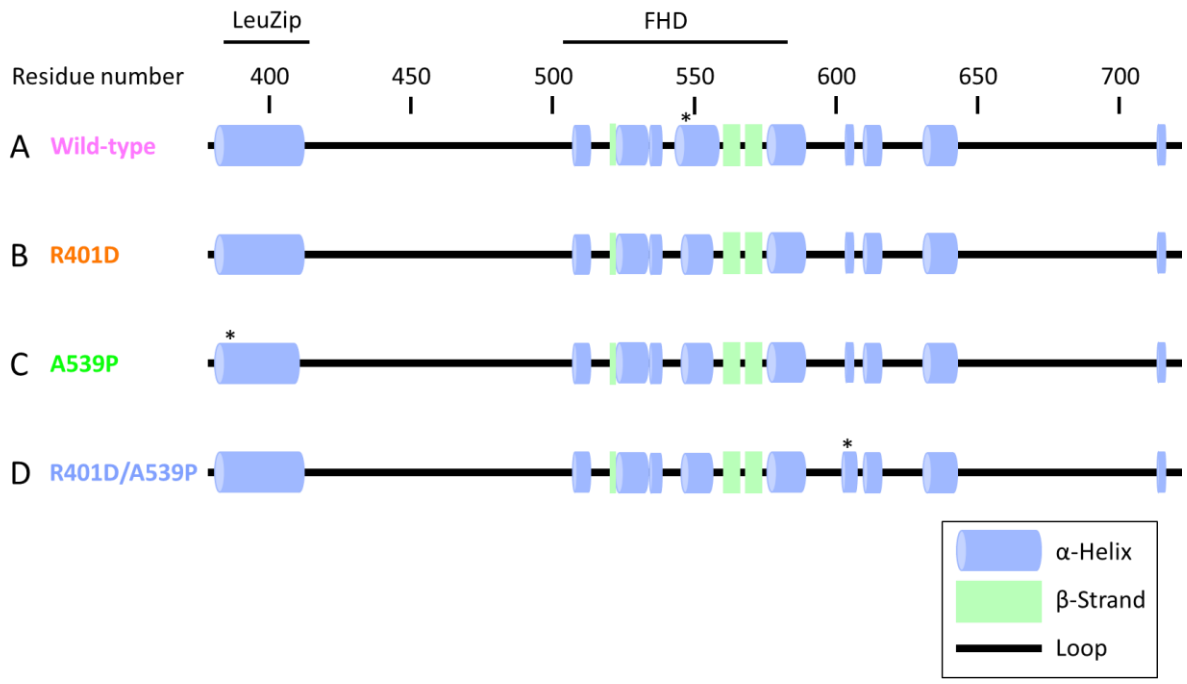


Figure 3.1: Schematic representation of secondary structure predictions of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). The approximate arrangements of α-helices, β-strands, and loops are shown based on protein secondary structure predictions. Residues are numbered relative to the full-length protein. All FOXP2 (LeuZip - end) variants showed the same secondary structural arrangement, with only minor differences in the lengths of three α-helices (indicated by * above the α-helices). The proteins consisted predominantly of loops (almost 70%), with ~ 30% α-helical content and a small proportion (~ 3%) of β-strands. α-Helical and β-stranded regions corresponded approximately with the LeuZip (Arg382 to Met411) and FHD (Arg504 to Asp576). Secondary structure predictions were obtained using PSIPRED 4.0 (Jones, 1999).

Protein regions that lack secondary structures are said to be disordered. Knowledge of protein disorder is useful since disordered protein regions are involved in many biological functions, due to the conformational freedom of these regions, arising on account of the disorder (Nielsen and Mulder, 2019). The online DISOPRED3 tool was used to predict regions of intrinsic disorder within the proteins, based on the protein primary structures (Jones, 1999; Jones and Cozzetto, 2015). Each amino acid residue was annotated with a confidence score based on the difference in the probabilities of that residue forming part of a disordered or ordered region (Basu *et al.*, 2017; Jones and Cozzetto, 2015). All DISOPRED3 confidence scores fall between 0 and 1, and any amino acid with a confidence score of 0.5 or higher was considered to form part of a disordered region (Basu *et al.*, 2017; Jones and Cozzetto, 2015). Figure 3.2 shows the results of the predictions. All FOXP2 (LeuZip - end) variants were predicted to be more than 60% disordered, and regions predicted to be disordered coincided with the loops predicted by PSIPRED 4.0. Wild-type FOXP2 (LeuZip - end) was 64% disordered, R401D FOXP2 (LeuZip - end) was 62% disordered, A539P FOXP2 (LeuZip - end) was 66% disordered and R401D/A539P FOXP2 (LeuZip - end) was 63% disordered. The ordered regions of the proteins corresponded approximately with the LeuZip and the FHD, and the remaining regions were predicted to be intrinsically disordered. As with the secondary structure predictions, similarities between the proteins, with respect to predicted regions of intrinsic disorder, could be due to DISOPRED3 generating the same multiple sequence alignment for all the FOXP2 (LeuZip - end) variants. Nevertheless, the data provided a useful overview of the disorder of the FOXP2 (LeuZip - end) proteins.

In addition to predicting intrinsically disordered regions, DISOPRED3 provided predictions of binding sites amongst these regions (Jones, 1999; Jones and Cozzetto, 2015). For all FOXP2 (LeuZip - end) variants, only the first 10 or 11 amino acid residues of each protein, comprising of the additional residues which are not derived from the FOXP2 sequence (Figure 2.1), were predicted to form part of a binding site (data not shown). DISOPRED3 predicts protein binding sites using a support vector machine that annotates amino acid residues based on evolutionary patterns and based on the amino acid compositions of putative binding sites (Jones and Cozzetto, 2015). A region is predicted to form a binding site only if it is disordered and it is shown to fold upon binding, and hence able to transition from disordered to ordered (Jones and Cozzetto, 2015). Based on this data, there are no such binding sites found within the disordered protein regions of FOXP2 (LeuZip - end).

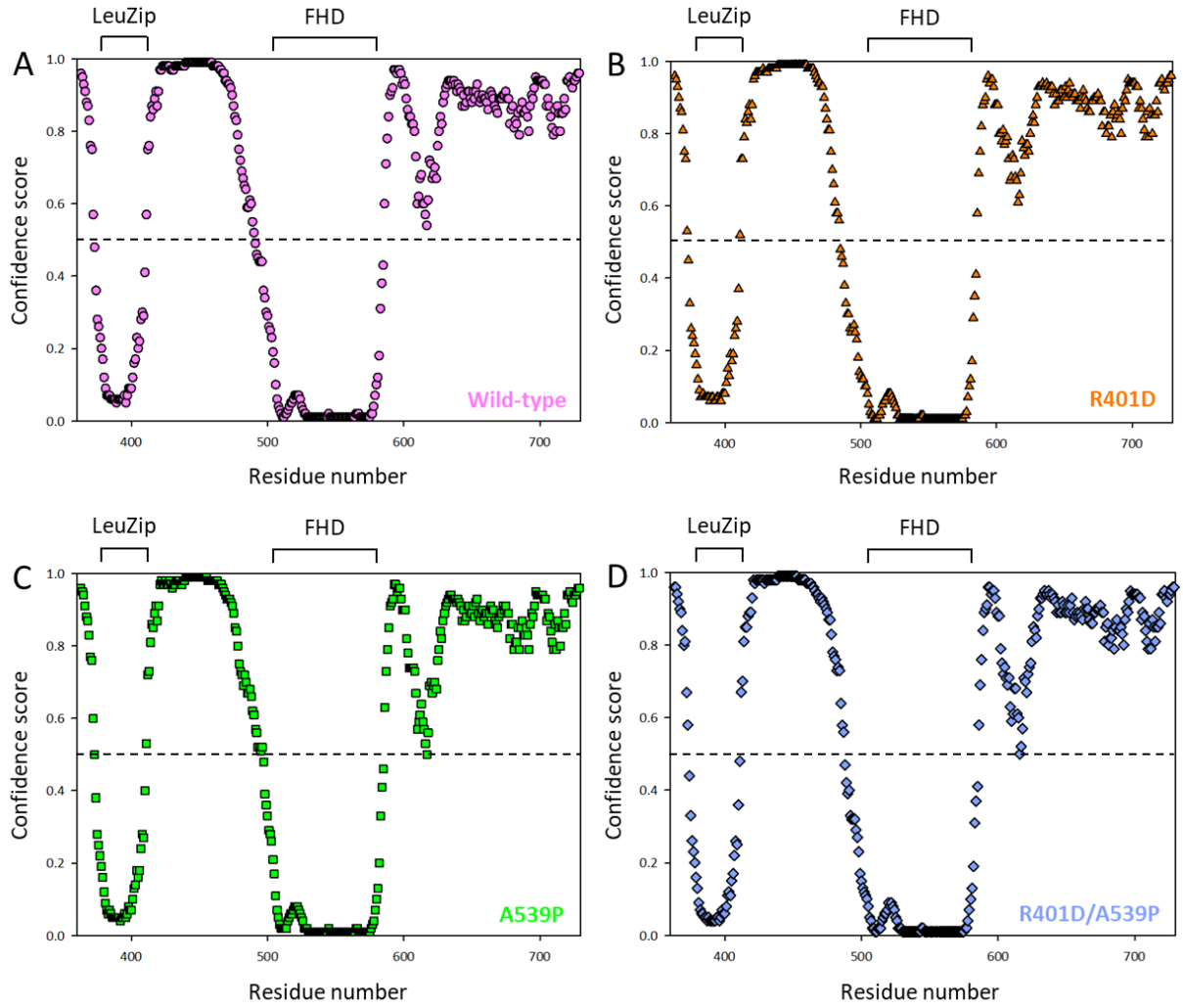


Figure 3.2: Disorder predictions of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). Regions with confidence scores of 0.5 (dashed line) or greater are disordered. Residues are numbered relative to the full-length protein. All FOXP2 (LeuZip - end) proteins showed large proportions of disorder, and showed approximately the same disorder pattern with respect to the amino acid sequences. Wild-type FOXP2 (LeuZip - end) **(A)** was predicted to be 64% disordered, R401D FOXP2 (LeuZip - end) **(B)** was predicted to be 62% disordered, A539P FOXP2 (LeuZip - end) **(C)** was predicted to be 66% disordered and R401D/A539P FOXP2 (LeuZip - end) **(D)** was predicted to be 63% disordered. Ordered regions correspond approximately with the LeuZip (Arg382 to Met411) and FHD (Arg504 to Asp576). Disorder predictions were obtained using DISOPRED3 (Jones and Cozzetto, 2015).

3.1.2 THREE-DIMENSIONAL PROTEIN STRUCTURE PREDICTIONS

The 3D or tertiary structure of a protein represents the next level of protein folding, after the secondary structure. Therefore, following secondary structure and disorder predictions, the 3D protein structures were investigated. The online I-TASSER server was used to predict the monomeric 3D structures of the FOXP2 (LeuZip - end) variants (Roy *et al.,* 2010; Yang and Zhang, 2015; Yang *et al.,* 2015). I-TASSER produced five models of each protein based on the amino acid sequences of the proteins. I-TASSER ranked each model with a C-score, which estimates the quality of the model based

on the threading template alignments and the parameters of the structure assembly during convergence (Roy *et al.,* 2010; Yang and Zhang, 2015; Yang *et al.,* 2015). C-scores of acceptable models fall between - 5 and 2, with a higher C-score indicating a higher confidence (Roy *et al.,* 2010; Yang and Zhang, 2015; Yang *et al.,* 2015). C-scores of all predicted models of all proteins were between - 3.90 and - 2.66, thus all predicted models were of acceptable quality. The models were energy minimised and the final energy minimised models were analysed with respect to the favourability of rotamers, backbone torsion angles, and covalent geometries using the online MolProbity application (Williams *et al.,* 2018). Based on a combination of the C-scores and MolProbity analyses, the best model was selected for each protein. C-scores of the selected wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) models were - 2.75, - 2.9, - 2.91, and - 2.83, respectively.

The MolProbity analyses of the selected models are shown in Table 3.1, with values indicated as good, acceptable, or unideal based on the MolProbity guidelines (Williams *et al.,* 2018). The clashscore is the number of steric clashes per 1000 atoms. Clashscores of wild-type and R401D FOXP2 (LeuZip - end) models were in the 98th percentile, the clashscore of the A539P FOXP2 (LeuZip - end) model was in the 97th percentile and the clashscore of the R401D/A539P FOXP2 (LeuZip - end) model was in the 99th percentile, with the best clashscores constituting the 100th percentile. Therefore, the clashscores of the protein models were good. Rotamers refer to conformational isomers of amino acids with flexible side chains. A rotamer is favoured when it is in the lowest possible energy conformation. Ideally, there should be more than 98% favoured rotamers and less than 0.3% poor rotamers. Models of the FOXP2 (LeuZip - end) proteins all had fewer favoured rotamers and more poor rotamers than ideal. Ramachandran outliers and Ramachandran favoured torsional angles were not within the ideal contours, as these should be less than 0.05% and greater than 98%, respectively. The FOXP2 (LeuZip - end) structures were based on computational predictions and not from experimentally derived data, therefore the poor rotamers and Ramachandran outliers could not be improved further since there was no experimental data to fit the structures to, so attempts to improve these parameters would have introduced bias into the structures. A MolProbity score is a single score describing the clashscore, rotamers and Ramachandran analysis, and it is normalised to the same scale that is used to describe the resolution of structures solved by X-ray crystallography. Ideally, MolProbity scores should lie in the 66th percentile or higher, with scores falling below the 33rd percentile being unideal. The MolProbity scores of the FOXP2 (LeuZip - end) variants were within the acceptable range (between the 33rd and 66th percentiles).

Table 3.1: Analyses of best models of FOXP2 (LeuZip - end) variants. I-TASSER was used to predict 3D models of the FOXP2 (LeuZip - end) variants (Roy *et al.,* 2010; Yang and Zhang, 2015; Yang *et al.,* 2015). After energy minimisation, the models were analysed using MolProbity (Williams *et al.,* 2018). Together with the I-TASSER C-scores, the results of the analyses were used to select the best models, and the analyses for the best models are shown here. Values highlighted in green are good, values highlighted in yellow are acceptable if justified, and values highlighted in rose are unideal, according to the MolProbity guidelines.

	Wild-type FOXP2 (LeuZip - end)	**R401D FOXP2 (LeuZip - end)**	**A539P FOXP2 (LeuZip - end)**	**R401D/A539P FOXP2 (LeuZip - end)**
Clashscore	3	2.83	3.35	2.29
Poor rotamers (%)	3.67	8.87	7.01	7.62
Favoured rotamers (%)	87.16	80.73	78.35	78.05
Ramachandran outliers (%)	3.01	9.56	6.56	6.83
Ramachandran favoured (%)	79.51	70.22	69.40	69.95
MolProbity score	2.26	2.63	2.61	2.52
Cβ deviations > 0.25 Å (%)	1.42	3.68	2.83	3.97
Bad bonds (%)	0	0	0	0
Bad angles (%)	1.25	1.48	1.43	1.60
Twisted peptides (%)	0.54	4.36	3.27	3.81
CaBLAM outliers (%)	14.3	17.3	14	13.2
CA geometry outliers (%)	3.85	4.95	6.32	7.69

Cβ describes the conformation of amino acid side chains relative to the protein backbone. A Cβ deviation occurs when a clash arises between a side chain and the backbone. There should ideally be no Cβ deviations in a protein structure, but deviations of less than 5% are acceptable, so the Cβ deviations of the proteins were within the acceptable range. There were no bad bonds in any of the FOXP2 (LeuZip - end) structures. Bad angles were greater than 0.5% for all FOXP2 (LeuZip - end) structures, which is unideal, as there should be less than 0.1% bad angles with up to 0.5% being acceptable. Attempts to correct bad angles were unsuccessful and resulted in increases in the clashscores and MolProbity scores of all structures. An ideal protein structure has no twisted peptides

and twisted peptides of up to 0.1% are acceptable. Twisted peptides were not within the acceptable range for any of the proteins. CaBLAM (Cα Based Low-resolution Annotation Method) describes expected secondary structures based on Cα coordinates, and outliers should ideally be less than 1%. CaBLAM outliers of up to 5% are acceptable, therefore CaBLAM outliers of all FOXP2 (LeuZip - end) structures were higher than the acceptable limit. CA geometry refers to the geometry of the protein backbone with respect to Cα. There should be less than 0.5% CA geometry outliers, and values of up to 1% are acceptable. All FOXP2 (LeuZip - end) structures had a higher proportion of outliers than is generally allowed. Twisted peptides, CaBLAM outliers and CA geometry outliers could not be improved without introducing bias into the structures, due to a lack of experimentally derived structure data.

Although the FOXP2 (LeuZip - end) models did not meet a number of criteria, according to the MolProbity analyses, this was expected due to I-TASSER predicting 3D protein structures using fragments from multiple templates, rather than a single template (Roy *et al.,* 2010; Yang and Zhang, 2015; Yang *et al.,* 2015). The use of multiple templates can result in unfavourable local topologies, especially for targets that do not have similar homologous templates available, but this improves the accuracy over models based on single templates (Roy *et al.,* 2010; Yang and Zhang, 2015; Yang *et al.,* 2015). Indeed, I-TASSER is widely recognised as the most accurate tool for 3D protein structure prediction (Croll *et al.*, 2019; Kryshtafovych *et al.*, 2018; Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Furthermore, the MolProbity score, which combines the clashscore, rotamer analysis, and Ramachandran analysis, was acceptable for all FOXP2 (LeuZip - end) models.

The R401D mutation of the LeuZip was designed based on the equivalent K251D mutation of the mFoxp3 LeuZip (Song *et al.,* 2012). In mFoxp3, Lys251 (equivalent to FOXP2 Arg401) forms a hydrogen bond with Glu248 of the same monomer (equivalent to FOXP2 Glu398), which positions Glu248 to form an intersubunit hydrogen bond with Gln234 of another mFoxp3 monomer (equivalent to FOXP2 Gln384) (Figure 1.14C) (Song *et al.,* 2012). Thus, the effect of the R401D mutation on this hydrogen bond network in the models was investigated. Since the models were monomeric, the intersubunit hydrogen bond between Glu398 and Gln384 could not be investigated directly. However, based on the mFoxp3 LeuZip, the hydrogen bond between Arg401 and Glu398 of a single monomer would form a key component of the hydrogen bond network, so this hydrogen bond was assessed (Song *et al.,* 2012). The models of the FOXP2 (LeuZip - end) structures showed that Arg401 formed hydrogen bonds with Glu398 in both wild-type and A539P FOXP2 (LeuZip - end) (Figure 3.3A and 3.3C). Mutation of Arg401 to aspartic acid disrupted this hydrogen bond and altered the position of the Glu398 side chain

(Figure 3.3B and 3.3D), which would be sufficient to disrupt the intersubunit hydrogen bond between Glu398 and Gln384.

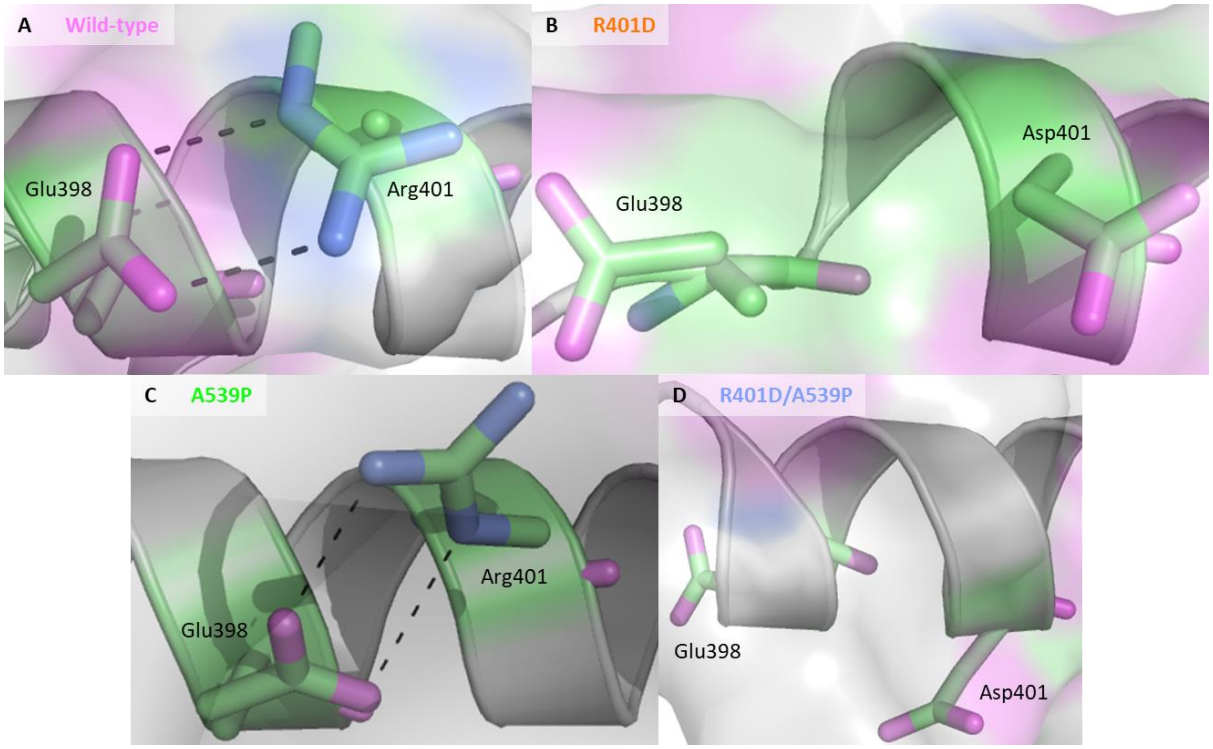


Figure 3.3: Effects of R401D (B), A539P (C) and R401D/A539P (D) mutations on LeuZip hydrogen bond network in predicted FOXP2 (LeuZip - end) protein structures. The Arg401 interaction of wild-type FOXP2 (LeuZip - end) is shown in **(A)** for comparison. Arg401 positions Glu398 correctly by hydrogen bonding (**A** and **C**), to facilitate the formation of an intersubunit hydrogen bond between Glu398 and Gln384 (not shown) which stabilises the LeuZip dimer. Incorporation of the R401D mutation is predicted to disrupt the intersubunit hydrogen bond, since the position of Glu398 is disrupted by the aspartic acid (**B** and **D**). Carbon atoms are green, oxygen atoms are pink, nitrogen atoms are blue, and hydrogen bonds are black dashed lines. Three-dimensional protein structure predictions were obtained using I-TASSER (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

Based on the hydrogen bond networks of mFoxp3, formation of the intersubunit hydrogen bond required for FOXP2 oligomerisation would depend on the accessibility of Gln384 (Song *et al.,* 2012). Therefore, SwissPdb Viewer v4.1.0 was used to assess the solvent-accessibility of Gln384 in the FOXP2 (LeuZip - end) models (Guex and Peitsch, 1997). Solvent-accessible amino acid residues were defined as residues with a minimum of 30% solvent-accessible surface area. The results showed that Gln384 was predicted to be solvent-accessible only in A539P FOXP2 (LeuZip - end), and buried in the other variants (Figure 3.4). However, inaccuracies could arise in the structures of the FOXP2 (LeuZip - end) variants since they are based on predictions. Additionally, proteins are dynamic and since the FOXP2 (LeuZip - end) proteins are predicted to be highly disordered, this would confer conformational

freedom to the proteins, enabling them to sample many conformations (Dunker *et al.*, 2008; Fisher and Stultz, 2011). The predicted structures represent only a single proposed conformation of each protein. Thus, the solvent-accessibility of Gln384 could fluctuate in a conformation-dependent manner.

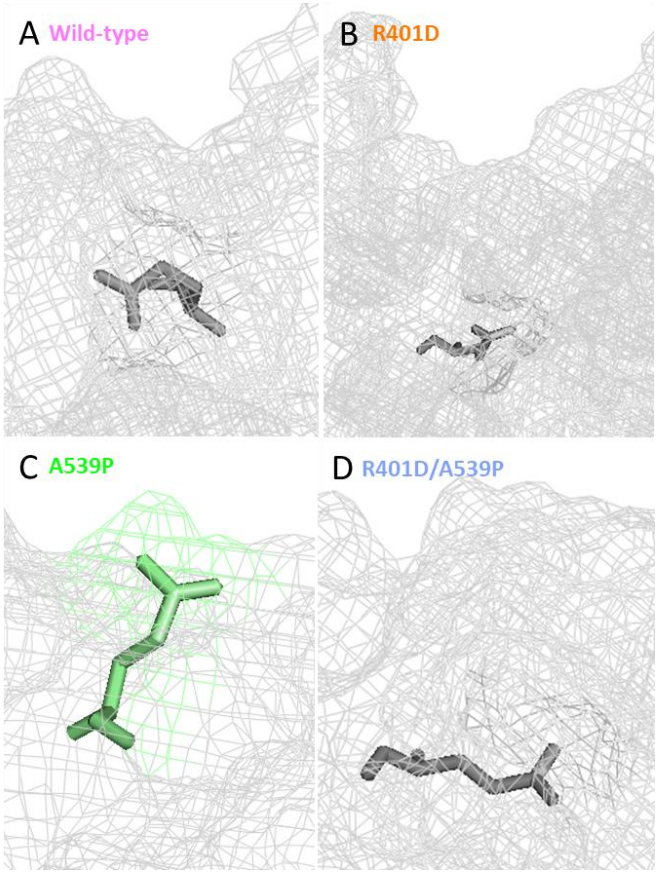


Figure 3.4: Solvent-accessibility of Gln384 in predicted structures of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). The FOXP2 (LeuZip - end) proteins are shown in surface mesh representation and Gln384 is shown in stick representation. Solvent-accessibility was defined as residues with a minimum of 30% solvent-accessible surface area. Gln384 is buried in wild-type, R401D and R401D/A539P FOXP2 (LeuZip - end) (dark grey), and solvent-accessible in A539P FOXP2 (LeuZip - end) (green). Three-dimensional protein structure predictions were obtained using I-TASSER (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

The effect of the R401D mutation on the overall structure of the LeuZip was then analysed. As seen in Figure 3.5A, although the wild-type FOXP2 (LeuZip - end) LeuZip was predominantly helical (80%), the LeuZip α-helix was segmented with a loop occurring between the two helical portions. This was surprising since LeuZip motifs usually form stable α-helices, and the A539P FOXP2 (LeuZip - end) LeuZip formed a single α-helix spanning the length of the region, as expected (Figure 3.5C) (Ellenberger, 1994; Landschulz *et al.,* 1988). In addition, the wild-type FOXP2 (LeuZip - end) LeuZip did not form a complete

α-helix in any of the five proposed models (data is only shown for the best model that was selected and not for the remaining four models). However, as with the solvent-accessibility of Gln384, inaccuracies could arise in these structures since they are based on predictions. Nevertheless, the R401D mutation caused a decrease in LeuZip helicity of ~ 15%, since the R401D and R401D/A539P LeuZips were 63% and 67% helical, respectively (Figure 3.5B and 3.5D). Although only monomeric 3D protein structures could be predicted, it is expected that the decreased helicity occurring as a result of the R401D mutation would affect oligomeric protein-protein interactions at the LeuZip interface.

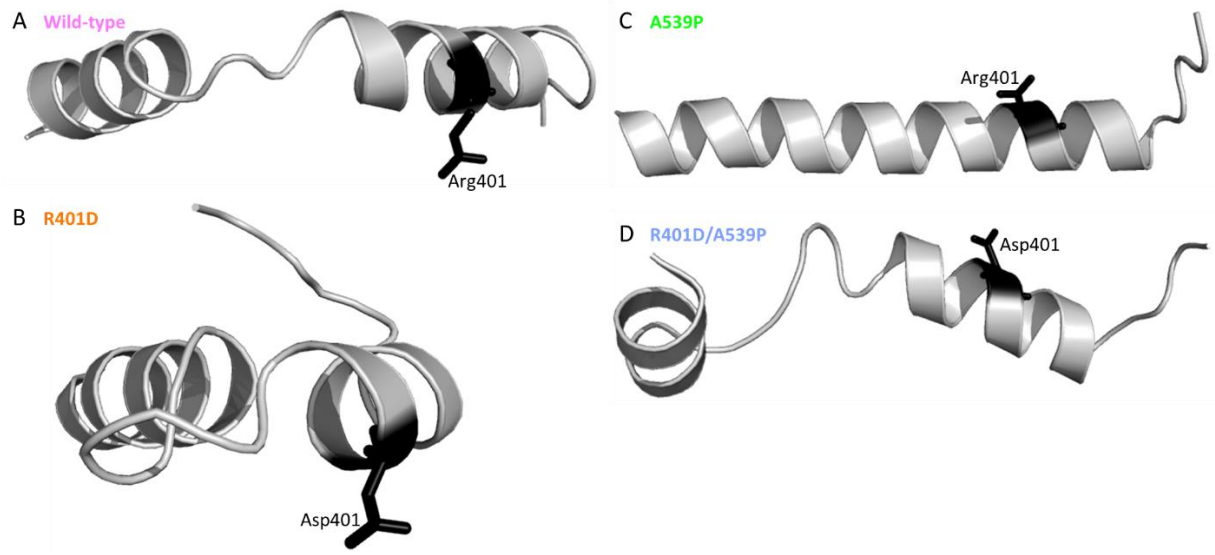


Figure 3.5: Comparison of predicted LeuZip structures of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). Although the wild-type FOXP2 (LeuZip - end) LeuZip did not form a complete α-helix, the LeuZips of wild-type and A539P FOXP2 (LeuZip - end) have high α-helical contents (80% and 87%, respectively). The R401D mutation of the LeuZip reduces the α-helical content of the LeuZips of both R401D and R401D/A539P FOXP2 (LeuZip - end) (63% and 67% α-helical, respectively). Arg401 and Asp401 are shown in stick representation (black). Three-dimensional protein structure predictions were obtained using I-TASSER (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

The A539P mutation of the FHD was designed based on the fact that the FHDs of FOXP proteins exhibit the unique ability to form domain-swapped dimers (Chen *et al.,* 2015; Chu *et al.,* 2011; Stroud *et al.,* 2006). FHD dimerisation has not been seen in any FOX protein outside the FOXP subfamily and has been attributed to the replacement of a conserved proline of the hinge loop by alanine (Ala539 in FOXP2) in FOXP subfamily members (Figure 1.10) (Chen *et al.,* 2015; Chu *et al.,* 2011; Stroud *et al.,* 2006). Mutation of Ala539 to a proline has previously been shown to prevent the formation of the FOXP2 FHD domain-swapped dimer (Morris and Fanucchi, 2016; Stroud *et al.*, 2006). A crystal structure of the isolated wild-type FOXP2 FHD has been solved, with FOXP2 bound to DNA in both monomeric and dimeric forms (Stroud *et al.,* 2006). Thus, the FHDs of the monomeric FOXP2

(LeuZip - end) models were compared to the monomeric FOXP2 FHD crystal structure, firstly to evaluate the similarity of the FHDs of the models to the crystal structure, and secondly to evaluate the effects, if any, of the A539P mutation on the structure of the FHD (Figure 3.6). Structures of the wild-type and R401D FOXP2 (LeuZip - end) FHDs were missing the short α-helix of wing 2, most likely due to an inaccuracy of the prediction, but aligned well with the crystal structure otherwise. The A539P mutation had no apparent effect on the fold of the FHD, since the FHDs of A539P and R401D/A539P FOXP2 (LeuZip - end) showed good alignments with that of the monomeric FOXP2 FHD crystal structure (root-mean-square deviation or RMSD of 0.388 Å and 0.374 Å, respectively). Due to the models being monomeric, the effects of the A539P mutation on domain swapping of the FHD could not be assessed.

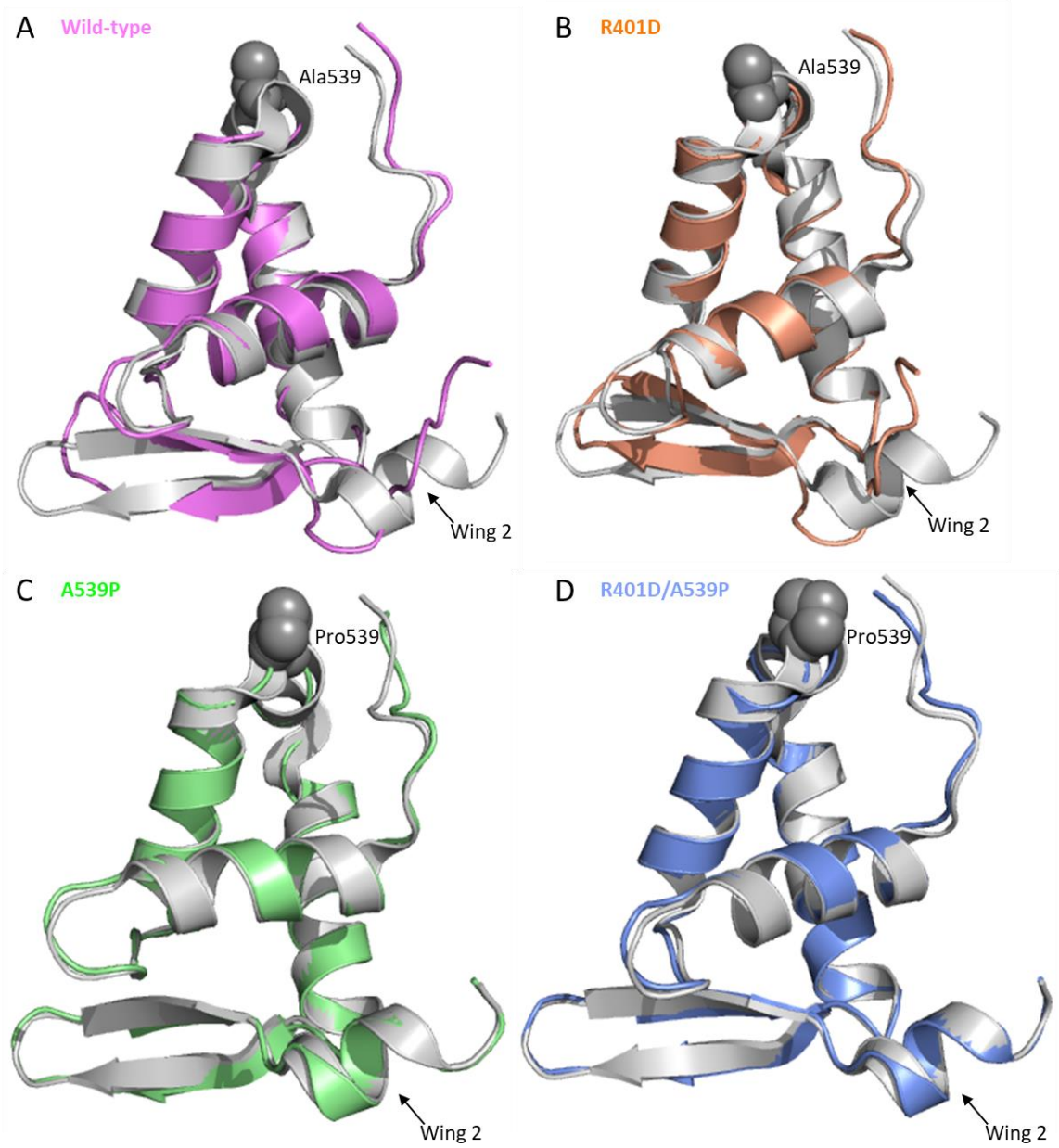


Figure 3.6: Alignment of predicted forkhead domain structures of FOXP2 (LeuZip - end) variants with the solved FOXP2 forkhead domain crystal structure. The solved FOXP2 FHD crystal structure (PDB 2A07) (Stroud *et al.,* 2006) is shown in light grey, and the FOXP2 (LeuZip - end) variants are shown in pink, orange, green, and blue as labelled. The DNA-binding FHD of R401D/A539P FOXP2 (LeuZip - end) **(D)** was the most similar to the solved FOXP2 FHD structure, and aligned with a RMSD of 0.374 Å. The FHD of wild-type FOXP2 (LeuZip - end) **(A)** was the least similar, and aligned with a RMSD of 1.089 Å. The FHDs of R401D **(B)** and A539P **(C)** FOXP2 (LeuZip - end) aligned with the solved FOXP2 FHD structure with RMSDs of 0.793 Å and 0.388 Å, respectively. The poorer fits of wild-type and R401D FOXP2 (LeuZip - end) are due to the structures missing the short α-helix of wing 2. Ala539 and Pro539 are shown in sphere representation (dark grey). The A539P mutation had no apparent effect on the fold of the FHD. Three-dimensional protein structure predictions were obtained using I-TASSER (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Alignments and images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

The overall folds of the FOXP2 (LeuZip - end) variants were then assessed. The monomeric models are shown in Figure 3.7. Upon visual inspection it was clear that all FOXP2 (LeuZip - end) variants were predicted to have a high proportion of looped/disordered regions. This agreed well with data obtained from secondary structure and disorder predictions, and was confirmed upon quantification, with the proportion of looped regions predicted to be more than 65% for all FOXP2 (LeuZip - end) variants. Wild-type and A539P FOXP2 (LeuZip - end) were 66% looped, R401D FOXP2 (LeuZip - end) was 76% looped and R401D/A539P FOXP2 (LeuZip - end) was 68% looped. According to disorder predictions, the ordered regions of the FOXP2 (LeuZip - end) proteins corresponded approximately with the LeuZip and FHD, and all other regions were unstructured (Figure 3.2). An assessment of the 3D protein structure models confirmed that most ordered regions occurred within the extended LeuZip α-helix and FHD. Although a few interspersed α-helices were observed, as with the disorder predictions, regions excluding the LeuZip and FHD were predominantly disordered (81% - 94% disorder in these regions for all FOXP2 (LeuZip - end) variants). The flexibility of the region between the LeuZip and FHD would allow for simultaneous protein-protein interactions of both interfaces during FOXP2 oligomerisation. Evidence suggests that the FOXP zinc finger and LeuZip form a single extended α-helix (Song *et al.,* 2012). The zinc finger is located N-terminal to the LeuZip, and the zinc finger-LeuZip region formed a continuous α-helix for all FOXP2 (LeuZip - end) variants. The solvent-accessible surface areas of the models were fairly similar, with wild-type FOXP2 (LeuZip - end) having the smallest surface area of 43 257 $Å^2$ and R401D/A539P FOXP2 (LeuZip - end) having the largest surface area of 43 404 $Å^2$. Thus, the R401D mutation of the LeuZip resulted in an increase in disorder and a small increase in surface area, the effects of which were seen in R401D/A539P FOXP2 (LeuZip - end) too. There are multiple FHD-DNA interactions that occur upon FOXP2 DNA binding, but the most noteworthy is the insertion of helix 3, the recognition helix, into the DNA major groove which relies on accessibility of helix 3 (Stroud *et al.,* 2006). Based on the solvent-accessible surface areas, most residues of helix 3 were buried in all FOXP2 (LeuZip - end) models. Helix 3 appeared to be obstructed by disordered loops, therefore FOXP2 DNA binding would require structural rearrangements of the flexible loops to enable insertion of helix 3 into the DNA major groove.

