

https://dspace.org/

Energetics and dynamics of the FOXP2 forkhead domain-DNA interaction

Morris, Gavin

2018

https://hdl.handle.net/10539/25843

Morris, Gavin (2018)Energetics and dynamics of the FOXP2 forkhead domain-DNA interaction, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/25843>

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

Energetics and Dynamics of the FOXP2 forkhead domain-DNA interaction

Gavin Morris

A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg in fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

I, Gavin Morris (student number: 448909), am a student registered for the degree of Doctor of Philosophy at the University of the Witwatersrand.

I hereby declare the following:

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



Gavin Morris

Date: 30/01/2018

-To my parents R. and C.A. Morris and my grandmother J.M. Ingram -

"Still round the corner there may wait,
A new road or secret gate
And though I oft have passed them by,
A day will come at last when I
Shall take the hidden paths that run
West of the Moon, East of the Sun."
~The Lord of the Rings, J.R.R. Tolkien

Abstract

The members of the forkhead box (FOX) family of transcription factors are key regulators in the development and metabolism of a wide variety of tissues in humans. The FOX transcription factors are classified by the presence of a canonical forkhead winged-helix DNA binding domain and are further divided into subfamilies based on sequence divergence of the forkhead domain. To date, only the FOXP subfamily is known to require dimerisation for transcriptional activity. Dimerisation of the FOXP subfamily members occurs at two distinct interfaces, a conserved leucine zipper domain, and through domain-swapping of a C-terminal forkhead domain. The role of the domain-swapped forkhead domain is unclear, although several attempts have been made to clarify this. Due to the orientation of the recognition motifs in the domain-swapped dimer, it has been speculated that it is capable of binding and congregating two distal promoter response elements, suggesting a role in cross-chromosomal gene co-regulation. The unique capability of the FOXP forkhead domain to dimerise is attributed to an evolutionary mutation (proline to alanine) that occurs in the hinge loop connecting the second and third α-helices. Further to this, the hinge loop has also been implicated in altering the specificity of the forkhead domain. Here, we aim to elucidate how the evolutionary proline to alanine mutation facilitates dimerisation and whether it has any role in defining the DNA binding specificity of the FOXP2 forkhead domain. To do this all experiments were conducted on an obligate monomeric mutant (A539P) and an engineered obligate dimeric mutant (F541C) FOXP2 forkhead domain in addition to the wild-type. High and low-resolution DNA binding studies involving electrophoretic mobility shift assays (EMSA), fluorescence polarisation (FP) studies and isothermal titration calorimetry (ITC) revealed that the FOXP2 forkhead domain preferentially binds to the FOXP2 consensus site as a monomer, despite having the capacity to form dimers in the absence of DNA. During these studies, a significant difference in the thermodynamic signatures of DNA binding was observed between the wild-type and A539P mutant FOXP2 forkhead domain. Further dissection of the

thermodynamic results revealed that the hinge loop mutation significantly alters the mechanism of DNA binding. The wild-type FOXP2 forkhead domain undergoes significant conformational changes upon DNA binding, shown by hydrogen-deuterium exchange mass spectrometry, in addition to making two additional contacts with the sugar-phosphate backbone of the consensus site. The large conformational changes incurred by DNA binding, stabilises the monomeric form of the FOXP2 forkhead domain and is indicative of a search-recognition conformational switch that is unique to the FOXP subfamily. Furthermore, *in vivo* studies, involving dual-luciferase reporter assays, show that dimerisation of the FOXP2 forkhead domain acts as a regulatory mechanism controlling the transcriptional activity of FOXP2. Together the work presented here proposes that the DNA binding by FOXP2, and by extension FOXP1 and 4, follows a monomeric pathway whereby FOXP2 translocate to the site of action as a monomer and in a context-dependent manner either dimerises or remains monomeric to fine-tune the regulation of target genes. This work provides the first detailed assessment of the energetics and dynamics that occur during DNA binding for not only the FOXP2 forkhead domain but any of the FOX forkhead domains. Furthermore, presented here is the first proposed mechanism of transcriptional regulation through the oligomeric state of the FOXP2 forkhead domain.

Research Outputs

ORIGINAL PUBLICATIONS

Publications forming part of PhD thesis:

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., Naiker, P., Dirr H.W. and Fanucchi S. The forkhead domain hinge loop plays a pivotal role in the DNA binding and transcriptional activity of FOXP2. Biological Chemistry. (Accepted for publication, in press).

Publications not forming part of thesis:

Morris*, G., Pahad*, N., Dirr, H.W. and Fanucchi, S. Divalent cations alter the DNA binding of the FOXP2 forkhead domain. (submitted for review).

*Joint first authors.

CONFERENCE PRESENTATIONS

24TH SOUTH AFRICAN SOCIETY OF BIOCHEMSTRY AND MOLECULAR BIOCHEMISTRY

POSTER PRESENTATION

TITLE: DESIGN AND CHARACTERISATION OF A COVALENT FOXP2 FORKHEAD DOMAIN DIMER

AUTHORS: GAVIN MORRIS AND SYLVIA FANUCCHI

6TH ANNUAL POSTGRADUATE SYMPOSIUM

POSTER PRESENTATION

TITLE: DESIGN AND CHARACTERISATION OF A COVALENT FOXP2 FORKHEAD DOMAIN DIMER

AUTHORS: GAVIN MORRIS AND SYLVIA FANUCCHI

EMBO PRACTICAL COURSE: BIOMOLECULAR INTERACTION ANALYSIS: FROM MOLECULE TO CELL 2016

POSTER PRESENTATION

TITLE: HINGE LOOP DYNAMICS OF THE FOXP2 FORKHEAD DOMAIN REGULATES DNA BINDING

AUTHORS: GAVIN MORRIS AND SYLVIA FANUCCHI

Acknowledgements

To my supervisor, Dr Sylvia Fanucchi, I would like to thank you for your guidance and support throughout my postgraduate endeavours. You provided the perfect amounts of guidance and lenience that allowed me to develop as a scientist. I have thoroughly enjoyed our long-winded discussions over new and exciting ideas (and sometimes silly ones). I am grateful that I had the opportunity to learn from such an inspiring, passionate and adventurous academic.

To my colleagues in the PSFRU, past and present, it was a privilege to work alongside you. Many of you made life in the lab a lot more enjoyable, from philosophical discussions to hilarious banter. Special thanks to Alison Williams, Donald Mahlanga, Gavin Owen, Gary Robertson, Roland Worth, Kimberley Wiid and Jake Zondagh.

To the FOXy team, I will not forget the times spent together optimising this, figuring out that and the excitement of new discoveries that came with it. Special thanks to Ashleigh Blane, Monare Thulo and Naadira Pahad.

To Prof. Heini Dirr, Prof. Yasien Sayed and Dr Ikechukwu Achilonu, your dedication to the training and development of young biochemists is astounding. Thank you for your time spent helping me in the lab, inspiring fresh ideas and providing a bedrock of knowledge and experience on which to grow.

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

Finally, to my parents, twin brother and significant other, my gratitude is immeasurable. You gave me the strength to continue when I felt most bleak. You provided comfort when I was distressed and most importantly you put your foot down at the times when I needed it.

Table of Contents

Declaration i

Abstract ii

Research outputs iv

Acknowledgements vi

List of abbreviations ix

List of figures x

1. Chapter 1 – Introduction 1

1.1 Transcriptional regulation of gene expression 1

1.2 Transcription factors 2

1.2.1 DNA binding domains 2

1.3 DNA binding and sequence recognition 6

1.3.1 Base and shape readout 6

1.3.2 Non-specific binding and DNA sliding 9

1.3.3 Conformational switching 12

1.3.4 Intrinsically disordered regions and fly-casting 13

1.3.5 Cooperativity and transcriptional synergy 15

1.4 Energetics of DNA binding 16

1.4.1 Enthalpy of binding 16

1.4.2 Entropy of binding 17

1.5 FOX transcription factors 18

1.5.1 FOX domain DNA binding 20

1.5.2 Clinical relevance of the FOX family 22

1.6 FOXP subfamily 23

1.6.1 FOXP forkhead domain swapping 25

1.6.2 FOXP dimerisation regulates transcriptional activity 28

1.7 FOXP2 29

1.7.1 FOXP2 structure 30

1.7.2 DNA binding by the FOXP2 forkhead domain 31

1.8 Problem identification 33

1.9 Objective and aims 35

2. Chapter 2 - A key evolutionary mutation enhances DNA binding of the FOXP2 forkhead domain 37

3. Chapter 3 – The forkhead domain hinge loop plays a pivotal role in the DNA binding and transcriptional activity of the FOXP2 38

4. Chapter 4 - General discussion 39

4.1 FOXP2 forkhead domain dimers 39

4.2 The hinge loop alters the specificity and affinity of the FOXP2 forkhead domain 40

4.3 Conformational switching in the FOXP2 forkhead domain 43

4.3 In vivo mechanism 47

4.4 Limitations of the mutations used in the experiments 49

4.5 Conclusion 50

5. Appendix A – Methodology 52

6. References 65

List of abbreviations

bZIP	Basic-leucine zipper domain
CD	Circular dichroism
CRR	*cis*-regulatory regions
DBP	DNA binding protein
DBD	DNA binding domain
DNA	Deoxyribonucleic acid
ΔC_p	Partial heat capacity change
ΔG	Gibbs free energy
ΔH	Enthalpy change
ΔS	Entropy change
EMSA	Electrophoretic mobility shift assay
FHD	Forkhead domain
FOX	Forkhead box
FOXP2	Forkhead box P2
FP	Fluorescence polarisation
HDXMS	Hydrogen-deuterium exchange mass spectrometry
HTH	Helix-turn-helix motif
ID	Intrinsically disordered
ITC	Isothermal titration calorimetry
IPEX	Immunodysregulation Polyendocrinopathy enteropathy, x-linked

$\boldsymbol{K_a}$	Equilibrium association constant
$\boldsymbol{K_d}$	Equilibrium dissociation constant
RE	Response element
TF	Trascription factor
TRE	*trans*-regulatory elements
TSS	Transcriptional start site
wHTH	Winged helix-turn-helix motif
ZnF	Zinc-finger motif

Where applicable, the IUPAC-IUBMB one- and three-letter abbreviations for the 20 standard amino acids and 4 standard DNA nucleotides were used.

List of figures

Figure 1: Cartoon representations of DNA binding domains. 5

Figure 2: Cartoon respresentation of base and shape readout. 8

Figure 3: DNA translocation by TFs. 11

Figure 4: Multiple sequence alignment of the forkhead domain for several FOX family members. 19

Figure 5: The FOX transcription factors bind a highly conserved promoter sequence. 21

Figure 6: Domain topology schematic of the FOXP subfamily members. 24

Figure 7: The FOXP forkhead domain exhibits the propensity to form domain-swapped dimers. 27

Figure 8: Comparison of DNA contacts made by the FOXP2 forkhead domain monomer and the domain-swapped dimer. 32

Figure 9: Cartoon representation of the FOXP2 forkhead domain divided into the two subdomains. 44

Figure 10: Switching between search and recognition modes in the forkhead domain may be controlled by the composition and inherent flexibility of the hinge loop region. 46

Figure 11: Proposed mechanism for differential gene expression controlled by the oligomeric state of the FOXP2 forkhead domain. 49

Chapter 1

Introduction

1.1 Transcriptional regulation of gene expression

Cellular response to external and internal stimuli relies upon tightly regulated expression of appropriate gene products. Transcription is the first step in gene expression. A precise combination of specific *cis*-regulatory regions (CRR) and *trans*-regulatory elements (TRE) results in temporally defined transcription of specific genes during fundamental life processes. Transcriptional regulation plays a significant role in numerous cellular pathways including cell cycle progression, differentiation and development (Levine, 2010). The *trans*-regulatory elements (TREs) are DNA regions that control the expression of distal genes by encoding a subset of DNA binding proteins (DBP) known as transcription factors. Transcription factors (TFs) recognise and interact with specific short DNA sequences (response elements) located within CRRs that are responsible for the transcriptional regulation of an adjacent gene. CRRs can host multiple different response elements to which a variety of specific TFs can bind (Levine, 2010). Transcription factors serve to facilitate or hinder the recruitment of the pre-initiation transcriptional complex components at the transcription start site (TSS) of a gene. The pre-initiation transcriptional complex comprises of a large multi-component complex composed of several general and specific transcription factors, cofactors and RNA polymerase II, which together forms the biological machinery responsible for transcribing a gene into pre-mRNA (Luse, 2013). In addition to TFs, several other factors play a role in assembly of the pre-initiation transcriptional complex (the biological machinery responsible for transcribing a gene into pre-mRNA): transcription factor post-translational modification, epigenetic modification of CRRs and chromatin structure (Pan *et al*.,2010; Todeschini *et al*., 2014). The mechanism by which the transcription of a single gene is regulated is context-dependent and can require the

arrangement of a multi-component transcription factor complex with several different post-translational modification states and the presence of several co-factors (Todeschini *et al*., 2014). At the heart of these gene regulatory networks lies the fundamental ability of transcription factors to rapidly locate and form a stable complex with the appropriate response element.

1.2 Transcription factors

1.2.1 DNA binding domains

Transcription factors are multidomain proteins that interact with a specific DNA sequence through a DNA-binding domain (DBD), a specialised structural element capable of interacting with the nucleotide bases and sugar-phosphate backbone of the DNA. A wide variety of TFs share common DBD motifs and can be classified into categories and further subdivided into TF families based on similarities in their DBDs (Luscombe *et al*., 2000). At least 1390 TF encoding genes (approximately 6% of all genes) have been identified in the *Homo sapiens* genome (Vaquerizas *et al*., 2009). However, a large percentage (approximately 47%) of these TF encoding genes remain unannotated and further experimental work is required before they can be classified (Vaquerizas *et al*., 2009). There are four predominant DBD groups; (i) zinc-coordinating (such as the zinc fingers), (ii) zipper-type (such as the basic-leucine zippers DBD), (iii) helix-loop-helix and (iv) helix-turn-helix (Luscombe *et al*., 2000; Vaquerizas *et al*., 2009; Gray *et al*., 2004).

The zinc finger motif is the most common DBD amongst eukaryotic transcription factor DBDs (Vaquerizas *et al*., 2009; Gray *et al*., 2004). The C_2H_2 zinc finger motif, the most common form of the zinc finger domain, is characterised by a 30-amino acid stretch of 9 repeated amino acids containing two invariant cysteine and two invariant histidine residues responsible for coordinating a Zinc ion cofactor (Figure 1; Miller *et al*., 1985). The zinc-coordination complex maintains a scaffold that allows the exposure of amino

acids sidechains along the single alpha helix in the domain to the nucleotide bases in the DNA major groove facilitating the necessary contacts required for specific sequence recognition (Figure 1; Klug *et al*., 2010). Unlike other DBDs which typically lie perpendicular to the fibre axis of DNA, the zinc finger motif inserts into the DNA double helix following the helical trajectory of the major groove, allowing several zinc fingers to be strung together in order to recognise DNA sequences of variable length (Klug, 2010).

The leucine zipper domain requires dimerisation for regulatory activity. A leucine zipper subunit consists of a single extended alpha helix composed of 60-80 amino acids that associate with a second leucine zipper subunit to form a dimeric structure capable of sequence specific DNA binding (Figure 1; Alber, 1992). The leucine zipper dimers are held together by regularly interspersed leucine residues in the N-terminal of the DBD and DNA sequence recognition occurs through the basic C-terminal region of the helix. Because of this, leucine zipper containing TFs can form homo- and hetero-typic dimers providing an additional layer of regulation through subunit combination dependent sequence recognition (Alber, 1992).

The helix-loop-helix DBD resembles the basic leucine zipper DBD. The helix-loop-helix (HLH) DBD also consists of an extended alpha helical structure except for the insertion of a loop in the middle of the helix (Massari and Murre, 2000). The C-terminal helix of the HLH domain consists of predominantly basic amino acids and is inserted into the major groove of the DNA. The loop stabilises the protein-DNA complex through additional contacts with the DNA backbone and major groove nucleotides (Massari and Murre, 2000). Like the leucine zipper DBD, the helix-loop-helix DBD form homo- and heterotypic dimers through an N-terminal leucine zipper facilitating combinatorial fine tuning of the recognised binding site (Massari and Murre, 2000).