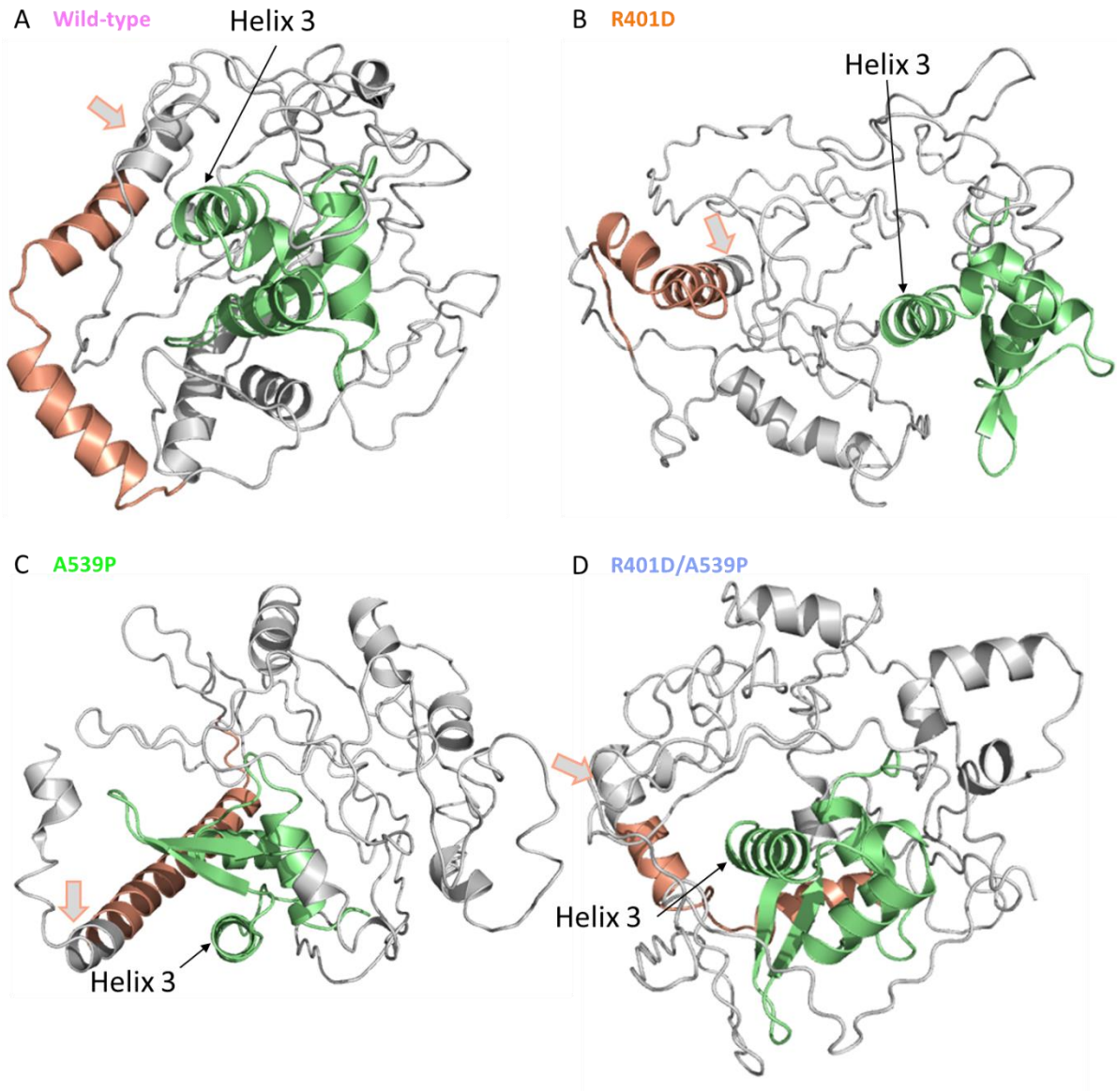


Figure 3.7: Predicted three-dimensional structures of wild-type (A), R401D (B), A539P (C) and R401D/A539P FOXP2 (D) (LeuZip - end). Structures of the monomeric FOXP2 (LeuZip - end) variants are depicted with LeuZips in orange and DNA-binding FHDs in green. Zinc fingers, located N-terminal to the LeuZips (grey arrows), formed uninterrupted α-helices with the LeuZips. All variants displayed large proportions of disordered loops (at least 65%), and regions excluding the LeuZip and FHD (grey) were predominantly looped (> 81% for all variants), with the exception of a few interspersed α-helices. The solvent-accessible surface areas of wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) were 43 257 Å^2, 43 375 Å^2, 43 366 Å^2, and 43 404 Å^2, respectively. The R401D mutation appeared to cause an increase in disorder, and a small increase in surface area, the effects of which were observed in R401D/A539P FOXP2 (LeuZip - end) too. Helix 3 was obstructed in all FOXP2 (LeuZip - end) variants, predominantly by flexible loops, necessitating structural rearrangements for DNA binding. Three-dimensional protein structure predictions were obtained using I-TASSER (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

The folds of the FOXP2 (LeuZip - end) variants were compared, and models of R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) were aligned with that of the wild-type. The alignments are depicted in Figure 3.8. The mutants showed poor alignment to the wild-type, with RMSDs of greater than 17 Å for all alignments. The R401D mutation of the LeuZip appeared to cause the largest structural rearrangements to the protein, the effects of which were seen in R401D/A539P FOXP2 (LeuZip - end) too. However, poor alignments of all mutants with the wild-type could be attributed to the large proportion of disordered regions observed in all variants, since these regions lack identifiable structures and would therefore not align.

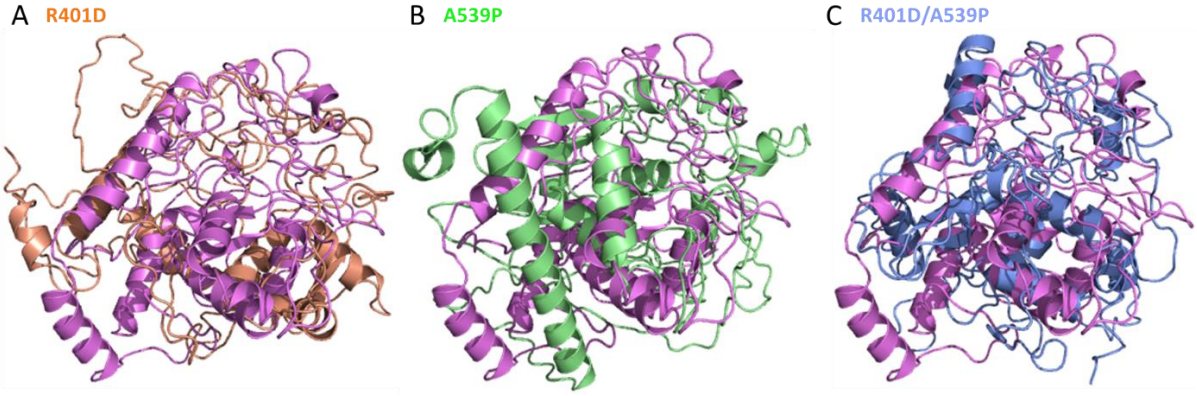


Figure 3.8: Alignment of predicted R401D (A), A539P (B) and R401D/A539P (C) FOXP2 (LeuZip - end) structures with the predicted structure of wild-type FOXP2 (LeuZip - end). Wild-type FOXP2 (LeuZip - end) is shown in pink, and the FOXP2 (LeuZip - end) mutants are shown in orange, green, and blue as labelled. Structures of the FOXP2 (LeuZip - end) mutants aligned poorly with that of the wild-type, presumably due to the large proportions of disordered regions present in all variants. A539P FOXP2 (LeuZip - end) was most similar to wild-type FOXP2 (LeuZip - end) and aligned with a RMSD of 17.6 Å. R401D and R401D/A539P FOXP2 (LeuZip - end) aligned with RMSDs of 20.9 Å and 20.3 Å, respectively. Three-dimensional protein structure predictions were obtained using I-TASSER (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Alignments and images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

3.2 PROTEIN PREPARATION

In silico methods provided insight pertaining to the secondary and tertiary structures of the FOXP2 (LeuZip - end) proteins. However, further experimentation was required to fully investigate the LeuZip and FHD interfaces, and the role of each interface in the structure and DNA binding of FOXP2 (LeuZip - end). Therefore, *in vitro* methods were used to obtain a thorough characterisation of the structure and DNA binding of the FOXP2 (LeuZip - end) proteins. In order to conduct downstream studies on the FOXP2 (LeuZip - end) proteins, it was first necessary to isolate a pure, soluble sample of each protein.

3.2.1 MUTAGENESIS

Since the R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) were novel constructs, preparation of these three mutants required site-directed mutagenesis of a FOXP2 (LeuZip - end)-encoding plasmid, to create the mutants. A pET-11a-wild-type FOXP2 (LeuZip - end) plasmid was used as the template to create the R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) mutants. Following mutagenesis, Sanger DNA sequencing confirmed the incorporation of the mutations (Figure 3.9). All mutations were successful.

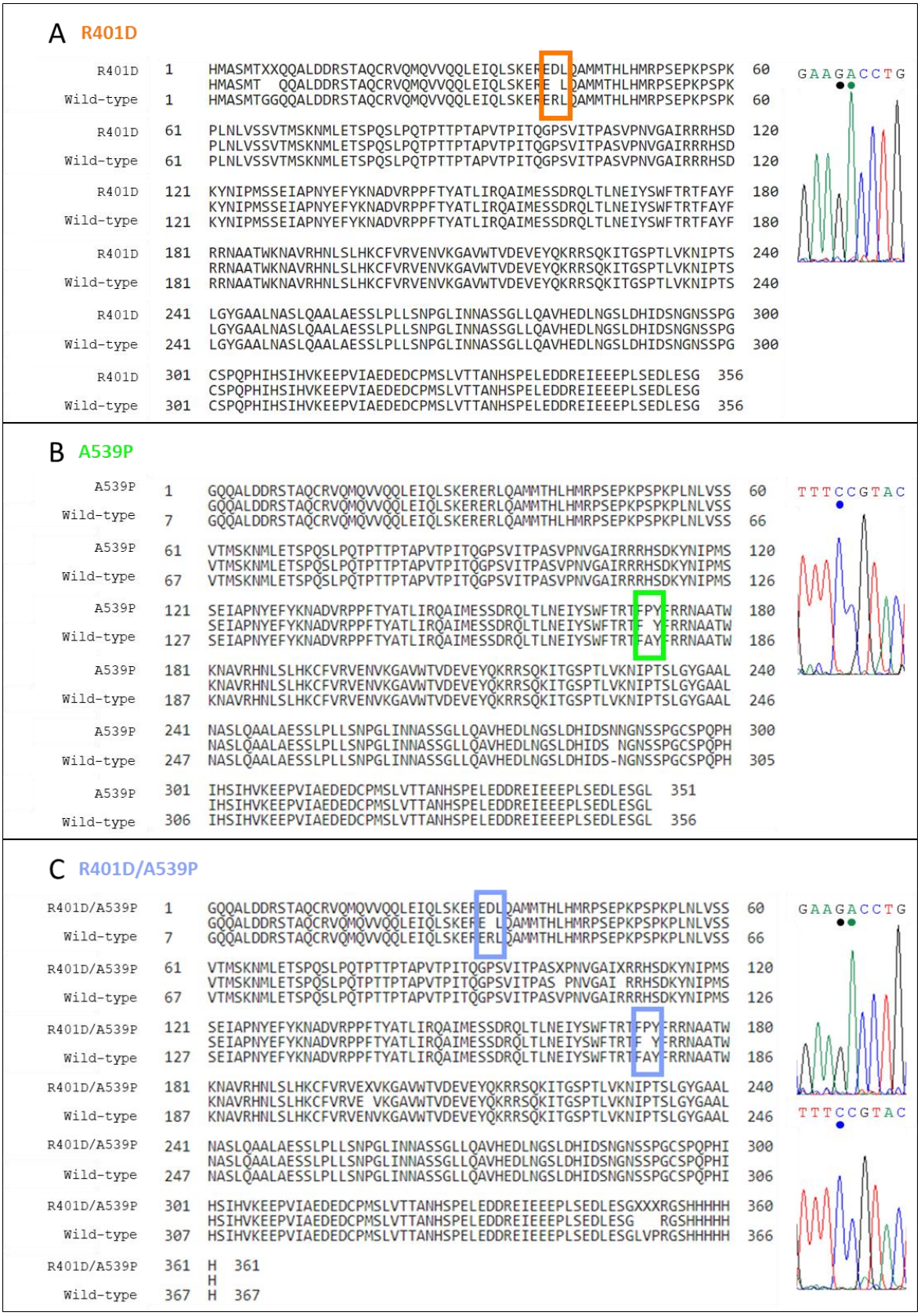


Figure 3.9: Confirmation of incorporation of R401D (A), A539P (B) and R401D/A539P (C) mutations of FOXP2 (LeuZip - end) by Sanger DNA sequencing. Sequence traces are shown (right-hand side) with the sites of mutation indicated by circles located above the traces and below the corresponding sequences. The R401D mutation required a double base change whereas the A539P mutation required a single base change. DNA sequences were translated to protein sequences using the

ExPASy translate tool (Gasteiger *et al.,* 2005) and aligned with the wild-type FOXP2 (LeuZip - end) sequence using Protein BLAST (Altschul *et al.,* 1990) (left-hand side). The boxes indicate the presence of the R401D, A539P and R401D/A539P mutations. Sequence numbering is relative to FOXP2 (LeuZip - end). Regions with poor signal (X) occurred at the ends of the sequences.

3.2.2 PROTEIN EXPRESSION TRIALS

In order to obtain sufficient protein to conduct downstream studies protein expression trials were conducted on R401D, A539P and R401D/A539P FOXP2 (LeuZip - end). The trials were used to screen different temperatures and durations of overexpression, and IPTG (inducer of overexpression) concentrations, to identify conditions for optimal overexpression of soluble proteins. As seen in Figure 3.10, the highest yields of soluble proteins were observed at 20 °C and 0.2 mM IPTG for all three mutants. R401D FOXP2 (LeuZip - end) overexpression peaked after 5 hours, while overexpression of A539P and R401D/A539P FOXP2 (LeuZip - end) peaked after 22 hours. R401D FOXP2 (LeuZip - end) overexpression was soluble in T7 Express pLysS Competent *E. coli* cells, whereas A539P and R401D/A539P FOXP2 (LeuZip - end) overexpression was soluble in BL21 (DE3) Competent *E. coli* cells. Although significant amounts of protein were observed in the insoluble fractions, purification from the soluble fractions yielded sufficient protein to conduct downstream studies so recovering protein from the insoluble fractions was not necessary. Wild-type FOXP2 (LeuZip - end) overexpression was conducted according to the conditions that were previously optimised (30 °C, 0.5 mM IPTG, 4 hours) (Thulo, 2019). All mutants were overexpressed at a lower IPTG concentration and lower temperature than the wild-type.

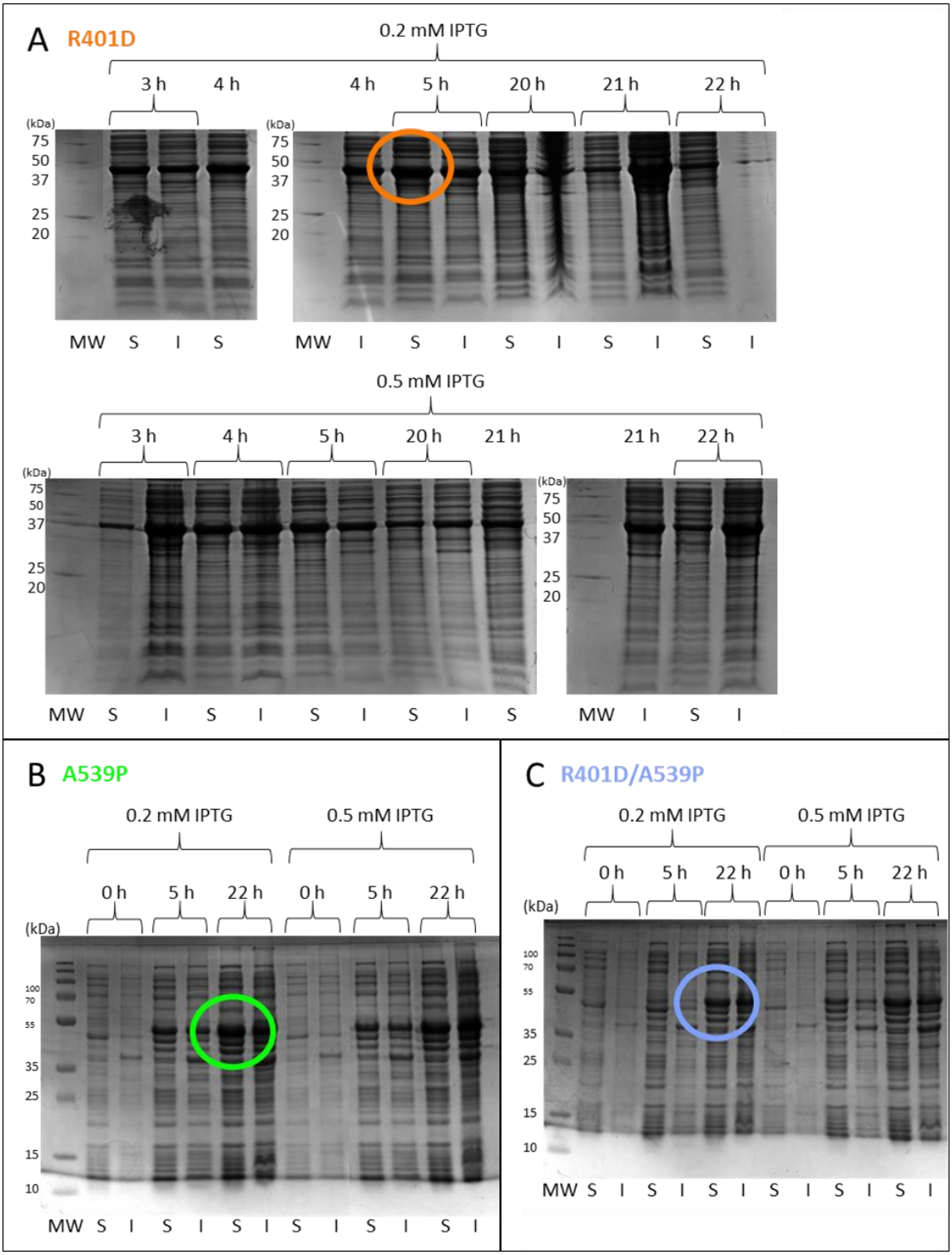


Figure 3.10: R401D (A), A539P (B) and R401D/A539P (C) FOXP2 (LeuZip - end) expression trials. Expression trials were conducted to optimise expression conditions in order to achieve the highest yields of soluble proteins. Trials were conducted at 0.2 mM and 0.5 mM IPTG, with samples collected at different time points (up to 22 hours) as indicated above the images. Results shown are for trials conducted at 20 °C. Soluble (S) and insoluble (I) fractions of each sample were separated as indicated below the images. The molecular weight markers (MW) are labelled according to the band sizes (left-hand side of each image). Ideal expression conditions are indicated in the circles (refer to Table 2.2).

3.2.3 PROTEIN PURIFICATION

All FOXP2 (LeuZip - end) variants were overexpressed with C-terminal His-tags. The presence of the His-tags allowed for simple purification from soluble cell lysate fractions by IMAC. The IMAC elution profiles are shown in Figure 3.11. Following the flow-through of unbound contaminant proteins, the salt/detergent wash step resulted in an increase in conductance and absorbance. Elution of the proteins with a high concentration of imidazole resulted in the final peaks, which were the sharpest peaks observed on all elution profiles. Protein yields were typically the highest for A539P FOXP2 (LeuZip - end) and the lowest for R401D/A539P FOXP2 (LeuZip - end). On average, the protein yields per litre of *E. coli* cell culture were approximately 12 mg for wild-type and R401D FOXP2 (LeuZip - end), 47 mg for A539P FOXP2 (LeuZip - end) and 6.5 mg for R401D/A539P FOXP2 (LeuZip - end).

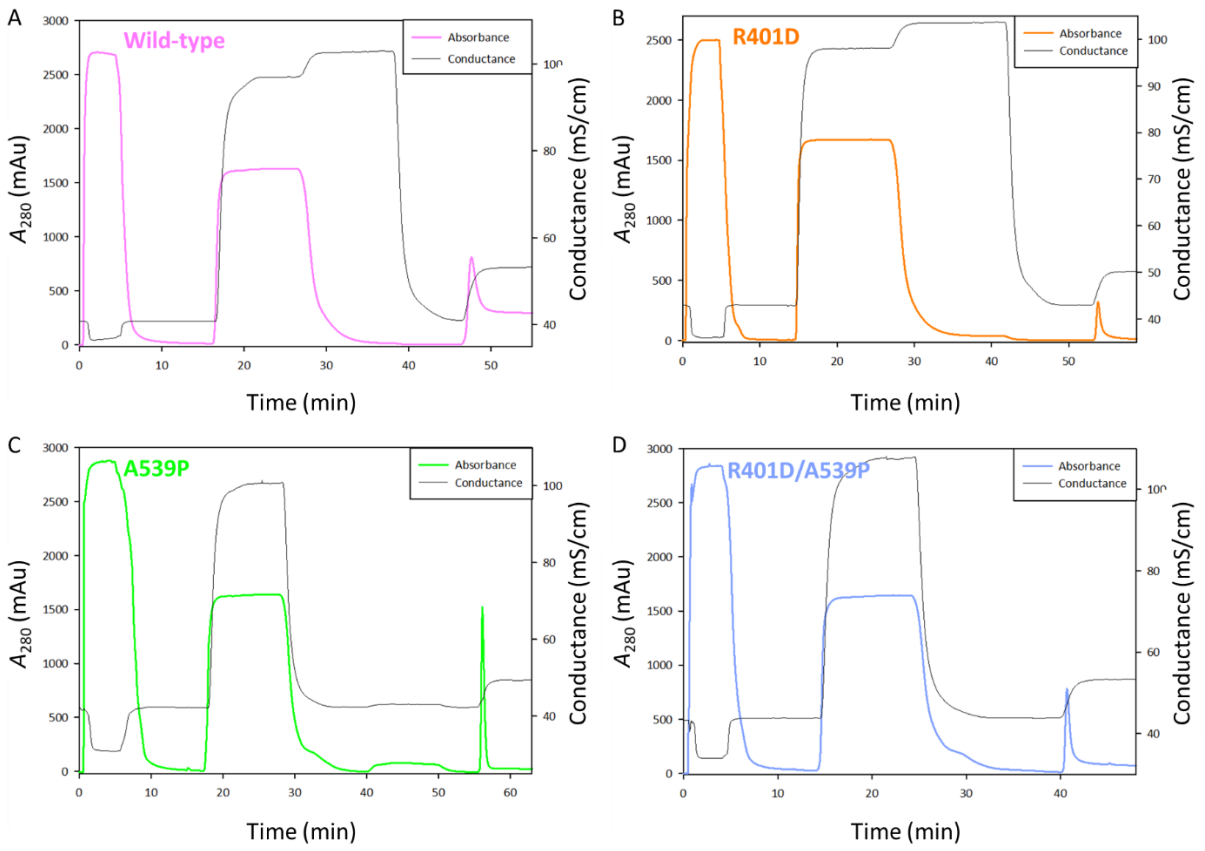


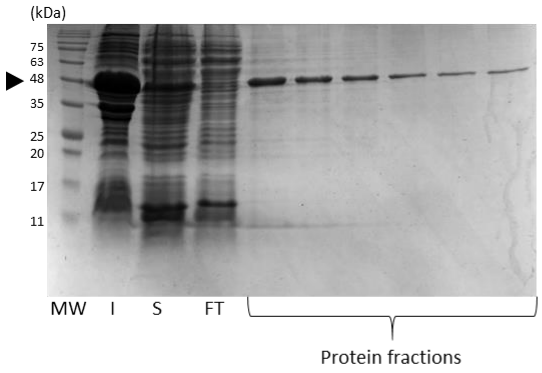
Figure 3.11: Immobilised metal-affinity chromatography purification of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). The absorbance at 280 nm is shown in pink, orange, green, and blue for each variant as labelled, and the conductance is shown in black. The initial peaks in absorbance corresponding with drops in conductance are due to the flow-through of unbound contaminant proteins. The second peaks in conductance and absorbance (± 15 minutes) are due to the conductance of NaCl and the absorbance of Triton™ X-100 in the salt/detergent wash. The imidazole wash step is most visible in the absorbance elution profile of A539P FOXP2 (LeuZip - end) (C), appearing as a flat peak (± 40 minutes). The proteins were eluted using 500 mM imidazole, resulting in the final peaks (≥ 40 minutes).

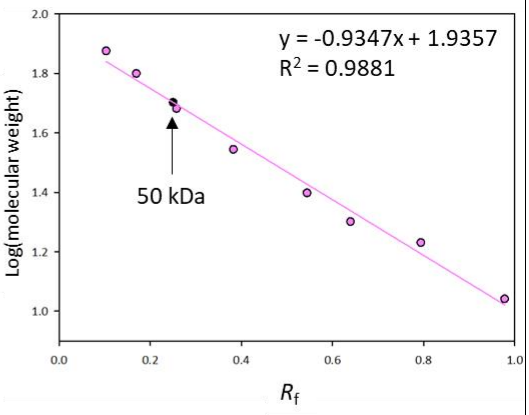
3.2.4 ASSESSMENT OF PROTEIN PURITY AND CONCENTRATION

Various quality control measures were taken to ensure that data obtained was reliable and reproducible. One such measure was the assessment of protein purity. Samples of the IMAC purification steps, including the eluted protein fractions, were collected and resolved using SDS-PAGE (Figure 3.12). Overexpression of the FOXP2 (LeuZip - end) variants was observed in both insoluble and soluble fractions of the cell lysates. Resolved samples of unbound proteins which eluted through the columns immediately (flow-through) showed that all FOXP2 (LeuZip - end) proteins that were loaded onto the columns bound to the columns, since no bands are seen corresponding with the FOXP2 (LeuZip - end) proteins. The identities of the pure protein bands were confirmed by sequencing using liquid chromatography-mass spectrometry. Protein bands of wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) corresponded with 50 kDa, 47 kDa, 48 kDa and 47 kDa, respectively. The molecular weights calculated using the online ExPASy ProtParam tool (Gasteiger *et al.,* 2005) based on the amino acid sequences were ~ 41 kDa for all FOXP2 (LeuZip - end) variants. Thus, the FOXP2 (LeuZip - end) variants all migrated as larger proteins than expected. This could be attributed to the FOXP2 (LeuZip - end) proteins having a higher proportion of basic amino acid residues in comparison with the proteins of the molecular weight marker, resulting in an increased apparent molecular weight, or to post-translational modifications of the FOXP2 (LeuZip - end) proteins, resulting in increased molecular weights compared to the expected molecular weights.

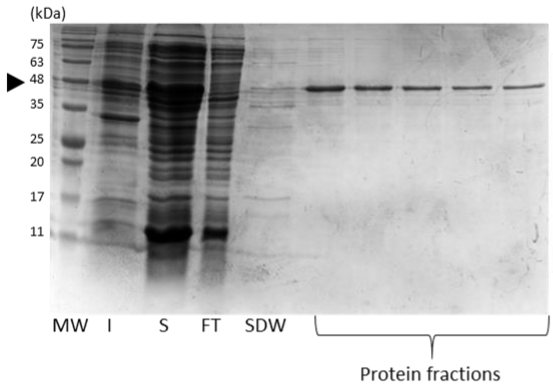
Protein concentration and purity with respect to DNA and RNA contamination were assessed by absorbance spectroscopy. An absorbance spectrum of each protein was measured between 240 nm and 350 nm (Figure 3.13). The maximum absorbance of all FOXP2 (LeuZip - end) variants occurred at 280 nm, which correlates with the maximum absorbance of tryptophan and tyrosine residues, and confirms the presence of protein in the samples. The absorbance measurements at 280 nm were used to calculate the protein concentrations based on the Beer-Lambert law. DNA and RNA absorb UV light at 260 nm, therefore the absorbance at 260 nm was used to ensure that samples were sufficiently free of nucleic acid contamination. All FOXP2 (LeuZip - end) variants had an A_{280}/A_{260} of 1.6 - 1.7. Absorbance spectra showed no turbidity at 340 nm indicating that the samples were sufficiently free of protein aggregates.

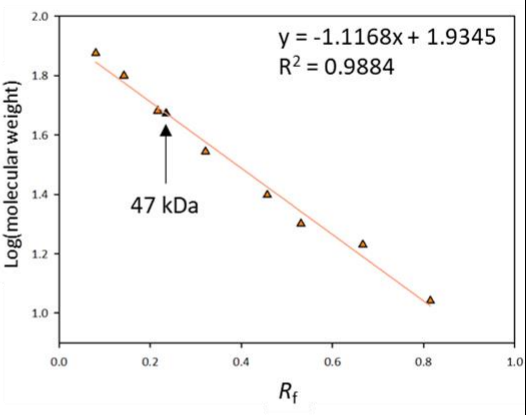
A Wild-type



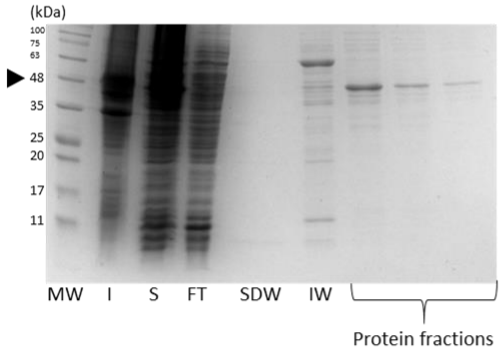


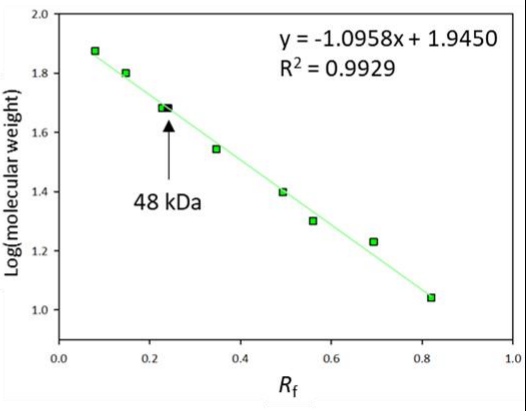
B R401D



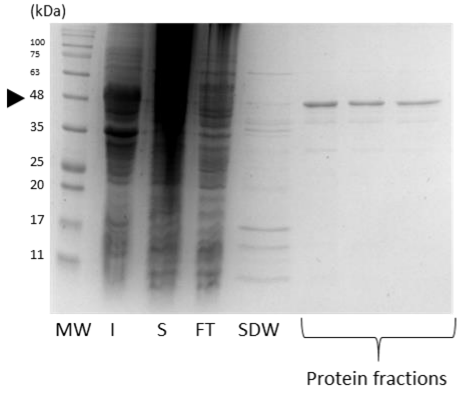


C A539P





D R401D/A539P



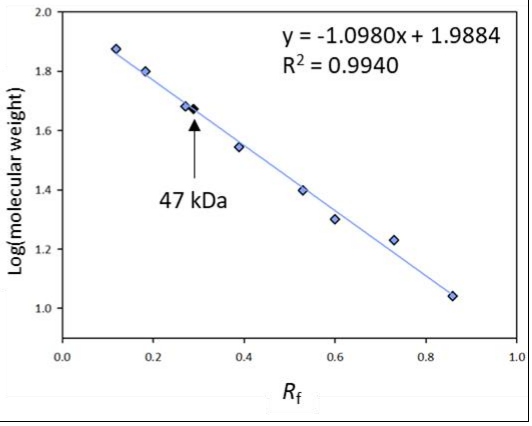


Figure 3.12: Glycine SDS-PAGE analyses of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end) purifications. Samples were resolved using 12.5% resolving gels (left-hand side). Bands of interest corresponding with the FOXP2 (LeuZip - end) proteins are indicated with triangles (left-hand side of the gels). Samples included unbound contaminant proteins (FT or flow-through), the salt/detergent wash (SDW) and the imidazole wash (IW). Protein overexpression was observed in both soluble (S) and insoluble (I) cell lysate fractions. All target proteins bound to the columns efficiently since no FOXP2 (LeuZip - end) proteins were detected in the FTs. Contaminant proteins were removed by the SDW and IW, as seen in samples of the SDW and IW. Eluted protein fractions were at least 90% pure when quantified by densitometry. The identities of the protein fractions were confirmed by sequencing using liquid chromatography-mass spectrometry. The molecular weight markers (MW) are labelled according to the band sizes (left-hand side of each gel image). Molecular weight markers were used to construct standard curves (right-hand side) by plotting the logarithm of each molecular weight against the relative mobility (R_f) for that band. Points representing the FOXP2 (LeuZip - end) variants are shown in black and labelled according to the calculated molecular weights.

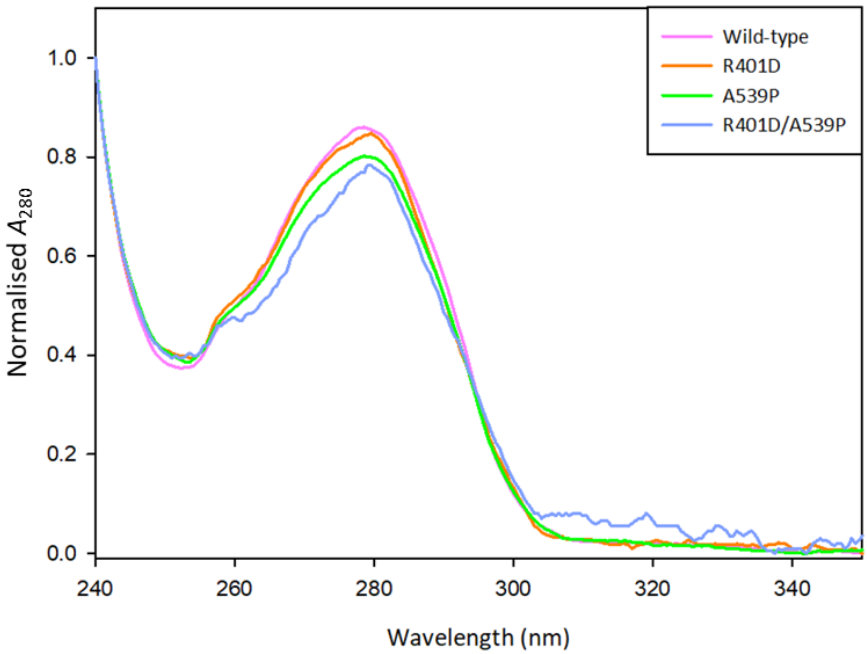


Figure 3.13: Absorbance spectra of wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end). Data points were normalised between 0 and 1. Absorbance maxima are observed at 280 nm for all variants, predominantly due to tryptophan residues in the protein samples. Noise in the absorbance spectrum of R401D/A539P FOXP2 (LeuZip - end) is due to a low protein concentration resulting in a weak signal. The purity of the protein samples with respect to DNA and RNA contamination was sufficiently high, with A_{280}/A_{260} ratios of 1.6 - 1.7. Turbidity at 340 nm was negligible indicating that there were no detectable protein aggregates present in any samples.

3.2.5 Confirmation of DNA binding

In order for FOXP2 to function as a TF it needs to bind to DNA. The cognate DNA sequence to which the FOXP2 DNA-binding FHD binds specifically and with high affinity has been previously identified (Nelson *et al.*, 2013; Webb *et al.*, 2017). After assessing protein purity and concentration, a further quality control measure was required to ensure that all FOXP2 (LeuZip - end) proteins were correctly

folded. This was achieved by testing the ability of the proteins to bind the cognate DNA sequence. The DNA sequence was synthesised as double-stranded oligonucleotides by Integrated DNA Technologies (USA) with random flanking sequences (see 2.3.1 Cognate DNA sequence). EMSAs are a quick and easy way of observing DNA binding in proteins, and were used to confirm the DNA binding of the FOXP2 (LeuZip - end) variants (Figure 3.14). All FOXP2 (LeuZip - end) variants were able to bind DNA which was confirmed by the appearance of DNA-protein complex bands that were shifted relative to free DNA bands, and all variants were therefore correctly folded.

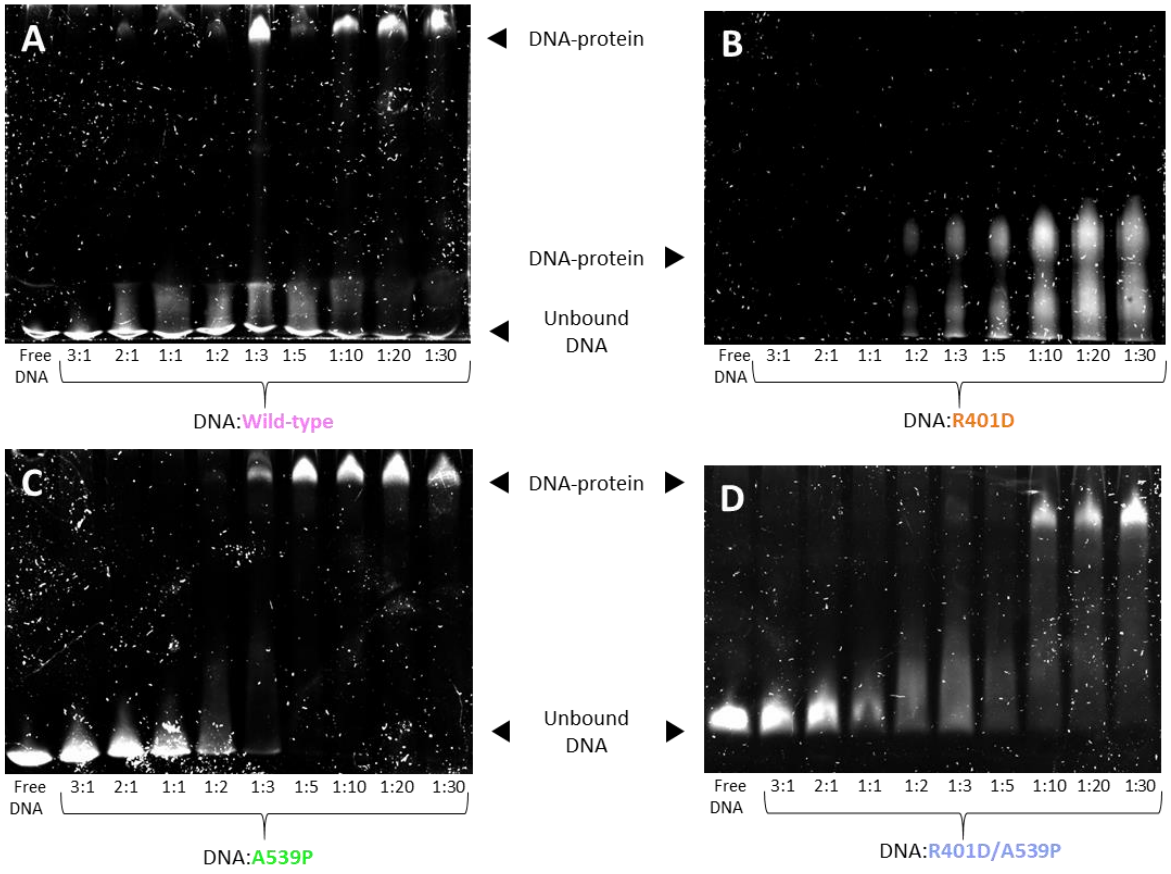


Figure 3.14: Electrophoretic mobility shift assays confirming DNA binding of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). A DNA concentration of 0.2 µM was used. The first lane of each assay contained free DNA, and the concentration ratio of protein to DNA was subsequently increased in each lane from 3:1 to 1:30. Sample compositions are indicated below each lane. Lower bands represent unbound DNA. The presence of the upper bands which represent DNA-protein complexes confirms that the FOXP2 (LeuZip - end) variants were all folded correctly, and hence able to bind DNA. The R401D FOXP2 (LeuZip - end) EMSA was slightly overrun, but DNA-protein complexes were visible, confirming that the protein was folded correctly.