The helix-turn-helix (HTH) group of TFs, comprising of 16 unique families, exhibit the greatest structural diversity of the major DBD groups (Luscombe *et al*., 2000). The canonical HTH motif comprises of a three α-helical bundle interconnected by short random coils (Figure 1; for a comprehensive review see: Aravind *et al*., 2005). The arrangement of the helical elements forms a stable hydrophobic core from which the DNA recognition helix, typically the third α-helix, is presented to the DNA major groove for sequence recognition. Several variations of the HTH motif exist including the winged HTH, the ribbon-helix-helix and the tetra-helical bundle (Aravind *et al*., 2005). The winged HTH (wHTH) DBD is a prominent extension of the HTH motif whereby the loops connecting the helices are extended to form large random coils or 2-stranded β-sheet wings (Figure 1; Aravind *et al*., 2005). The wing extensions make additional contacts with the DNA in the form of DNA backbone contacts or insertion of side chains within the minor groove adjacent to the binding site of the recognition helix providing additional selectivity and stability (Brennan, 1993). One of the largest wHTH TF families is the forkhead box (FOX) superfamily estimated to be the sixth largest group of annotated transcription factors in humans (Vaquerizas *et al*., 2009).

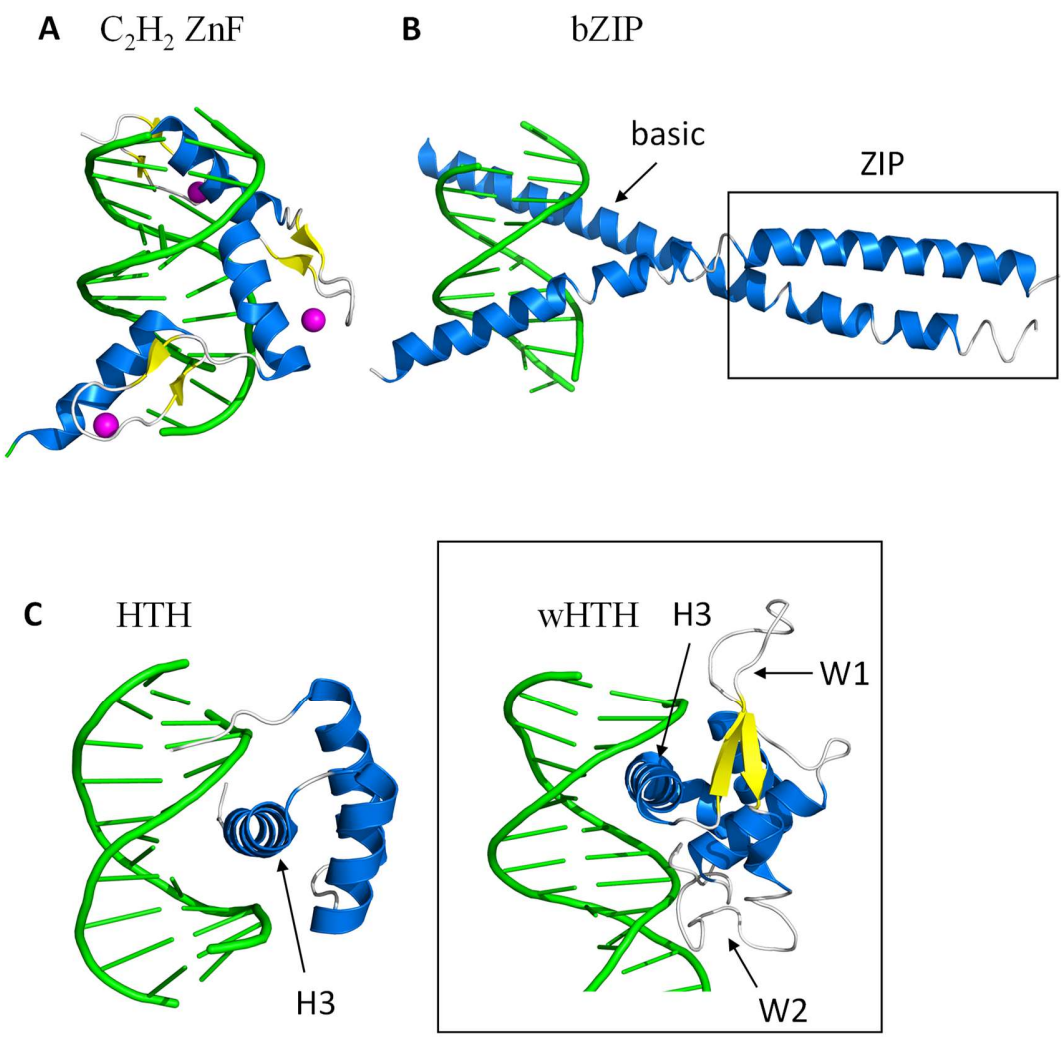


Figure 1: Cartoon representations of DNA binding domains. Cartoon representations of the three most common eukaryotic DNA binding motifs: (A) Zinc finger (ZnF, the zinc atoms are highlighted in purple), (B) leucine zipper (bZIP; comprised of a basic domain and leucine zipper, ZIP), (C) helix-turn-helix (HTH) and the winged helix variant of the HTH motif (wHTH; inset). The recognition helix (H3) of the HTH and wHTH DBDs are inserted into the DNA major groove. The two wings (W1 and W2), after which the wHTH is named, make additional DNA backbone contacts with the DNA. Models were rendered in Pymol™ v 0.99 (DeLano Scientific, 2006) using the PDB codes: 4X9J (Egr-1 zinc finger), 1T2K (ATF2-Jun leucine zipper dimer), 1E3O (Oct-1 helix-turn-helix) and 1VTN (FOXA1 winged helix-turn-helix), respectively.

1.3 DNA binding and sequence recognition

The control of gene expression is a complex process regulated by the speed of response element location by TFs, TF specificity and the energetics that govern it, and nonlinear interactions including transcription factor oligomerisation and cooperativity (Smolen *et al*, 2000).

1.3.1 Base and shape readout

Despite the large structural diversity of TF DBDs all are capable of DNA sequence discrimination. All TFs inherently have an affinity for associating with DNA; however, the interaction requirements for formation of a stable protein-DNA complex at a specific DNA sequence (specificity) varies substantially both between and within TF families (Svingen and Tonissen, 2006; Nakagawa *et al*., 2013). Initially, specificity was thought to derive exclusively from the formation of distinct hydrogen bonding patterns between residue side chains in the DBD and the Watson-Crick base pairs in the major/minor groove, termed direct readout (Seeman *et al*., 1976; Viswamitra *et al*., 1982). Although the large majority of protein-DNA crystal structures display some form of direct readout, it has become apparent that direct readout is insufficient in fully explaining DBD specificity (Otwinowski *et al.*, 1988; Rohs *et al*., 2010). DNA binding proteins must navigate a complex electrostatic field governed by cation concentration gradients and sequence-dependent DNA morphology while reading the DNA sequence to find a target consensus site (Srinivasan *et al*., 2009; Rohs *et al*., 2009). A full description of specificity requires a complex electrostatic complementarity with the DNA sugar-phosphate backbone and induced conformational changes in addition to direct readout of the Watson-Crick base pairs to facilitate a tight shape complementarity between the DNA and DBD (Shakked *et al.*, 1989, Rohs *et al*., 2009; Pabo and Sauer, 1984).

Broadly, specificity involves two distinct sets of interactions. Firstly, hydrogen bonding and van der Waals contacts occur between proteins residue side chains and the nitrogenous base pairs (base readout) and secondly, tight shape complementarity exists between the concave DNA groove of the binding site and the convex DBD (shape readout) (Figure 2). Both sets of interactions are necessary for complete specificity of a protein-DNA interaction (Harrington, 1992; Noy *et al*., 2016). Base readout relies predominantly on the formation of sequence specific hydrogen bonds between the donor and acceptor groups on the nitrogenous bases in the major or minor groove of Watson-Crick base pairs, the residue sidechains and backbone amide groups of the recognition element of the DNA binding protein. Transcription factors read the hydrogen bonding patterns with residue sidechains and even with amide backbone moieties (Rohs *et al*., 2010). Specificity is fine-tuned by the directionality of the hydrogen bond and the formation of bidentate hydrogen bonds between a single residue side chain and one or two base pair moieties (Coulocheri *et al*., 2007).

Shape readout relies more on the culmination of a large number of van der Waals contacts with moieties within the grooves of the DNA duplex as well as the presentation of charged groups along the DNA backbone with counter charged or polar groups in neighbouring regions of the DBD. The hydrogen bonds, salt bridges and charge-charge interactions formed between the phosphate-sugar backbone and the basic amino acid residues is an essential component of protein-DNA interactions (Luscombe *et al*., 2001; Privalov *et al*., 2011; Strauch, 2001). Once a binding site has been located the strong electrostatic interactions act to stabilise the complex, emphasised by the strong dependence of DNA binding affinity on the ionic strength of the solution (Strauch, 2001; Privalov *et al*., 2011).

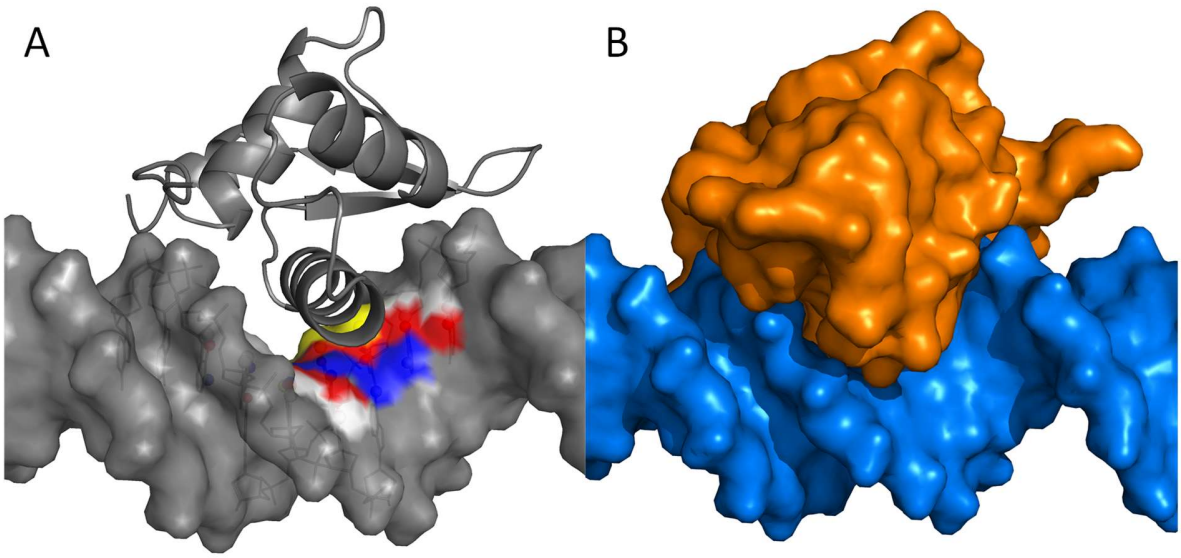


Figure 2: Cartoon respresentation of base and shape readout. (A) Base readout involves hydrogen bonds (blue) and van der Waals (red and yellow) contacts between the recognition helix and the nitrogenous bases of the recognition sequence. (B) The entire DBD (orange) is required for complete shape readout of the consensus site. The entire major groove of the binding site is encompassed by the DBD. Base contacts were determined for the FOXP2 forkhead domain bound to a consensus DNA site (PDB codes: 2A07) using the PyMol$^{(c)}$ plugin PDIViz 1.2.2 (Ribeiro *et al*., 2015) and rendered in PyMol^{TM} v0.99 (DeLano Scientific, 2006).

The entire DBD motif plays a role in accurate DNA sequence discrimination (Strauch, 2001). Regions of the protein surrounding the recognition element play two fundamental roles. Firstly, they provide the structural architecture that facilitates the correct orientation of the recognition element within the major or minor groove of the DNA promoting specific base-residue hydrogen bonds and contacts (Strauch, 2001). Secondly, the surrounding regions are rich in positively charged or polar residues that result in strong electrostatic interactions with the counter charge of the DNA backbone (Rohs *et al*., 2009). This requirement is readily observed in the studies of the forkhead domain whereby alterations in regions adjacent to the recognition helix (architectural

changes) or removal of a small C-terminal loop (sugar-phosphate backbone contacts) significantly reduces the DNA binding affinity of the protein (Pierrou *et al*., 1994; Overdier *et al*., 1994; Murphy *et al*., 2004; Boura *et al*., 2007).

1.3.2 Non-specific binding and DNA sliding

Considering the size of the human genome (~6 billion base pairs) and the size of a typical enhancer or promoter element (7-10 base pairs) it is very likely that the concentration of similar but non-specific sites of the same length far exceeds the concentration of specific sites for a given TF (Badis *et al*., 2009; Wunderlich and Mirny, 2009). The high number of non-specific sites combined with the low number of the TF present in the nucleus has prompted the evolution of several unique mechanisms employed by eukaryotic TFs and other eukaryotic DNA binding proteins (DBPs) to facilitate rapid site location and discrimination (Simicevic *et al*., 2013; Slutsky and Mirny, 2004; Kolomeisky, 2010).

DNA binding proteins can find target sequences with remarkable speed and their association rates are up to 2 orders of magnitude higher than is explained by theoretical 3D diffusion limits (Givaty and Levy, 2009). The association rates of several DBPs, for example the Lac repressor, have been determined *in vitro* and fall in the range of 1 x $10^{8\text{-}10}$ $M^{-1}s^{-1}$ (Riggs *et al*., 1970). To accomplish these high speeds, DBPs possess the unique ability to non-specifically bind to DNA and slide along the phosphate-sugar backbone in a facilitated 1D diffusion interspersed with small correlated hops and intersegmental transfers (Figure 3) (Von Hippel and Berg, 1989). TFs must find a balance between making the necessary contacts to remain associated with the DNA in order to undergo facilitated diffusion but not too many as to become trapped at non-specific or similar but invalid response elements, a problem known as the speed-stability paradox (Slutsky and Mirny, 2004).

The combination of sliding and hopping allows the DNA binding protein to locate and distinguish a specific binding site on a timescale of 1-10s within the cell (Givaty and Levy, 2009). The sliding, a stochastic process driven by thermal diffusion, has been observed, using NMR paramagnetic relaxation enhancement studies as well as by single molecule fluorescence microscopic imaging (Iwahara and Clore, 2006; Takayama and Clore, 2006; Blainey *et al*., 2006; Wang *et al*., 2006; Gormoan and Greene, 2008). These studies also reveal that DBPs slide along DNA following the helical trajectory of the DNA duplex while maintaining close proximity between the DNA groove (major and/or minor) and the secondary structural element responsible for specific sequence recognition (Iwahara *et al.,* 2006; Iwahara and Clore, 2006). The sliding mechanism and non-specific interaction between the DBD and DNA is dominated by non-specific electrostatic charge interactions rather than base and phosphate specific hydrogen bonds, allowing for sequence agnostic mobility of the protein during translocation (Givaty and Levy, 2009; von Hippel and Berg, 1989). These electrostatic interactions are strongly dependent on the ionic strength of the solution and theoretical studies have shown that increasing the concentration of counter ions in solution increases the propensity of the DBP to hop along the DNA backbone (Givaty and Levy, 2009). Moderate ionic strength (below 100 mM) was found to promote an optimal ratio of sliding to hopping (20% and 80%, respectively) to minimise the search time of the SAP-1 ETS DBD (Givaty and Levy, 2009). During sliding, the recognition element of the DBP scans the nitrogenous bases within the major and/or minor groove of the DNA duplex (Iwahara and Clore, 2006).