3.3 CHARACTERISATION OF PROTEIN STRUCTURE AND STABILITY

According to the *in silico* studies, there were no major differences in the secondary structures of the proteins overall, and only minor differences in the folds of the proteins. However, inaccuracies may arise in the results of these studies, since they are based on predictions, and since the predictions rely on multiple sequence alignments which could reduce the sensitivity of these methods to the effects of mutations, due to the likelihood of similarities between the multiple sequence alignments of the wild-type and mutant proteins. Furthermore, the predictions could not assess any changes occurring upon DNA binding. Therefore, the protein structures were examined using *in vitro* techniques. The results of the *in vitro* structural characterisations were compared with those obtained from *in silico* studies, for a thorough characterisation of the structure of each FOXP2 (LeuZip - end) variant. Additionally, the *in vitro* experiments allowed for the examination of the oligomeric states, thermal stabilities, and dynamics of the proteins.

Solvent-accessible cysteine residues could allow for the formation of intra- or intermolecular disulfide bonds, which are capable of altering the tertiary and quaternary structures of proteins (Song *et al.*, 2007). The FOXP2 (LeuZip - end) proteins each contain four cysteine residues. Therefore, prior to conducting downstream studies, SwissPdb Viewer v4.1.0 was used to assess the solvent-accessibility of the cysteine residues in the FOXP2 (LeuZip - end) models (Guex and Peitsch, 1997). Solvent-accessible cysteine residues were defined as residues with a minimum of 30% solvent-accessible surface area. The results are shown in Figure 3.15. Based on the 3D protein structure predictions, Cys685 of wild-type FOXP2 (LeuZip - end), Cys381 of A539P and R401D/A539P FOXP2 (LeuZip - end), and Cys662 of R401D/A539P FOXP2 (LeuZip - end) were solvent-accessible. However, as with the solvent-accessibility of Gln384, it should be noted that due to the dynamic nature of proteins, the FOXP2 (LeuZip - end) proteins are likely to sample many conformations, especially since the high proportions of disordered regions would confer conformational freedom to the FOXP2 (LeuZip - end) proteins (Dunker *et al.*, 2008; Fisher and Stultz, 2011). The predicted structures represent only a single proposed monomeric conformation of each protein. Thus, the solvent-accessibility of the cysteine residues could fluctuate in a conformation-dependent manner. Nevertheless, based on the predicted structures, the FOXP2 (LeuZip - end) proteins could form intra- or intermolecular disulfide bonds. In order to investigate the effects of any disulfide bond formation, *in vitro* studies were conducted under both non-reducing and reducing conditions where relevant.

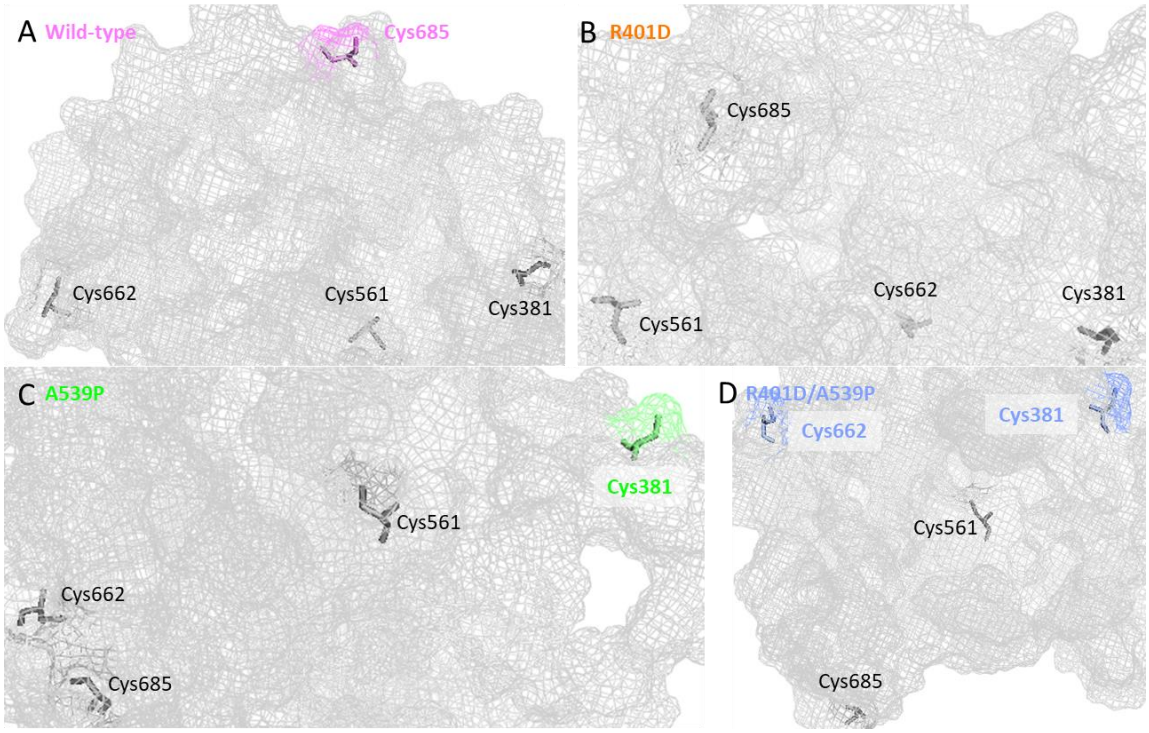


Figure 3.15: Solvent-accessibility of cysteine residues in predicted structures of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). The FOXP2 (LeuZip - end) proteins are shown in surface mesh representation and cysteine residues are shown in stick representations. Solvent-accessibility was defined as residues with a minimum of 30% solvent-accessible surface area. Cys685 is solvent-accessible in wild-type FOXP2 (LeuZip - end) (pink), Cys381 is solvent-accessible in A539P FOXP2 (LeuZip - end) (green) and R401D/A539P FOXP2 (LeuZip - end) (blue), and Cys662 is solvent-accessible in R401D/A539P FOXP2 (LeuZip - end) (blue). All other cysteine residues of all FOXP2 (LeuZip - end) variants are buried (dark grey). Three-dimensional protein structure predictions were obtained using I-TASSER (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

3.3.1 CIRCULAR DICHROISM SPECTROPOLARIMETRY

The secondary structure predictions and 3D structure predictions provided insight into probable protein secondary structures, but the accuracy of the results obtained depends on the accuracy of the predictions. Therefore, far-UV CD spectropolarimetry was used as an additional method to assess the secondary structures of the FOXP2 (LeuZip - end) proteins. Spectra were recorded on 5 µM protein in the absence and presence of 2.5 µM DNA, under reducing conditions (for non-reducing conditions see Appendix A in Chapter 5 Appendices).

The typical spectrum of an α-helical protein has a double negative band at 208 nm and 222 nm, and a positive band near 190 nm (Greenfield, 2006a). The typical spectrum of a disordered protein has low ellipticities at most wavelengths, with a negative band at approximately 195 nm (Greenfield, 2006a). All spectra had clear negative bands at 222 nm, both in the absence and presence of DNA (Figure 3.16).

Although the spectra were unlike the typical spectrum of an α-helical protein, the negative bands at 222 nm can only be attributed to α-helices since no other secondary structure produces this band (Greenfield, 2006a). At wavelengths below and above 222 nm, the spectra had low ellipticities and were close to 0 deg·cm^{2}·dmol^{-1}, which is typical of disordered proteins (Greenfield, 2006a). The bands at 208 nm and 190 nm that are typical of α-helical proteins did not occur in any of the spectra, suggesting that the FOXP2 (LeuZip - end) proteins had a large proportion of disordered regions, which presumably masked the appearance of these bands. Thus, for each protein, the combination of secondary structural elements (α-helices and loops) produced a CD spectrum that is a cross between the characteristic spectrum of each element.

Therefore, based on a visual assessment of the spectra, the proteins consisted predominantly of disordered regions, with a small proportion of α-helices, both in the absence and presence of DNA, and DNA binding did not appear to have any substantial effect on the secondary structures of the proteins. The results obtained were in agreement with secondary and 3D protein structure predictions, which predicted the proteins to be predominantly disordered with a few α-helical regions, most of which occurred within the LeuZip and FHD (Figure 3.1, 3.2, and 3.7). Although secondary structure predictions and 3D structure predictions of all FOXP2 (LeuZip - end) proteins predicted two or three β-strands within the FHD, according to the predictions the β-strand content was low (~ 3% for all FOXP2 (LeuZip - end) variants) and was therefore not expected to contribute to the far-UV CD spectra.

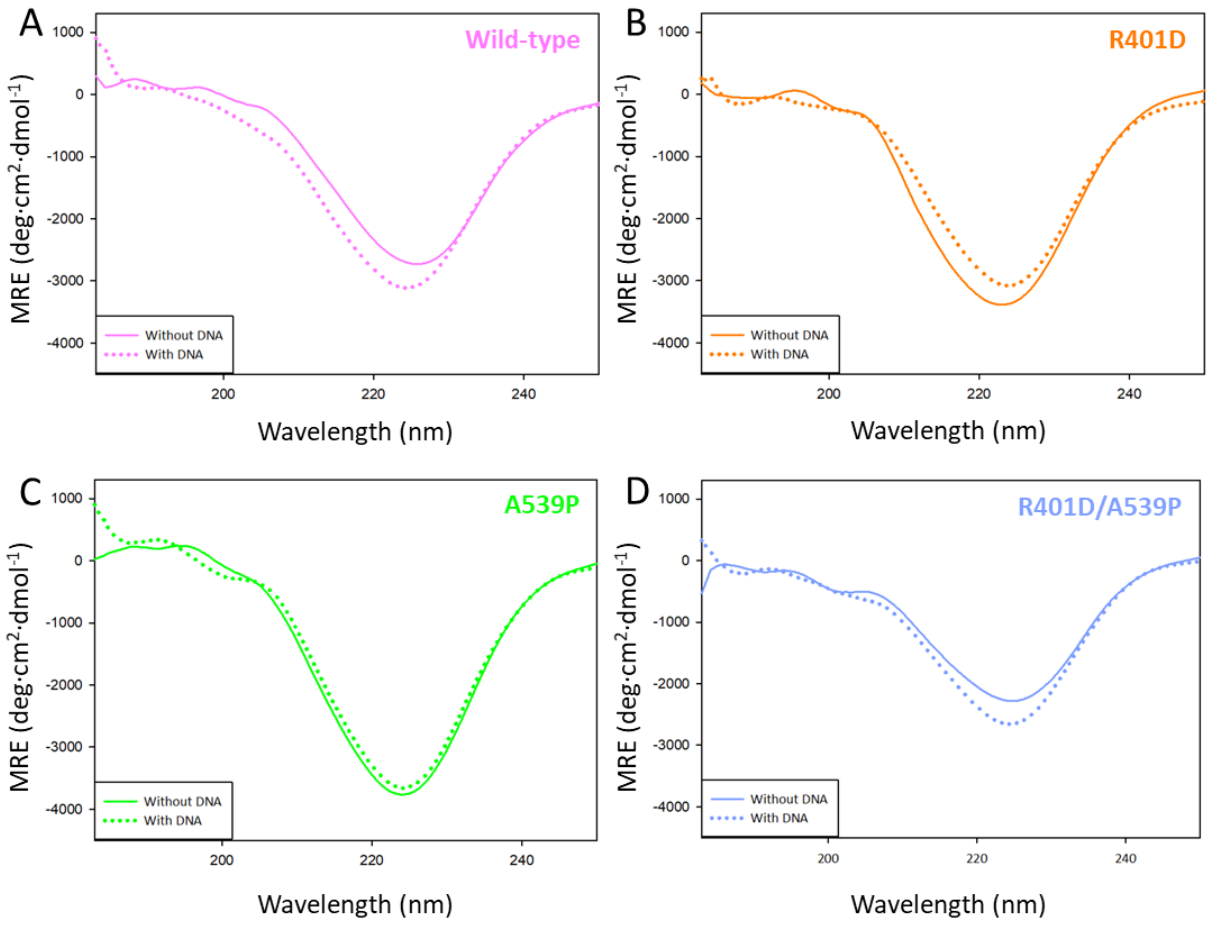


Figure 3.16: Circular dichroism spectra of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). Spectra were measured on 5 µM protein, in the absence (solid lines) and presence (dotted lines) of 2.5 µM DNA at 2 mM DTT. Data points with a high tension of over 600 V were removed. Based on DichroWeb analyses, wild-type FOXP2 (LeuZip - end) showed a small decrease in β-strands ($p < 0.05$) and a small increase in unordered regions ($p < 0.01$) upon DNA binding (Whitmore and Wallace, 2004, 2008). No other significant changes occurred upon DNA binding of all FOXP2 (LeuZip - end) variants.

The CD spectra were dissected using DichroWeb to quantify the proportion of each secondary structural element present in each protein (Whitmore and Wallace, 2004, 2008). A summary of the DichroWeb data is shown in Figure 3.17. The DichroWeb data showed a very low α-helical content (~ 4%), with a β-strand content of ~ 34% and a combined content of turns and unordered regions (loops) of ~ 64% for the FOXP2 (LeuZip - end) proteins. It is likely that the DichroWeb algorithm underestimated the proportion of α-helices in the FOXP2 (LeuZip - end) proteins, as a consequence of the missing positive band near 190 nm that is typical of α-helical proteins. However, overlooking the discrepancy, according to the DichroWeb analyses there were no significant changes to the secondary structures of R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) upon DNA binding. DNA binding resulted in a minor loss of β-strands and minor increase in disorder in wild-type FOXP2 (LeuZip - end),

which can be attributed to structural rearrangements occurring upon DNA binding. In the absence of DNA, R401D and A539P FOXP2 (LeuZip - end) each had a slightly higher proportion of α-helices and a slightly lower proportion of β-strands than the wild-type. In addition, the A539P mutant was more disordered than wild-type FOXP2 (LeuZip - end), in the absence of DNA. R401D/A539P FOXP2 (LeuZip - end) showed no significant differences to the wild-type in the absence of DNA and a slightly higher proportion of β-strands in the presence of DNA.

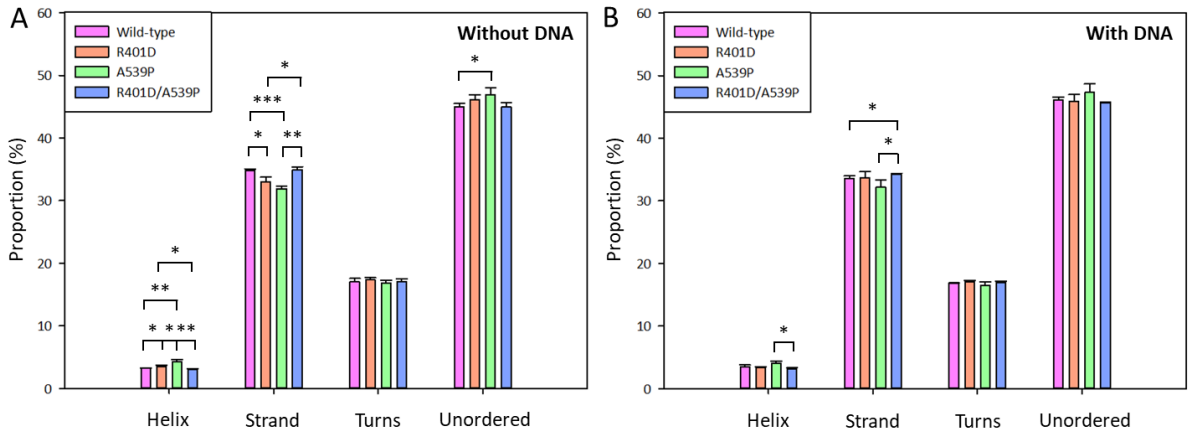


Figure 3.17: Analyses of secondary structural elements of FOXP2 (LeuZip - end) variants in the absence (A) and presence (B) of DNA. Quantitative analyses of data obtained from far-UV CD spectropolarimetry at 2 mM DTT indicating the proportion of α-helices, β-strands, turns, and unordered regions in each FOXP2 (LeuZip - end) variant. Statistical significance of differences between the proportions of secondary structural elements are indicated by two-sample t-tests (*p < 0.05, ** p < 0.01, and *** p < 0.001). Analyses were computed using DichroWeb (Whitmore and Wallace, 2004, 2008).

A summary of the DichroWeb data is shown in Table 3.2. Summaries of data obtained from secondary structure predictions and 3D protein structure predictions (with respect to secondary structural content) are included for comparison. Results of secondary and 3D structure predictions were fairly similar, predicting an α-helical content of ~ 30% and loop content of ~ 70% for all FOXP2 (LeuZip - end) variants, with a small proportion of β-strands (~ 3%). The predicted 3D structure of A539P FOXP2 (LeuZip - end) was an exception as it was less α-helical (21%) and more looped (76%). DichroWeb data showed a similar proportion of loops (~ 63%) compared with secondary and 3D structure predictions. However, DichroWeb appeared to underestimate α-helical content and overestimate β-strand content.

Table 3.2: Comparison of secondary structural evaluations of FOXP2 (LeuZip - end) proteins. The results of DichroWeb analyses of far-UV CD spectropolarimetry data (in the absence of DNA, under reducing conditions), PSIPRED 4.0 protein secondary structure predictions, and I-TASSER predictions of 3D protein structures, with respect to secondary structural content, are summarised for comparison.

		DichroWeb [1]	PSIPRED [2]	I-TASSER [3]
Wild-type FOXP2	α-Helix (%)	3	29	32
(LeuZip - end)	β-strand (%)	35	3	2
	Loop (%)	62*	68	66
A539P FOXP2	α-Helix (%)	4	30	21
(LeuZip - end)	β-strand (%)	33	3	3
	Loop (%)	63*	67	76
R401D FOXP2	α-Helix (%)	4	29	31
(LeuZip - end)	β-strand (%)	32	3	3
	Loop (%)	64*	69	66
R401D/A539P FOXP2	α-Helix (%)	3	30	29
(LeuZip - end)	β-strand (%)	35	3	3
	Loop (%)	62*	67	68

* DichroWeb classifies turns and unordered regions as distinct secondary structures. These are combined as "loop".

[1] Whitmore and Wallace, 2004, 2008; [2] Jones, 1999 ; [3] Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015

3.3.2 INTRINSIC TRYPTOPHAN FLUORESCENCE SPECTROSCOPY

The folds or tertiary structures of the FOXP2 (LeuZip - end) proteins were studied by intrinsic tryptophan fluorescence spectroscopy. This allowed for assessing the effects of the R401D, A539P and R401D/A539P mutations on the global protein structure, and hence the contributions of the LeuZip and FHD interfaces to the protein fold. Furthermore, intrinsic tryptophan fluorescence spectroscopy was used to assess any changes to the protein tertiary structures occurring upon DNA binding, specifically with regards to the environments of the tryptophan residues.

The FOXP2 (LeuZip - end) proteins each contain three tryptophan residues. Although all three FOXP2 (LeuZip - end) tryptophan residues are located within the FHD, regions outside of the FHD could come in close proximity to the tryptophan residues depending on the 3D folds of the proteins, especially when not bound to DNA. Therefore, intrinsic tryptophan fluorescence could be affected by global changes in protein fold, which can affect the local environments of the tryptophan residues. Based on the 3D protein structure predictions, the tryptophan residues of the FOXP2 (LeuZip - end) proteins

interact with regions outside of the FHD. Since charged amino acid residues are known to quench intrinsic tryptophan fluorescence, some instances of charged amino acid residues that are not within the FHD but are in close proximity to the FOXP2 (LeuZip - end) tryptophan residues are shown in Figure 3.18 (Lakowicz, 2006b).

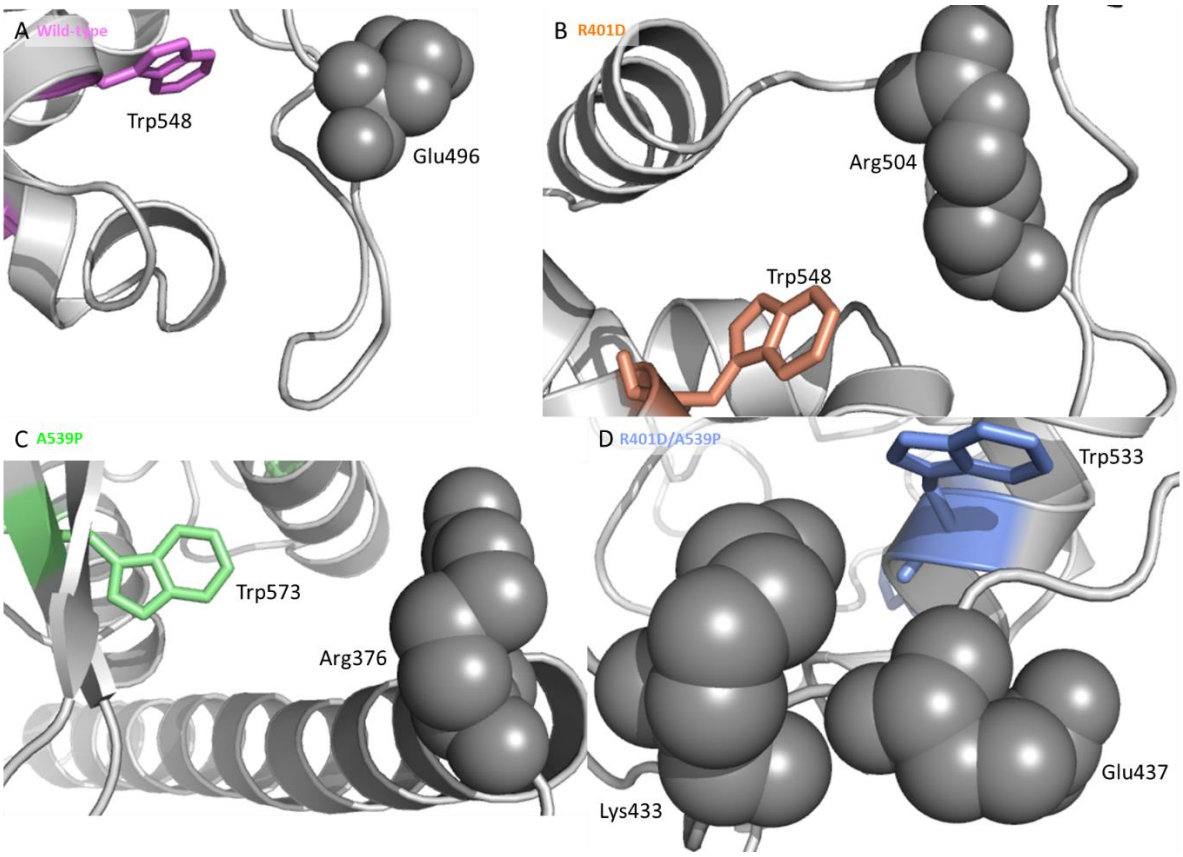


Figure 3.18: Charged residues not located within the forkhead domain could affect intrinsic tryptophan fluorescence. **(A)** Trp548 of wild-type FOXP2 (LeuZip - end) (pink) comes in close proximity to the region between the LeuZip and FHD. Glu496 is located in this region. **(B)** Trp548 of R401D FOXP2 (LeuZip - end) (orange) comes in close proximity to the region N-terminal to the FHD. Arg504 is located in this region, immediately N-terminal of helix 1. **(C)** Trp573 of A539P FOXP2 (LeuZip - end) (green) comes in close proximity to the LeuZip. Arg376 is located immediately N-terminal of the LeuZip. **(D)** Trp533 of R401D/A539P (blue) FOXP2 (LeuZip - end) comes in close proximity to the region between the LeuZip and FHD. Lys433 and Glu437 are located in this region. All charged residues are shown in sphere representation (dark grey). Three-dimensional protein structure predictions were obtained using I-TASSER (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

Intrinsic tryptophan fluorescence spectroscopy was conducted using 5 µM protein in the absence and presence of 2.5 µM DNA, under reducing conditions (for non-reducing conditions see Appendix B in Chapter 5 Appendices). The spectrum of each protein had a double peak both in the absence and presence of DNA (Figure 3.19A). The characteristic double peak can be attributed to differences in the environments of the individual tryptophan residues. Peaks at lower wavelengths (~ 329 nm) imply a

more hydrophobic environment such as a hydrophobic pocket on the protein interior, whereas peaks at higher wavelengths (~ 339 nm) imply a more hydrophilic environment and could occur due to fluorescence of tryptophan residues that are solvent-exposed (Vivian and Callis, 2001). Since the peaks were similar for all FOXP2 (LeuZip - end) variants, there were no major differences in the environments of the tryptophan residues between the variants. Therefore, the FOXP2 (LeuZip - end) proteins all exhibited similar folds, in the absence of DNA.

Fluorescence quenching occurred upon DNA binding of all FOXP2 (LeuZip - end) proteins (Figure 3.19A). This was expected since DNA is known to quench fluorescence through photoinduced electron transfer (Lakowicz, 2006a). During the transfer, DNA donates electrons to tryptophan residues, especially via guanine bases which are good electron donors (Lakowicz, 2006a). Comparisons between the maximum intensities at 339 nm of the DNA-bound proteins normalised relative to the unbound proteins were plotted as a bar graph (Figure 3.19B). In the presence of DNA, the normalised fluorescence intensities of R401D and A539P FOXP2 (LeuZip - end) were significantly higher than that of the wild-type. Therefore, differences were observed in the environments of the tryptophan residues between the DNA-bound proteins, possibly due to differences in the DNA-protein contacts formed for each protein.

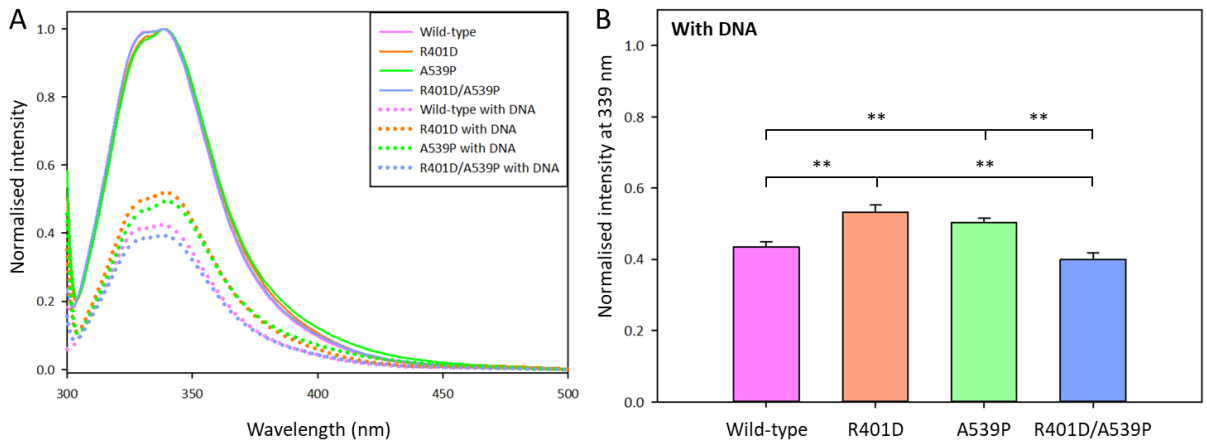


Figure 3.19: Intrinsic tryptophan fluorescence spectra of FOXP2 (LeuZip - end) variants (A) and comparison of the emission maxima in the presence of DNA (B). Spectra were measured on 5 µM protein, in the absence (solid lines) and presence (dotted lines) of 2.5 µM DNA at 2 mM DTT **(A)**. Spectra showed no differences in the absence of DNA. Fluorescence intensities of all FOXP2 (LeuZip - end) variants were quenched significantly upon DNA binding ($p < 0.01$ for R401D FOXP2 (LeuZip - end), and $p < 0.001$ for all other variants). Data points were normalised between 0 and 1 relative to the unbound proteins, and the emission maxima at 339 nm were compared in the presence of DNA **(B)**. Statistical significance of differences between the normalised emission maxima are indicated by two-sample t-tests (*$p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$).

3.3.3 SIZE EXCLUSION CHROMATOGRAPHY

Since the R401D and A539P mutations were designed to disrupt homotypic associations of the LeuZip and FHD, respectively, it was necessary to probe the effects that these mutations had on the quaternary structure of FOXP2 (LeuZip - end). Wild-type and A539P FOXP2 (LeuZip - end) were each predicted to have one solvent-accessible cysteine residue while R401D/A539P FOXP2 (LeuZip - end) was predicted to have two solvent-accessible cysteine residues (Figure 3.15). The presence of exposed cysteine residues could allow for the formation of intermolecular disulfide bonds, thereby affecting the apparent protein quaternary structure. Therefore, experiments were conducted at 0 mM and 2 mM DTT to observe the effects, if any, of disulfide bonding on the protein quaternary structures (for non-reducing conditions see Appendix C in Chapter 5 Appendices).

The size exclusion chromatography column was calibrated with standards of known molecular weights (Figure 3.20). Standards were prepared and used in mixtures according the manufacturer's recommendations. The elution volume (V_e) of each standard was determined and the column void volume (V_o) was determined using Blue Dextran. The logarithm of the molecular weight of each standard was plotted against the ratio of V_e to V_o to produce a calibration curve. This allowed for the sizes of the FOXP2 (LeuZip - end) proteins to be estimated relative to the sizes of the standards, using the equation of the calibration curve. The equation was y = - 1.5821x + 4.4606 and the R^2 was 0.9905.

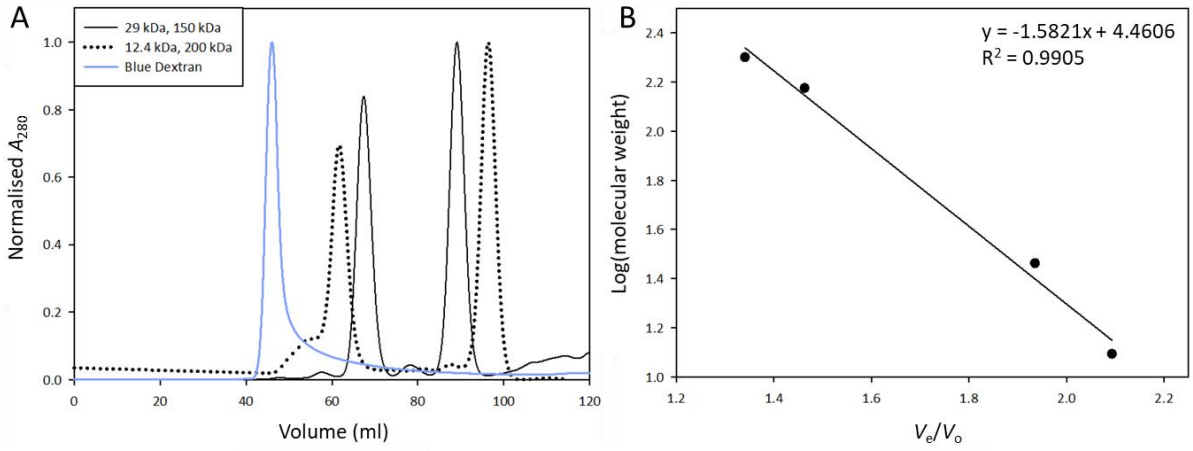


Figure 3.20: Size exclusion chromatography calibration using the Gel Filtration Markers Kit for Protein (Sigma-Aldrich, USA). (A) Elution profiles of the standards: Data points were normalised between 0 and 1. The void volume (V_o) of the size exclusion chromatography column was found to be 46.05 ml using Blue Dextran. The column was calibrated with standards at 12.4 kDa (cytochrome C), 29 kDa (carbonic anhydrase), 150 kDa (alcohol dehydrogenase), and 200 kDa (β-amylase). Elution profiles (**A**) were used to construct a calibration curve **(B)** by plotting the logarithms of the molecular weights against the ratios of the elution volumes (V_e) to V_o. The curve was fit with an R^2 of 0.9905 to the equation y = - 1.5821 x + 4.4606.

The quaternary structures of the FOXP2 (LeuZip - end) variants were assessed by size exclusion chromatography, to determine the relative sizes and numbers of species present. It is expected that the hydrodynamic volumes of the FOXP2 (LeuZip - end) proteins would be larger than globular proteins of the same molecular weights, on account of the large proportions of disordered regions observed. Although this could lead to overestimates of the protein sizes, data obtained from size exclusion chromatography would be useful for comparative purposes, especially with respect to oligomeric states. Furthermore, since all FOXP2 (LeuZip - end) proteins exhibited similar proportions of disordered regions, the effects of hydrodynamic volume would be similar for all variants. Therefore, the effects of the LeuZip and FHD mutations on the oligomeric state of FOXP2 (LeuZip - end), and hence the contribution of each interface to the protein quaternary structure, could be evaluated through size exclusion chromatography.

Samples of each protein were prepared, under reducing conditions, and loaded onto the column. The elution volume was determined for each peak and the equation of the calibration curve was used to calculate the corresponding sizes (Figure 3.21). A peak was seen in the void volume for all FOXP2 (LeuZip - end) variants, which was much more pronounced for R401D and R401D/A539P FOXP2 (LeuZip - end). This peak could appear due to the formation of a super-oligomer or large protein masses, which form more readily upon the exposure of buried LeuZip hydrophobic patches caused by incorporation of the R401D mutation. Furthermore, although it is a single peak it could represent a number of super-oligomers or protein masses of different sizes, which elute as a single peak in the void volume because each of the super-oligomers or protein masses is too large to enter the porous column resin. The peak was not due to amorphous aggregation or precipitation, since the peak was assessed using absorbance spectroscopy and there was no evidence of turbidity at 340 nm.

Ignoring the peak in the void volume, wild-type FOXP2 (LeuZip - end) consisted of a single peak at 248 kDa. The R401D mutation of the LeuZip caused a shift to 125 kDa, therefore homotypic LeuZip associations were disrupted by the R401D mutation. A539P FOXP2 (LeuZip - end) was fairly similar to wild-type FOXP2 (LeuZip - end), with a major peak at 266 kDa, therefore disrupting the FHD dimer interface alone was insufficient to cause any substantial change to the quaternary structure of the protein overall. However, a small peak was observed at 96 kDa, which increased in proportion substantially when the A539P mutation was combined with the R401D mutation in R401D/A539P FOXP2 (LeuZip - end). R401D/A539P FOXP2 (LeuZip - end) thus consisted of two peaks at 137 kDa and 97 kDa. The small peak observed at ~ 86 ml for R401D/A539P FOXP2 (LeuZip - end) was an artefact of

the column, since tests on the peak including SDS-PAGE and absorbance spectroscopy showed no evidence of there being protein in the peak.

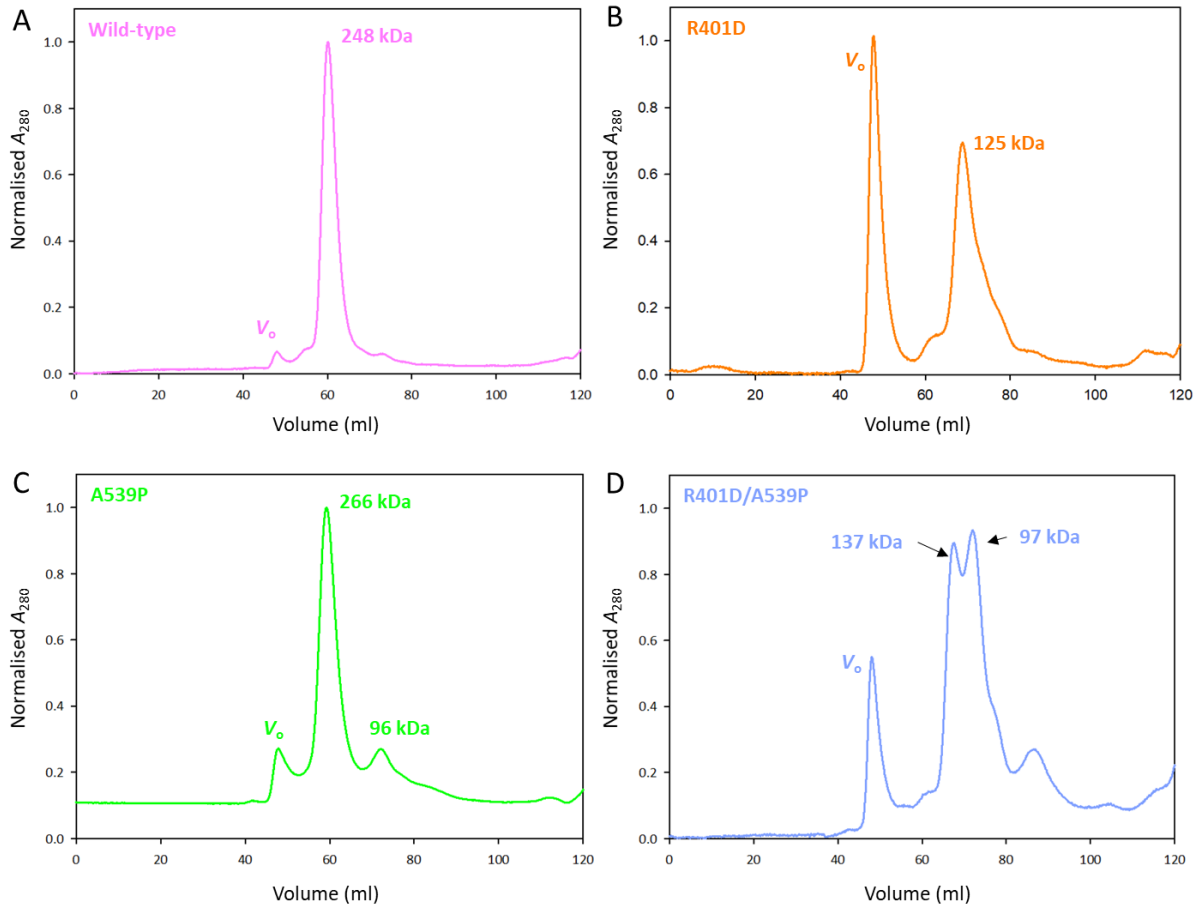


Figure 3.21: Size exclusion chromatography elution profiles of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). Samples were prepared with 25 - 35 µM protein at 2 mM DTT. Data points were normalised between 0 and 1. Wild-type FOXP2 (LeuZip - end) **(A)** consisted of a single major peak at 248 kDa. The R401D mutation **(B)** caused a shift to 125 kDa. The A539P mutation **(C)** did not cause any substantial changes on its own, with the major peak of A539P FOXP2 (LeuZip - end) corresponding to 266 kDa. However, combined with the R401D mutation **(D)**, the effects of the A539P mutation were observed with the appearance of a pronounced smaller species at 97 kDa. The peaks in the void volume (V_o) could be due to large protein masses, since these peaks are much more pronounced in R401D and R401D/A539P FOXP2 (LeuZip - end), which could be on account of incorporation of the R401D mutation enhancing the formation of large protein masses by the exposure of buried LeuZip hydrophobic patches.

Due to the FOXP2 (LeuZip - end) variants being disordered and flexible proteins, it is assumed that these proteins adopt extended, relatively unstructured folds. The apparent sizes of the various FOXP2 (LeuZip - end) oligomers as determined by size exclusion chromatography may be larger than the actual sizes, since the size exclusion chromatography column was calibrated using globular protein standards and the FOXP2 (LeuZip - end) variants are non-globular. In addition, the formation of a

higher order oligomer could magnify the effects of the disorder and flexibility from each monomer, and further increase the apparent size due to an increase in the hydrodynamic radius. As a result, a description of the oligomeric state of each species observed is speculative. Nevertheless, if one were to use the calculated molecular weights based on the amino acid sequences, and the apparent sizes of the proteins derived from SDS-PAGE, there is a possibility that wild-type FOXP2 (LeuZip - end) exists as a hexamer. The R401D mutation caused the disruption of the hexamer into trimers and the A539P mutation, especially when combined with the R401D mutation, resulted in the formation of FOXP2 (LeuZip - end) dimers. Therefore, A539P FOXP2 (LeuZip - end) consisted predominantly of a hexameric species (with an insignificant proportion of dimer), and R401D/A539P FOXP2 (LeuZip - end) consisted of a mixture of trimers and dimers. Thus, the R401D mutation of the LeuZip did not result in a monomeric species, despite the equivalent K251D mutation of mFoxp3 resulting in a monomeric species (Song *et al.*, 2012). Although this may seem surprising, it can be explained by the position of Arg401 in the LeuZip helix – Arg401 occurs at position *g* (Figure 1.7), a non-core position. This finding reveals that although the Arg401-Glu398-Gln384 hydrogen bond network is important for FOXP2 homo-oligomerisation, it contributes to the formation of higher order oligomers rather than the association of monomers. The fact that R401D FOXP2 (LeuZip - end) was not exclusively monomeric was useful, since this allowed for the effects of the A539P mutation to be apparent in R401D/A539P FOXP2 (LeuZip - end).

Attempts to separate the species by collecting the size exclusion chromatography peaks separately were unsuccessful. This was the case particularly for the 125 kDa species of R401D FOXP2 (LeuZip - end) and the 137 kDa and 97 kDa species of R401D/A539P FOXP2 (LeuZip - end), as the peak in the void volume re-appeared immediately (data not shown). Separation attempts were complicated further by an inconsistency in the resolution of the 137 kDa and 97 kDa R401D/A539P FOXP2 (LeuZip - end) peaks, causing difficulties in the separation of the species.

3.3.4 ACRYLAMIDE FLUORESCENCE QUENCHING

Acrylamide fluorescence quenching was used to study the folds and flexibilities of the FOXP2 (LeuZip - end) proteins. This allowed for the analysis of the effects of the R401D, A539P and R401D/A539P mutations on the flexibility of the FOXP2 (LeuZip - end) protein. In addition, quenching studies conducted in the absence and presence of DNA allowed for the analysis of the effects of DNA binding on the flexibility of each FOXP2 (LeuZip - end) variant. Acrylamide fluorescence quenching depends on the solvent-accessibility of the tryptophan residues (Eftink and Ghiron, 1981). Therefore, prior to conducting the experiment, SwissPdb Viewer v4.1.0 was used to assess the

solvent-accessibility of the tryptophan residues in the FOXP2 (LeuZip - end) models (Guex and Peitsch, 1997). Solvent-accessible tryptophan residues were defined as residues with a minimum of 30% solvent-accessible surface area. The results are shown in Figure 3.22. When not bound to DNA, Trp533 of wild-type FOXP2 (LeuZip - end) is predicted to be solvent-accessible. All tryptophan residues of R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) were predicted to be buried. As previously mentioned, it is worth noting that intrinsically disordered proteins are dynamic and are able to sample many conformations (Dunker *et al.*, 2008; Fisher and Stultz, 2011). Therefore, the FOXP2 (LeuZip - end) proteins would constitute an ensemble of structures, and the predicted structures represent only a single proposed conformation of each protein. Thus, the true solvent-accessibility of the tryptophan residues could differ from the predicted solvent-accessibility based on the models.

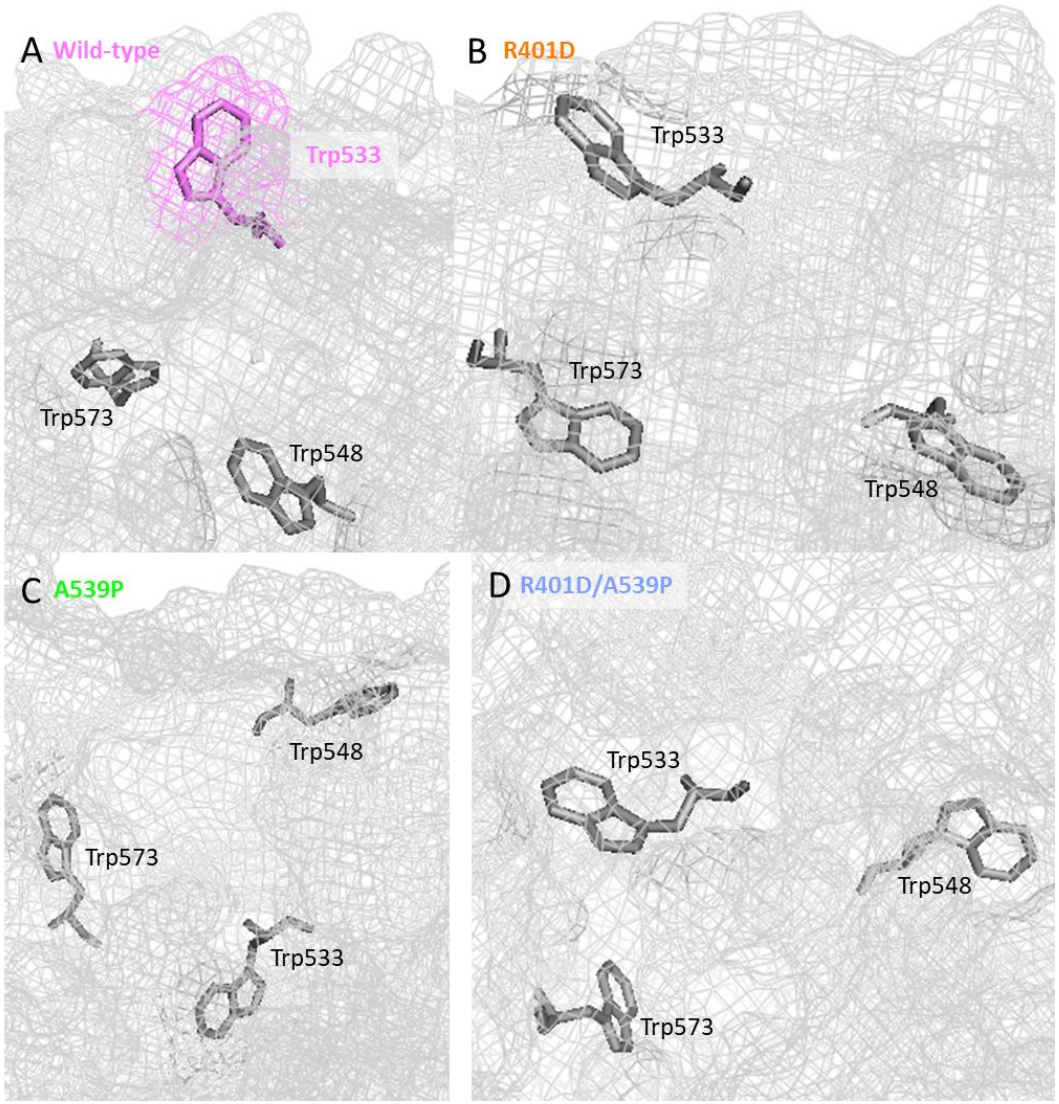


Figure 3.22: Solvent-accessibility of tryptophan residues in predicted structures of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). The FOXP2 (LeuZip - end) proteins are shown in surface mesh representation and tryptophan residues are shown in stick representation. Solvent-accessibility was defined as residues with a minimum of 30% solvent-accessible surface area. All three FOXP2 (LeuZip - end) tryptophan residues are located within the DNA-binding FHD. Trp533 is solvent-accessible in wild-type FOXP2 (LeuZip - end) (pink). All other tryptophan residues of all FOXP2 (LeuZip - end) variants are buried (dark grey). Three-dimensional protein structure predictions were obtained using I-TASSER (Roy *et al.*, 2010; Yang and Zhang, 2015; Yang *et al.*, 2015). Images were rendered using The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC.

Stern-Volmer plots (Figure 3.23) were produced by measuring the fluorescence of 5 µM protein while increasing the concentration of acrylamide from 0 M to 0.25 M. Fluorescence was monitored at an emission wavelength of 339 nm since this was the wavelength of maximum fluorescence intensity for all FOXP2 (LeuZip - end) proteins (see 3.3.2 Intrinsic tryptophan fluorescence spectroscopy). Fluorescence quenching was studied in the absence and presence of 2.5 µM DNA, under reducing conditions. In its DNA-bound form, a protein is typically less flexible and more shielded from the

solvent than in its unbound form due to the formation of DNA-protein contacts which restrict the motion of the protein and exclude the solvent (Hauser *et al.*, 2016; Huang *et al.*, 2019; Jones *et al.*, 1999; Lee *et al.*, 2008; Mathiasen *et al.*, 2016; Morris *et al.*, 2018a; Prabakaran *et al.*, 2006; Zandarashvili *et al.*, 2012). The effects of the reduced flexibility and shielding were evident as DNA binding reduced the fluorescence quenching for all FOXP2 (LeuZip - end) variants. However, the decrease in quenching observed in R401D/A539P FOXP2 (LeuZip - end) was not statistically significant. This suggests that the tryptophan residues of R401D/A539P FOXP2 (LeuZip - end) are less shielded when the protein is in its DNA-bound form relative to its unbound form, in comparison with the other variants. Furthermore, the normalised fluorescence intensity of DNA-bound R401D/A539P FOXP2 (LeuZip - end) was lower than the other FOXP2 (LeuZip - end) proteins, according to intrinsic tryptophan fluorescence spectroscopy (Figure 3.19). Typically, the quantum yield of tryptophan fluorescence decreases in a more polar environment, implying that the tryptophan residues of the DNA-bound R401D/A539P FOXP2 (LeuZip - end) were, indeed, less shielded and more solvent-exposed in comparison with the other proteins (Lotte *et al.,* 2004). Taken together, these results show that R401D/A539P FOXP2 (LeuZip - end) could differ with respect to the DNA-protein contacts formed upon DNA binding, or with respect to DNA-binding affinity.

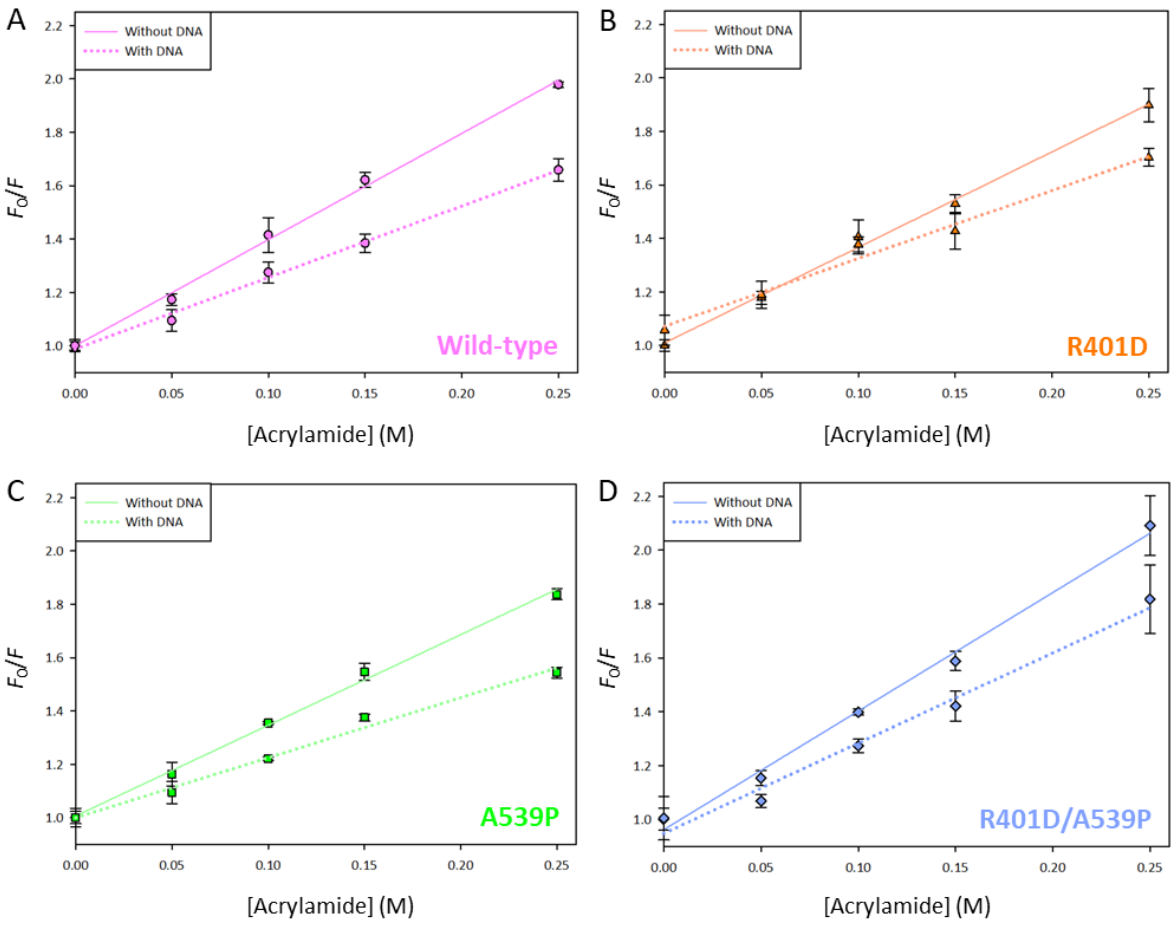


Figure 3.23: Stern-Volmer plots for acrylamide fluorescence quenching of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). Each protein was maintained at a concentration of 5 µM, and fluorescence quenching at 339 nm was monitored between 0 M and 0.25 M acrylamide, in the absence (solid lines) and presence (dotted lines) of 2.5 µM DNA at 2 mM DTT. Stern-Volmer plots were used to calculate Stern-Volmer constants. The Stern-Volmer constant of R401D/A539P FOXP2 (LeuZip - end) did not change significantly upon DNA binding ($p > 0.05$). The Stern-Volmer constants of all other FOXP2 (LeuZip - end) variants decreased significantly upon DNA binding ($p < 0.001$ for wild-type FOXP2 (LeuZip - end), $p < 0.05$ for R401D FOXP2 (LeuZip - end), and $p < 0.01$ for A539P FOXP2 (LeuZip - end)).

The Stern-Volmer plots were used to calculate a Stern-Volmer constant for each protein in its unbound and DNA-bound state (Figure 3.24). The Stern-Volmer constants of A539P FOXP2 (LeuZip - end) was significantly lower than those of the wild-type in the unbound and DNA-bound forms, so the tryptophan residues of A539P FOXP2 (LeuZip - end) were less solvent-accessible than those of the wild-type. This could be due to a difference in the conformation of A539P FOXP2 (LeuZip - end) in comparison with the wild-type, which resulted in the increased surface area predicted by the 3D protein models (Figure 3.7) and increased hydrodynamic radius observed by size exclusion chromatography (Figure 3.21). In the absence of DNA, the Stern-Volmer constant of R401D/A539P FOXP2 (LeuZip - end) was significantly higher than that of the wild-type, suggesting that the tryptophan residues of this mutant were more solvent-accessible than those of the wild-type.

Wild-type and R401D/A539P FOXP2 (LeuZip - end) adopt different quaternary structures (Figure 3.21), which could account for the increased Stern-Volmer constant of the mutant. Despite differences in quaternary structures, the Stern-Volmer constants of R401D FOXP2 (LeuZip - end) did not differ significantly from those of the wild-type in both unbound and DNA-bound states. Thus, although the tryptophan residues of wild-type FOXP2 (LeuZip - end), according to 3D protein structure predictions, were more solvent-accessible than those of the mutants (Figure 3.22), Stern-Volmer constants revealed that the tryptophan residues of R401D/A539P FOXP2 (LeuZip - end) were the most solvent-accessible. However, the solvent-accessibility of the tryptophan residues could differ between monomeric and oligomeric states, and the models represent only a single proposed monomeric conformation for each protein.

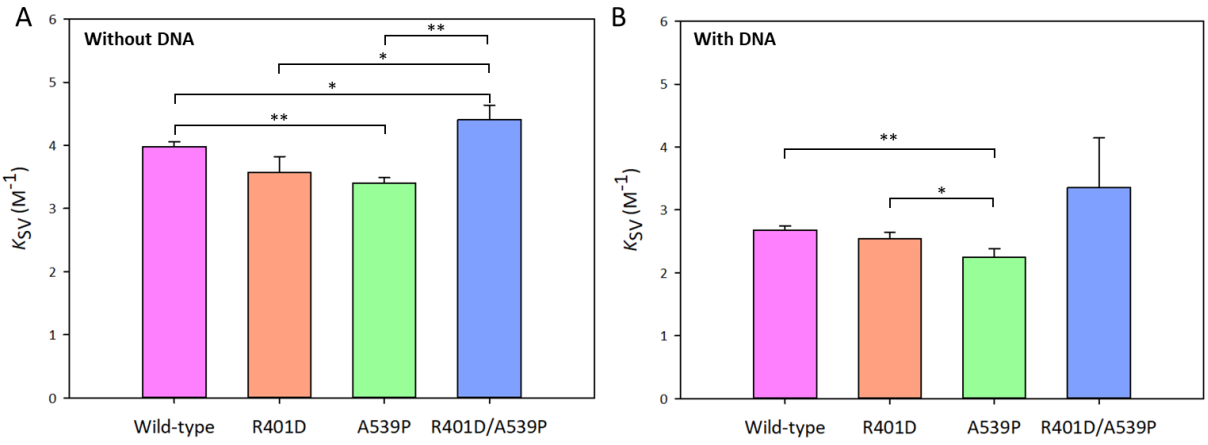


Figure 3.24: Comparison of Stern-Volmer constants for fluorescence quenching of FOXP2 (LeuZip - end) variants in the absence (A) and presence (B) of DNA. Stern-Volmer constants were 3.98 ± 0.11 M^{-1}, 3.57 ± 0.13 M^{-1}, 3.40 ± 0.09 M^{-1} and 4.40 ± 0.18 M^{-1}, for wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) in the absence of DNA, respectively. The Stern-Volmer constants in the presence of DNA were 2.68 ± 0.11 M^{-1}, 2.54 ± 0.15 M^{-1}, 2.25 ± 0.10 M^{-1} and 3.35 ± 0.23 M^{-1}, for wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end), respectively. Statistical significance of differences between Stern-Volmer constants are indicated by two-sample t-tests (*$p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$).

3.3.5 PROTEIN THERMAL UNFOLDING

The relative stabilities of the FOXP2 (LeuZip - end) proteins were evaluated using far-UV CD spectropolarimetry to monitor thermal unfolding. This was used to probe the effects of the R401D, A539P and R401D/A539P mutations on the thermal stability of FOXP2 (LeuZip - end). A comparison of the thermal stabilities of the proteins allowed for evaluating the importance of the LeuZip and FHD interfaces for structural stability. Furthermore, this enabled a comparison of the thermal stabilities of the different oligomers observed for each FOXP2 (LeuZip - end) variant by size exclusion chromatography, and hence the stabilities of the interactions involved in the formation of each oligomer. The CD spectrum of each FOXP2 (LeuZip - end) protein had a negative band at 222 nm

(Figure 3.16). CD spectropolarimetry was monitored on 5 µM (Figure 3.25A) and 15 µM (Figure 3.25B) protein at 222 nm from 20 °C to 80 °C. Two protein concentrations were used to investigate the effects, if any, of concentration on thermal stability. Thermal unfolding was conducted under non-reducing conditions due to the thermal instability of DTT, thus reference is made to the structural characterisations of the proteins under non-reducing conditions (see Appendices A, B, and C in Chapter 5 Appendices).

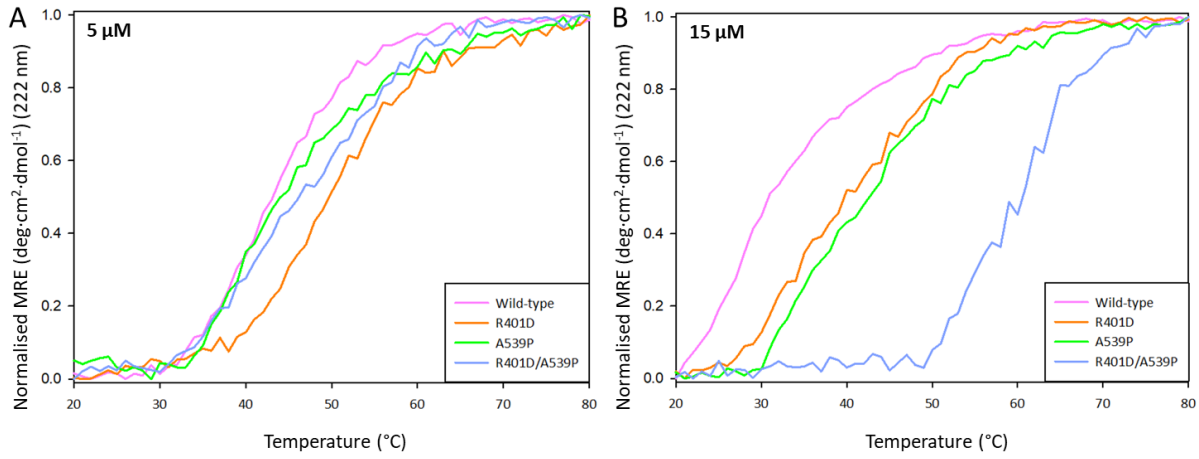


Figure 3.25: Thermal unfolding of wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end). Circular dichroism was monitored at 222 nm from 20 °C to 80 °C, on 5 µM **(A)** and 15 µM **(B)** protein. Data points were normalised between 0 and 1 deg·cm²·dmol⁻¹. A high tension of below 600 V was ensured. The unfolding curves were fairly similar at a protein concentration of 5 µM. At 15 µM protein, wild-type FOXP2 (LeuZip - end) was the least thermally stable, since unfolding of the wild-type commenced immediately at 20 °C. At the same concentration, R401D/A539P FOXP2 (LeuZip - end) was the most thermally stable, since unfolding of this mutant was initiated at a high temperature (~ 50 °C). The differences in thermal stability could be on account of differences in the oligomeric states of the variants, since wild-type FOXP2 (LeuZip - end) forms a large species of approximately hexameric size, whereas R401D/A539P FOXP2 (LeuZip - end) is predominantly an approximately dimeric species.

At a protein concentration of 5 µM, the unfolding curves were fairly similar. At a protein concentration of 15 µM, there was more variation in the unfolding curves. Unfolding of wild-type FOXP2 (LeuZip - end) commenced immediately at 20 °C, and the wild-type exhibited the lowest thermal stability. Unfolding of R401D/A539P FOXP2 (LeuZip - end), on the other hand, was initiated close to 50 °C, and this mutant exhibited the highest thermal stability. Furthermore, unfolding of R401D/A539P FOXP2 (LeuZip - end) exhibited the most cooperativity, with a more distinct pre-transition region relative to the other FOXP2 (LeuZip - end) variants. Overall, the order of stabilities was R401D/A539P > A539P ≥ R401D > wild-type, arranged from most to least stable. Since there were only minor differences in the protein secondary structures (see Figure A1 and A2 in Chapter 5 Appendices), and no apparent differences in the protein tertiary structures (see Figure B1 in Chapter 5 Appendices),

differences in stability are most likely attributable to differences in the oligomeric states of the FOXP2 (LeuZip - end) proteins (see Figure C1 in Chapter 5 Appendices). Thus, it is plausible that the large FOXP2 (LeuZip - end) hexamer is less thermally stable than the smaller oligomers. Unfolding of the wild-type hexamer is initiated at lower temperatures, as the hexamer is first broken down into smaller oligomers, finally resulting in a dimer and perhaps eventually a monomer which then unfolds completely and loses all secondary structure. R401D/A539P FOXP2 (LeuZip - end) already exists predominantly as dimers which are more thermally stable. The unfolding of dimers, perhaps first to monomers which then undergo further unfolding, is only initiated at higher temperatures (~ 50 °C). Furthermore, the unfolding of dimers is more cooperative than the unfolding of higher order oligomers, which could be attributed to the interactions involved in the folding of the monomer and formation of the dimer being stronger than those involved in the formation of higher order oligomers (van Holde *et al.*, 1998). Although predominantly hexameric, A539P FOXP2 (LeuZip - end) has a small proportion of dimer which is not present in R401D FOXP2 (LeuZip - end) and could contribute to the slightly higher thermal stability seen in A539P FOXP2 (LeuZip - end). Thermal unfolding resulted in complete protein aggregation at both 5 µM and 15 µM protein concentrations.

An increase in protein concentration increases macromolecular crowding, a phenomenon that has opposing effects on proteins. Macromolecular crowding can result in the destabilisation of proteins as a consequence of excluded-volume effects, resulting in aggregation (Jarrett and Lansbury Jr, 1993; Minton, 1981, 2000; Wang *et al.,* 2012; Zimmerman and Minton, 1993). In other instances, macromolecular crowding can increase protein chemical and thermal stability since the unfolded state occupies a larger volume than the folded state, therefore macromolecular crowding enhances the stability of the native state (Christiansen *et al.,* 2010; Homouz *et al.,* 2008, 2009; Perham *et al.,* 2007). The effect that macromolecular crowding has on a protein is affected by the size of the protein and the magnitude of conformational changes required to transition between unfolded and native states (Christiansen *et al.,* 2010; Dong *et al.,* 2010). Interestingly, the stabilisation or destabilisation of proteins under crowded conditions also depends on protein-protein and protein-DNA interactions of the unfolded state, in comparison with the native state (Danielsson *et al.*, 2015). It was observed that the thermal stabilities of wild-type, R401D and A539P FOXP2 (LeuZip - end) decreased with an increase in protein concentration, whereas the thermal stability of R401D/A539P FOXP2 (LeuZip - end) increased with an increase in protein concentration. The opposing effects could be explained by the sizes of species formed for each FOXP2 (LeuZip - end) variant. The initial breakdown of the larger oligomers (wild-type, R401D and A539P FOXP2 (LeuZip - end)), presumably into smaller oligomers which eventually form a monomer before unfolding completely, could entail smaller conformational

changes in comparison with the complete unfolding and loss of secondary structure of the monomer itself. Furthermore, a shift was observed in the thermal stability of R401D FOXP2 (LeuZip - end) relative to the other variants between 5 µM and 15 µM, since this variant was the most thermally stable protein at 5 µM and the third most thermally stable protein at 15 µM. This could be due to the protein-protein interactions of the unfolded and native state differing for R401D FOXP2 (LeuZip - end) compared to the other FOXP2 (LeuZip - end) variants. The observation that the R401D mutant differs in its protein-protein interactions in its native state, since it readily forms large protein masses to a greater extent than any of the other variants (see Figure C1 in Chapter 5 Appendices), could hint at different protein-protein interactions for this mutant in the unfolded state as well.

3.4 DNA-BINDING STUDIES

Like other TFs, FOXP2 exerts its function of controlling gene expression through interacting with and binding to DNA. Based on structural characterisations, DNA binding appeared to influence some aspects of secondary and tertiary structure, as well as the dynamics, of the FOXP2 (LeuZip - end) variants. Furthermore, differences were observed between the wild-type and mutant FOXP2 (LeuZip - end) proteins when bound to DNA. These findings suggest that FOXP2 homo-oligomerisation mediated via the LeuZip and FHD interfaces could influence DNA binding. Therefore, DNA-binding studies were conducted to assess the effects of the R401D, A539P and R401D/A539P mutations on the interactions of FOXP2 (LeuZip - end) with DNA.

3.4.1 ELECTROPHORETIC MOBILITY SHIFT ASSAYS

Size exclusion chromatography was used to study the quaternary structures of the FOXP2 (LeuZip - end) proteins, and showed differences between the proteins, especially in the R401D and R401D/A539P mutants compared to the wild-type (Figure 3.21). Since size exclusion chromatography could only be used to evaluate the oligomeric states of the unbound proteins, EMSAs were used to probe the oligomeric states of the proteins when bound to DNA. This was used to compare the number of DNA-bound species present for each protein with the number of unbound species observed through size exclusion chromatography. Samples containing mixtures of DNA and protein were prepared and resolved by electrophoresis. Unbound DNA migrated more quickly than bound DNA, and bound DNA migrated at different speeds depending on the overall size of the DNA-protein complex. This allowed for the separation of DNA-protein complexes of different sizes representing DNA bound by protein in different oligomeric states. EMSAs of wild-type and A539P FOXP2 (LeuZip - end) were conducted under non-reducing conditions. EMSAs of R401D and R401D/A539P FOXP2 (LeuZip - end) were conducted under both non-reducing and reducing conditions, since the quaternary structures of these

proteins differed substantially under non-reducing and reducing conditions when studied by size exclusion chromatography (for non-reducing conditions see Appendix D in Chapter 5 Appendices).

Images of the EMSAs are shown in Figure 3.26. EMSAs of wild-type, R401D and A539P FOXP2 (LeuZip - end) showed only two band positions; a single position representing a single species of protein in complex with DNA and a second position representing free DNA, which migrated more quickly and was therefore lower on each gel. The DNA-protein complexes of R401D/A539P FOXP2 (LeuZip - end) resolved as double bands, which indicated the presence of two DNA-bound protein species. Bands representing free DNA occurred lower on the gel. EMSA results were in agreement with size exclusion chromatography results, since both techniques showed a single major species for wild-type, R401D and A539P FOXP2 (LeuZip - end), and two major species for R401D/A539P FOXP2 (LeuZip - end). Although the sizes of the species observed by the EMSAs could not be determined, it is plausible that the same protein species form in both unbound and DNA-bound configurations.

There was no evidence of the size exclusion chromatography peaks that occurred in the void volume, and represented large protein masses, on the EMSAs. This could be due to a number of reasons. The first possibility is that the peaks occurring in the void volume could represent the sum of many small peaks representing a myriad of super-oligomers or protein masses of different sizes that were each too large to be resolved by size exclusion chromatography. In this instance, the protein masses would have separated on the EMSAs and if there were only very small amounts of each protein mass present, the signal from each protein mass would be too low and undetectable. The second possibility is that the protein masses were simply too large to enter the pores of the gels. The third possibility is that, in the presence of DNA, the super-oligomers or protein masses dissociate into smaller species such as dimers, trimers, and hexamers, and therefore form part of the DNA-protein complexes that were visualised on the EMSAs.

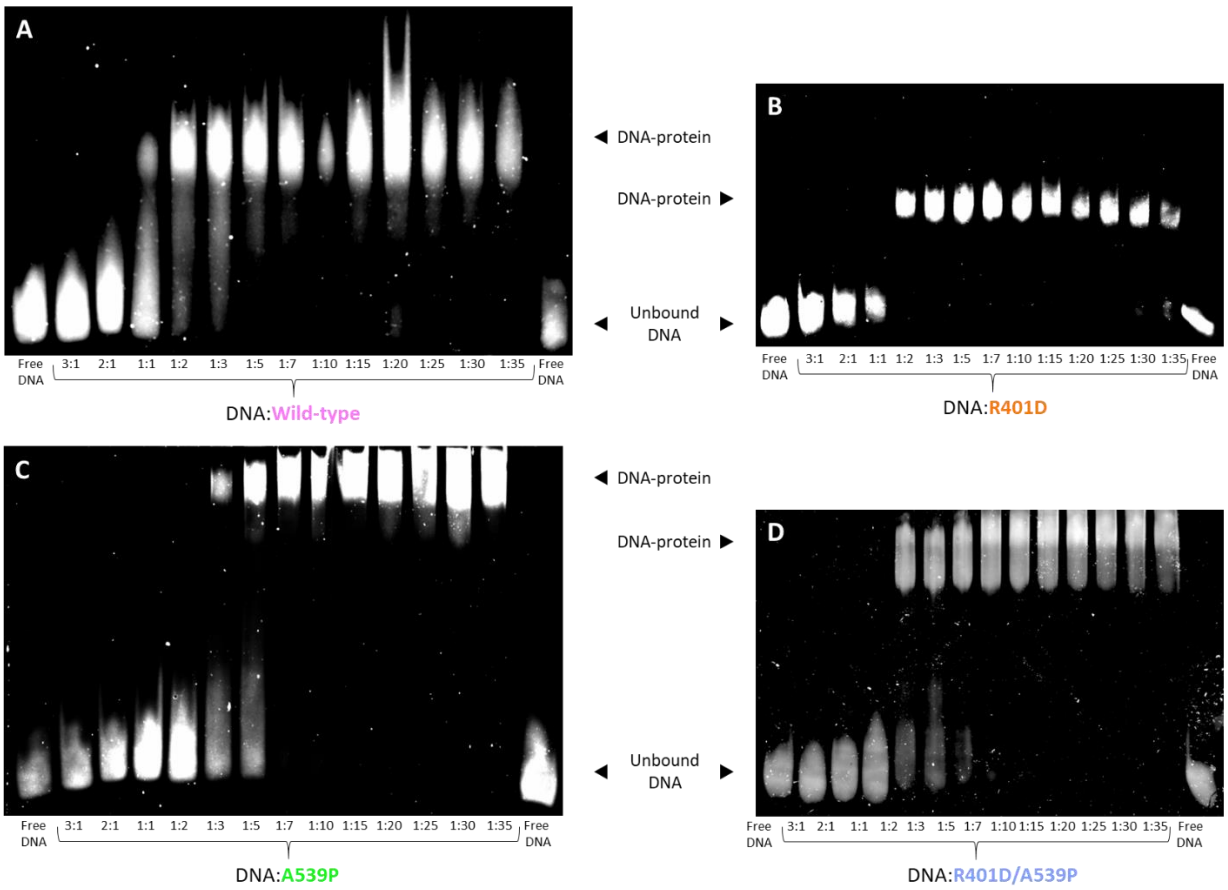


Figure 3.26: Electrophoretic mobility shift assays comparing the number of DNA-bound species present of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end). A DNA concentration of 0.5 µM was used. The first and last lanes of each assay contained free DNA, and the concentration ratio of protein to DNA was increased in the remaining lanes from 3:1 to 1:35. Sample compositions are indicated below each lane. Wild-type and A539P FOXP2 (LeuZip - end) samples were prepared at 0 mM DTT, while R401D and R401D/A539P FOXP2 (LeuZip - end) samples were prepared at 2 mM DTT. Lower bands represent unbound DNA and upper bands represent DNA-protein complexes. Wild-type, R401D and A539P FOXP2 (LeuZip - end) each bound DNA as a single species, whereas the double bands seen for R401D/A539P FOXP2 (LeuZip - end) were indicative of two species.

Under non-reducing conditions, the DNA-protein complex was seen at a DNA:protein ratio of 1:1 for wild-type, 1:2 for R401D, 1:3 for A539P and 1:5 for R401D/A539P FOXP2 (LeuZip - end) (see Figure D1 in Chapter 5 Appendices for R401D and R401D/A539P FOXP2 (LeuZip - end)). Therefore, the DNA-binding affinities can be estimated to be wild-type > R401D > A539P > R401D/A539P, arranged from strongest to weakest affinity. However, although an EMSA can be used to investigate and quantify DNA-binding affinity (Heffler *et al.*, 2012; Perumal *et al.*, 2015), it is a low resolution technique, so a higher resolution fluorescence anisotropy DNA-binding assay was used to quantify the DNA-binding affinities of the proteins.

3.4.2 FLUORESCENCE ANISOTROPY DNA-BINDING ASSAYS

The results of EMSAs and size exclusion chromatography confirmed that although the R401D, A539P and R401D/A539P mutations had effects on the quaternary structure of FOXP2 (LeuZip - end), the FOXP2 (LeuZip - end) variants were all able to bind DNA. EMSAs suggested that the mutations could affect the DNA-binding affinity of FOXP2 (LeuZip - end), so fluorescence anisotropy assays were conducted to quantify the DNA-binding affinities of the proteins. DNA-binding assays were conducted under non-reducing conditions, to allow for comparison with previous assays conducted on the isolated FOXP2 FHD in the same environment (for reducing conditions see Appendix E in Chapter 5 Appendices).

Labelled DNA (with a 5-carboxy-X-rhodamine fluorescent marker at the 5' end) at a concentration of 200 nM was combined with the proteins at concentrations of up to 6000 nM. The resulting fluorescence anisotropy signals were measured and used to plot DNA-binding isotherms (Figure 3.27). The K_Ds for DNA binding were 347 ± 31.6 nM, 460 ± 30.0 nM, 611 ± 45.4 nM and 778 ± 33.3 nM, for wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end), respectively. Therefore, the arrangement of the DNA-binding affinities is wild-type > R401D > A539P > R401D/A539P, from strongest to weakest affinity. This was in agreement with EMSA results. Statistical analyses showed that the decrease in DNA-binding affinity due to the R401D mutation was less significant than the decrease in DNA-binding affinity observed for A539P and R401D/A539P FOXP2 (LeuZip - end), relative to the wild-type. Therefore, disrupted hexameric LeuZip associations had a less significant effect on DNA-binding affinity, whereas the prevention of domain swapping of the DNA-binding FHD decreased DNA-binding affinity more significantly. The combined effects of both mutations had the most significant effect on DNA-binding affinity, since R401D/A539P FOXP2 (LeuZip - end) had the weakest affinity for DNA.

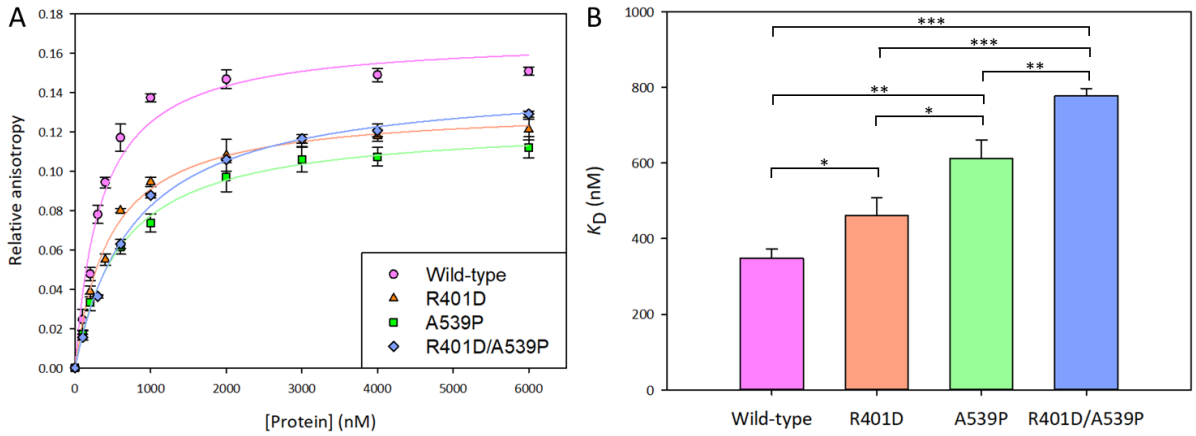


Figure 3.27: Fluorescence anisotropy DNA-binding assays of wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end). (A) DNA-binding isotherms: The concentration of DNA was maintained at 200 nM, and the protein concentration was increased from 0 nM to 6000 nM. Data points were normalised relative to the anisotropy of free DNA. Isotherms were fit with a single-site model. **(B)** Dissociation constants derived from the isotherms showed that wild-type FOXP2 (LeuZip - end) bound DNA with the highest affinity (K_D = 347 ± 31.6 nM) and R401D/A539P FOXP2 (LeuZip - end) bound DNA with the lowest affinity (K_D = 778 ± 33.3 nM). DNA-binding affinities of R401D FOXP2 (LeuZip - end) (K_D = 460 ± 30.0 nM) and A539P FOXP2 (LeuZip - end) (K_D = 611 ± 45.4 nM) were intermediate to that of wild-type and R401D/A539P FOXP2 (LeuZip - end). Statistical significance of differences between K_Ds are indicated by two-sample t-tests (*p < 0.05, ** p < 0.01, and *** p < 0.001).