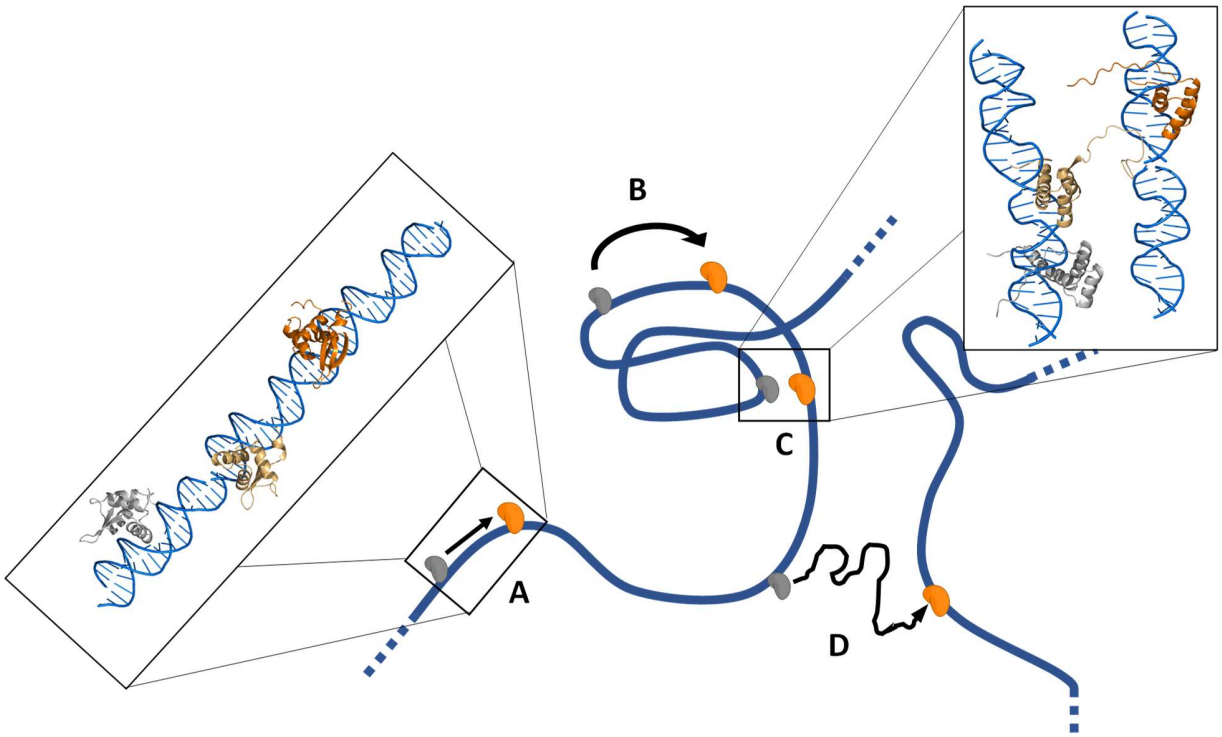


Figure 3: DNA translocation by TFs. (A) DNA binding proteins can non-specifically bind to DNA and slide along the DNA backbone following the helical trajectory of the duplex. This process is Brownian in nature and is driven by thermal diffusion (Iwahara *et al*., 2006). The protein remains associated with the DNA through moderate electrostatic interactions while maintaining close contact between the recognition motif and the bases of the appropriate DNA groove. (B) During sliding DNA, binding proteins can intermittently hop over short segments of the DNA allowing the protein to sample, at random, genomic DNA faster than simply by sliding. (C) Intrinsically disordered regions inherent in many eukaryotic transcription factors facilitate a bridged ("Monkey bar") transfer between adjacent DNA duplexes (Vuzman *et al*., 2012). (D) DNA binding proteins can detach from DNA completely and transfer to a distal DNA strand through 3D diffusion.

1.3.3 Conformational switching

One model on how DBPs can rapidly discriminate DNA sequences while translocating along DNA, termed conformational switching, proposes that DBDs can switch between a search and a complex-stable recognition conformation (Slutsky and Mirny, 2004; Zhou, 2011). In this model, the search conformation maximises DNA translocation through electrostatic backbone contacts but is unsuitable for the formation of a stable complex through base contacts due to insufficient hydrogen bonds.The recognition conformation acts to promote an active stable protein-DNA complex through base specific hydrogen bonds and tight shape complementarity with the bound groove.

Theoretical and NMR studies confirm that conformational switching between search and stable conformations occurs in the Egr-1 zinc finger protein, HOXD9 HTH DBD and acts to minimise both false positives and search times for the DBD (Iwahara and Clore, 2006; Iwahara and Levy, 2013; Zandarashvili *et al*., 2015). The DNA binding affinity of this DBD was shown to be distinct for either the non-specific search mode (K_a, the association constant, in the range of 1 x 10^6 M/s) or the specific recognition mode of DNA binding (K_a, the association constant, in the range of 1 x 10^9 M/s) (Zandarashvili *et al*., 2015). Characterisation of the search and stable conformations of DBDs using NMR and computational models revealed that the predominant factor in determining the distribution of the protein population between the search and recognition modes is highly dependent on the protein-DNA interactions as well as the intra/interdomain interactions of the DBD (Zandarashvili *et al*., 2015; Iwahara and Levy, 2013). For conformation switching to occur, the necessary contacts must be made to facilitate an optimal trajectory down the proteins folding funnel, overcoming the transition state energies (Shoemaker *et al*., 2000). By altering the inter and intramolecular interactions in Egr-1 with mutagenesis, Zandarashvili *et al.*, were able to shift the proportion of Egr-1 from predominantly residing in the search mode to predominantly residing in the recognition mode, reducing the translocation rate of the

DBD and increasing the specificity. This provides a model for evolutionary mutations in the DBD of similar TF family members that alter sequence specificity and affinity.

1.3.4 Intrinsically disordered regions and fly-casting

DNA binding domains exhibit a conformational flexibility that allows rapid sampling of the DNA sequence as the protein translocates along the DNA backbone. To ensure accurate discrimination of binding sites in CRRs, eukaryotic transcription factors employ a plethora of energetic checkpoints throughout the course of association and translocation. DNA binding proteins can contain several intrinsically disordered (ID) regions that act as energetic brakes only refolding when contacts are made with a highly specific sequence of DNA bases (Levy *et al*., 2007; Dragan *et al*., 2004; Spolar and Record, 1994; Guo *et al*., 2012). In TFs these ID regions are predominantly linker regions and tails at physiological temperature that provide additional contacts to fine tune the specificity of the protein for the appropriate RE (Vuzman *et al*, 2012; Guo *et al*., 2012; Wright and Dyson, 1999). The inherent flexibility of unstructured regions is influenced by the local environment including pH and ionic strength; any posttranslational modifications the protein may have; and interactions with other proteins and macromolecules (Wright and Dyson, 1999). Due to induced folding of unstructured motifs upon DNA binding and a reduction in the motional freedom of both protein and DNA, the spontaneous process of DNA binding by eukaryotic transcription factors is often accompanied by a significant loss of systemic entropy (Wright and Dyson, 1999). Such a large loss in entropy is compensated for by the formation of novel hydrogen bonds (enthalpic compensation) as well as exclusion of counter-ions and water molecules from the nascent complex interface (non-systemic entropic compensation). The entropic cost of induced folding results in minimal free energy of binding for most eukaryotic transcription factors, such that only the most favourable interactions proceed spontaneously, conserving some sequence specificity (Wright and Dyson, 1999).

ID regions also play a critical role in DBD translocation along DNA. They can act as an electrostatic antenna for the DNA, pulling the DBD toward the DNA in a co-dependent binding-refolding process (Rentzeperis *et* al., 1999; Shoemaker *et al*., 2000). This mechanism, known as fly-casting, is a phenomenon widely observed in eukaryotic transcription factors and acts as an important control mechanism to regulate activity. By remaining partially unfolded, DNA binding proteins increase the probability of making minimal contact with a proximal DNA duplex, increasing the rate of site binding up to 1.6 times according to theoretical models (Shoemaker *et al*., 2000). Fly-casting acts to kinetically regulate gene expression; the difference in binding rates controlled by refolding rates of different DBDs plays a pivotal role in ensuring transcription factors do not dwell for extensive periods at non-specific sites or in non-productive complexes (Kohler *et al*., 1999; Rentzeperis *et al*.,1999; Markovitz and Levy, 2009). The activity of several TFs (perhaps the most well-known of these is the bZIP family of leucine zipper TFs) is controlled by the formation of homo- and heterotypic dimers. Interestingly, such TFs have been shown to prefer binding to DNA non-specifically as monomers and form an active dimeric complex at a target binding site via the fly-casting mechanism (Kohler *et al*., 1999; Rentzeperis *et al*., 1999).

To summarise, fly-casting and conformation switching phenomena act synergistically to enhance site location, discrimination and activity. Firstly, partial folding reduces the time required for a nascently translated TF to become an active component of the gene regulatory network. It does this by providing stage-wise assisted folding by dividing the folding free energy over several smaller transition states (at DNA association and then again at recognition site) instead of fully folding in a non-productive environment (for example, the cytosol); secondly, by remaining partially unfolded in solution increases the DNA detection radius of the TF, improving the probability of finding DNA to which it can non-specifically bind and begin translocation. Once a partially folded DBP non-specifically binds to DNA the rate of translocation is much faster for the smaller monomeric component. Finally, incomplete folding of the DBD prevents non-productive protein-protein interactions at non-specific sites or in the cytosol; because

of this the search space for dimerisation partners is reduced to the pseudo-one dimensional domain of the DNA duplex. (Levy *et al*., 2007; Shoemaker *et al*., 2000; Marcovitz and Levy, 2009).

1.3.5 Cooperativity and transcriptional synergy

Due to the high probability of finding non-functional response element sequences outside of *cis*-regulatory regions, many TFs cooperate to ensure functional CRR placement. Dimerisation of eukaryotic transcription factors is one method employed to synergistically improve both specificity and functional site identification by effectively increasing the length of the functional RE (Amoutzias *et al*., 2008). A number of well-known transcription factor families function, and are regulated, through the formation of homo- and heterotypic dimers, for example, the bZIP leucine zipper family whereby the formation of Jun-ATF 2 heterodimers favours activation of genes controlled by the ENK-2 response element as opposed to the AP-1 sites typically bound by Jun homodimers (Hai and Curran, 1991).

TFs can also cooperate to improve affinity for low affinity sites for other TFs that would not otherwise reside at the RE alone. Cooperativity, the act of a bound TF influencing the affinity of another TF for an adjacent RE, can occur between TF of the same species or between those from completely different families (Todeschini *et al*., 2014). Cooperativity provides multiple benefits in reducing gene expression noise in the cell. Firstly, cooperative binding between the same species of TFs provides a buffer against random fluctuations of TF concentration, allowing even poorly expressed TFs to exert the necessary gene regulation (Todeschini *et al*., 2014). Cooperative interactions also act as additional level of control in fine tuning gene expression through combination-dependent activity of TFs. Interactions between TFs of different species can change the degree to which a gene is expressed by improving or reducing the affinity for the transcriptional machinery or other general transcription factors (Todeschini *et al*., 2014). Transcriptional synergy can be homo- (cooperation between the same species of

TF) and heterotypic (between different TFs) depending on the cellular context. Transcriptional synergy is an important requirement for accurate response to internal and external stimuli as it provides additional requirements for the activation of a particular gene, critical when one considers the degree of signal noise that occurs in the cellular environment (Todeschini *et al*., 2014). Naturally transcriptional synergy is not the last level of control in gene regulation: post-translational modifications, DNA accessibility and DNA modification are fundamental components in the finely tuned gene expression system.

1.4 Energetics of DNA binding

Studying protein-DNA interactions requires a combination of thermodynamic and structural data (Rohs *et al.,* 2009). DNA binding proteins display a wide array of thermodynamic strategies in site identification and binding (Jen-Jacobson *et al*., 2000a). The thermodynamic signature of macromolecular associations provides extensive insight into the nature of the interactions that may not be immediately evident in static structural data (for example, data obtained from crystal structures). Furthermore, dissection of the enthalpic and entropic terms into contributing components is required for a complete insight into the physical forces that govern the interaction.

1.4.1 Enthalpy of binding

The enthalpic component of DNA binding by proteins originates from changes in intra and intermolecular interactions formed during the process of DNA binding (Jen-Jacobson *et al*., 2000a). Opposing the enthalpic gains from the protein-DNA contacts are the enthalpic requirements for desolvation of the DNA grooves and sugar-phosphate backbone (Privalov *et al*.,2007; Jen-Jacobson *et al*., 2000a). However, to date no direct correlation between ΔH and the degree of desolvation or the extent of surface burial has been determined that encompasses all protein-DNA systems (Privalov *et al*., 2007; Jen-Jacobson *et al*., 2000). Additionally, steric strain induced by unfavourable adopted

positions of side chains in both the DBD and DNA during formation of the protein-DNA interface can result in a significant enthalpic cost (Privalov *et al*.,2007; Jen-Jacobson *et al*., 2000a). The temperature dependence of the enthalpic component of binding is another useful parameter (ΔC_p) when studying protein-DNA interactions. This can provide insight into the types of interactions that occur in the protein-DNA interface, the degree of conformational adjustments undergone by the protein and DNA during complex formation and the desolvation area of the interface (Privalov *et al*., 2007; Prabhu and Sharp, 2005; Dragan *et al*., 2004; Ha *et al*., 1989; Spolar and Record; 1994; Jen-Jacobson *et al*., 2000b).

1.4.2 Entropy of binding

DNA is a polyanionic copolymer composed of two associated strands of nucleotides connected by phosphodiester bonds. The high charge density along the DNA phosphate-sugar backbone results in a cation gradient approximated by a non-linear Poisson-Boltzmann distribution (Srinivasan *et al*., 2009). The entropic component of binding comprises of cratic entropy of mixing associated with the exclusion of counter-ions and water from the protein-DNA interface together with conformational changes in both the protein and DNA binding partners to accommodate shape complementarity (deHaseth *et al*., 1977; Manning *et al*., 1978; Record *et al*., 1978). Cation absorption to the phosphate backbone of DNA shields the negative charges from the basic amino acid rich protein motifs. Exclusion of the cations from the DNA backbone requires energetic input to be excluded from the interface so that the necessary charge-charge interactions can occur. Therefore, the affinity of any protein-DNA interaction is inherently tightly coupled to the ionic strength of the solution (deHaseth *et al*., 1977; Manning *et al*., 1978; Record *et al*., 1978). This phenomenon can be exploited by performing binding studies at increasing ionic strength which can be used to dissect the entropic term of the interaction into salt-dependent (electrostatic) and salt-independent (non-electrostatic) components (Privalov *et al*., 2011). The electrostatic component reflects polar and coulombic interactions with the phosphate backbone of the DNA and is minimally

affected by the DNA sequence (Dragan *et al*., 2004). Conversely, the non-electrostatic component is determined by the base specific interactions (van der Waals contacts, hydrogen bonding etc.) and therefore can be used as an indicator of the specificity of the protein binding (Privalov *et al*., 2011; Dragan *et al*., 2004).

A full analysis of the enthalpic term (particularily the partial heat capacity) and the entropic term (salt-dependent studies) can provide detailed insight into the molecular driving forces involved in DNA binding of a protein of interest. In addition, by coupling the thermodynamic studies with mutagenesis, structural and dynamics (for example NMR and HDXMS) characterisation studies it is possible to elucidate a detailed image of the protein-DNA system of interest.

1.5 FOX transcription factors

The forkhead box (FOX) proteins are a specific TF family classified by the presence of a canonical forkhead DNA binding domain (Figure 4). The FOX domain is a subgroup of the winged-HTH (wHTH) supergroup owing to two extended C-terminal loops that form a two-stranded anti-parallel β-sheet (known as the first wing) capped by a fourth α-helix or an extended loop known as the second wing (SCOP classification number: 46832; Clark *et al*., 1993). The forkhead domain was first identified in the *Drosophila melanogaster Forkhead* gene mutants and in the mammalian HNF-3α TF (FOXA1) (Weigel and Jackle, 1990). Since the discovery of the FOX domain, FOX TFs have been identified in three eukaryotic kingdoms not including plants (animals, fungi and opisthokonts) stressing their importance in animal evolution (Shimeld *et al.,* 2010; Kaestner *et al*., 2000). Attempts at identifying the evolutionary origin have dated the first appearance of the FOX domain to be the common ancestor of bilaterians, cnidarians and placozoans (Shimeld *et al.,* 2010). To date, the FOX family has been extended to encompass over one hundred transcription factors, in Chordata, classified into nineteen subfamilies (FOXA-S) based on the divergence of the FOX domain sequence (Kaufmann and Knochel, 1996; Hannenhalli and Kaestner, 2009).