Chapter 4: Discussion

Oligomerisation is a common means of regulating TF activity, and hence downstream gene expression (Funnell and Crossley, 2012). TFs associate to form both homo- and hetero-oligomers (Funnell and Crossley, 2012). The simplest TF oligomer is a dimer, but large TF oligomers do exist; a TF octamer has been observed through X-ray crystallography and a TF hexadecamer has been detected *in vitro* (Chen *et al.*, 2001; Eliahoo *et al.*, 2015; Funnell and Crossley, 2012). TF oligomerisation has diverse effects; for example, oligomerisation is often a prerequisite for DNA binding and hence activity or, in some cases, each TF binding partner dictates binding specificity or controls the switch between transcriptional activation and repression or, in other cases still, oligomer formation regulates nuclear trafficking (Bujalka *et al.,* 2013; Grandori *et al.,* 2000; Hinde *et al.,* 2016; Kim *et al.,* 2017; Li *et al.,* 2013b; Lüscher, 2001; Vitelli *et al.,* 2000). For FOXP TFs in particular, oligomerisation regulates the downstream effects of transcriptional activity and the inhibition of oligomerisation has been shown to prevent FOXP function (Chae *et al.,* 2006; Li *et al.,* 2007, 2004; Lopes *et al.,* 2006; Lozano *et al.,* 2017; Ren *et al.,* 2019; Rudra *et al.,* 2012). There is evidence of the existence of oligomers as large as tetramers for all FOXP proteins (Li *et al.*, 2007; Mendoza, 2011; Song *et al.*, 2012).

The focus of this research was to explore the oligomerisation interfaces of FOXP2, and hence of the FOXP proteins. Previous studies have emphasised the importance of the FOXP LeuZip for homo-oligomerisation, and for hetero-oligomerisation amongst FOXP subfamily members (Chae *et al.,* 2006; Li *et al.,* 2004; Wang *et al.,* 2003). However, the propensity of the FOXP FHD to form domain-swapped dimers, which is unique to the FOXP subfamily of FOX TFs, could further contribute to FOXP oligomerisation (Bandukwala *et al.,* 2011; Chu *et al.,* 2011; Stroud *et al.,* 2006). Key residues for FOXP3 LeuZip oligomerisation and FOXP2 FHD dimerisation have previously been identified (Morris and Fanucchi, 2016; Perumal *et al.,* 2015; Song *et al.,* 2012; Stroud *et al.,* 2006). Based on these residues, this research aimed to use point mutations at each interface, to investigate the relative contributions of the LeuZip and FHD to the oligomerisation of FOXP2. The R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) variants used in this study are novel constructs, and allowed for a thorough characterisation of FOXP2 homo-oligomerisation.

There are different factors that determine the propensity for oligomerisation, and a number of forces that drive oligomerisation. Large, multi-domain proteins have evolved due to the biological need for a small number of proteins to be able to fulfil a large number of cellular functions (Jelesarov and Karshikoff, 2008). However, with increasing protein size comes an increasing probability of

translational errors, since estimates of translational errors are as frequent as one error every 1000 - 10 000 codons (Ogle and Ramakrishnan, 2005; Parker, 1989). The formation of oligomers from smaller monomers decreases translational errors whilst enabling the same number of proteins to perform many more cellular functions (Jelesarov and Karshikoff, 2008). Other factors that often, but not always, drive oligomerisation are increased thermodynamic stability of the oligomer compared to monomers, and macromolecular crowding since the oligomer occupies a smaller volume than isolated monomers (Bookchin *et al.,* 1999; Hashimoto *et al.,* 2011; Lindner and Ralston, 1997; Marianayagam *et al.,* 2004; Rivas *et al.,* 2001; Snoussi and Halle, 2005; Wilf and Minton, 1981; Zimmerman and Trach, 1988). However, depending on the size of the oligomer, macromolecular crowding can result in decreased oligomer stability and protein aggregation due to excluded-volume effects (Christiansen *et al.,* 2010; Dong *et al.,* 2010; Jarrett and Lansbury Jr, 1993; Minton, 2000; Wang *et al.,* 2012; Wilf and Minton, 1981; Zimmerman and Minton, 1993).

4.1 OLIGOMERISATION OF FOXP2

Wild-type FOXP2 (LeuZip - end) readily formed an oligomer of approximately hexameric size, with no monomeric species present (Figure 3.21). The R401D mutation of the LeuZip disrupted the formation of the large hexamer, resulting in smaller oligomers of approximately trimeric size. The A539P mutation of the FHD, when combined with the R401D LeuZip mutation, partially disrupted formation of the trimer, resulting in a mixture of species of approximately trimeric and dimeric sizes. It cannot be stated with certainty which of the LeuZip and FHD interfaces was involved in the formation of each oligomer (dimer, trimer, and hexamer), since mixtures of oligomeric states were observed, hence the oligomerisation of each FOXP2 (LeuZip - end) variant could occur due to a single interface or due to the combined effects of LeuZip and FHD interactions. Furthermore, dimeric, trimeric, and hexameric interactions may occur via LeuZips or via domain swapping (Anckar and Sistonen, 2011; Gotte *et al.*, 1999; Grandori *et al.*, 2000; Hai and Curran, 1991; Hirota *et al.*, 2010; Li *et al.*, 2013b; Liu *et al.*, 2001, 2002; Lüscher, 2001; Nair and Burley, 2003; Rabindran *et al.*, 1993; Sawaya *et al.*, 1999; Singleton *et al.*, 2000).

The opposing effects of macromolecular crowding on oligomerisation were apparent, since the thermal stability of R401D/A539P increased with an increasing protein concentration, due to the presence of a dimer (Figure 3.25). In contrast, the thermal stabilities of wild-type, R401D and A539P FOXP2 (LeuZip - end) decreased with an increasing protein concentration, due to the presence of the larger trimer and hexamer, in the non-reducing environment (see Figure C1 in Chapter 5 Appendices). The low stability of the wild-type hexamer suggests a transient rather than permanent association

(Nooren and Thornton, 2003). In contrast, the relatively high stability of the dimer implies that the dimer could be a stable and more permanent intermediate in the formation and dissociation of the hexamer (Nooren and Thornton, 2003).

The thermal stability of R401D/A539P FOXP2 (LeuZip - end) was particularly high, with an estimated denaturation midpoint of approximately 60 °C at a protein concentration of 15 μM (Figure 3.25). In a previous study, thermal unfolding of the wild-type FOXP2 FHD begins at temperatures below 20 °C, whereas the A539P FOXP2 FHD unfolds at higher temperatures, with a denaturation midpoint of approximately 50 °C (Morris and Fanucchi, 2016). As evident from the crystal structure, the FOXP2 FHD itself is highly structured and predominantly α-helical, whereas the FOXP2 (LeuZip - end) variants were largely disordered, based on far-UV circular dichroism spectropolarimetry (Figure 3.16 and A1 in Chapter 5 Appendices), as well as disorder predictions (Figure 3.2) and 3D protein structure predictions (Figure 3.7). Disordered proteins typically have a high proportion of polar and charged amino acid residues, which contribute to enhanced protein solubility and result in disordered proteins exhibiting resilience to both elevated and lowered temperatures (Tantos *et al.*, 2009; Uversky, 2002b). Indeed, all FOXP2 (LeuZip - end) proteins had a combined proportion of polar and charged amino acid residues of over 50% (Figure 2.1). Therefore, the increased thermal stability of the FOXP2 (LeuZip - end) dimer compared to the isolated FHD can be attributed to the high proportion of intrinsic disorder.

Often, intrinsically disordered protein regions adopt a partially folded conformation upon oligomerisation, a process that is necessary for protein function and lowers the folding free energy barrier through the coupling of folding with thermodynamically favourable protein-protein interactions (Huang and Liu, 2009; Uversky, 2002a). This was not observed for FOXP2 (LeuZip - end) homo-oligomerisation since disorder predictions and 3D protein structure predictions of the monomer, and far-UV circular dichroism spectropolarimetric evaluations of the dimer, trimer, and hexamer, showed essentially equal proportions of disordered regions for the monomer in comparison with the oligomers (Figure 3.2 and Table 3.2). Furthermore, there were no significant differences in the proportion of disordered regions between the dimer and higher order oligomers (Figure 3.17). Intrinsically disordered proteins often form the hubs of protein interaction networks by forming transient oligomers with multiple binding partners (Dunker *et al.,* 2005; Kim *et al.,* 2008; Stein *et al.,* 2009). Thus, although there was no decrease in disorder upon homo-oligomerisation, the folding or partial folding of FOXP2 (LeuZip - end) disordered regions could occur upon hetero-oligomerisation, since the high proportion of intrinsic disorder observed in FOXP2 (LeuZip - end) is ideal for

hetero-oligomerisation and would enable binding to multiple binding partners. Indeed, in addition to interactions with FOXP subfamily members, FOXP2 interacts with many TFs, cofactors, and other proteins, and there are likely many more FOXP2 binding partners that are yet to be identified (Chokas *et al.,* 2010; Deriziotis *et al.,* 2014; Estruch *et al.,* 2016, 2018; Li *et al.,* 2004; Tanabe *et al.,* 2011; Wu *et al.,* 2006; Zhou *et al.,* 2008). The multitude of FOXP2 protein binding partners, together with a high proportion of intrinsically disordered regions, is suggestive of a "fly-casting" binding mechanism; the conformational freedom of FOXP2 would lower the free energy barrier – and thus reduce the number of interactions required with the binding partner, and increase the rate – of oligomerisation (Huang and Liu, 2009).

DISOPRED3 was used to predict intrinsically disordered regions in the FOXP2 (LeuZip - end) variants, and to detect protein binding sites within the disordered regions (Jones and Cozzetto, 2015). DISOPRED3 detects protein binding sites based on regions that are able to transition from disordered to ordered, and hence fold, upon binding (Jones and Cozzetto, 2015). The predictions showed that only the LeuZip and FHD regions are ordered (Figure 3.2), and there were no predicted binding sites within the disordered protein regions (data not shown). The disorder predictions agreed well with data obtained from far-UV circular dichroism spectropolarimetry, as well as 3D structure predictions, with respect to the overall proportion of disordered regions (Table 3.2). Although binding sites that do not fold upon binding could exist amongst the disordered regions of FOXP2 (LeuZip - end), there is currently no evidence of such binding sites. Therefore, the LeuZip and FHD are the only regions known to form the stable interfaces of protein-protein and protein-DNA interactions in FOXP2 (LeuZip - end). These interfaces shall be discussed further.

4.2 THE LEUCINE ZIPPER INTERFACE

A number of FOXP disease-linked mutations have been mapped to the zinc finger-LeuZip coiled-coil, emphasising the importance of this region for FOXP function. A FOXP1 mutation resulting in deletion of the zinc finger-LeuZip region was detected in a patient diagnosed with nonsyndromic intellectual disability, resulting in global developmental delays and language impairments (Hamdan *et al.,* 2010). Numerous FOXP3 LeuZip mutations have been identified in IPEX patients. These include L242P, and deletion mutations of Glu251 and Lys250 – equivalent to Arg401 that was mutated for this study (Chatila *et al.,* 2000; Gambineri *et al.,* 2008; Song *et al.,* 2012; Wildin *et al.,* 2002; Zuo *et al.,* 2007b).

The FOXP2 LeuZip is crucial for homo-oligomerisation, since the R401D mutation of the LeuZip was sufficient to alter the oligomeric state significantly (Figure 3.21). The FOXP LeuZip is required not only

for homo-oligomerisation but for hetero-oligomerisation amongst FOXP subfamily members as well (Chae *et al.,* 2006; Li *et al.,* 2004; Wang *et al.,* 2003). Generally, homo-oligomerisation interfaces have a high proportion of hydrophobic residues, and homo-oligomers are more permanent, whereas hetero-oligomer interfaces typically have a high proportion of polar residues which interact through hydrogen bonding, and hetero-oligomers are more transient (Jones and Thornton, 1996). As previously mentioned, the increased thermal stability of R401D/A539P FOXP2 (LeuZip - end) relative to the wild-type suggests a more permanent nature of the FOXP2 (LeuZip - end) dimer, and implies that the dimer interface is predominantly hydrophobic since elevated temperatures increase the strength and stability of hydrophobic interactions (Uversky, 2002b). Indeed, an analysis of the FOXP2 LeuZip reveals that the core positions (*a* and *d*), which would form the dimer interface, are occupied by predominantly hydrophobic residues whereas positions *e* and *g*, which would be involved in the formation of oligomers larger than dimers, are occupied by predominantly polar and charged residues (Figure 1.7A).

Based on monomeric 3D structure predictions, the R401D mutation disrupted hydrogen bonding between Arg401 and Glu398 (Figure 3.3), and thus had the equivalent effect as the K251D mutation of mFoxp3 on this key component of the LeuZip hydrogen bond network (Song *et al.,* 2012). Furthermore, incorporation of the R401D mutation decreased the helicity of the LeuZip (Figure 3.5) and increased the overall disorder of FOXP2 (LeuZip - end) (Figure 3.7). However, although the R401D mutation had a significant effect on the oligomeric state of FOXP2 (LeuZip - end), R401D FOXP2 (LeuZip - end) associated to form a trimer (Figure 3.21) and was not monomeric. Based on these findings, it is clear that Arg401 is not as crucial to FOXP2 homo-oligomerisation as the equivalent Lys252 is to FOXP3 homo-oligomerisation (Song *et al.,* 2012). However, previous studies have shown that mutations of the FOXP2 LeuZip alone can result in a monomer, emphasising the significance of the LeuZip for FOXP2 homo-oligomerisation (Häußermann *et al.,* 2019; Li *et al.,* 2007, 2004). Therefore, Arg401 could be more pertinent to the formation of FOXP2 hetero-oligomers, or higher order homo-oligomers, especially since it is located at position *g* (Figure 1.7), a non-core position. FOXP3 Lys252, equivalent to FOXP2 Arg401, is subjected to acetylation which has been suggested to play a role in regulating the oligomerisation of FOXP3 (Song *et al.,* 2012). Modification of arginine residues usually occurs by methylation and not acetylation, since the arginine guanidinium group is strongly basic with a pK_a of 13.8 and is therefore not susceptible to substitution by an acetyl group (Fitch *et al.,* 2015). Arginine methylation is a common post-translational modification identified in the regulation of many cellular processes, the inhibition of which has been associated with many diseases such as cancer (Blanc and Richard, 2017). Although arginine methylation does not affect its positive

charge, the addition of a methyl group removes a hydrogen atom, thereby interrupting hydrogen bonding, and increases the hydrophobic character of the arginine side chain (Guccione and Richard, 2019). Therefore, in addition to other regulatory pathways, FOXP1/2/4 could employ methylation of the conserved LeuZip arginine residue (Arg401 in FOXP2) as a physiological means of regulating the formation of FOXP hetero-oligomers or higher order homo-oligomers, though further investigation is required to confirm this.

Previous studies have suggested that homo- and hetero-oligomerisation mediated by the FOXP LeuZip is necessary for DNA binding and transcriptional activity (Chae *et al.,* 2006; Li *et al.,* 2004). Indeed, despite the R401D mutation not disrupting homo-oligomerisation completely, since R401D FOXP2 (LeuZip - end) was not monomeric, this mutation resulted in a decrease in DNA-binding affinity (Figure 3.27), emphasising the importance of LeuZip-mediated associations for DNA binding. However, the decrease was less significant than the decrease observed due to the A539P mutation of the FHD.

4.3 THE FORKHEAD DOMAIN INTERFACE AND DNA BINDING

The importance of the FOXP FHD is clear given that it is the DBD. Furthermore, a number of FOXP disease-linked mutations have been mapped to the FHD. Patients with the R525Q mutation of the FOXP1 FHD show symptoms of relapsing-remitting fevers, urinary tract defects, global developmental delays, language impairments, macrocephaly, and autism (Bekheirnia *et al.,* 2017; Johnson *et al.,* 2018; Myers *et al.,* 2017). The FOXP2 R553H mutation linked to DVD occurs within the FHD (Lai *et al.,* 2001). FOXP3 mutations of the FHD are associated with IPEX; for example the FOXP3 A384T mutation has been detected in IPEX patients (Wildin *et al.,* 2001).

According to 3D protein structure predictions, the FHD DNA recognition helix (helix 3) is buried in all FOXP2 (LeuZip - end) variants (Figure 3.7), necessitating conformational changes to expose helix 3 and allow insertion of helix 3 into the major groove for DNA binding. Since the predicted protein structures are for monomers, the possibility exists that conformational changes occurring during homo- and hetero-oligomerisation could result in the exposure of helix 3. Nevertheless, the large proportion of intrinsically disordered protein regions could be influential in FOXP2 (LeuZip - end) DNA binding, through lowering the free energy barrier for structural rearrangements and conferring conformational freedom to the protein, allowing for the exposure of helix 3. Furthermore, the intrinsically disordered regions could allow for the “fly-casting” mechanism of DNA binding, thus increasing the rate of DNA binding (Shoemaker *et al.,* 2000).

Based on the FHD crystal structure, upon domain swapping, the relatively exposed Trp533 becomes buried and Trp548 forms part of the core of the FHD dimer interface (Stroud *et al.,* 2006). Additionally, 3D protein structure predictions show that the FOXP2 (LeuZip - end) tryptophan residues are able to interact with regions outside of the FHD in the absence of DNA (Figure 3.18). Thus, intrinsic tryptophan fluorescence could be affected by domain swapping and by the global protein fold when FOXP2 (LeuZip - end) is unbound. However, intrinsic tryptophan fluorescence is expected to predominantly be influenced by DNA-protein interactions when FOXP2 (LeuZip - end) is bound since the FHD interacts closely with DNA. Trp573, in particular, interacts directly with the DNA backbone (Stroud *et al.,* 2006). Stern-Volmer constants indicate that the solvent-accessibility of the FOXP2 (LeuZip - end) tryptophan residues decreases upon DNA binding (Figure 3.23 and 3.24). Therefore, the FHD itself appears to be restricted rather than flexible when bound to DNA. This is consistent with a previous study of the isolated FHD which found a decrease in disordered regions and an increase in α-helical content upon DNA binding (Webb *et al.,* 2017). The decrease in the Stern-Volmer constant of R401D/A539P FOXP2 (LeuZip - end) upon DNA binding was not significant, which could be due to the significantly weaker DNA-binding interactions of this variant in comparison with the other variants (Figure 3.27), since a decreased FHD flexibility is associated with a high DNA-binding affinity (Webb *et al.,* 2017). However, despite the increased order of the FHD upon DNA binding, the FOXP2 (LeuZip - end) variants, particularly the wild-type, appear to undergo a slight increase in global disorder upon DNA binding (Figure 3.16 and A1 in Chapter 5 Appendices). Although the decreased FHD flexibility upon DNA binding would result in an entropic loss, the global increase in disorder would result in an increase in entropy, therefore the DNA binding process would still be entropically favourable. Furthermore, the global increase in disorder could increase the propensity for oligomerisation, implying a sequential mechanism whereby the binding of FOXP2 to a *cis*-regulatory element results in the subsequent recruitment and binding of binding partners.

Intrinsic tryptophan fluorescence spectra of wild-type and A539P FOXP2 (LeuZip - end) showed no significant differences in the absence of DNA (Figure 3.19). Since all FOXP2 (LeuZip - end) tryptophan residues are located within the FHD, this suggests that the A539P mutation of the FHD had no significant effect on the structure of the FHD itself. Predicted monomeric 3D structures of wild-type and A539P FOXP2 (LeuZip - end) showed similar folds in the FHD, and both FHDs aligned well with the monomeric FHD crystal structure (Figure 3.6). Therefore, any differences observed between A539P FOXP2 (LeuZip - end) and the wild-type were due to prevented domain swapping of the FHD. Despite the prevention of FHD domain swapping, A539P FOXP2 (LeuZip - end) still existed predominantly as a hexamer and the A539P mutation only had an effect on the oligomeric state of the protein when

present with the R401D mutation of the LeuZip, in which case the A539P mutation resulted in a partial shift from trimer to dimer (Figure 3.21). Therefore, although domain swapping of the FHD does affect the overall oligomeric state of FOXP2 (LeuZip - end), the effect of FHD domain swapping is not as significant as interactions of the LeuZip. Despite A539P FOXP2 (LeuZip - end) adopting the same hexameric quaternary structure as the wild-type, the prevention of FHD domain swapping affected DNA binding and resulted in a significant loss in DNA-binding affinity (Figure 3.27). This indicated that the formation of the FOXP2 (LeuZip - end) hexamer specifically is not significant for DNA binding and other oligomeric states could be allowed.

A previous study suggested that the unique flexibility of the FOXP FHD and propensity for dimerisation through domain swapping, arising due to the substitution of the conserved FOX hinge loop proline with alanine (Ala539 in FOXP2), is an evolutionary mutation that conferred increased DNA-binding affinity and decreased specificity to FOXP subfamily members (Morris and Fanucchi, 2016). However, a comparison of FOX DNA-binding affinities suggests that FOXP subfamily members have a lower DNA-binding affinity compared with other FOX proteins (Table 4.1), including FOXA and FOXO proteins. The FOXM1 FHD has a weaker affinity for DNA than all FOX proteins listed in Table 4.1, including all FOXP2 variants (Littler *et al.,* 2010). However, full-length FOXM1 has a high affinity for DNA with a K_D of 214.5 nM (Xiang *et al.,* 2017). Furthermore, DNA-binding assays for both the FOXM1 FHD and full-length FOXM1 were conducted under reducing conditions, which could decrease the apparent affinity for DNA, as seen in this study (Figure 3.27 and E1 in Chapter 5 Appendices). It is worth noting that all FOXA, FOXO and FOXM proteins listed in Table 4.1 have the conserved hinge loop proline residue. In the isolated FOXP2 FHD, fluorescence anisotropy DNA-binding assays suggest that the A539P mutation does not affect DNA binding (Morris and Fanucchi, 2016). However, K_Ds obtained through isothermal titration calorimetry, which is a more sensitive technique, have shown that even in the isolated FHD the A539P mutation results in a significant decrease in DNA-binding affinity (Morris and Fanucchi, 2016). Interestingly, for all FOX proteins, the isolated FHDs appear to bind DNA more weakly than longer constructs, emphasising the importance of regions outside of the FHD for enhancing FOX DNA binding.

Table 4.1: Comparison of dissociation constants (K_Ds) of FOX DNA binding obtained from fluorescence anisotropy assays. The table is arranged from highest to lowest DNA-binding affinity. Protein constructs that were used in this study are indicated in boldface. Identifiable domain(s) included in non-full-length constructs are listed (NLS is nuclear localisation signal). For some FOX proteins, the DNA-binding affinity has been quantified for the isolated FHD, as well as for a longer construct including other domains in addition to the FHD. This is indicated with triangles of the same colour (rose coloured triangles (▶) for FOXO4, blue coloured triangles (▶) for FOXM1, pink coloured triangles (▶) for FOXP2, and green coloured triangles (▶) for A539P FOXP2). Assays that were conducted under reducing conditions are indicated, due to the differences observed in DNA-binding affinities between non-reducing and reducing conditions in this study.

	Protein	Domain(s)	K_D (nM)
	FOXA1	Full-length	33.48 ± 10.76 [1]
▶	FOXO4	FHD; NLS	34 ± 6.0 [2]
	FOXO1 (Reducing)	FHD	93 ± 9.0 [3]
▶	FOXO4	FHD	188 ± 10.0 [4]
▶	FOXM1 (Reducing)	Full-length	214.5 ± 248.9 [5]
	FOXO3 (Reducing)	FHD; NLS (partial)	295 ± 26.0 [6]
▶	**FOXP2**	LeuZip; FHD	347 ± 31.6
	R401D FOXP2	LeuZip; FHD	460 ± 30.0
▶	**A539P FOXP2**	LeuZip; FHD	611 ± 45.4
	R401D/A539P FOXP2	LeuZip; FHD	778 ± 33.3
▶	A539P FOXP2	FHD	878 ± 66.2 [7]
▶	FOXP2	FHD	892 ± 49.9 [7]
▶	FOXM1 (Reducing)	FHD	7000 ± 300 [8]

[1] Aung *et al.,* 2014; [2] Silhan *et al.,* 2009; [3] Zhang *et al.,* 2002; [4] Boura *et al.,* 2007; [5] Xiang *et al.,* 2017; [6] Tsai *et al.,* 2007; [7] Morris and Fanucchi, 2016; [8] Littler *et al.,* 2010

4.4 OLIGOMERISATION AND DNA BINDING PATHWAY

The oligomerisation and DNA binding of oligomeric TFs typically proceeds by one of two classical recognised pathways (Figure 4.1). The "monomer pathway", based on the General Control Nondepressible 4 (GCN4) TF, proceeds by DNA binding of a monomer, followed by protein-protein interactions with additional monomers to form an oligomer. In some scenarios, binding of the first TF enables the recruitment and binding of a second TF to the adjacent *cis*-regulatory element (Ma and Ptashne, 1988; Ptashne and Gann, 1997). The "dimer pathway", or in the case of FOXP2 the "oligomer pathway", is based on λ-repressor and proceeds by DNA binding of a pre-assembled oligomer (Berger *et al.,* 1998; Ptashne *et al.,* 1980). Evaluations of the monomer pathway have shown that this pathway improves DNA-binding specificity by reducing non-specific protein-protein interactions and increasing the rate at which the TF is able to scan DNA and locate the appropriate *cis*-regulatory element (Kohler *et al.,* 1999). On the other hand, the dimer pathway is favourable for controlling the number of TF monomers, since this number fluctuates drastically in the monomer pathway, so implementation of the dimer pathway is a mechanism of regulating genetic noise, which applies not only to TF dimers but to higher order oligomers as well (Bandukwala *et al.,* 2011; Ghim and Almaas, 2008).

As shown in previous studies, LeuZip-mediated associations are essential for FOXP DNA binding, therefore it is unlikely that the FOXP TFs bind DNA as monomers (Li *et al.,* 2007, 2004). In addition, based on the results of this study, it is evident that FOXP2 (LeuZip - end) forms a homo-oligomer in the absence of DNA *in vitro*, therefore the presence of DNA is not essential to oligomer formation. However, the increased proportion of intrinsically disordered regions upon DNA binding suggests that DNA binding has an allosteric effect and increases the propensity for protein-protein interactions, specifically through the "fly-casting" mechanism. Furthermore, although very similar proportions of the R401D/A539P FOXP2 (LeuZip - end) dimer and trimer were observed in the absence of DNA by size exclusion chromatography (Figure 3.21), the equilibrium appeared to shift in favour of the larger trimer when bound to DNA on an EMSA (Figure 3.26), particularly at higher protein concentrations that were more similar to the concentration used for size exclusion chromatography (the highest protein concentration used for the EMSA was 17.5 µM, whereas the concentration used for size exclusion chromatography was 30 µM). Thus it remains somewhat unclear as to whether FOXP2 follows the monomer or dimer/oligomer pathway.

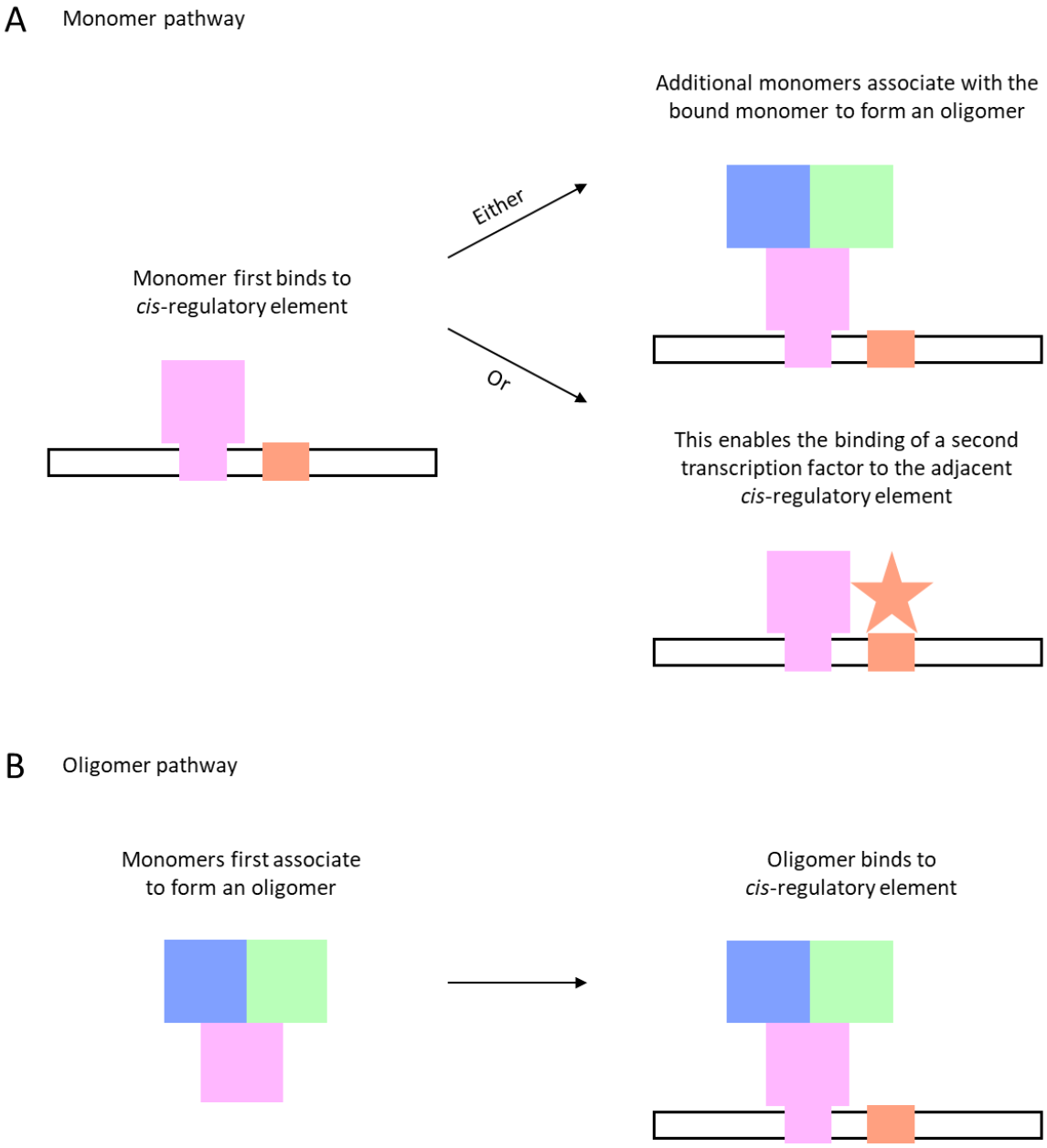


Figure 4.1: Assembly and DNA binding pathways of oligomeric transcription factors. In the monomer pathway **(A)**, a TF monomer binds to the *cis*-regulatory element. Additional monomers then associate with the DNA-bound monomer to form a functional oligomer. In some cases, binding of the monomer enables the recruitment and binding of a second TF to the adjacent *cis*-regulatory element. In the oligomer (or dimer) pathway **(B)**, TF monomers first assemble to form an oligomer, which then binds to the *cis*-regulatory element.

For some TFs, oligomerisation and DNA binding appear to deviate from the classical monomer and dimer/oligomer pathways. An example is the hypochlorite-responsive TF that undergoes oligomerisation in the absence of DNA to form a dodecamer, suggesting that DNA binding proceeds via the dimer/oligomer pathway (Gebendorfer *et al.,* 2012). However, the dodecamer is unable to bind DNA and dissociates into smaller DNA-binding oligomers in the presence of DNA (Gebendorfer *et*

al., 2012). Interestingly, the dissociation of the dodecamer does not depend on the presence of the target site and occurs even in the presence of random DNA sequences, indicating that dissociation is required to scan DNA and locate the target site (Gebendorfer *et al.,* 2012). Another example is TFEB that binds DNA as a homo- or heterodimer, but associates with another TFEB dimer to form a tetramer in its unbound form (Fisher *et al.,* 1991). Thus, for both the hypochlorite-responsive TF and TFEB, the higher order oligomers appear to exist as a storage form when not bound to DNA. On the other hand, reports regarding the oligomerisation and DNA binding of the well-studied p53 TF are contradictory. In solution, p53 readily forms a tightly bound dimer of dimers (Brandt *et al.,* 2009; Nicholls *et al.,* 2002). The formation of the p53 oligomer is necessary for DNA binding, and the binding of the p53 oligomer to DNA results in the recruitment of other TFs and proteins (Beckerman and Prives, 2010; Halazonetis and Kandil, 1993; Shaulian *et al.,* 1993). Based on these findings it would appear that p53 follows a dimer/oligomer pathway, but there is evidence to suggest that the formation of the tetramer is induced by DNA binding (Weinberg *et al.,* 2004). In addition, the protein-protein interactions differ between the unbound and DNA-bound homotetramers (Veprintsev *et al.,* 2006). Thus, it is plausible that p53 exists as a tetramer when not bound to DNA, as a storage form, which then dissociates into a smaller species (dimer or monomer) in order to scan DNA and locate the target site more efficiently, and then reassociates to form a tetramer of a different configuration upon binding DNA. Therefore, the possibility exists that, like the examples of the hypochlorite-responsive TF, TFEB, and p53, the oligomerisation and DNA binding of FOXP2 deviates from the classical monomer and oligomer pathways.

It is possible that FOXP2 forms a large oligomer (such as a hexamer) in the absence of DNA as a storage form. The large FOXP2 oligomer could dissociate into a smaller oligomer in the presence of DNA, in order to scan DNA and locate the target site. Upon binding DNA, FOXP2 rapidly recruits and associates with the appropriate binding partners to form either homotypic or heterotypic oligomers, or both. This would allow for simultaneous domain swapping of the FHD, only when bound to the appropriate partner since not all heterotypic binding partners are compatible with the FHD for domain swapping. The formation of heterotypic protein-protein interactions after rather than before DNA binding would increase the rate of recruitment and binding through limiting the search space to the DNA backbone. In the absence of heterotypic binding partners, DNA-bound FOXP2 (LeuZip - end) could only form homo-oligomers which were observed by EMSAs (Figure 3.26). The DNA-scanning oligomer could be a dimer, based on the stability of the R401D/A539P FOXP2 (LeuZip - end) mutant. Furthermore, based on Stern-Volmer constants, wild-type FOXP2 (LeuZip - end) experienced the largest conformational change upon DNA binding, followed by the A539P mutant and then the R401D mutant. The

R401D/A539P mutant did not experience a significant conformational change upon DNA binding. This could be explained by the quaternary structure of each FOXP2 (LeuZip - end) variant, with wild-type and A539P FOXP2 (LeuZip - end) existing as large oligomers of approximately hexameric size that would require large conformational changes to enable dissociation for DNA binding. Conversely, the R401D/A539P mutant exists predominantly as oligomers of dimeric and trimeric size, and only the trimers would require small conformational changes to facilitate DNA binding.

4.5 PHYSIOLOGICAL ROLE OF FOXP2 OLIGOMERISATION

Apart from enabling interactions with multiple binding partners in heterotypic associations, oligomerisation could enhance FOXP2 function in a number of ways. Homo-oligomerisation of FOXP2 could play a physiological role in enhancing DNA-binding affinity, since wild-type FOXP2 (LeuZip - end) had the strongest affinity for DNA (Figure 3.27). DNA-binding assays were conducted under non-reducing conditions and in the same environment, DNA binding quenched the intrinsic tryptophan fluorescence of wild-type FOXP2 (LeuZip - end) the most (see Figure B1 in Chapter 5 Appendices). This could be due to the high affinity of the wild-type for DNA resulting in more protein molecules being bound in comparison with the mutants, giving rise to more quenching, since the extent of intrinsic tryptophan fluorescence quenching can be proportional to the quantity of DNA-bound protein (Carpenter *et al.*, 2001). Alternatively, the quenching of wild-type FOXP2 (LeuZip - end) could be due to stabilising DNA-protein interactions that were unable to form in the mutants. TF oligomerisation often contributes to DNA bending or looping, or to the formation of the transcription-initiation complex through augmenting the wrapping of DNA around the complex (Vámosi and Rueda, 2018). This is possible since TF oligomerisation mediates simultaneous interactions with distal DNA sites (Vámosi and Rueda, 2018). Thus, in addition to enhanced DNA-binding affinity, FOXP2 oligomerisation could contribute to these processes. Crystal structures of the FOXP2 and FOXP3 FHD domain-swapped dimers support this hypothesis by showing that the FHD dimer is unable to simultaneously bind two adjacent DNA sites due to the positioning of the DNA recognition helices (Bandukwala *et al.,* 2011; Stroud *et al.,* 2006). As seen in Figure 4.2, the FOXP3 domain-swapped dimer is capable of binding two non-adjacent DNA sites simultaneously. Even in the absence of domain swapping, for example in heterotypic oligomers that do not allow for domain swapping to occur, or if domain swapping is dependent on the *cis*-regulatory element, FOXP2, and hence FOXP, oligomerisation could play a physiological role in enabling simultaneous interactions with distal DNA sites. Finally, hetero-oligomerisation of FOXP2 could improve DNA-binding affinity and specificity relative to homo-oligomers of FOXP2, like the Jun-ATF2 heterodimer which has improved specificity relative to the promiscuous Jun homodimer (Pan and Nussinov, 2011).

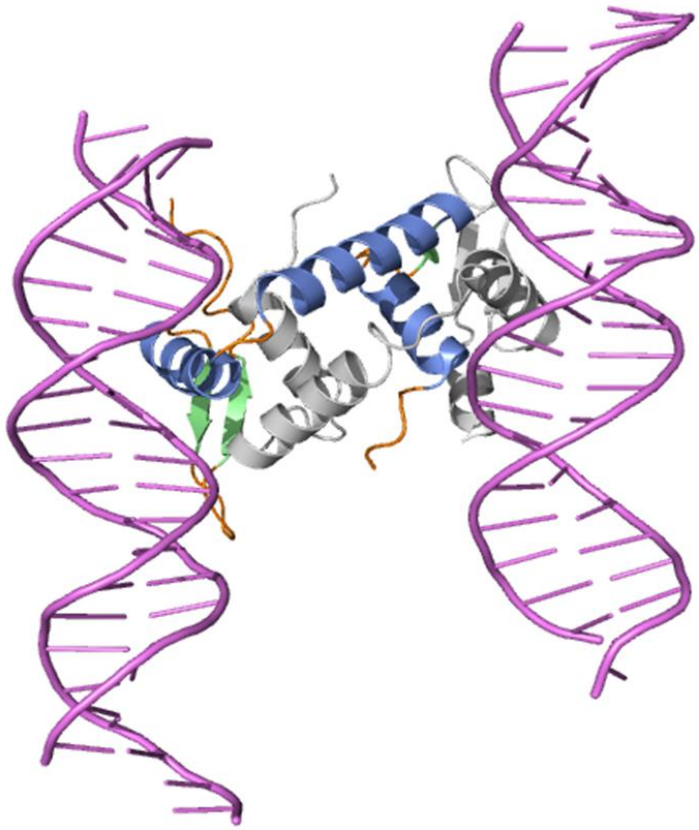


Figure 4.2: Domain swapping of the FOXP forkhead domain allows for simultaneous binding at distal DNA sites. Each monomer of the dimer is able to interact with DNA through insertion of the recognition helix (helix 3) at the major groove, at distal DNA sites due to the relative positions of helix 3 (one monomer is depicted in grey). Simultaneous binding at distal DNA sites could contribute to DNA bending or looping, or to the formation of the transcription-initiation complex through augmenting the wrapping of DNA around the complex. α-Helices are blue, β-strands are green, loops are orange, and DNA is pink. The image was rendered on The PyMOL Molecular Graphics System, Version 0.99, Schrödinger, LLC (PDB 3QRF) (Bandukwala *et al.,* 2011).

Thus, there is sufficient evidence that FOXP homo- and hetero-oligomerisation is an important regulator of FOXP transcriptional activity. The extent to which FOXP oligomerisation affects gene expression could be more far-reaching than current evidence suggests, and further studies are necessary to gain a full understanding of FOXP protein-protein interactions. The possibility of the existence of multiple FOXP oligomeric states, as observed for homotypic FOXP3 associations and which could exist for heterotypic FOXP interactions, would permit multidimensional control of FOXP function (Li *et al.,* 2007). Furthermore, the control of FOXP oligomeric states through post-translational modifications could provide a neat mechanism of fine-tuning the regulation of FOXP-controlled gene expression.