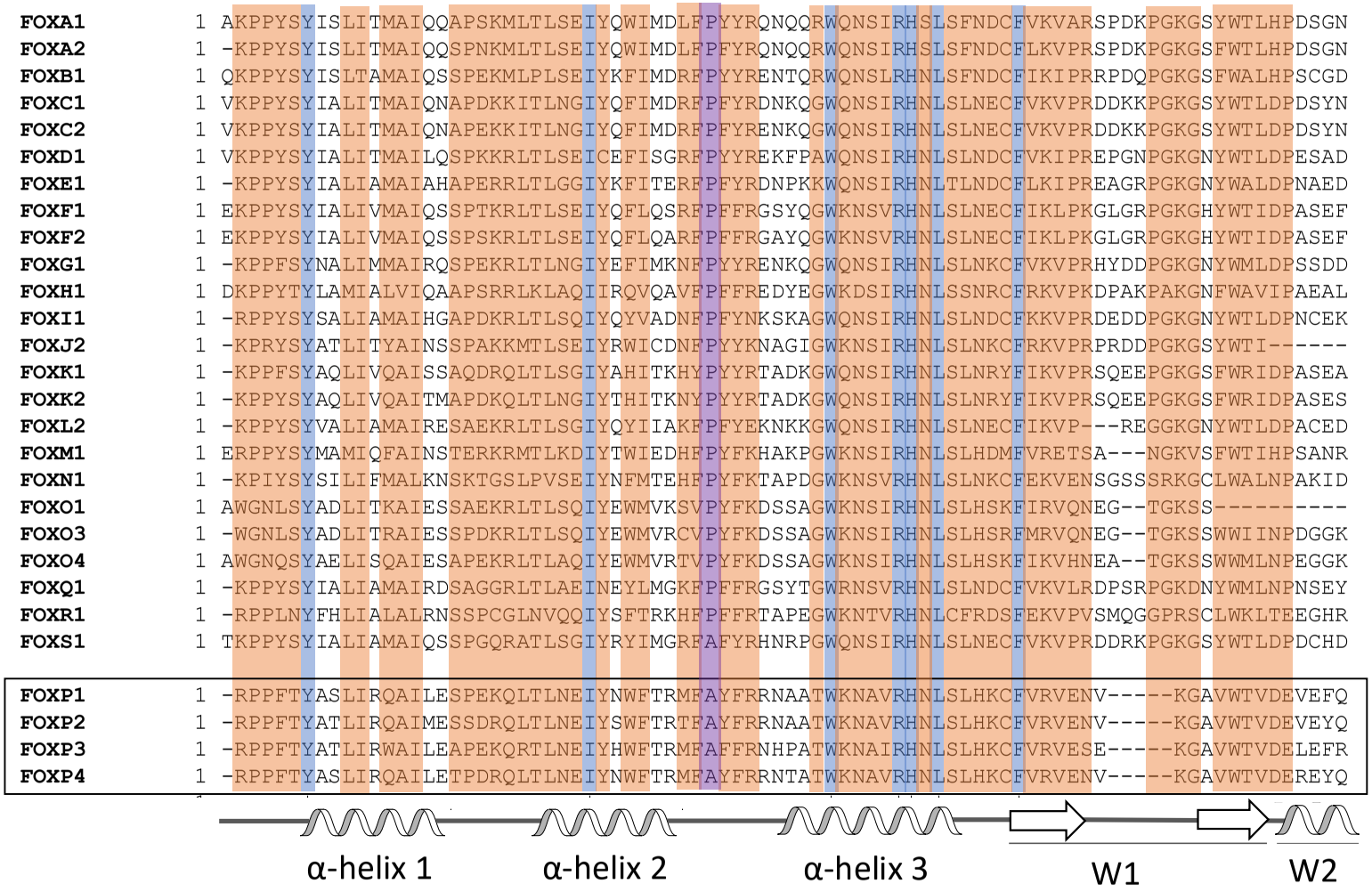


Figure 4: Multiple sequence alignment of the forkhead domain for several FOX family members. Approxiamtely 60 % of the amino acids are shared or replaced by a similarly charged or polar residue (orange) and almost 7 % are completely conserved (blue). A secondary structure topology map is shown below the alignment. The hinge loop appears between α-helix 2 and α-helix 3 (the recognition helix of the FOX proteins). Significant variation in the sequence is observed in the wing regions (W1 and W2) which form a two stranded β-sheet that primarily interacts with the DNA backbone. The FOXP subfamily members are the only known FOX proteins capable of forming domain-swapped homo- and heterotypic dimers purported to be due to the mutation of the highly conserved proline located in the middle of the hinge loop (purple).

1.5.1 FOX domain DNA binding

The FOX domain displays a surprisingly high degree of conservation across the entire cast of family members, particularly in the secondary structural element responsible for sequence recognition, the third α-helix (Figure 5). It is likely due to this that many of the FOX TFs recognise and act on a highly conserved core consensus sequence, 5'-RYAAAYA-3' (Pierrou *et al*., 1994; Overdier *et al*., 1994; Kaufmann and Knochel, 1995; Hannenhalli and Kaestner, 2009). However, many FOX subfamilies are also capable of binding to secondary sequences different from the FOX consensus sequence, for example 5'-AHAACA-3' (FOXA1), 5'-ACGC-3'(FoxN1 and FoxN4) and 5'-CCCGATAG-3' (FOXP2) (Overdier *et al*., 1994; Shlake *et al*., 1997; Luo *et al*., 2012; Webb *et al*., 2017).

The specificity of the forkhead domain is altered directly by primary sequence differences in regions that form a part of the protein-DNA interface as well as regions that do not interactwith the DNA. From crystallographic structure analysis it can be seen that the protein-DNA interactions with the consensus DNA sequence differ primarily between the residues of the N-terminal of helix 1, the C-terminal of helix 2 and the two wings due to the sequence variation in these regions among the FOX family members (Figure 4 and Figure 5). The wings of the forkhead domain, displaying the greatest sequence variation across the FOX family, are known to alter both DNA binding affinity and specificity by making numerous contacts with the DNA sugar-phosphate backbone and minor groove contacts (Obsil and Obsilova, 2008; Cirillo and Zaret, 2007; Murphy *et al*., 2004).