According to the most recent cancer statistics, breast cancer is the most common cancer in women both locally and globally, and prostate cancer is the most common cancer in South African men and

the second most common cancer in all men (Bray *et al.,* 2018; Cancer Association of South Africa, 2014). FOXP2 has been associated with both breast cancer and prostate cancer, and many other diseases, and is thus a promising target for the development of novel diagnostics and therapeutics (Cuiffo and Karnoub, 2016; Cuiffo *et al.,* 2014; Stumm *et al.,* 2013). However, it is only with a thorough understanding of FOXP2, how it interacts with other proteins and with DNA, and how it functions, and with the identification of key residues for FOXP2 protein-protein and protein-DNA interactions, that new strategies can be designed and implemented to effectively and specifically detect and target dysregulated FOXP2 in diseased states.

CONCLUSION

FOX TFs are a large TF family that regulate a diverse range of processes throughout the human body. The FOXP subfamily exhibit the unique ability to form homotypic dimers and higher order oligomers, a phenomenon that has not been observed in any other FOX subfamily. FOXP oligomerisation occurs via two interfaces, the LeuZip and DNA-binding FHD. Although most previous studies have emphasised the LeuZip as the key protein-protein interface, the contribution of the FOXP FHD to oligomerisation has not been thoroughly investigated. In this study of FOXP2 (LeuZip - end), by use of the R401D LeuZip mutation and A539P FHD mutation, the homo-oligomerisation interfaces were investigated and the effects of the mutations on the structure and DNA binding of FOXP2 were evaluated. The R401D LeuZip mutation is a novel mutation. The A539P mutation has previously only been studied in the FHD, so this was the first study evaluating the effects of prevented FHD domain swapping on FOXP2 oligomerisation, in the presence of the LeuZip interface.

On its own, the A539P FHD mutation did not alter the oligomeric state of FOXP2 (LeuZip - end) substantially. Although the R401D LeuZip mutation altered the FOXP2 (LeuZip - end) oligomeric state significantly, it did not result in a monomeric species. This allowed for the effects of the A539P mutation to be evident in the R401D/A539P mutant, where the A539P mutation further disrupted FOXP2 (LeuZip - end) homo-oligomerisation. The R401D/A539P mutant was resilient to elevated temperatures, which could be due to the stability of the R401D/A539P FOXP2 (LeuZip - end) dimer relative to the larger R401D FOXP2 (LeuZip - end) trimer, and wild-type and A539P FOXP2 (LeuZip - end) hexamer, suggesting that the dimer is a more stable intermediate in the formation of higher order oligomers. Disrupting FHD domain swapping resulted in a more significant decrease in DNA-binding affinity than the R401D mutation of the LeuZip, perhaps because the R401D mutation did not completely eliminate LeuZip interactions. Interestingly, the formation of a hexameric species specifically did not appear to contribute to DNA binding, since although wild-type and A539P FOXP2 (LeuZip - end) were both predominantly hexameric, the DNA-binding affinity of the A539P mutant was significantly lower than that of the wild-type. Furthermore, wild-type and A539P FOXP2 (LeuZip - end) appeared to experience the largest conformational changes upon DNA binding, and the DNA binding-induced conformational changes of R401D/A539P FOXP2 (LeuZip - end) were not significant. Taken together, the findings of this study suggest that FOXP2 higher order oligomers could dissociate in the presence of DNA, possibly into dimers, to scan DNA and locate the appropriate *cis*-regulatory element. Subsequent to DNA binding, FOXP2 re-associates to form higher order homo- and/or hetero-oligomers. This work shows that although the LeuZip is crucial for FOXP2

homo-oligomerisation, the FHD also contributes to the association of FOXP2 monomers. The possible existence of multiple oligomeric states could regulate FOXP2 DNA binding and transcriptional activity.

Chapter 5: Appendices

APPENDIX A: CIRCULAR DICHROISM SPECTROPOLARIMETRY (NON-REDUCING CONDITIONS)

Far-UV CD spectropolarimetry was conducted under non-reducing conditions (see 3.3.1 Circular dichroism spectropolarimetry for reducing conditions). This was useful for gaining insight into the protein secondary structures under non-reducing conditions, to aid interpretation of protein thermal unfolding and fluorescence anisotropy DNA-binding assays, which were conducted under non-reducing conditions. Spectra were recorded on 5 µM protein in the absence and presence of 2.5 µM DNA, without DTT. All spectra (Figure A1) were similar to those observed under reducing conditions, with the occurrence of slight negative bands at ~ 208 nm, especially in the absence of DNA. DNA binding had more significant effects on the protein secondary structures under non-reducing conditions, since the spectra differed more between unbound and DNA-bound states. As with data obtained under reducing conditions, a visual assessment of the spectra suggests that the proteins consisted predominantly of disordered regions, with a small proportion of α-helices, both in the absence and presence of DNA.

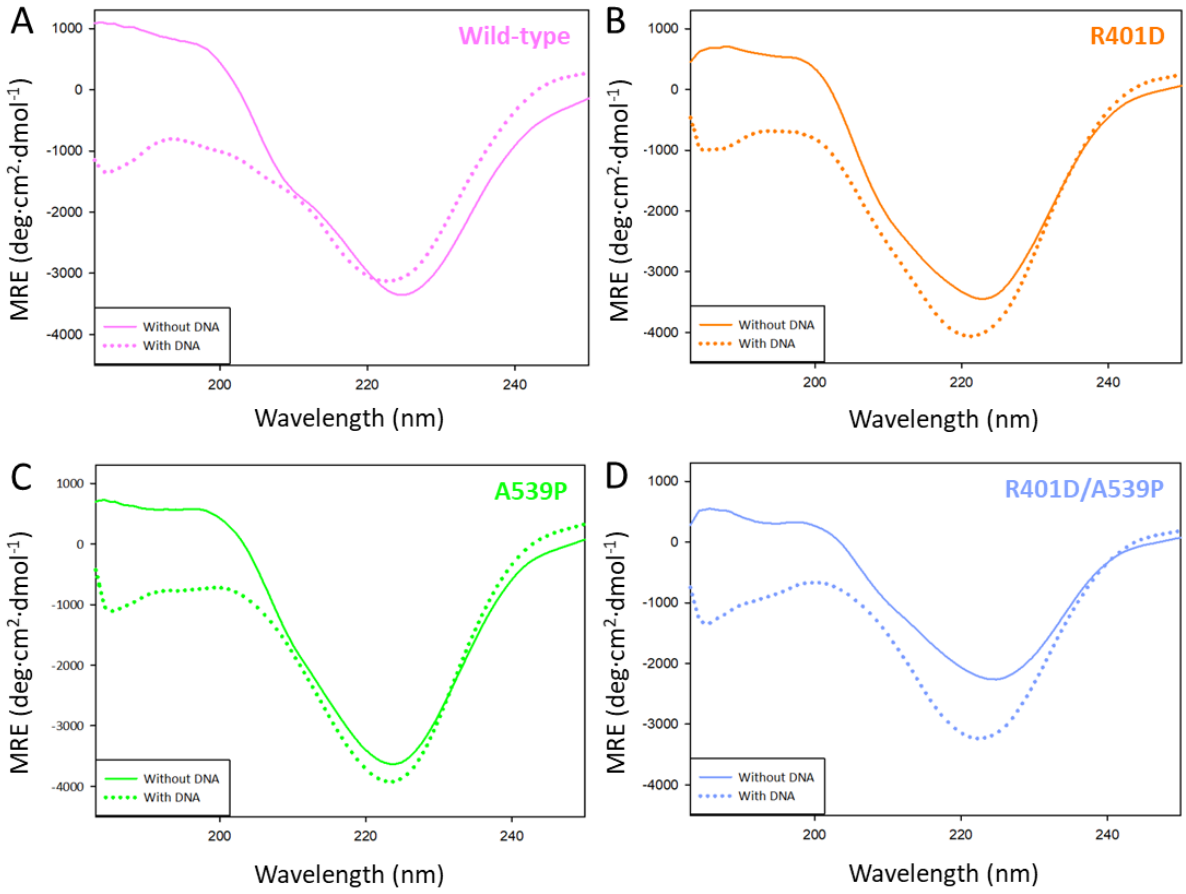


Figure A1: Circular dichroism spectra of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end), measured under non-reducing conditions. Spectra were measured on 5 µM protein, in the absence (solid lines) and presence (dotted lines) of 2.5 µM DNA at 0 mM DTT. Data points with a high tension of over 600 V were removed. Based on DichroWeb analyses, wild-type FOXP2 (LeuZip - end) showed a decrease in α-helical content ($p < 0.05$) upon DNA binding. R401D ($p < 0.05$), A539P ($p < 0.01$) and R401D/A539P ($p < 0.01$) FOXP2 (LeuZip - end) decreased in β-strands upon DNA binding. All FOXP2 (LeuZip - end) variants decreased significantly in turns and increased significantly in unordered regions upon DNA binding ($p < 0.05$ for wild-type FOXP2 (LeuZip - end); $p < 0.01$ for R401D FOXP2 (LeuZip - end), for A539P FOXP2 (LeuZip - end) turns and for R401D/A539P FOXP2 (LeuZip - end) unordered regions; $p < 0.001$ for A539P FOXP2 (LeuZip - end) unordered regions and for R401D/A539P FOXP2 (LeuZip - end) turns).

The CD spectra were dissected using DichroWeb to quantify the proportion of each secondary structural element present in each protein (Whitmore and Wallace, 2004, 2008). A summary of the DichroWeb data is shown in Figure A2. As previously explained (see 3.3.1 Circular dichroism spectropolarimetry), it is likely that the DichroWeb algorithm underestimated the proportion of α-helices in the FOXP2 (LeuZip - end) proteins. The DichroWeb data was very similar to that obtained under reducing conditions (α-helical content of ~ 4%, β-strand content of ~ 34%, and content of turns/unordered regions of ~ 61%). Under non-reducing conditions, DNA binding caused a decrease in α-helical content in wild-type FOXP2 (LeuZip - end), a decrease in β-strands in R401D, A539P and R401D/A539P FOXP2 (LeuZip - end), and a decrease in turns and increase in unordered regions in all

FOXP2 (LeuZip - end) variants. These changes can be attributed to structural rearrangements occurring upon DNA binding. The R401D mutation of the LeuZip caused a decrease in α-helical content in the absence of DNA, a decrease in disorder in the absence of DNA, and an increase in turns both in the absence and presence of DNA. A539P FOXP2 (LeuZip - end) had a lower proportion of α-helices in the absence of DNA, and a lower proportion of β-strands in the presence of DNA, compared to the wild-type. In the absence of DNA, R401D/A539P FOXP2 (LeuZip - end) had a lower proportion of α-helices and disorder, and a higher proportion of β-strands than the wild-type, with no significant differences observed in the presence of DNA. Overall, the secondary structures of the proteins were similar under non-reducing and reducing conditions.

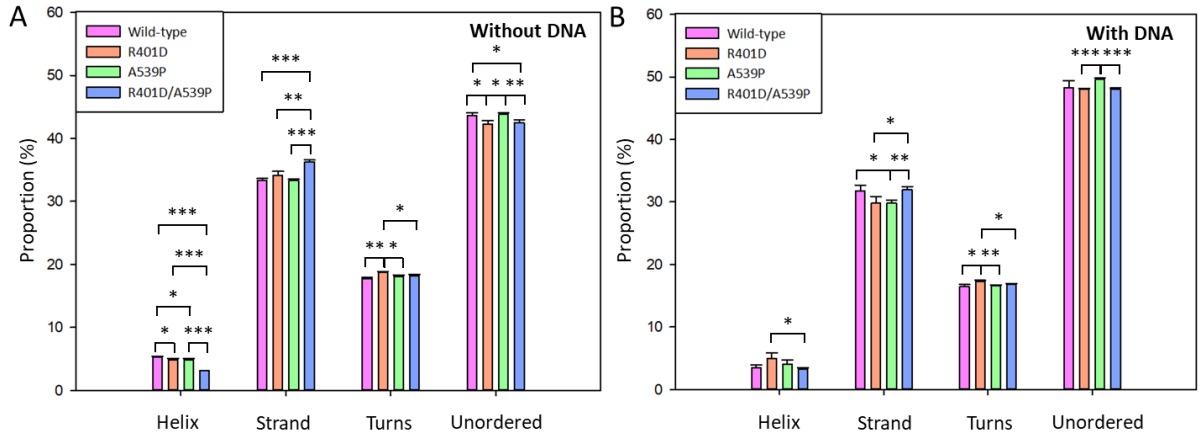


Figure A2: Analyses of secondary structural elements of FOXP2 (LeuZip - end) variants in the absence (A) and presence (B) of DNA, under non-reducing conditions. Quantitative analyses of data obtained from far-UV CD spectropolarimetry at 0 mM DTT indicating the proportion of α-helices, β-strands, turns, and unordered regions in each FOXP2 (LeuZip - end) variant. Statistical significance of differences between the proportions of secondary structural elements are indicated by two-sample t-tests (*$p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$). Analyses were computed using DichroWeb (Whitmore and Wallace, 2004, 2008).

APPENDIX B: INTRINSIC TRYPTOPHAN FLUORESCENCE SPECTROSCOPY (NON-REDUCING CONDITIONS)

Intrinsic tryptophan fluorescence spectroscopy was conducted under non-reducing conditions (see 3.3.2 Intrinsic tryptophan fluorescence spectroscopy for reducing conditions). This was useful for gaining insight into the protein folds under non-reducing conditions, to aid interpretation of protein thermal unfolding and fluorescence anisotropy DNA-binding assays, which were conducted under non-reducing conditions. Spectra were measured on 5 µM protein in the absence and presence of 2.5 µM DNA, without DTT. Much like under reducing conditions, the spectrum of each protein had a double peak both in the absence and presence of DNA (Figure B1A), and fluorescence quenching occurred upon DNA binding of all FOXP2 (LeuZip - end) proteins (Figure B1A). Comparisons between the maximum intensities at 339 nm of the DNA-bound proteins normalised relative to the unbound proteins were plotted as a bar graph (Figure B1B). In the presence of DNA, the normalised fluorescence intensities of R401D, A539P and R401D/A539P FOXP2 (LeuZip - end) were significantly higher than that of the wild-type. Therefore, the environments of the tryptophan residues differed, possibly due to differences in the DNA-protein contacts formed for each protein.

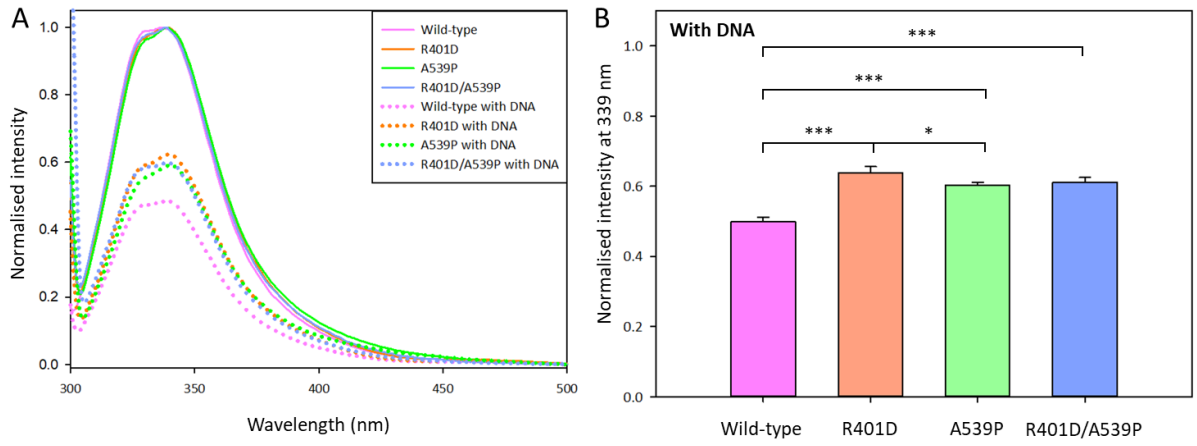


Figure B1: Intrinsic tryptophan fluorescence spectra of FOXP2 (LeuZip - end) variants (A) and comparison of the emission maxima in the presence of DNA (B), measured under non-reducing conditions. Spectra were measured on 5 µM protein, in the absence (solid lines) and presence (dotted lines) of 2.5 µM DNA at 0 mM DTT **(A)**. Spectra showed no differences in the absence of DNA. Fluorescence intensities of all FOXP2 (LeuZip - end) variants were quenched significantly upon DNA binding ($p < 0.01$ for wild-type and R401D FOXP2 (LeuZip - end), and $p < 0.001$ for A539P and R401D/A539P FOXP2 (LeuZip - end)). Data points were normalised between 0 and 1 relative to the unbound proteins, and the emission maxima at 339 nm were compared in the presence of DNA **(B)**. Statistical significance of differences between the normalised emission maxima are indicated by two-sample t-tests (*$p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$).

Overall, the intrinsic tryptophan fluorescence of the proteins was similar under non-reducing and reducing conditions with the exception of R401D/A539P FOXP2 (LeuZip - end). Under non-reducing conditions, the normalised fluorescence intensity of DNA-bound R401D/A539P FOXP2 (LeuZip - end) was higher than that of the wild-type whereas under reducing conditions, the normalised fluorescence intensity of DNA-bound R401D/A539P FOXP2 (LeuZip - end) was lower than that of the wild-type. This could be due to a difference in protein fold between the two environments, especially since R401D/A539P FOXP2 (LeuZip - end) was predicted to have a solvent-accessible cysteine residue (Figure 3.15). R401D and A539P FOXP2 (LeuZip - end) showed a higher normalised fluorescence intensity relative to the wild-type when bound to DNA, under both non-reducing and reducing conditions.

APPENDIX C: SIZE EXCLUSION CHROMATOGRAPHY (NON-REDUCING CONDITIONS)

Size exclusion chromatography was conducted under non-reducing conditions (see 3.3.3 Size exclusion chromatography for reducing conditions). This was useful for gaining insight into the protein oligomeric states under non-reducing conditions, to aid interpretation of protein thermal unfolding and fluorescence anisotropy DNA-binding assays, which were conducted under non-reducing conditions. Samples of each protein were prepared and loaded onto the column. The elution volume was determined for each peak and the equation of the calibration curve (Figure 3.20) was used to calculate the corresponding sizes. The elution profiles are shown in Figure C1.

A small shoulder appeared in the peak of wild-type FOXP2 (LeuZip - end) that did not occur under reducing conditions. Based on the oligomeric states assigned to the peaks under reducing conditions, this shoulder could represent a small proportion of a disulfide linked octamer. There were no major differences in the quaternary structure of A539P FOXP2 (LeuZip - end) between reducing and non-reducing conditions. Ignoring the peak in the void volume, R401D FOXP2 (LeuZip - end) was predominantly trimeric, which was observed under reducing conditions, but a hexameric species formed that was not present under reducing conditions. R401D/A539P FOXP2 (LeuZip - end) consisted predominantly of a dimeric species with a small proportion of hexamer. Under reducing conditions, a trimeric species was seen instead of the R401D/A539P FOXP2 (LeuZip - end) hexamer, so the formation of the hexamer was due to disulfide bonding.

Therefore, under non-reducing conditions disulfide bonding contributed to the formation of higher order oligomers that were not observed under reducing conditions. The differences were most significant for R401D and R401D/A539P FOXP2 (LeuZip - end). Thus, the R401D mutation appears to expose buried cysteine residues, the effects of which were apparent in R401D/A539P FOXP2 (LeuZip - end). This was unexpected since R401D FOXP2 (LeuZip - end) was not predicted to have any solvent-accessible cysteine residues, whereas wild-type, A539P and R401D/A539P FOXP2 (LeuZip - end) were predicted to have solvent-accessible cysteine residues (Figure 3.15). However, as previously mentioned, inaccuracies could arise in the structures of the FOXP2 (LeuZip - end) variants since they are based on predictions. In addition, the structures represent only a single proposed conformation adopted by each protein, but experimentally derived data such as size exclusion chromatography data applies to the average of a multitude of conformations adopted by each protein (Bernadó and Blackledge, 2010). The solvent-accessibility of the cysteine residues could fluctuate depending on the conformation of each protein.

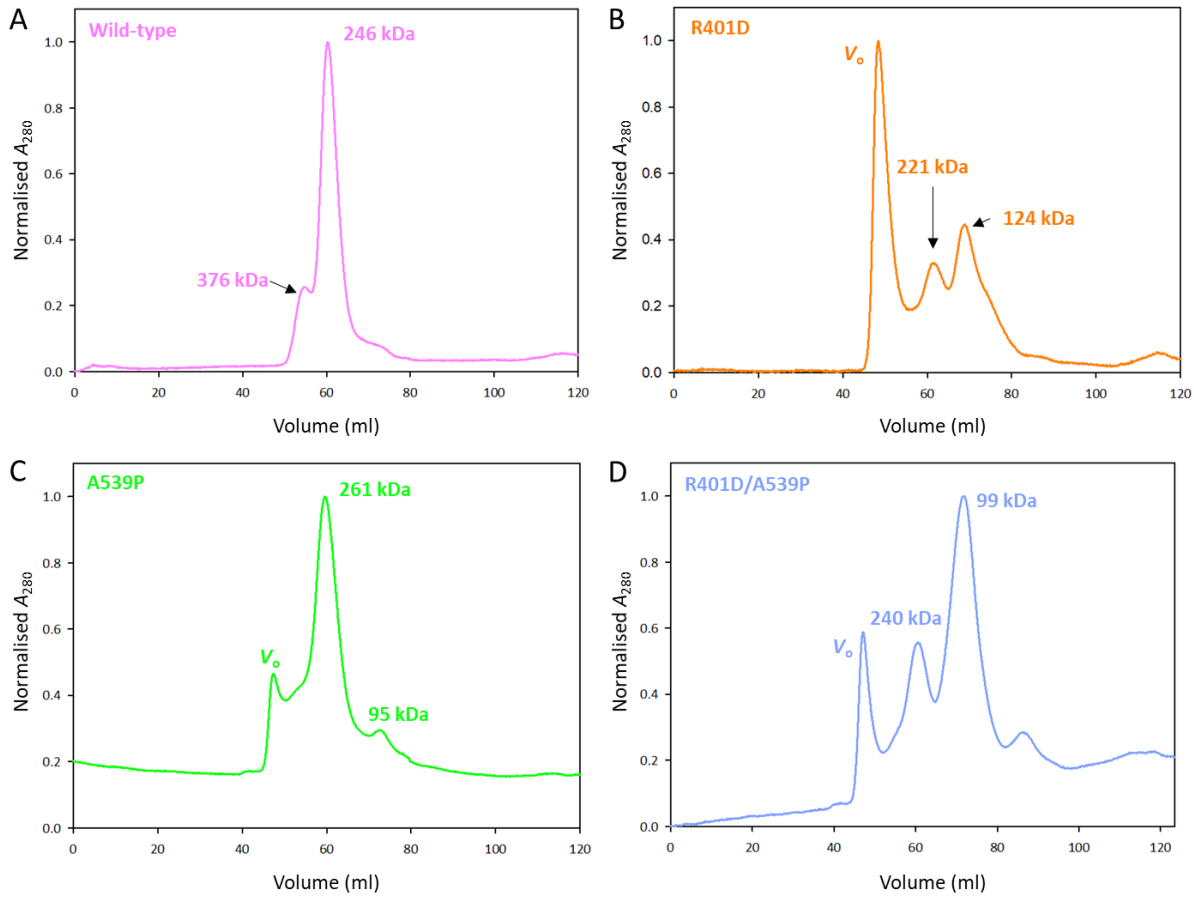


Figure C1: Size exclusion chromatography elution profiles of wild-type (A), R401D (B), A539P (C) and R401D/A539P (D) FOXP2 (LeuZip - end), under non-reducing conditions. Samples were prepared with 25 - 35 µM protein at 0 mM DTT. Data points were normalised between 0 and 1. Wild-type FOXP2 (LeuZip - end) **(A)** consisted of a single major peak at 246 kDa. The R401D mutation caused a new peak to appear at 124 kDa **(B)**. The A539P mutation did not cause any substantial changes on its own, with the major peak of A539P FOXP2 (LeuZip - end) corresponding to 261 kDa **(C)**. However, combined with the R401D mutation, the effects of the A539P mutation were observed with the appearance of a smaller species at 99 kDa **(D)**. The peaks in the void volume (V_o) could be due to large protein masses, since these peaks are much more pronounced in R401D and R401D/A539P FOXP2 (LeuZip - end), which could be on account of incorporation of the R401D mutation enhancing the formation of large protein masses by the exposure of buried LeuZip hydrophobic patches.

APPENDIX D: ELECTROPHORETIC MOBILITY SHIFT ASSAYS (NON-REDUCING CONDITIONS)

EMSAs of R401D and R401D/A539P FOXP2 (LeuZip - end) were conducted under non-reducing conditions (see 3.4.1 Electrophoretic mobility shift assays for reducing conditions). This enabled comparison with size exclusion chromatography data obtained under non-reducing conditions. Images of the EMSAs are shown in Figure D1. For both proteins, the DNA-protein complexes resolved as double bands, which indicated the presence of two DNA-bound protein species. Bands representing free DNA occurred lower on the gel. EMSA results were in agreement with size exclusion chromatography results (Figure C1), since both techniques showed two major species for R401D and R401D/A539P FOXP2 (LeuZip - end) under non-reducing conditions. As with EMSAs conducted under reducing conditions, there was no evidence of the size exclusion chromatography peaks that occurred in the void volume on the EMSAs.

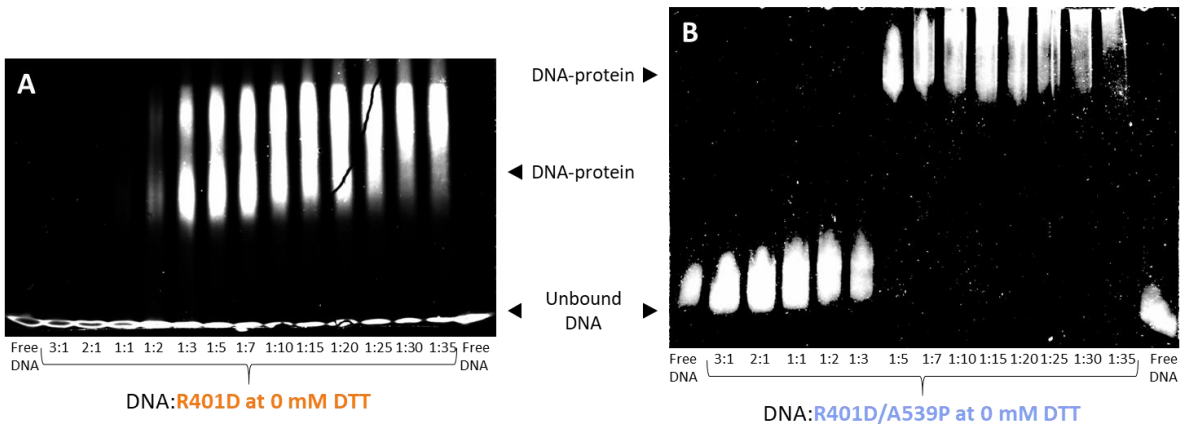


Figure D1: Electrophoretic mobility shift assays comparing the number of DNA-bound species present of R401D (A) and R401D/A539P (B) FOXP2 (LeuZip - end), under non-reducing conditions. A DNA concentration of 0.5 µM was used. The first and last lanes of each assay contained free DNA, and the concentration ratio of protein to DNA was increased in the remaining lanes from 3:1 to 1:35. Sample compositions are indicated below each lane. Samples were prepared at 0 mM DTT. Lower bands represent unbound DNA and upper bands represent DNA-protein complexes. R401D and R401D/A539P FOXP2 (LeuZip - end) both bound DNA as two species, as seen from the double bands.

APPENDIX E: FLUORESCENCE ANISOTROPY DNA-BINDING ASSAYS (REDUCING CONDITIONS)

Fluorescence anisotropy assays were conducted to assess the DNA-binding affinities of the proteins under reducing conditions, since structural characterisations and acrylamide fluorescence quenching – which evaluated the effects of DNA binding on the proteins with respect to structure and dynamics – were conducted under reducing conditions. Thus, DNA-binding affinities were compared in both environments, and it was ensured that the relative affinities of the FOXP2 (LeuZip - end) proteins for DNA were equivalent under non-reducing and reducing conditions (see 3.4.2 Fluorescence anisotropy DNA-binding assays for non-reducing conditions).

Labelled DNA (with a 5-carboxy-X-rhodamine fluorescent marker) at a concentration of 200 nM was combined with the proteins at concentrations of up to 6000 nM. The resulting fluorescence anisotropy signals were measured and used to plot DNA-binding isotherms (Figure E1). The K_Ds for DNA binding were 591 ± 78.5 nM, 815 ± 113 nM, 890 ± 71.3 nM and 1065 ± 126 nM, for wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end), respectively. Therefore, the DNA-binding affinities were wild-type > R401D > A539P > R401D/A539P, arranged from strongest to weakest affinity. This was in agreement with results obtained under non-reducing conditions (Figure 3.27).

Although the DNA-binding affinities of the FOXP2 (LeuZip - end) variants relative to each other were equivalent under non-reducing and reducing conditions, the addition of DTT resulted in an approximate 1.4 - 1.8 fold decrease in DNA-binding affinity for all proteins. Like other thiols, DTT is capable of generating hydroxyl radicals and other oxidative species, which can create DNA nicks and cause damage to DNA bases (Aruoma *et al.,* 1991; Hanna and Mason, 1992; Misra and Fridovich, 1971; Reed and Douglas, 1989, 1991). While previous studies suggested that the effects of DTT on DNA are copper-dependent, a more recent study showed that DTT is capable of creating DNA nicks in double-stranded DNA without any added copper, at concentrations of 0.1 - 10 mM DTT (Fjelstrup *et al.,* 2017; Misra and Fridovich, 1971; Reed and Douglas, 1991). Thus, the decreased DNA-binding affinity observed under reducing conditions could be due to DTT-induced DNA damage, resulting in a loss of FOXP2 DNA binding.

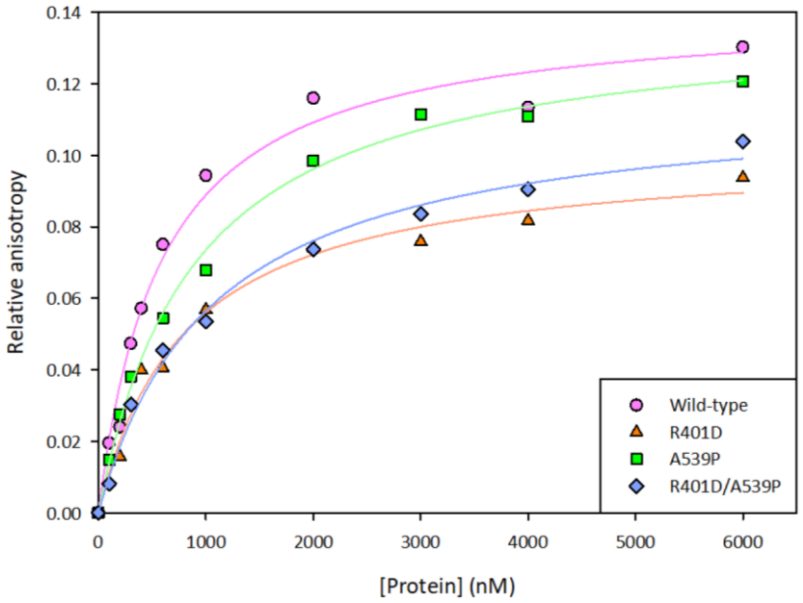


Figure E1: Fluorescence anisotropy DNA-binding assays of wild-type, R401D, A539P and R401D/A539P FOXP2 (LeuZip - end), under reducing conditions. The concentration of DNA was maintained at 200 nM, and the protein concentration was increased from 0 nM to 6000 nM. Samples were prepared at 2 mM DTT. Data points were normalised relative to the anisotropy of free DNA. DNA-binding isotherms were fit with a single-site model. Dissociation constants derived from the isotherms showed that wild-type FOXP2 (LeuZip - end) bound DNA with the highest affinity (K_D = 591 ± 78.5 nM) and R401D/A539P FOXP2 (LeuZip - end) bound DNA with the lowest affinity (K_D = 1065 ± 126 nM). DNA-binding affinities of R401D FOXP2 (LeuZip - end) (K_D = 815 ± 113 nM) and A539P FOXP2 (LeuZip - end) (K_D = 890 ± 71.3 nM) were intermediate to that of wild-type and R401D/A539P FOXP2 (LeuZip - end).

Chapter 6: References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., *et al.* (2000). The genome sequence of Drosophila melanogaster. Science (80-.). *287*, 2185–2195.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., Ribas de Pouplana, L., Martínez-Castilla, L., and Yanofsky, M.F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. U. S. A. *97*, 5328–5333.

Amoutzias, G., Bornberg-Bauer, E., Oliver, S., and Robertson, D. (2006). Reduction/oxidation-phosphorylation control of DNA binding in the bZIP dimerization network. BMC Genomics *7*, 107–117.

Amoutzias, G.D., Robertson, D.L., Oliver, S.G., and Bornberg-Bauer, E. (2004). Convergent evolution of gene networks by single-gene duplications in higher eukaryotes. EMBO Rep. *5*, 274–279.

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

Anckar, J., and Sistonen, L. (2011). Regulation of HSF1 function in the heat stress response: implications in aging and disease. Annu. Rev. Biochem. *80*, 1089–1115.

Aoki, K., Suzuki, K., Sugano, T., Tasaka, T., Nakahara, K., Kuge, O., Omori, A., and Kasai, M. (1995). A novel gene, Translin, encodies a recombination hotspot binding protein associated with chromosomal translocations. Nat. Genet. *10*, 167–174.

Aoki, K., Suzuki, K., Ishida, R., and Kasai, M. (1999). The DNA binding activity of Translin is mediated by a basic region in the ring-shaped structure conserved in evolution. FEBS Lett. *443*, 363–366.

Aoyagi, N., and Wassarman, D.A. (2000). Genes encoding Drosophila melanogaster RNA polymerase II general transcription factors: Diversity in TFIIA and TFIID components contributes to gene-specific transcriptional regulation. J. Cell Biol. *150*, 45–50.

Aravind, L., and Koonin, E. V (1999). DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res. *27*, 4658–4670.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M., and Iyer, L. (2005). The many faces of the helix-turn-helix domain: Transcription regulation and beyond. FEMS Microbiol. Rev. *29*, 231–262.

Artavanis-Tsakonas, S., Rand, M.D., and Lake, R.J. (1999). Notch signaling: Cell fate control and signal integration in development. Science (80-.). *284*, 770–776.

Aruoma, O.I., Halliwell, B., Gajewski, E., and Dizdaroglu, M. (1991). Copper-ion-dependent damage to the bases in DNA in the presence of hydrogen peroxide. Biochem. J. *273*, 601–604.

Aung, K.M.M., New, Y.S., Hong, S., Sutarlie, L., Lim, M.G.L., Tan, S.K., Cheung, E., and Su, X. (2014). Studying forkhead box protein A1–DNA interaction and ligand inhibition using gold nanoparticles, electrophoretic mobility shift assay, and fluorescence anisotropy. Anal. Biochem. *448*, 95–104.

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., *et al.* (2009). Diversity and complexity in DNA recognition by transcription factors. Science (80-.). *324*, 1720–1723.

Bahadur, R.P., Chakrabarti, P., Rodier, F., and Janin, J. (2003). Dissecting subunit interfaces in homodimeric proteins. Proteins Struct. Funct. Genet. *53*, 708–719.

Bailey, P.C., Martin, C., Toledo-Ortiz, G., Quail, P.H., Huq, E., Heim, M.A., Jakoby, M., Werber, M., and Weisshaar, B. (2003). Update on the basic helix-loop-helix transcription factor gene family in Arabidopsis thaliana. Plant Cell *15*, 2497–2502.

Balli, D., Zhang, Y., Snyder, J., Kalinichenko, V. V, and Kalin, T. V (2011). Endothelial cell-specific deletion of transcription factor FoxM1 increases urethane-induced lung carcinogenesis. Cancer Res. *71*, 40–50.

Baltimore, D. (2001). Our genome unveiled. Nature *409*, 814–816.

Bandukwala, H.S., Wu, Y., Feurer, M., Chen, Y., Barbosa, B., Ghosh, S., Stroud, J.C., Benoist, C., Mathis, D., Rao, A., *et al.* (2011). Structure of a domain-swapped FOXP3 dimer on DNA and its function in regulatory T cells. Immunity *34*, 479–491.

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. Curr. Opin. Biotechnol. *10*, 411–421.

Banham, A.H., Beasley, N., Campo, E., Fernandez, P.L., Fidler, C., Gatter, K., Jones, M., Mason, D.Y., Prime, J.E., Trougouboff, P., *et al.* (2001). The FOXP1 winged helix transcription factor is a novel candidate tumor suppressor gene on chromosome 3p. Cancer Res. *61*, 8820–8829.

Barducci, A., Bonomi, M., Prakash, M.K., and Parrinello, M. (2013). Free-energy landscape of protein oligomerization from atomistic simulations. Proc. Natl. Acad. Sci. U. S. A. *110*, 4708–4713.

Barth, H.G., Boyes, B.E., and Jackson, C. (1994). Size exclusion chromatography. Anal. Chem. *66*, 595–620.

Barton, S., Jacak, R., Khare, S.D., Ding, F., and Dokholyan, N. V. (2007). The length dependence of the polyQ-mediated protein aggregation. J. Biol. Chem. *282*, 25487–25492.

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*, 20654–20664.

Basu, S., Söderquist, F., and Wallner, B. (2017). Proteus: a random forest classifier to predict disorder-to-order transitioning binding regions in intrinsically disordered proteins. J. Comput. Aided. Mol. Des. *31*, 453–466.

Baxevanis, A.D., and Vinson, C.R. (1993). Interactions of coiled coils in transcription factors: where is the specificity? Curr. Opin. Genet. Dev. *3*, 278–285.

Beamer, L.J., and Pabo, C.O. (1992). Refined 1.8 Å crystal structure of the λ repressor-operator complex. J. Mol. Biol. *227*, 177–196.

Beckerman, R., and Prives, C. (2010). Transcriptional regulation by p53 and p73. Cold Spring Harb. Perspect. Biol. *2*, 1–18.

Beckett, D. (2001). Regulated assembly of transcription factors and control of transcription initiation. J. Mol. Biol. *314*, 335–352.

Bekheirnia, M.R., Bekheirnia, N., Bainbridge, M.N., Gu, S., Akdemir, Z.H.C., Gambin, T., Janzen, N.K., Jhangiani, S.N., Muzny, D.M., Michael, M., *et al.* (2017). Whole-exome sequencing in the molecular diagnosis of individuals with congenital anomalised of kidney and urinary tract and identification of a new causative gene. Genet. Med. *19*, 412–420.

Benayoun, B.A., and Veitia, R.A. (2009). A post-translational modification code for transcription factors: sorting through a sea of signals. Trends Cell Biol. *19*, 189–197.

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

Bennett, M.J., Choe, S., and Eisenberg, D. (1994). Domain swapping: Entangling alliances between proteins. Proc. Natl. Acad. Sci. U. S. A. *91*, 3127–3131.

Berardi, A., Quilici, G., Spiliotopoulos, D., Corral-Rodriguez, M.A., Martin-Garcia, F., Degano, M., Tonon, G., Ghitti, M., and Musco, G. (2016). Structural basis for PHDVC5HCHNSD1–C2HRNizp1 interaction: implications for Sotos syndrome. Nucleic Acids Res. *44*, 3448–3463.

Berger, C., Piubelli, L., Haditsch, U., and Rudolf Bosshard, H. (1998). Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor. FEBS Lett. *425*, 14–18.

Bernadó, P., and Blackledge, M. (2010). Proteins in dynamic equilibrium. Nature *468*, 1046–1048.

Bertrand, S., Brunet, F.G., Escriva, H., Parmentier, G., Laudet, V., and Robinson-Rechavi, M. (2004). Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol. Biol. Evol. *21*, 1923–1937.

Bhat, U.G., Halasi, M., and Gartel, A.L. (2009). Thiazole antibiotics target FoxM1 and induce apoptosis in human cancer cells. PLoS One *4*, 1–7.

Blanc, R.S., and Richard, S. (2017). Arginine methylation: The coming of age. Mol. Cell *65*, 8–24.

Blane, A., and Fanucchi, S. (2015). Effect of pH on the structure and DNA binding of the FOXP2 forkhead domain. Biochemistry *54*, 4001–4007.

Blane, A., Dirr, H.W., and Fanucchi, S. (2018). A phosphomimetic study implicates Ser557 in regulation of FOXP2 DNA binding. Protein J. *37*, 311–323.

Bledsoe, R.K., Montana, V.G., Stanley, T.B., Delves, C.J., Apolito, C.J., McKee, D.D., Consler, T.G., Parks, D.J., Stewart, E.L., Willson, T.M., *et al.* (2002). Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. Cell *110*, 93–105.

Blount, A.L., Schmidt, K., Justice, N.J., Vale, W.W., Fischer, W.H., and Bilezikjian, L.M. (2009). FoxL2 and Smad3 coordinately regulate follistatin gene transcription. J. Biol. Chem. *284*, 7631–7645.

De Bodt, S., Raes, J., Van De Peer, Y., and Theißen, G. (2003). And then there were many: MADS goes genomic. Trends Plant Sci. *8*, 475–483.

Bookchin, R.M., Balazs, T., Wang, Z., Josephs, R., and Lew, V.L. (1999). Polymer structure and solubility of deoxyhemoglobin S in the presence of high concentrations of volume-excluding 70-kDa dextran. Effects of non-S hemoglobins and inhibitors. J. Biol. Chem. *274*, 6689–6697.

Boura, E., Silhan, J., Herman, P., Vecer, J., Sulc, M., Teisinger, J., Obsilova, V., and Obsil, T. (2007). Both the N-terminal loop and wing W2 of the forkhead domain of transcription factor Foxo4 are important for DNA binding. J. Biol. Chem. *282*, 8265–8275.

Brandt, T., Petrovich, M., Joerger, A.C., and Veprintsev, D.B. (2009). Conservation of DNA-binding specificity and oligomerisation properties within the p53 family. BMC Genomics *10*, 1–14.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. *68*, 394–424.

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

Brown, R.S. (2005). Zinc finger proteins: getting a grip on RNA. Curr. Opin. Struct. Biol. *15*, 94–98.

Brownlie, P., Ceska, T., Lamers, M., Romier, C., Stier, G., Teo, H., and Suck, D. (1997). The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. Structure *5*, 509–520.

Buchan, D.W.A., and Jones, D.T. (2019). The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res. *47*, W402–W407.

Buchan, D.W.A., Ward, S.M., Lobley, A.E., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2010). Protein annotation and modelling servers at University College London. Nucleic Acids Res. *38*, 563–568.

Bujalka, H., Koenning, M., Jackson, S., Perreau, V.M., Pope, B., Hay, C.M., Mitew, S., Hill, A.F., Lu, Q.R., Wegner, M., *et al.* (2013). MYRF is a membrane-associated transcription factor that autoproteolytically cleaves to directly activate myelin genes. PLoS Biol. *11*, 1–15.

Burley, S.K., and Roeder, R.G. (1996). Biochemistry and structural biology of transcription factor IID (TFIID). Annu. Rev. Biochem. *65*, 769–799.

Butland, S.L., Devon, R.S., Huang, Y., Mead, C.-L., Meynert, A.M., Neal, S.J., Lee, S. Sen, Wilkinson, A., Yang, G.S., Yuen, M.M.S., *et al.* (2007). CAG-encoded polyglutamine length polymorphism in the human genome. BMC Genomics *8*, 1–18.

Canadillas, J.M.P., Tidow, H., Freund, S.M. V, Rutherford, T.J., Ang, H.C., and Fersht, A.R. (2006). Solution structure of p53 core domain: Structural basis for its instability. Proc. Natl. Acad. Sci. U. S. A. *103*, 2109–2114.

Carpenter, M.L., Oliver, A.W., and Kneale, G.G. (2001). Analysis of DNA-protein interactions by intrinsic fluorescence. Methods Mol. Biol. *148*, 491–502.

Castaneda, M., Chen, L., Pradhan, L., Li, S., Zein, R., Lee, Y., Lim, H.-S., Nam, H.-J., and Lee, J. (2018). A forkhead box protein C2 inhibitor: targeting epithelial-mesenchymal transition and cancer metastasis. ChemBioChem *19*, 1359–1364.

Castrillon, D.H., Miao, L., Kollipara, R., Horner, J.W., and DePinho, R.A. (2003). Suppression of ovarian follicle activation in mice by the transcription factor Foxo3a. Science (80-.). *301*, 215–218.

Chae, W.-J., Henegariu, O., Lee, S.-K., and Bothwell, A.L.M. (2006). The mutant leucine-zipper domain impairs both dimerization and suppressive function of Foxp3 in T cells. Proc. Natl. Acad. Sci. U. S. A. *103*, 9631–9636.

Chakerian, A.E., Tesmer, V.M., Manly, S.P., Brackett, J.K., Lynch, M.J., Hoh, J.T., and Matthews, K.S. (1991). Evidence for leucine zipper motif in lactose repressor protein. J. Biol. Chem. *266*, 1371–1374.

Chakrabarti, P., and Janin, J. (2002). Dissecting protein-protein recognition sites. Proteins Struct. Funct. Genet. *47*, 334–343.

Chamberlain, N.L., Driver, E.D., and Miesfeld, R.L. (1994). The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res. *22*, 3181–3186.

Charng, W.L., Karaca, E., Coban Akdemir, Z., Gambin, T., Atik, M.M., Gu, S., Posey, J.E., Jhangiani, S.N., Muzny, D.M., Doddapaneni, H., *et al.* (2016). Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. BMC Med. Genomics *9*, 1–14.

Chatila, T.A., Blaeser, F., Ho, N., Lederman, H.M., Voulgaropoulos, C., Helms, C., and Bowcock, A.M. (2000). JM2, encoding a fork head–related protein, is mutated in X-linked autoimmunity–allergic

disregulation syndrome. J. Clin. Invest. *106*, 75–81.

Chen, H., and Privalsky, M.L. (1995). Cooperative formation of high-order oligomers by retinoid X receptors: An unexpected mode of DNA recognition. Proc. Natl. Acad. Sci. U. S. A. *92*, 422–426.

Chen, S.-Y., and Van Der Meer, B.W. (1993). Theory of two-photon induced fluorescence anisotropy decay in membranes. Biophys. J. *64*, 1567–1575.

Chen, S., Rosner, M.H., and Calvo, J.M. (2001). Leucine-regulated self-association of leucine-responsive regulatory protein (Lrp) from Escherichia coli. J. Mol. Biol. *312*, 625–635.

Chen, Y., Chen, C., Zhang, Z., Liu, C., Johnson, M.E., Espinoza, C.A., Edsall, L.E., Ren, B., Zhou, X.J., Grant, S.F.A., *et al.* (2015). DNA binding by FOXP3 domain-swapped dimer suggests mechanisms of long-range chromosomal interactions. Nucleic Acids Res. *43*, 1268–1282.

Chiravuri, M., Lee, H., Mathieu, S.L., and Huber, B.T. (2000). Homodimerization via a leucine zipper motif is required for enzymatic activity of quiescent cell proline dipeptidase. J. Biol. Chem. *275*, 26994–26999.

Chokas, A.L., Trivedi, C.M., Lu, M.M., Tucker, P.W., Li, S., Epstein, J.A., and Morrisey, E.E. (2010). Foxp1/2/4-NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of interleukin. J. Biol. Chem. *285*, 13304–13313.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J. V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science (80-.). *325*, 834–840.

Christiansen, A., Wang, Q., Samiotakis, A., Cheung, M.S., and Wittung-Stafshede, P. (2010). Factors defining effects of macromolecular crowding on protein stability: An in vitro/in silico case study using cytochrome c. Biochemistry *49*, 6519–6530.

Chu, Y., Chang, C., Shiu, J., Chang, Y., Chen, C., and Chuang, W. (2011). Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: Insight into its domain swapping and DNA binding. Protein Sci. *20*, 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*, 412–420.

Cohen, S.X., Moulin, M., Hashemolhosseini, S., Kilian, K., Wegner, M., and Muller, C.W. (2003). Structure of the GCM domain-DNA complex: a DNA-binding domain with a novel fold and mode of target site recognition. EMBO J. *22*, 1835–1845.

Croll, T.I., Sammito, M.D., Kryshtafovych, A., and Read, R.J. (2019). Evaluation of template-based modeling in CASP13. Proteins *87*, 1113–1127.

Cuatrecasas, P., Wilchek, M., and Anfinsen, C.B. (1968). Selective enzyme purification by affinity chromatography. Proc. Natl. Acad. Sci. U. S. A. *61*, 636–643.

Cuiffo, B.G., and Karnoub, A.E. (2016). Silencing FOXP2 in breast cancer cells promotes cancer stem cell traits and metastasis. Mol. Cell. Oncol. *3*, e1019022.

Cuiffo, B.G., Campagne, A., Bell, G.W., Lembo, A., Orso, F., Lien, E.C., Bhasin, M.K., Raimo, M., Hanson, S.E., Marusyk, A., *et al.* (2014). MSC-regulated microRNAs converge on the transcription factor FOXP2 and promote breast cancer metastasis. Cell Stem Cell *15*, 762–774.

Dai, L., and Zhou, Y. (2011). Characterizing the existing and potential structural space of proteins by large-scale multiple loop permutations. J. Mol. Biol. *408*, 585–595.

Danielsson, J., Mu, X., Lang, L., Wang, H., Binolfi, A., Theillet, F.X., Bekei, B., Logan, D.T., Selenko, P., Wennerström, H., *et al.* (2015). Thermodynamics of protein destabilization in live cells. Proc. Natl. Acad. Sci. U. S. A. *112*, 12402–12407.

Deng, H., Jia, Y., and Zhang, Y. (2018). Protein structure prediction. Int. J. Mod. Phys. B *32*, 1–18.

Deng, Y., Liu, J., Zheng, Q., Eliezer, D., Kallenbach, N.R., and Lu, M. (2006). Antiparallel four-stranded coiled coil specified by a 3-3-1 hydrophobic heptad repeat. Structure *14*, 247–255.

Deng, Y., Liu, J., Zheng, Q., Li, Q., Kallenbach, N.R., and Lu, M. (2008). A heterospecific leucine zipper tetramer. Chem. Biol. *15*, 908–919.

Deppmann, C.D., Alvania, R.S., and Taparowsky, E.J. (2006). Cross-species annotation of basic leucine zipper factor interactions: insight into the evolution of closed interaction networks. Mol. Biol. Evol. *23*, 1480–1492.

Deriziotis, P., Graham, S.A., Estruch, S.B., and Fisher, S.E. (2014). Investigating protein-protein interactions in live cells using bioluminescence resonance energy transfer. J. Vis. Exp. 1–15.

Dong, H., Qin, S., and Zhou, H.-X. (2010). Effects of macromolecular crowding on protein conformational changes. PLoS Comput. Biol. *6*, e1000833.

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

Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M., and Uversky, V.N. (2005). Flexible nets: The roles of intrinsic disorder in protein interaction networks. FEBS J. *272*, 5129–5148.

Dunker, A.K., Oldfield, C.J., Meng, J., Romero, P., Yang, J.Y., Chen, J.W., Vacic, V., Obradovic, Z., and Uversky, V.N. (2008). The unfoldomics decade: An update on intrinsically disordered proteins. BMC Genomics *9*, 1–26.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. *6*, 197–208.

Eftink, M.R., and Ghiron, C.A. (1981). Fluorescence quenching studies with proteins. Anal. Biochem. *114*, 199–227.

Eftink, M.R., Selva, T.J., and Wasylewski, Z. (1987). Studies of the efficiency and mechanism of fluorescence quenching reactions using acrylamide and succinimide as quenchers. Photochem. Photobiol. *46*, 23–30.

Eliahoo, E., Marx, A., Manor, H., and Alian, A. (2015). A novel open-barrel structure of octameric translin reveals a potential RNA entryway. J. Mol. Biol. *427*, 756–762.

Ellenberger, T. (1994). Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. Curr. Opin. Struct. Biol. *4*, 12–21.

Emery, B., Agalliu, D., Cahoy, J.D., Watkins, T.A., Dugas, J.C., Mulinyawe, S.B., Ibrahim, A., Ligon, K.L., Rowitch, D.H., and Barres, B.A. (2009). Identification of myelin-gene regulatory factor as a critical transcriptional regulator required for CNS myelination. Cell *138*, 172–185.

Enokizono, Y., Konishi, Y., Nagata, K., Ouhashi, K., Uesugi, S., Ishikawa, F., and Katahira, M. (2005). Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. J. Biol. Chem. *280*, 18862–18870.

Estruch, S.B., Graham, S.A., Chinnappa, S.M., Deriziotis, P., and Fisher, S.E. (2016). Functional characterization of rare FOXP2 variants in neurodevelopmental disorder. J. Neurodev. Disord. *8*, 1–19.

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. Hum. Mol. Genet. *27*, 1212–1227.

Evans, R.F., Kuntz, R.R., Volkert, W.A., and Ghiron, C.A. (1978). Flash photolysis of N-acetyl-L-tryptophanamide: The relationship between radical yields and fluorescence quenching. Photochem. Photobiol. *27*, 511–515.

Favicchio, R., Dragan, A.I., Kneale, G.G., and Read, C.M. (2009). Fluorescence spectroscopy and anisotropy in the analysis of DNA-protein interactions. Methods Mol. Biol. *543*, 589–611.

Ferre-D' Amare, A.R., Pognonec, P., Roeder, R.G., and Burley, S.K. (1994). Structure and function of the b/HLH/Z domain of USF. EMBO J. *13*, 180–189.

Finkenstadt, P.M., Jeon, M., and Baraban, J.M. (2002). Trax is a component of the Translin-containing RNA binding complex. J. Neurochem. *83*, 202–210.

Fisher, C.K., and Stultz, C.M. (2011). Constructing ensembles for intrinsically disordered proteins. Curr. Opin. Struct. Biol. *21*, 426–431.

Fisher, D.E., Carr, C.S., Parent, L.A., and Sharp, P.A. (1991). TFEB has DNA-binding and oligomerization properties of a unique helix-loop-helix/leucine-zipper family. Genes Dev. *5*, 2342–2352.

Fitch, C.A., Platzer, G., Okon, M., Garcia-Moreno, B.E., and McIntosh, L.P. (2015). Arginine: Its pKa

value revisited. Protein Sci. *24*, 752–761.

Fjelstrup, S., Andersen, M.B., Thomsen, J., Wang, J., Stougaard, M., Pedersen, F.S., Ho, Y.-P., Hede, M.S., and Knudsen, B.R. (2017). The effects of dithiothreitol on DNA. Sensors *17*, 1–10.

Flachsbart, F., Caliebe, A., Kleindorp, R., Blanché, H., Von Eller-Eberstein, H., Nikolaus, S., Schreiber, S., and Nebel, A. (2009). Association of FOX03A variation with human longevity confirmed in German centenarians. Proc. Natl. Acad. Sci. U. S. A. *106*, 2700–2705.

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

Fried, M.G. (1989). Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. Electrophoresis *10*, 366-376.

Friedman, J.R., and Kaestner, K.H. (2006). The Foxa family of transcription factors in development and metabolism. Cell. Mol. Life Sci. *63*, 2317–2328.

Friedman, A.M., Fischmann, T.O., and Steitz, T.A. (1995). Crystal structure of lac repressor core tetramer and its implications for DNA looping. Science (80-.). *268*, 1721–1727.

Frueh, D.P., Goodrich, A., Mishra, S., and Nichols, S. (2013). NMR methods for structural studies of large monomeric and multimeric proteins. Curr. Opin. Struct. Biol. *23*, 734–739.

Funnell, A.P.W., and Crossley, M. (2012). Homo- and heterodimerization in transcriptional regulation. In Protein Dimerization and Oligomerization in Biology, J.M. Matthews, ed. (New York: Springer), pp. 105–121.

Gambineri, E., Perroni, L., Passerini, L., Bianchi, L., Doglioni, C., Meschi, F., Bonfanti, R., Sznajer, Y., Tommasini, A., Lawitschka, A., *et al.* (2008). Clinical and molecular profile of a new series of patients with immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome: Inconsistent correlation between forkhead box protein 3 expression and disease severity. J. Allergy Clin. Immunol. *122*, 1105–1112.

Gamsjaeger, R., Liew, C.K., Loughlin, F.E., Crossley, M., and Mackay, J.P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. Trends Biochem. Sci. *32*, 63–70.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server (New York City: Humana Press).

Gaston, K., Bell, A., Busby, S., and Fried, M. (1992). A comparison of the DNA bending activities of the DNA binding proteins CRP and TFIID. Nucleic Acids Res. *20*, 3391–3396.

Gauthier, M., and Degnan, B.M. (2008). The transcription factor NF-κB in the demosponge Amphimedon queenslandica: insights on the evolutionary origin of the Rel homology domain. Dev. Genes Evol. *218*, 23–32.

Gebendorfer, K.M., Drazic, A., Le, Y., Gundlach, J., Bepperling, A., Kastenmüller, A., Ganzinger, K.A.,

Braun, N., Franzmann, T.M., and Winter, J. (2012). Identification of a hypochlorite-specific transcription factor from Escherichia coli. J. Biol. Chem. *287*, 6892–6903.

Ghim, C.-M., and Almaas, E. (2008). Genetic noise control via protein oligomerization. BMC Syst. Biol. *2*, 94.

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

Godzik, A., Jambon, M., and Friedberg, I. (2007). Computational protein function prediction: Are we making progress? Cell. Mol. Life Sci. *64*, 2505–2511.

Gong, X., Jia, M., Ruan, Y., Shuang, M., Liu, J., Wu, S., Guo, Y., Yang, J., Ling, Y., Yang, X., *et al.* (2004). Assocation between the FOXP2 gene and autistic disorder in Chinese population. Am. J. Med. Genet. *127*, 113–116.

Gopnik, M., and Crago, M.B. (1991). Familial aggregation of a developmental language disorder. Cognition *39*, 1–50.

Gorsky, M.K., Burnouf, S., Dols, J., Mandelkow, E., and Partridge, L. (2016). Acetylation mimic of lysine 280 exacerbates human Tau neurotoxicity in vivo. Sci. Rep. *6*, 1–12.

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

Grandori, C., Cowley, S.M., James, L.P., and Eisenman, R.N. (2000). The Myc/Max/Mad network and the transcriptional control of cell behaviour. Annu. Rev. Cell Dev. Biol. *16*, 653–699.

Greenfield, N.J. (1999). Applications of circular dichroism in protein and peptide analysis. TrAC - Trends Anal. Chem. *18*, 236–244.

Greenfield, N.J. (2006a). Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. *1*, 2876–2890.

Greenfield, N.J. (2006b). Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat. Protoc. *1*, 2527–2535.

Greer, E.L., and Brunet, A. (2005). FOXO transcription factors at the interface between longevity and tumor suppression. Oncogene *24*, 7410–7425.

Greer, E.L., and Brunet, A. (2008). FOXO transcription factors in ageing and cancer. Acta Physiol. *192*, 19–28.

Grosberg, A.Y., Nguyen, T.T., and Shklovskii, B.I. (2002). Colloquium: The physics of charge inversion in chemical and biological systems. Rev. Mod. Phys. *74*, 329–345.

Guccione, E., and Richard, S. (2019). The regulation, functions and clinical relevance of arginine methylation. Nat. Rev. Mol. Cell Biol. *20*, 642–657.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for

comparative protein modeling. Electrophoresis *18*, 2714–2723.

Hai, T., and Curran, T. (1991). Cross family dimerization of transcriptional factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proc. Natl. Acad. Sci. U. S. A. *88*, 3720–3724.

Halazonetis, T.D., and Kandil, A.N. (1993). Conformational shifts propagate from the oligomerization domain of p53 to its tetrameric DNA binding domain and restore DNA binding to select p53 mutants. EMBO J. *12*, 5057–5064.

Hall, T.M.T. (2005). Multiple modes of RNA recognition by zinc finger proteins. Curr. Opin. Struct. Biol. *15*, 367–373.

Hamdan, F.F., Daoud, H., Rochefort, D., Piton, A., Gauthier, J., Langlois, M., Foomani, G., Dobrzeniecka, S., Krebs, M.-O., Joober, R., *et al.* (2010). De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. Am. J. Hum. Genet. *87*, 671–678.

Han, J., Gu, W., and Hecht, N.B. (1995). Testis-brain RNA-binding protein, a testicular translational regulatory RNA-binding protein, is present in the brain and binds to the 3' untranslated regions of transported brain mRNAs. Biol. Reprod. *53*, 707–717.

Hanna, P.M., and Mason, R.P. (1992). Direct evidence for inhibiton of free radical formation from Cu(I) and hydrogen peroxide by glutathione and other potential ligands using the EPR spin-trapping technique. Arch. Biochem. Biophys. *295*, 205–213.

Hannenhalli, S., and Kaestner, K.H. (2009). The evolution of Fox genes and their role in development and disease. Nat. Rev. Genet. *10*, 233–240.

Harbury, P.B., Zhang, T., Kim, P.S., and Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science (80-.). *262*, 1401–1407.

Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y., Wang, L., Lupyan, D., Dahlgren, M.K., Knight, J.L., *et al.* (2016). OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. J. Chem. Theory Comput. *12*, 281–296.

Hashimoto, K., Nishi, H., Bryant, S., and Panchenko, A.R. (2011). Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. Phys. Biol. *8*, 1–15.

Hashizume, H., Shiraki, M., and Imahori, K. (1967). Study of circular dichroism of proteins and polypeptides in relation to their backbone and side chain conformations. J. Biochem. *62*, 543–551.

Hauser, K., Essuman, B., He, Y., Coutsias, E., Garcia-Diaz, M., and Simmerling, C. (2016). A human transcription factor in search mode. Nucleic Acids Res. *44*, 63–74.

Häußermann, K., Young, G., Kukura, P., and Dietz, H. (2019). Dissecting FOXP2 oligomerization and DNA binding. Angew. Chemie Int. Ed. *58*, 7662–7667.

Heffler, M.A., Walters, R.D., and Kugel, J.F. (2012). Using electrophoretic mobility shift assays to measure equilibrium dissociation constants: GAL4-p53 binding DNA as a model system. Biochem.

Mol. Biol. Educ. *40*, 383–387.

Hentze, N., Le Breton, L., Wiesner, J., Kempf, G., and Mayer, M.P. (2016). Molecular mechanism of thermosensory function of human heat shock transcription factor Hsf1. Elife *5*, 1–24.

Hermans, K.G., Boormans, J.L., Gasi, D., van Leenders, G.J.H.L., Jenster, G., Verhagen, P.C.M.S., and Trapman, J. (2009). Overexpression of prostate-specific TMPRSS2(exon 0)-ERG fusion transcripts corresponds with favorable prognosis of prostate cancer. Clin. Cancer Res. *15*, 6398–6404.

Hill, R.B., Raleigh, D.P., Lombardi, A., and Degrado, W.F. (2000). De novo design of helical bundles as models for understanding protein folding and function. Acc. Chem. Res. *33*, 745–754.

Hinde, E., Pandzic, E., Yang, Z., Ng, I.H.W., Jans, D.A., Bogoyevitch, M.A., Gratton, E., and Gaus, K. (2016). Quantifying the dynamics of the oligomeric transcription factor STAT3 by pair correlation of molecular brightness. Nat. Commun. *7*, 1–14.

Hirota, S., Hattori, Y., Nagao, S., Taketa, M., Komori, H., Kamikubo, H., Wang, Z., Takahashi, I., Negi, S., Sugiura, Y., *et al.* (2010). Cytochrome c polymerization by successive domain swapping at the C-terminal helix. Proc. Natl. Acad. Sci. U. S. A. *107*, 12854–12859.

van Holde, K.E., Johnson, W.C., and Ho, P.S. (1998). Principles of Physical Biochemistry (New Jersey: Prentice-Hall).

Hollier, B.G., Tinnirello, A.A., Werden, S.J., Evans, K.W., Taube, J.H., Sarkar, T.R., Sphyris, N., Shariati, M., Kumar, S. V., Battula, V.L., *et al.* (2013). FOXC2 expression links epithelial-mesenchymal transition and stem cell properties in breast cancer. Cancer Res. *73*, 1981–1992.

Homouz, D., Perham, M., Samiotakis, A., Cheung, M.S., and Wittung-Stafshede, P. (2008). Crowded, cell-like environment induces shape changes in aspherical protein. Proc. Natl. Acad. Sci. U. S. A. *105*, 11754–11759.

Homouz, D., Stagg, L., Wittung-Stafshede, P., and Cheung, M.S. (2009). Macromolecular crowding modulates folding mechanism of α/β protein apoflavodoxin. Biophys. J. *96*, 671–680.

Huang, Y., and Liu, Z. (2009). Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: A critical assessment of the “fly-casting” mechanism. J. Mol. Biol. *393*, 1143–1159.

Huang, K., Xhani, S., Albrecht, A. V., Ha, V.L.T., Esaki, S., and Poon, G.M.K. (2019). Mechanism of cognate sequence discrimination by the ETS-family transcription factor ETS-1. J. Biol. Chem. *294*, 9666–9678.

Jackson, B.C., Carpenter, C., Nebert, D.W., and Vasiliou, V. (2010). Update of human and mouse forkhead box (FOX) gene families. Hum. Genomics *4*, 345–352.

Jacob, E., Pucshansky, L., Zeruya, E., Baran, N., and Manor, H. (2004). The human protein translin specifically binds single-stranded microsatellite repeats, d(GT)n, and G-strand telomeric repeats, d(TTAGGG)n: A study of the binding parameters. J. Mol. Biol. *344*, 939–950.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., and Parcy, F. (2002). bZIP transcription factors in Arabidopsis. Trends Plant Sci. *7*, 106–111.

Jarrett, J.T., and Lansbury Jr, P.T. (1993). Seeding "one-dimensional crystallization" of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie? Cell *73*, 1055–1058.

Jelesarov, I., and Karshikoff, A. (2008). Meet to fold: The peculiar folding of oligomeric protens. Biotechnol. Biotechnol. Equip. *22*, 598–605.

Jia, W.Z., Yu, T., An, Q., Yang, H., Zhang, Z., Liu, X., and Xiao, G. (2016). MicroRNA-190 regulates FOXP2 genes in human gastric cancer. Onco. Targets. Ther. *9*, 3643–3651.

Jin, C., and Liao, X. (1999). Backbone dynamics of a winged helix protein and its DNA complex at different temperatures: Changes of internal motions in Genesis upon binding to DNA. J. Mol. Biol. *292*, 641–651.

Johnson, T.B., Mechels, K., Anderson, R.E., Cain, J.T., Sturdevant, D.A., Braddock, S., Pinz, H., Wilson, M.A., Landsverk, M., Roux, K.J., *et al.* (2018). Characterization of a recurrent missense mutation in the forkhead DNA-binding domain of FOXP1. Sci. Rep. *8*, 1–8.

Jolly, C., and Morimoto, R.I. (2000). Role of the heat shock response and molecular chaperones in oncogenesis and cell death. J. Natl. Cancer Inst. *92*, 1564–1572.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. *292*, 195–202.

Jones, D.T., and Cozzetto, D. (2015). DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. Bioinformatics *31*, 857–863.

Jones, D.T., and Ward, J.J. (2003). Prediction of disordered regions in proteins from position specific score matrices. Proteins Struct. Funct. Genet. *53*, 573–578.

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

Jones, S., and Thornton, J.M. (1997). Analysis of protein-protein interaction sites using surface patches. J. Mol. Biol. *272*, 121–132.

Jones, S., van Heyningen, P., Berman, H.M., and Thornton, J.M. (1999). Protein-DNA interactions: a structural analysis. J. Mol. Biol. *287*, 877–896.

Jonsson, H., and Peng, S.L. (2005). Forkhead transcription factors in immunology. Cell. Mol. Life Sci. *62*, 397–409.

Jorgensen, W.L., and Tirado-Rives, J. (1988). The OPLS potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. *110*, 1657–1666.

Jorgensen, W.L., Maxwell, D.S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc.

118, 11225–11236.

Jovin, T.M., Bartholdi, M., and Vaz, W.L.C. (1981). Rotational diffusion of biological macromolecules by time-resolved delayed luminescence (phosphorescence, fluorescence) anisotropy. Ann. N. Y. Acad. Sci. *366*, 176–196.

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

Kasai, M., Matsuzaki, T., Katayanagi, K., Omori, A., Maziarz, R.T., Strominger, J.L., Aoki, K., and Suzuki, K. (1997). The translin ring specifically recognizes DNA ends at recombination hot spots in the human genome. J. Biol. Chem. *272*, 11402–11407.

Kaufmann, E., and Knöchel, W. (1996). Five years on the wings of fork head. Mech. Dev. *57*, 3–20.

Kaufmann, E., Muller, D., and Knöchel, W. (1995). DNA recognition site analysis of Xenopus winged helix proteins. J. Mol. Biol. *248*, 239–254.

Khoury, G.A., Baliban, R.C., and Floudas, C.A. (2011). Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. Sci. Rep. *1*, 1–5.

Kim, D., Choi, J.O., Fan, C., Shearer, R.S., Sharif, M., Busch, P., and Park, Y. (2017). Homo-trimerization is essential for the transcription factor function of Myrf for oligodendrocyte differentiation. Nucleic Acids Res. *45*, 5112–5125.

Kim, I.-M., Ackerson, T., Ramakrishna, S., Tretiakova, M., Wang, I.-C., Kalin, T. V., Major, M.L., Gusarova, G.A., Yoder, H.M., Costa, R.H., *et al.* (2006). The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer. Cancer Res. *66*, 2153–2161.

Kim, I.S., Sinha, S., de Crombrugghe, B., and Maity, S.N. (1996). Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule. Mol. Cell. Biol. *16*, 4003–4013.

Kim, P.M., Sboner, A., Xia, Y., and Gerstein, M. (2008). The role of disorder in interaction networks: A structural analysis. Mol. Syst. Biol. *4*, 1–7.

Kishore, D., Kundu, S., and Kayastha, A.M. (2012). Thermal, chemical and pH induced denaturation of a multimeric β-galactosidase reveals multiple unfolding pathways. PLoS One *7*, 50380.

Klaus, M., Prokoph, N., Girbig, M., Wang, X., Huang, Y.-H., Srivastava, Y., Hou, L., Narasimhan, K., Kolatkar, P.R., Francois, M., *et al.* (2016). Structure and decoy-mediated inhibition of the SOX18/Prox1-DNA interaction. Nucleic Acids Res. *44*, 3922–3935.

Klug, A. (1999). Zinc finger peptides for the regulation of gene expression. J. Mol. Biol. *293*, 215–218.

Koenning, M., Jackson, S., Hay, C.M., Faux, C., Kilpatrick, T.J., Willingham, M., and Emery, B. (2012).

Myelin gene regulatory factor is required for maintenance of myelin and mature oligodendrocyte identity in the adult CNS. J. Neurosci. *32*, 12528–12542.

Koh, K.P., Sundrud, M.S., and Rao, A. (2009). Domain requirements and sequence specificity of DNA binding for the forkhead transcription factor FOXP3. Public Libr. Sci. *4*, 1–9.

Kohler, J.J., Metallo, S.J., Schneider, T.L., and Schepartz, A. (1999). DNA specificity enhanced by sequential binding of protein monomers. Proc. Natl. Acad. Sci. U. S. A. *96*, 11735–11739.

Konopka, G., Bomar, J.M., Winden, K., Coppola, G., Jonsson, Z.O., Gao, F., Peng, S., Preuss, T.M., Wohlschlegel, J.A., and Geschwind, D.H. (2009). Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature *462*, 213–217.

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

Koswatta, T.J., Samaraweera, P., and Sumanasinghe, V.A. (2011). A simple comparison between specific protein secondary structure prediction tools. Trop. Agric. Res. *23*, 91–98.

Kryshtafovych, A., Monastyrskyy, B., Fidelis, K., Moult, J., Schwede, T., and Tramontano, A. (2018). Evaluation of the template-based modeling in CASP12. Proteins *86*, 321–334.

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

Lai, C.S., 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*, 519–523.

Lai, E., Clark, K.L., Burley, S.K., and Darnell, J.E. (1993). Hepatocyte nuclear factor 3/fork head or “winged helix” proteins: A family of transcription factors of diverse biologic function. Proc. Natl. Acad. Sci. U. S. A. *90*, 10421–10423.

Laissue, P. (2019). The forkhead-box family of transcription factors: key molecular players in colorectal cancer pathogenesis. Mol. Cancer *18*, 1–13.

Lakowicz, J.R. (2006a). Principles of Fluorescence Spectroscopy (New York: Springer Science).

Lakowicz, J.R. (2006b). Protein Fluorescence. In Principles of Fluorescence Spectroscopy, (New York: Springer Science), pp. 529–575.

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

Lander, E., Linton, L., Birren, B., Nusbaum, C., Zody, M., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

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 (80-.). *240*, 1759–1764.

Lapidus, L.J. (2017). Protein unfolding mechanisms and their effects on folding experiments. F1000Research *6*, 1–8.

Larsen, T.A., Olson, A.J., and Goodsell, D.S. (1998). Morphology of protein-protein interfaces. Structure *6*, 421–427.

Latchman, D.S. (1993). Transcription factors: an overview. Int. J. Biochem. Cell Biol. *29*, 417–422.

Lee, T., and Young, R. (2000). Transcription of eukayotic protein-coding genes. Annu. Rev. Genet. *34*, 77–137.

Lee, T., and Young, R. (2014). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251.

Lee, C.S., Friedman, J.R., Fulmer, J.T., and Kaestner, K.H. (2005). The initiation of liver development is dependent on Foxa transcription factors. Nature *435*, 944–947.

Lee, K.S., Bumbaca, D., Kosman, J., Setlow, P., and Jedrzejas, M.J. (2008). Structure of a protein-DNA complex essential for DNA protection in spores of Bacillus species. Proc. Natl. Acad. Sci. U. S. A. *105*, 2806–2811.

Leonard, M. (1997). New packing materials for protein chromatography. J. Chromatogr. B Biomed. Appl. *699*, 3–27.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147–151.

Levy, D.E., and Marié, I.J. (2012). STATus report on tetramers. Immunity *36*, 553–555.

Li, B., Samanta, A., Song, X., Iacono, K.T., Brennan, P., Chatila, T.A., Roncador, G., Banham, A.H., Riley, J.L., Wang, Q., *et al.* (2007). FOXP3 is a homo-oligomer and a component of a supramolecular regulatory complex disabled in the human XLAAD/IPEX autoimmune disease. Int. Immunol. *19*, 825–835.

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. Mol. Cell. Biol. *24*, 809–822.

Li, X., Wang, W., Wang, J., Malovannaya, A., Xi, Y., Li, W., Guerra, R., Hawke, D.H., Qin, J., and Chen, J. (2015). Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. Mol. Syst. Biol. *11*, 775.

Li, Y., Zhang, X., and Cao, D. (2013a). The role of shape complementarity in the protein-protein interactions. Sci. Rep. *3*, 3–9.

Li, Z., Park, Y., and Marcotte, E.M. (2013b). A bacteriophage tailspike domain promotes self-cleavage of a human membrane-bound transcription factor, the myelin regulatory factor MYRF. PLoS Biol. *11*, 1–14.

Lindner, R.A., and Ralston, G.B. (1997). Macromolecular crowding: Effects on actin polymerisation.

Biophys. Chem. *66*, 57–66.

Lipfert, J., Doniach, S., Das, R., and Herschlag, D. (2014). Understanding nucleic acid-ion interactions. Annu. Rev. Biochem. *83*, 813–841.

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

Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors.

Liu, J., Zheng, Q., Deng, Y., Li, Q., Kallenbach, N.R., and Lu, M. (2007). Conformational specificity of the lac repressor coiled-coil tetramerization domain. Biochemistry *46*, 14951–14959.

Liu, L.-L., Zhang, D.-H., Mao, X., Zhang, X.-H., and Zhang, B. (2014). Over-expression of FoxM1 is associated with adverse prognosis and FLT3-ITD in acute myeloid leukemia. Biochem. Biophys. Res. Commun. *446*, 280–285.

Liu, Y., Gotte, G., Libonati, M., and Eisenberg, D. (2001). A domain-swapped RNase A dimer with implications for amyloid formation. Nat. Struct. Biol. *8*, 211–214.

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

van Loosdregt, J., Vercoulen, Y., Guichelaar, T., Gent, Y.Y.J., Beekman, J.M., van Beekum, O., Brenkman, A.B., Hijnen, D.-J., Mutis, T., Kalkhoven, E., *et al.* (2010). Regulation of Treg functionality by acetylation-mediated Foxp3 protein stabilization. Blood *115*, 965–974.

van Loosdregt, J., Brunen, D., Fleskens, V., Pals, C.E.G.M., Lam, E.W.F., and Coffer, P.J. (2011). Rapid temporal control of foxp3 protein degradation by sirtuin-1. PLoS One *6*, 1–10.

Lopes, J.E., Torgerson, T.R., Schubert, L.A., Anover, S.D., Ocheltree, E.L., Ochs, H.D., and Ziegler, S.F. (2006). Analysis of FOXP3 reveals multiple domains required for its function as a transcriptional repressor. J. Immunol. *177*, 3133–3142.

Lotte, K., Plessow, R., and Brockhinke, A. (2004). Static and time-resolved fluorescence investigations of tryptophan analogues - A solvent study. Photochem. Photobiol. Sci. *3*, 348–359.

Low, L.Y., Hernandez, H., Robinson, C. V, O'Brien, R., Grossmann, J.G., Ladbury, J.E., and Luisi, B. (2002). Metal-dependent folding and stability of nuclear hormone receptor DNA-binding domains. J. Mol. Biol. *319*, 87–106.

Lozano, T., Gorraiz, M., Lasarte-Cía, A., Ruiz, M., Rabal, O., Oyarzabal, J., Hervás-stubbs, S., Llopiz, D., Sarobe, P., Prieto, J., *et al.* (2017). Blockage of FOXP3 transcription factor dimerization and FOXP3/AML1 interaction inhibits T regulatory cell activity: sequence optimization of a peptide inhibitor. Oncotarget *8*, 71709–71724.

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. Mech. Dev. *2*, S197–S202.

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. Nature *352*, 497–505.

Luo, H., Jin, K., Xie, Z., Qiu, F., Li, S., Zou, M., Cai, L., Hozumi, K., Shima, D.T., and Xiang, M. (2012). Forkhead box N4 (Foxn4) activates Dll4-Notch signaling to suppress photoreceptor cell fates of early retinal progenitors. Proc. Natl. Acad. Sci. U. S. A. *109*, 553–562.

Lüscher, B. (2001). Function and regulation of the transcription factors of the Myc/Max/Mad network. Gene *277*, 1–14.

Ma, J., and Ptashne, M. (1988). Converting a eukaryotic transcriptional inhibitor into an activator. Cell *55*, 443–446.

Ma, X., Xu, L., Gavrilova, O., and Mueller, E. (2014). Role of forkhead box protein A3 in age-associated metabolic decline. Proc. Natl. Acad. Sci. U. S. A. *111*, 14289–14294.

Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. J. Comput. Aided. Mol. Des. *27*, 221–234.

Madureira, P.A., Varshochi, R., Constantinidou, D., Francis, R.E., Coombes, R.C., Yao, K.-M., and Lam, E.W.F. (2006). The forkhead box M1 protein regulates the transcription of the estrogen receptor α in breast cancer cells. J. Biol. Chem. *281*, 25167–25176.

Mao, Y.-J., Sheng, X.-R., and Pan, X.-M. (2007). The effects of NaCl concentration and pH on the stability of hyperthermophilic protein Ssh10b. BMC Biochem. *8*, 1–8.

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

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

Mathiasen, L., Valentini, E., Boivin, S., Cattaneo, A., Blasi, F., Svergun, D.I., and Bruckmann, C. (2016). The flexibility of a homeodomain transcription factor heterodimer and its allosteric regulation by DNA binding. FEBS J. *283*, 3134–3154.

Matthews, J.M., and Sunde, M. (2002). Zinc fingers — folds for many occasions. IUBMB Life *54*, 351–355.

Mayr, F., Schutz, A., Doge, N., and Heinemann, U. (2012). The Lin28 cold-shock domain remodels pre-

let-7 microRNA. Nucleic Acids Res. *40*, 7492–7506.

Mendoza, E. (2011). FoxP1, FoxP2 and FoxP4 in the song control system of zebra finches: molecular interactions and relevance for vocal learning. Freie Universität.

de Mendoza, A., Sebé-Pedrós, A., Sestak, M.S., Matejcic, M., Torruella, G., Domazet-Loso, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. U. S. A. *110*, 4858–4866.

Metrick, M.A., Temple, J.E., and MacDonald, G. (2013). The effects of buffers and pH on the thermal stability, unfolding and substrate binding of RecA. Biophys. Chem. *184*, 29–36.

Midic, U., Oldfield, C., Dunker, A.K., Obradovic, Z., and Uversky, V.N. (2009). Unfoldomics of human genetic diseases: illustrative examples of ordered and intrinsically disordered members of the human diseasome. Protein Pept. Lett. *16*, 1533–1547.

Miller, M., Shuman, J.D., Sebastian, T., Dauter, Z., and Johnson, P.F. (2003). Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein α*. J. Biol. Chem. *278*, 15178–15184.

Miller, Y., Ma, B., and Nussinov, R. (2010). Zinc ions promote Alzheimer AB aggregation via population shift of polymorphic states. Proc. Natl. Acad. Sci. U. S. A. *107*, 9490–9495.

Minton, A.P. (1981). Excluded volume as a determinant of macromolecular structure and stability. Biopolymers *20*, 2093–2120.

Minton, A.P. (2000). Implications of macromolecular crowding for protein assembly. Curr. Opin. Struct. Biol. *10*, 34–39.

Misra, H.P., and Fridovich, I. (1971). The generation of superoxide radical during the autoxidation of ferredoxins. J. Biol. Chem. *247*, 6886–6890.

Monastyrskyy, B., Fidelis, K., Moult, J., Tramontano, A., and Kryshtafovych, A. (2011). Evaluation of disorder predictions in CASP9. Proteins Struct. Funct. Bioinforma. *79*, 107–118.

Monastyrskyy, B., Kryshtafovych, A., Moult, J., Tramontano, A., and Fidelis, K. (2014). Assessment of protein disorder region predictions in CASP10. Proteins Struct. Funct. Bioinforma. *82*, 127–137.

Morris, G., and Fanucchi, S. (2016). A key evolutionary mutation enhances DNA binding of the FOXP2 forkhead domain. Biochemistry *55*, 1959–1967.

Morris, G., Stoychev, S., Naicker, P., Dirr, H.W., and Fanucchi, S. (2018a). The forkhead domain hinge-loop plays a pivotal role in DNA binding and transcriptional activity of FOXP2. Biol. Chem. *399*, 881–893.

Morris, G., Pahad, N., Dirr, H.W., and Fanucchi, S. (2018b). A conserved cation binding site in the DNA binding domain of forkhead box transcription factors regulates DNA binding by FOXP2. Arch. Biochem. Biophys. *657*, 56–64.

Moult, J., Pedersen, J.T., Judson, R., and Fidelis, K. (1995). A large-scale experiment to assess protein structure prediction methods. Proteins Struct. Funct. Genet. *23*, ii–v.

Mukherjee, A., and Bagchi, B. (2006). Solvent frictional forces in the rotational diffusion of proteins in water. Curr. Sci. *91*, 1208–1216.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.

Myatt, S.S., and Lam, E.W.-F. (2007). The emerging roles of forkhead box (Fox) proteins in cancer. Nat. Rev. Cancer *7*, 847–859.

Myers, A., du Souich, C., Yang, C.L., Borovik, L., Mwenifumbo, J., Rupps, R., Study, C.A.U.S.E.S., Lehman, A., and Boerkoel, C.F. (2017). FOXP1 haploinsufficiency: Phenotypes beyond behavior and intellectual disability? Am. J. Med. Genet. *173A*, 3172–3181.

Nair, S.K., and Burley, S.K. (2003). X-Ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors. Cell *112*, 193–205.

Nakagawa, S., Gisselbrecht, S.S., Rogers, J.M., Hartl, D.L., and Bulyk, M.L. (2013). DNA-binding specificity changes in the evolution of forkhead transcription factors. Proc. Natl. Acad. Sci. U. S. A. *110*, 12349–12354.

Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K.T., and Takahashi, T. (2006). Characterization of the class IV homeodomain-leucine zipper gene family in Arabidopsis. Plant Physiol. *141*, 1363–1375.

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 Res. *41*, 5991–6004.

Nicholls, C.D., McLure, K.G., Shields, M.A., and Lee, P.W.K. (2002). Biogenesis of p53 involves cotranslational dimerization of monomers and posttranslational dimerization of dimers. Implications on the dominant negative effect. J. Biol. Chem. *277*, 12937–12945.

Nielsen, J.T., and Mulder, F.A.A. (2019). Quality and bias of protein disorder predictors. Sci. Rep. *9*, 1–12.

Nielsen, A.L., Jorgensen, P., Lerouge, T., Cervino, M., Chambon, P., and Losson, R. (2004). Nizp1, a novel multitype zinc finger protein that interacts with the NSD1 histone lysine methyltransferase through a unique C2HR motif. Mol. Cell. Biol. *24*, 5184–5196.

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

Nikolov, D.B., Chen, H., Halay, E.D., Hoffmannt, A., Roeder, R.G., and Burley, S.K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. Proc. Natl. Acad. Sci. U. S.

A. *93*, 4862–4867.

Nooren, I.M.A., and Thornton, J.M. (2003). Diversity of protein-protein interactions. EMBO J. *22*, 3486–3492.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. *302*, 205–217.

O'Neil, K.T., Hoess, R.H., and DeGrado, W.F. (1990). Design of DNA-binding peptides based on the leucine zipper motif. Science (80-.). *249*, 774–778.

O'Shea, E.K., Klemm, J.D., Kim, P.S., and Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science (80-.). *254*, 539–544.

Obsil, T., and Obsilova, V. (2008). Structure/function relationships underlying regulation of FOXO transcription factors. Oncogene *27*, 2263–2275.

Obsil, T., and Obsilova, V. (2011). Structural basis for DNA recognition by FOXO proteins. Biochim. Biophys. Acta *1813*, 1946–1953.

Ogle, J.M., and Ramakrishnan, V. (2005). Structural insights into translational fidelity. Annu. Rev. Biochem. *74*, 129–177.

Ostashevskii, I.Y. (1972). Connective between parameters of fluorescence and protein structure. Mol. Biol. *6*, 1–9.

Overdier, D.G., Porcella, A., and Costa, R.H. (1994). The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. Mol. Cell. Biol. *14*, 2755–2766.

Ozkan, S.B., Wu, G.A., Chodera, J.D., and Dill, K.A. (2007). Protein folding by zipping and assembly. Proc. Natl. Acad. Sci. U. S. A. *104*, 11987–11992.

Pan, Y., and Nussinov, R. (2011). The role of response elements organizaiton in transcription factor selectivity: The IFN-B enhanceosome example. PLoS Comput. Biol. *7*, 1–14.

Parker, J. (1989). Errors and alternatives in reading the universal genetic code. Microbiol. Rev. *53*, 273–298.

Pastore, A., Martin, S.R., Politou, A., Kondapalli, K.C., Stemmler, T., and Temussi, P.A. (2007). Unbiased cold denaturation: low- and high-temperature unfolding of yeast frataxin under physiological conditions. J. Am. Chem. Soc. *129*, 5374–5375.

Patel, L., Abate, C., and Curran, T. (1990). Altered protein conformation on DNA binding by Fos and Jun. Nature *347*, 572–575.

Pawlak, M., Lefebvre, P., and Staels, B. (2012). General molecular biology and architecture of nuclear receptors. Curr. Top. Med. Chem. *12*, 486–504.

Perham, M., Stagg, L., and Wittung-Stafshede, P. (2007). Macromolecular crowding increases

structural content of folded proteins. FEBS Lett. *581*, 5065–5069.

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. Protein J. *34*, 111–121.

Perutz, M.F., Johnson, T., Suzuki, M., and Finch, J.T. (1994). Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. Proc. Natl. Acad. Sci. U. S. A. *91*, 5355–5358.

Pierrou, S., Hellqvist, M., Samuelsson, L., Enerbäck, S., and Carlsson, P. (1994). Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. EMBO J. *13*, 5002–5012.

Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. Nature *258*, 598–599.

Prabakaran, P., Siebers, J.G., Ahmad, S., Gromiha, M.M., Singarayan, M.G., and Sarai, A. (2006). Classification of Protein-DNA complexes based on structural descriptors. Structure *14*, 1355–1367.

Privalov, P.L., Griko, Y. V., Venyaminov, S.Y., and Kutyshenko, V.P. (1986). Cold denaturation of myoglobin. J. Mol. Biol. *190*, 487–498.

Provencher, S.W., and Glockner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. Biochemistry *20*, 33–37.

Psenakova, K., Kohoutova, K., Obsilova, V., Ausserlechner, M.J., Veverka, V., and Obsil, T. (2019). Forkhead domains of FOXO transcription factors differ in both overall conformation and dynamics. Cells *8*, 966.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577.

Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R., Meyer, B.J., Pabo, C.O., Roberts, T.M., and Sauer, R.T. (1980). How the λ repressor and cro work. Cell *19*, 1–11.

Rabindran, S.K., Haroun, R.I., Clos, J., Wisniewski, J., and Wu, C. (1993). Regulation of heat shock factor trimer formation: role of a conserved leucine zipper. Science (80-.). *259*, 230–234.

Reed, C.J., and Douglas, K.T. (1989). Single-strand cleavage of DNA by Cu(II) and thiols: A powerful chemical DNA-cleaving system. Biochem. Biophys. Res. Commun. *162*, 1111–1117.

Reed, C.J., and Douglas, K.T. (1991). Chemical cleavage of plasmid DNA by glutathione in the presence of Cu(II) ions. The Cu(II)-thiol system for DNA strand scission. Biochem. J. *275*, 601–608.

Ren, J., Han, L., Tang, J., Liu, Y., Deng, X., Liu, Q., Hao, P., Feng, X., Li, B., Hu, H., *et al.* (2019). Foxp1 is critical for the maintenance of regulatory T-cell homeostasis and suppressive function. PLoS Biol. *17*, 1–29.

Rivas, G., Fernandez, J.A., and Minton, A.P. (2001). Direct observation of the enhancement of noncooperative protein self-assembly by macromolecular crowding: Indefinite linear self-association of bacterial cell division protein FtsZ. Proc. Natl. Acad. Sci. U. S. A. *98*, 3150–3155.

Roll, P., Vernes, S.C., Bruneau, N., Cillario, J., Ponsole-Lenfant, M., Massacrier, A., Rudolf, G., Khalife, M., Hirsch, E., Fisher, S.E., *et al.* (2010). Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. Hum. Mol. Genet. *19*, 4848–4860.

Roman, C., Matera, A.G., Cooper, C., Artandi, S., Blain, S., Ward, D.C., and Calame, K. (1992). mTFE3, an X-linked transcriptional activator containing basic helix-loop-helix and zipper domains, utilizes the zipper to stabilize both DNA binding and multimerization. Mol. Cell. Biol. *12*, 817–827.

Rousso, D.L., Pearson, C.A., Gaber, Z.B., Miquelajauregui, A., Li, S., Portera-Cailliau, C., Morrisey, E.E., and Novitch, B.G. (2012). Foxp-mediated suppression of N-cadherin regulates neuroepithelial character and progenitor maintenance in the CNS. Neuron *74*, 314–330.

Roy, S., and Kumar, V. (2014). A practical approach on SDS PAGE for separation of protein. Int. J. Sci. Res. *3*, 955–960.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. *5*, 725–738.

Rudra, D., DeRoos, P., Chaudhry, A., Niec, R., Arvey, A., Samstein, R.M., Leslie, C., Shaffer, S.A., Goodlett, D.R., and Rudensky, A.Y. (2012). Transcription factor Foxp3 and its protein partners form a complex regulatory network. Nat. Immunol. *13*, 1010–1019.

Ruvkun, G., and Hobert, O. (1998). The taxonomy of developmental control in Caenorhabditis elegans. Science (80-.). *282*, 2033–2041.

Salih, D.A.M., and Brunet, A. (2008). FoxO transcription factors in the maintenance of cellular homeostasis during aging. Curr. Opin. Cell Biol. *20*, 126–136.

Samanta, A., Li, B., Song, X., Bembas, K., Zhang, G., Katsumata, M., Saouaf, S.J., Wang, Q., Hancock, W.W., Shen, Y., *et al.* (2008). TGF-B and IL-6 signals modulate chromatin binding and promoter occupancy by acetylated FOXP3. Proc. Natl. Acad. Sci. U. S. A. *105*, 14023–14027.

Sanjuán, J., Tolosa, A., González, J.C., Aguilar, E.J., Pérez-Tur, J., Nájera, C., Moltó, M.D., and Frutos, R. De (2006). Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations. Psychiatr. Genet. *16*, 67–72.

Sawaya, M.R., Guo, S., Tabor, S., Richardson, C.C., and Ellenberger, T. (1999). Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. Cell *99*, 167–177.

Schlake, T., Schorpp, M., Nehls, M., and Boehm, T. (1997). The nude gene encodes a sequence-specific DNA binding protein with homologs in organisms that lack an anticipatory immune system. Proc. Natl. Acad. Sci. U. S. A. *94*, 3842–3847.

Schmid, F. (2001). Biological Macromolecules : Spectrophotometry Concentrations. Encycl. Life Sci. 1–4.

Seelig, J., and Schönfeld, H.-J. (2016). Thermal protein unfolding by differential scanning calorimetry

and circular dichroism spectroscopy Two-state model versus sequential unfolding. Q. Rev. Biophys. *49*, 1–24.

Seoane, J., Le, H.-V., Shen, L., Anderson, S.A., and Massague, J. (2004). Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. Cell *117*, 211–223.

Shaulian, E., Zauberman, A., Milner, J., Davies, E.A., and Oren, M. (1993). Tight DNA binding and oligomerization are dispensable for the ability of p53 to transactivate target genes and suppress transformation. EMBO J. *12*, 2789–2797.

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

Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. (2010). Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J. Chem. Theory Comput. *6*, 1509–1519.

Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. (2000). Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc. Natl. Acad. Sci. U. S. A. *97*, 8868–8873.

Shore, P., and Sharrocks, A.D. (1995). The MADS-box family of transcription factors. Eur. J. Biochem. *229*, 1–13.

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. J. Biol. Chem. *276*, 27488–27497.

Shu, W., Lu, M.M., Zhang, Y., Tucker, P.W., Zhou, D., and Morrisey, E.E. (2007). Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. Dev. Dis. *134*, 1991–2000.

Shuman, J.D., Vinson, C.R., and McKnight, S.L. (1990). Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. Science (80-.). *249*, 771–774.

Silhan, J., Vacha, P., Strnadova, P., Vecer, J., Herman, P., Sulc, M., Teisinger, J., Obsilova, V., and Obsil, T. (2009). 14-3-3 protein masks the DNA binding interface of forkhead transcription factor FOXO4. J. Biol. Chem. *284*, 19349–19360.

Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B.M., and Vervoort, M. (2007). Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol. Biol. *7*.

Sin, C., Li, H., and Crawford, D.A. (2014). Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. J. Mol. Neurosci. *55*, 437–448.

Singleton, M.R., Sawaya, M.R., Ellenberger, T., and Wigley, D.B. (2000). Crystal structure of T7 gene 4 ring indicates a mechanism for sequential hydrolysis of nucleotides. Cell *101*, 589–600.

Sinha, S., Maity, S.N., Lu, J., and De Crombrugghe, B. (1995). Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. Proc. Natl. Acad. Sci. U. S. A. *92*, 1624–1628.

Sinha, S., Kim, I.S., Sohn, K.Y., de Crombrugghe, B., and Maity, S.N. (1996). Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. Mol. Cell. Biol. *16*, 328–337.

Siper, P.M., De Rubeis, S., Trelles, M. del P., Durkin, A., Di Marino, D., Muratet, F., Frank, Y., Lozano, R., Eichler, E.E., Kelly, M., *et al.* (2017). Prospective investigation of FOXP1 syndrome. Mol. Autism *8*, 57.

Slabinski, L., Jaroszewski, L., Rodrigues, A.P.C., Rychlewski, L., Wilson, I.A., Lesley, S.A., and Godzik, A. (2007). The challenge of protein structure determination - lessons from structural genomics. Protein Sci. *16*, 2472–2482.

Smith, L., and Kelleher, N. (2013). Proteoform: a single term describing protein complexity. Nat. Methods *10*, 186–187.

Smolen, P., Baxter, A.D., and Byrne, H.J. (2000). Mathematical modeling of gene networks. Neuron *26*, 567–580.

Snoussi, K., and Halle, B. (2005). Protein self-association induced by macromolecular crowding: A quantitative analysis by magnetic relaxation dispersion. Biophys. J. *88*, 2855–2866.

Solan, A., Ratia, K., and Fairman, R. (2002). Exploring the role of alanine in the structure of the Lac repressor tetramerization domain, a ferritin-like Alacoil. J. Mol. Biol. *317*, 601–612.

Song, J., Yuan, Z., Tan, H., Huber, T., and Burrage, K. (2007). Predicting disulfide connectivity from protein sequence using multiple sequence feature vectors and secondary structure. Bioinformatics *23*, 3147–3154.

Song, X., Li, B., Xiao, Y., Chen, C., Wang, Q., Liu, Y., Berezov, A., Xu, C., Gao, Y., Li, Z., *et al.* (2012). Structural and biological features of FOXP3 dimerization relevant to regulatory T cell function. Cell Rep. *1*, 665–675.

Spaeth, J.M., Hunter, C.S., Bonatakis, L., Guo, M., French, C.A., Slack, I., Hara, M., Fisher, S.E., Ferrer, J., Morrisey, E.E., *et al.* (2015). The FOXP1, FOXP2 and FOXP4 transcription factors are required for islet alpha cell proliferation and function in mice. Diabetologia *58*, 1836–1844.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Sreerama, N., and Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal. Biochem. *287*, 252–260.

Srivastava, A., Nagai, T., Srivastava, A., Miyashita, O., and Tama, F. (2018). Role of computational methods in going beyond x-ray crystallography to explore protein structure and dynamics. Int. J. Mol. Sci. *19*, 3401–3423.

Stein, A., Pache, R.A., Bernadó, P., Pons, M., and Aloy, P. (2009). Dynamic interactions of proteins in complex networks: A more structured view. FEBS J. *276*, 5390–5405.

van Stokkum, I.H., Spoelde, H.J., Bloemendal, M., van Grondelle, R., and Groen, F.C. (1990). Estimation of protein secondary structure and error analysis from circular dichroism spectra. Anal. Biochem. *15*, 110–118.

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*, 159–166.

Stumm, L., Burkhardt, L., Steurer, S., Simon, R., Adam, M., Becker, A., Sauter, G., Minner, S., Schlomm, T., Sirma, H., *et al.* (2013). Strong expression of the neuronal transcription factor FOXP2 is linked to an increased risk of early PSA recurrence in ERG fusion-negative cancers. J. Clin. Pathol. *66*, 563–568.

Sundberg, R.J., and Martin, R.B. (1974). Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological systems. Chem. Rev. *74*, 471–517.

Tacal, O., and Ozer, I. (2002). A comparison between SDS-PAGE and size exclusion chromatography as analytical methods for determining product composition in protein conjugation reactions. J. Biochem. Biophys. Methods *52*, 161–168.

Takahashi, H., Takahashi, K., and Liu, F.-C. (2009). FOXP genes, neural development, speech and language disorders. Adv. Exp. Med. Biol. *665*, 117–129.

Takahashi, K., Liu, F.-C., Hirokawa, K., and Takahashi, H. (2008a). Expression of Foxp4 in the developing and adult rat forebrain. J. Neurosci. Res. *86*, 3106–3116.

Takahashi, K., Liu, F.-C., Oishi, T., Mori, T., Higo, N., Hayashi, M., Hirokawa, K., and Takahashi, H. (2008b). Expression of FOXP2 in the developing monkey forebrain: Comparison with the expression of the genes FOXP1, PBX3, and MEIS2. J. Comp. Neurol. *509*, 180–189.

Talavera, D., Robertson, D.L., and Lovell, S.C. (2011). Characterization of protein-protein interaction interfaces from a single species. PLoS One *6*, 1–9.

Tanabe, Y., Fujita, E., and Momoi, T. (2011). FOXP2 promotes the nuclear translocation of POT1, but FOXP2 (R553H), mutation related to speech-language disorder, partially prevents it. Biochem. Biophys. Res. Commun. *410*, 593–596.

Tanase, M., Zolla, V., Clement, C.C., Borghi, F., Urbanska, A.M., Rodriguez-Navarro, J.A., Roda, B., Zattoni, A., Reschiglian, P., Cuervo, A.M., *et al.* (2015). Hydrodynamic size-based separation and characterization of protein aggregates from total cell lysates. Nat. Protoc. *10*, 134–148.

Tang, H., Guo, Q., Zhang, C., Zhu, J., Yang, H., Zou, Y.-L., Yan, Y., Hong, D., Sou, T., and Yan, X.-M. (2010).

identification of an intermediate signature that marks the inital phases of the colorectal adenoma-carcinoma transition. Int. J. Mol. Med. *26*, 631–641.

Tantos, A., Friedrich, P., and Tompa, P. (2009). Cold stability of intrinsically disordered proteins. FEBS Lett. *583*, 465–469.

Tao, R., de Zoeten, E.F., Özkaynak, E., Chen, C., Wang, L., Porrett, P.M., Li, B., Turka, L.A., Olson, E.N., Greene, M.I., *et al.* (2007). Deacetylase inhibition promotes the generation and function of regulatory T cells. Nat. Med. *13*, 1299–1307.

Teramitsu, I., Kudo, L.C., London, S.E., Geschwind, D.H., and White, S.A. (2004). Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. J. Neurosci. *24*, 3152–3163.

Teufel, A., Wong, E.A., Mukhopadhyay, M., Malik, N., and Westphal, H. (2003). FoxP4, a novel forkhead transcription factor. Biochim. Biophys. Acta *1627*, 147–152.

Thulo, M. (2019). Structural and functional characterisation of FOXP1 and FOXP2 near-full length variants and their protein-protein interactions. University of the Witwatersrand.

Titeca, K., Lemmens, I., Tavernier, J., and Eyckerman, S. (2019). Discovering cellular protein-protein interactions: Technological strategies and opportunities. Mass Spectrom. Rev. *38*, 79–111.

Tootle, T.L., and Rebay, I. (2005). Post-translational modifications influence transcription factor activity: A view from the ETS superfamily. BioEssays *27*, 285–298.

Tsai, K.L., Huang, C.-Y., Chang, C.-H., Sun, Y.-J., Chuang, W.-J., and Hsiao, C.-D. (2006). Crystal structure of the human FOXK1a-DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins. J. Biol. Chem. *281*, 17400–17409.

Tsai, K.L., Sun, Y.-J., Huang, C.-Y., Yang, J.-Y., Hung, M.-C., and Hsiao, C.-D. (2007). Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. Nucleic Acids Res. *35*, 6984–6994.

Urreizti, R., Damanti, S., Esteve, C., Franco-Valls, H., Castilla-Vallmanya, L., Tonda, R., Cormand, B., Vilageliu, L., Opitz, J.M., Neri, G., *et al.* (2018). A de novo FOXP1 truncating mutation in a patient originally diagnosed as C syndrome. Sci. Rep. *8*, 1–6.

Uversky, V.N. (2002a). Natively unfolded proteins: A point where biology waits for physics. Protein Sci. *11*, 739–756.

Uversky, V.N. (2002b). What does it mean to be natively unfolded? Eur. J. Biochem. *269*, 2–12.

Vámosi, G., and Rueda, D. (2018). DNA bends the knee to transcription factors. Biophys. J. *114*, 2253–2254.

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., and Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language

disorder. Proc. Natl. Acad. Sci. U. S. A. *92*, 930–933.

Veprintsev, D.B., Freund, S.M.V., Andreeva, A., Rutledge, S.E., Tidow, H., Pérez Cañadillas, J.M., Blair, C.M., and Fersht, A.R. (2006). Core domain interactions in full-length p53 in solution. Proc. Natl. Acad. Sci. U. S. A. *103*, 2115–2119.

Vinkemeier, U., Moarefi, I., Darnell, J.E., and Kuriyan, J. (1998). Structure of the amino-terminal protein interaction domain of STAT-4. Science (80-.). *279*, 1048–1052.

Vitelli, L., Condorelli, G., Lulli, V., Hoang, T., Luchetti, L., Croce, C.M., and Peschle, C. (2000). A pentamer transcriptional complex including tal-1 and retinoblastoma protein downmodulates c-kit expression in normal erythroblasts. Mol. Cell. Biol. *20*, 5330–5342.

Vivian, J.T., and Callis, P.R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. Biophys. J. *80*, 2093–2109.

van der Vos, K.E., and Coffer, P.J. (2008). FOXO-binding partners: it takes two to tango. Oncogene *27*, 2289–2299.

Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K.H., Ang, S.-L., Wert, S., Stahlman, M.T., and Whitsett, J.A. (2005). Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. J. Biol. Chem. *280*, 13809–13816.

Wang, X., and Hayes, J.J. (2008). Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. Mol. Cell. Biol. *28*, 227–236.

Wang, Y., and Levy, D.E. (2006). C. elegans STAT: evolution of a regulatory switch. FASEB J. *20*, 1641–1652.

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

Wang, I.-C., Meliton, L., Ren, X., Zhang, Y., Balli, D., Snyder, J., Whitsett, J.A., Kalinichenko, V. V., and Kalin, T. V. (2009). Deletion of forkhead box M1 transcription factor from respiratory epithelial cells inhibits pulmonary tumorigenesis. PLoS One *4*, 1–13.

Wang, Y., Sarkar, M., Smith, A.E., Krois, A.S., and Pielak, G.J. (2012). Macromolecular crowding and protein stability. J. Am. Chem. Soc. *134*, 16614–16618.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. *337*, 635–645.

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. J. Biochem. *162*, 45–54.

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*, 645–658.

Weinberg, R.L., Veprintsev, D.B., and Fersht, A.R. (2004) Cooperative binding of tetrameric p53 to DNA. J. Mol. Biol. *341*, 1145-1159.

Weiss, M.A., Ellenberger, T., Wobbe, C.R., Lee, J.P., Harrison, S.C., and Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. Nature *347*, 575–578.

Whitmore, L., and Wallace, B.A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. *32*, W668–W673.

Whitmore, L., and Wallace, B.A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. Biopolymers *89*, 392–400.

Wildin, R.S., Ramsdell, F., Peake, J., Faravelli, F., Casanova, J.L., Buist, N., Levy-Lahad, E., Mazzella, M., Goulet, O., Perroni, L., *et al.* (2001). X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. Nat. Genet. *27*, 18–20.

Wildin, R.S., Smyk-Pearson, S., and Filipovich, A.H. (2002). Clinical and molecular features of the immunodysregulation, polyendocrinopathy, enteropathy, X linked (IPEX) syndrome. J. Med. Genet. *39*, 537–545.

Wilf, J., and Minton, A.P. (1981). Evidence for protein self-association induced by excluded volume Myoglobin in the presence of globular proteins. BBA - Protein Struct. *670*, 316–322.

Willcox, B.J., Donlon, T.A., He, Q., Chen, R., Grove, J.S., Yano, K., Masaki, K.H., Willcox, D.C., Rodriguez, B., and Curb, J.D. (2008). FOXO3A genotype is strongly associated with human longevity. Proc. Natl. Acad. Sci. U. S. A. *105*, 13987–13992.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., *et al.* (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315.

Wingelhofer, B., Neubauer, H.A., Valent, P., Han, X., Constantinescu, S.N., Gunning, P.T., Müller, M., and Moriggl, R. (2018). Implications of STAT3 and STAT5 signaling on gene regulation and chromatin remodeling in hematopoietic cancer. Leukemia *32*, 1713–1726.

Wingender, E., Schoeps, T., and Donitz, J. (2013). TFClass: an expandable hierarchical classification of human transcription factors. Nucleic Acids Res. *41*, D165–D170.

Wingender, E., Schoeps, T., Haubrock, M., Krull, M., and Dönitz, J. (2018). TFClass: Expanding the classification of human transcription factors to their mammalian orthologs. Nucleic Acids Res. *46*, D343–D347.

Wolfe, S.A., Nekludova, L., and Pabo, C.O. (2000). DNA recognition by Cys2His2 zinc finger proteins. Annu. Rev. Biophys. Biomol. Struct. *29*, 183–212.

Woody, R.W., and Koslowski, A. (2002). Recent developments in the electronic spectroscopy of amides and α-helical polypeptides. Biophys. Chem. *101–102*, 535–551.

Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signaling and regulation. Nat. Rev. Mol. Cell Biol. *16*, 18–29.

Wu, S., and Zhang, Y. (2007). LOMETS : A local meta-threading-server for protein structure prediction. Nucleic Acids Res. *35*, 3375–3382.

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A.D., Stroud, J.C., Bates, D.L., Guo, L., Han, A., Ziegler, S.F., *et al.* (2006). FOXP3 Controls Regulatory T Cell Function through Cooperation with NFAT. Cell *126*, 375–387.

Wu, Y., Dey, R., Han, A., Jayathilaka, N., Philips, M., Ye, J., and Chen, L. (2010). Structure of the MADS-box/MEF2 domain of MEF2A bound to DNA and its implication for myocardin recruitment. J. Mol. Biol. *397*, 520–533.

Wyrick, J.J., and Young, R.A. (2002). Deciphering gene expression regulatory networks. Curr. Opin. Genet. Dev. *12*, 130–136.

Xiang, Q., Tan, G., Jiang, X., Wu, K., Tan, W., and Tan, Y. (2017). Suppression of FOXM1 transcriptional activities via a single-stranded DNA aptamer generated by SELEX. Sci. Rep. *7*, 1–12.

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 Genet. *11*, 1–32.

Xu, D., and Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys. J. *101*, 2525–2534.

Yadav, M.K., Leman, L.J., Price, D.J., Brooks, C.L.I., Stout, D., and Ghadiri, M.R. (2006). Coiled-coils at the edge of configurational heterogeneity. Structural analyses of parallel and antiparallel homotetrameric coiled-coils reveal configurational sensitivity to a single solvent-exposed amino acid substitution. Biochemistry *45*, 4463–4473.

Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. *43*, W174–W181.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. Nat. Methods *12*, 7–8.

Yang, Y., Gao, J., Wang, J., Heffernan, R., Hanson, J., Paliwal, K., and Zhou, Y. (2018). Sixty-five years of the long march in protein secondary structure prediction: the final stretch? Brief. Bioinform. *19*, 482–494.

Yao, J., Liu, Y., Liang, H.G., Zhang, C., Zhu, J.Z., Qin, X., Sun, M., Qu, S.S., and Yu, Z.N. (2005). The effect of zinc(II) on the growth of E. coli studied by microcalorimetry. J. Therm. Anal. Calorim. *79*, 39–43.

Yoshida, Y., Wang, I.-C., Yoder, H.M., Davidson, N.O., and Costa, R.H. (2007). The forkhead box M1 transcription factor contributes to the development and growth of mouse colorectal cancer. Gastroenterology *132*, 1420–1431.

Yoshidome, T., and Kinoshita, M. (2009). Hydrophobicity at low temperatures and cold denaturation of a protein. Phys. Rev. E Stat. Nonlinear Soft Matter Phys. *79*, 30905.

Zandarashvili, L., Vuzman, D., Esadze, A., Takayama, Y., Sahu, D., Levy, Y., and Iwahara, J. (2012). Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. Proc. Natl. Acad. Sci. U. S. A. *109*, 1–9.

Zelent, B., Kusba, J., Gryczynski, I., Johnson, M.L., and Lakowicz, J.R. (1998). Time-resolved and steady-state fluorescence quenching of N-acetyl-l-tryptophanamide by acrylamide and iodide. Biophys. Chem. *73*, 53–75.

Zhang, X., and Darnell, J.E. (2001). Functional importance of Stat3 tetramerization in activation of the α2-macroglobulin gene. J. Biol. Chem. *276*, 33576–33581.

Zhang, Y., and Skolnick, J. (2004). SPICKER: A clustering approach to identify near-native protein folds. J. Comput. Chem. *25*, 865–871.

Zhang, J., Liang, Y., and Zhang, Y. (2011a). Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. Structure *19*, 1784–1795.

Zhang, X., Gan, L., Pan, H., Guo, S., He, X., Olson, S.T., Mesecar, A., Adam, S., and Unterman, T.G. (2002). Phosphorylation of serine 265 suppresses transactivation by FKHR (FOXO1) by multiple mechanisms. J. Biol. Chem. *277*, 45276–45284.

Zhang, Y., Kolinski, A., and Skolnick, J. (2003). TOUCHSTONE II: A new approach to ab initio protein structure prediction. Biophysical *85*, 1145–1164.

Zhang, Z., Witham, S., and Alexov, E. (2011b). On the role of electrostatics on protein-protein interactions. Phys. Biol. *8*, 1–18.

Zhao, Y., Zeng, C., Tarasova, N.I., Chasovskikh, S., Dritschilo, A., and Timofeeva, O.A. (2013). A new role for STAT3 as a regulator of chromatin topology. Transcription *4*, 227–231.

Zhou, H.X., and Qin, S. (2007). Interaction-site prediction for protein complexes: A critical assessment. Bioinformatics *23*, 2203–2209.

Zhou, Y., and Karplus, M. (1999). Interpreting the folding kinetics of helical proteins. Nature *401*, 400–403.

Zhou, B., Zhong, Q., Minoo, P., Li, C., Ann, D.K., Frenkel, B., Morrisey, E.E., Crandall, E.D., and Borok, Z. (2008). Foxp2 inhibits Nkx2.1-mediated transcription of SP-C via interactions with the Nkx2.1 homeodomain. Am. J. Respir. Cell Mol. Biol. *38*, 750–758.

Zhu, H. (2014). Targeting forkhead box transcription factors FOXM1 and FOXO in leukemia. Oncol. Rep.

32, 1327–1334.

Zhu, H. (2016). Forkhead box transcription factors in embryonic heart development and congenital heart disease. Life Sci. *144*, 194–201.

Zhu, C., Byers, K.J.R.P., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M. V., Radhakrishnan, M., *et al.* (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res. *19*, 556–566.

Zhu, X., Ahmad, S.M., Aboukhalil, A., Busser, B.W., Kim, Y., Tansey, T.R., Haimovich, A., Jeffries, N., Bulyk, M.L., and Michelson, A.M. (2012). Differential regulation of mesodermal gene expression by Drosophila cell type-specific Forkhead transcription factors. Development *139*, 1457–1466.

Zimmerman, S.B., and Minton, A.P. (1993). Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu. Rev. Biophys. Biomol. Struct. *22*, 27–65.

Zimmerman, S.B., and Trach, S.O. (1988). Effects of macromolecular crowding on the association of E.coli ribosomal particles. Nucleic Acids Res. *16*, 6309–6326.

Zou, C., Sun, K., Mackaluso, J.D., Seddon, A.E., Jin, R., Thomashow, M.F., and Shiu, S.-H. (2011). Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. *108*, 14992–14997.

Zuo, T., Liu, R., Zhang, H., Chang, X., Liu, Y., Wang, L., Zheng, P., and Liu, Y. (2007a). FOXP3 is a novel transcriptional repressor for the breast cancer oncogene SKP2. J. Clin. Invest. *117*, 3765–3773.

Zuo, T., Wang, L., Morrison, C., Chang, X., Zhang, H., Li, W., Liu, Y., Wang, Y., Liu, X., Chan, M.W.Y., *et al.* (2007b). FOXP3 is an X-linked breast cancer suppressor gene and an important repressor of the HER-2/ErbB2 oncogene. Cell *129*, 1275–1286.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Research Office

INSTITUTIONAL BIOSAFETY COMMITTEE
(R 14/16)

CLEARANCE CERTIFICATE **PROTOCOL NUMBER: 20180105Lab**

BRIEF DESCRIPTION OF APPLICATION:

Laboratory situated in room 411-434 URC Proten Structure-Fucntion Research Unit Gate House 4th Floor

APPLICANT: Prof Heini Dirr et al

SCHOOL/DEPARTMENT : Molecular and Cell Biology

DATE CONSIDERED: 29 January 2018

DECISION OF COMMITTEE: Approved unconditionally

These laboratories are considered to meet the standards of the Institutional Biosafety Committee (IBC), for BSL1 approval, as specified in the IBC's published Standard Operating procedures.
Any change to the type of biohazardous agents being handled should be reported to the IBC
DAFF- Reg no 39.2/University of the Witwatersrand-18/103 expiry date 31 January 2018 - 31 January 2021

1. This clearance certificate expires on 12 March 2023 and may be renewed on application.

DATE OF APPROVAL: 12 March 2018 **CHAIRPERSON:**

Professor / Name

Radiation and Health Physics Unit
University of Witwatersrand
Johannesburg, Republic of South Africa
Tel: +27 (0) 11 717 6331

DECLARATION OF APPLICANT:

To be completed in duplicate and **one copy** returned to the University of the Witwatersrand, Faculty of Health Sciences, Research Office, Office 301, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown,2193.

1. I have read, understood and accepted the approval conditions above
2. I note that the University Safety Officer, or his/her representative, may at any reasonable time inspect my laboratory or trial site to ensure compliance with current Health and Safety legislation. I undertake to offer my full co-operation in any such inspection.
3. I have read, understood and will comply with the *recommended standard operating procedures for the handling of biohazardous materials* posted at http://intranet.wits.ac.za/academic/uro/Pages/Institutional-Biosafety-Committee-(IBC).aspx

Signed: _____ **Date:** _____

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

agriculture,
forestry & fisheries

Department:
Agriculture, Forestry and Fisheries
REPUBLIC OF SOUTH AFRICA

CERTIFICATE OF REGISTRATION

Registration Number: **39.2/UNIVERSITY OF WITWATERSRAND -18/103**

It is hereby certified that the facility situated at:

UNIVERSITY OF WITWATERSRAND
SCHOOL OF MOLECULAR AND CELL BIOLOGY
EAST CAMPUS
GATE HOUSE ROOM 431
JOHANNESBURG
2050

has been registered in the name(s) of: **PROF H. W DIRR**

on behalf of: **UNIVERSITY OF WITWATERSRAND**

in terms of Regulation 8 of the Genetically Modified Organisms Act, 1997 (Act No. 15 of 1997) in respect of the following-

type of facility: **LABORATORY**

containment level: **ONE (1)**

for period: **31 JANUARY 2018 – 31 JANUARY 2021**

Date issued: **31 JANUARY 2018**

DEPT. AGRICULTURE, FORESTRY & FISHERIES
DIRECTORATE GENETIC RESOURCES
2018 -01- 31
PRIVATE BAG X973 PRETORIA 0001
REGISTRAR - GMO ACT

Registrar for Genetically Modified Organisms

Date