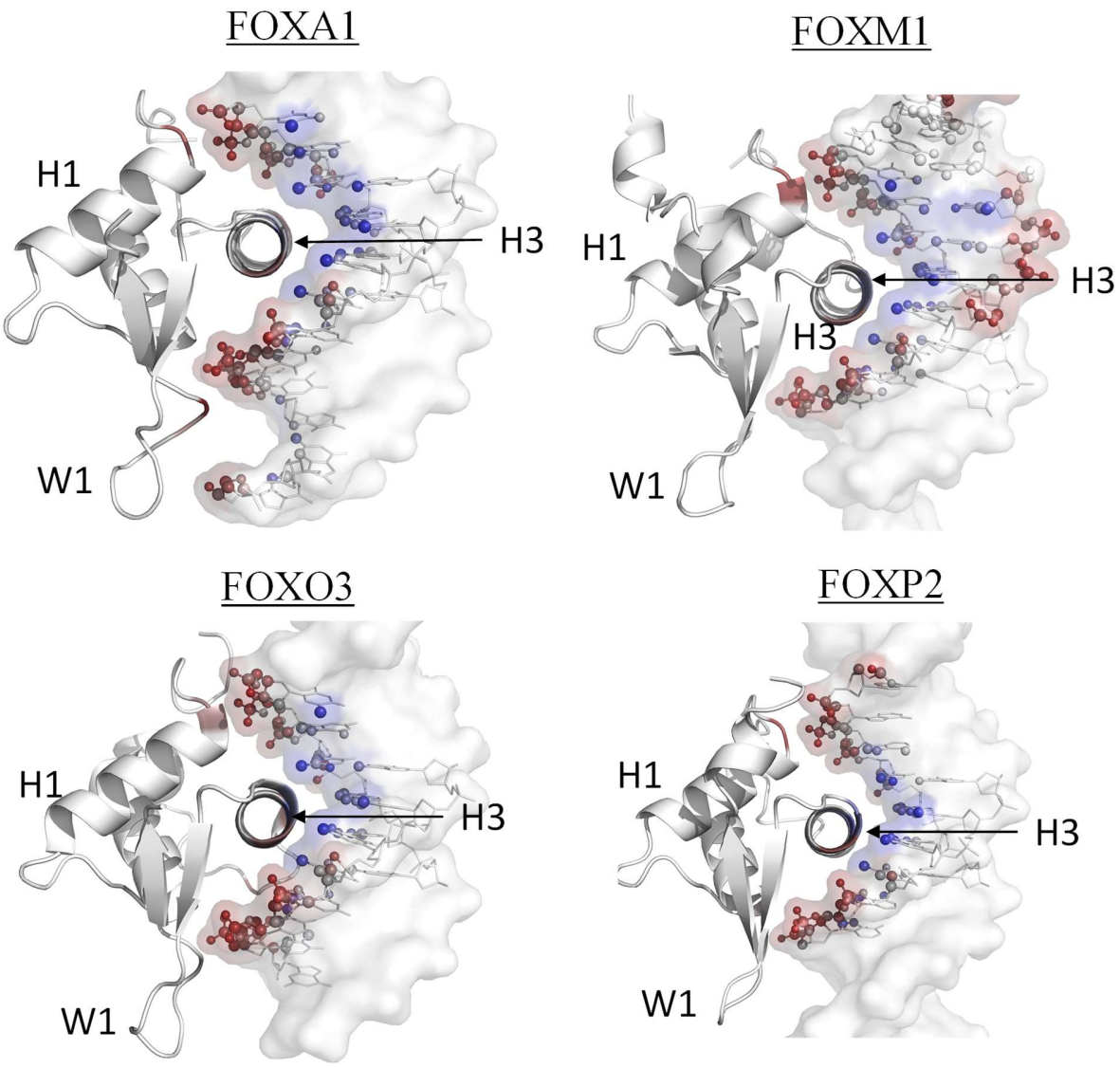


Figure 5: The FOX transcription factors bind a highly conserved promoter sequence. Cartoon representation of the co-crystal structures of FOXA2 (PDB code: 5X07), FOXM1 (PDB code: 3G73), FOXO3a (PDB code: 2UZK) and FOXP2 (PDB code: 2A07) bound to an oligonucleotide containing the same DNA motif (TGTTTAC). Base recognition occurs between residues of the recognition helix (H3) and the nitrogenous bases of the consensus FOX site (blue). The electrostatic and hydrophobic interactions between the N-terminal of helix 1 (H1), the C-terminal of helix 2 (overshadowed by W1) and the β-sheet wing (W1) and the sugar-phosphate backbone flanking the FOX consensus sequence are shown in red. Images were created using

Pymol v1.8. (DeLano Scientific, 2006) The interactions were analysed and coloured with the Pymol plugin PDIViz v1.2.2 (Ribeiro *et al*., 2015).

Notably, some regions that do not form a part of the forkhead domain-DNA interface have profound effects on the specificity of the FOX TFs. The disordered loop connecting the second and recognition helices, the hinge loop, has significant sequence variation across the FOX families and can alter the DNA binding specificity despite not being directly involved in the protein-DNA interface (Pierrou *et al*., 2994; Marsden *et al*., 1998; Overdier *et al*., 1994). NMR studies have shown that structural switching in the hinge loop region, between unstructured and helical, shortens and repositions the recognition helix in the major groove of the consensus binding site (Marsden *et al*., 1998). Taken together, sequence variations of the regions adjacent to those that are directly and indirectly involved in DNA binding, fine tune the forkhead domain-DNA interaction possibly resulting in reduced interference between co-expressed FOX proteins even though they share consensus binding site similarity.

1.5.2 Clinical relevance of the FOX family

The members of the FOX family of transcription factors have become clinically relevant targets owing to their overarching involvement in the progression and survival of a cancerous cell state in several tissues (Myatt and Lam, 2007). Several FOX members (FOXA, FOXC, FOXM, FOXO and FOXP) have been implicated in the initiation, survival and progression of a variety of cancers (Myatt and Lam, 2007). FOXM1, prohibitively expressed in proliferating cell lines, is a master regulator of G_1-S and G_2-S cell cycle progression controlling expression of systems that regulate cell metabolism, growth and replication including *Cks1*, *Cyclin B1* and *CDC25B* (Wang *et al*., 2005). It is no surprise that overexpression of FOXM1 has been implicated as a critical player in the initiation, progression and survival of several cancers (Koo *et al*., 2012; Laoukili *et al*., 2007; Pilarsky *et al*., 2004; Zhou *et al*., 2014;). Importantly, selective therapeutics targeting the forkhead domain of FOXM1 have shown promising potency in downregulating many of the genetic systems upregulated by aberrant

FOXM1expression in cancer cell cultures (Hegde *et al*., 2011; Gormally *et al*., 2014). Surprisingly, the FOXP subfamily members, notable for their role in the development of the brain and immune system, have become considerable roleplayers in the survival of diffuse large B-cell lymphomas, prostate and breast cancer (Cuiffo *et al*., 2014; Shikegawa *et al*., 2011; Gao *et al*., 2017; Takayama *et al*., 2008; Triulzi *et al*., 2013; Wong *et al*., 2016). Although difficult, novel therapeutics targeting the FOX proteins may prove to be an important step forward in the fight against cancer. This initiative promotes further understanding of the structural biology of specific forkhead domain-DNA interactions, essential to the design of novel therapeutics that target and control the transcriptional master regulators of a diseased cell state.

1.6 FOXP subfamily

The FOXP subfamily consists of four members (FOXP1-4) that share a homologous forkhead domain sequence. FOXP3 is involved in the differentiation of regulatory T cells that help to suppress the immune system whereas FOXP1/2 and 4 share overlapping roles in the development of the lung, gut and brain tissues (Sakaguchi *et al*., 2010; Shu *et al*., 2001; Lu *et al*., 2002; Lai *et al*., 2003; Chatila, 2000). Dysregulation of FOXP proteins results in a variety of congenital and non-communicable diseases. Non-functional FOXP3 is responsible for the auto-immune disease IPEX (immunodysregulation, polyendocrinopathy, enteropathy, X linked syndrome) (Bennett *et al*., 2001). Deletions of interstitial regions surrounding the *FOXP1* gene and monogenic mutations within the *FOXP1* gene have been implicated in several autism spectrum disorders, mental disorders (schizophrenia) and cognitive disabilities (neurodevelopmental delay and specific language impairment) (Sollis *et al*., 2016; Horn, 2012; Meerschaut *et al*., 2017). This led to the classification of a novel FOXP1-related intellectual disability syndrome (Meerschaut *et al*., 2017). Similarly, mutations in FOXP2 have been implicated in the development of a severe speech and language disorder and mental disorders including schizophrenia (Lai *et al*., 2001; Tolosa *et al*., 2010).

The FOXP subfamily members are the only known FOX proteins to form DNA binding dimers (Li *et al*., 2004). To date, two dimerisation interfaces have been identified in the FOXP subfamily members (Bandukwala *et al*., 2011, Li *et al*., 2004, Stroud *et al*., 2006; Chu *et al*., 2011). All the FOXP members contain a leucine zipper domain located N-terminal of the forkhead domain (Figure 6) that facilitates dimerisation and may be necessary for efficient DNA binding (Bandukwala *et al*., 2011; Li *et al*., 2004). Furthermore, through this domain FOXP1/2 and 4 can form heterotypic dimers (Li *et al*., 2004).

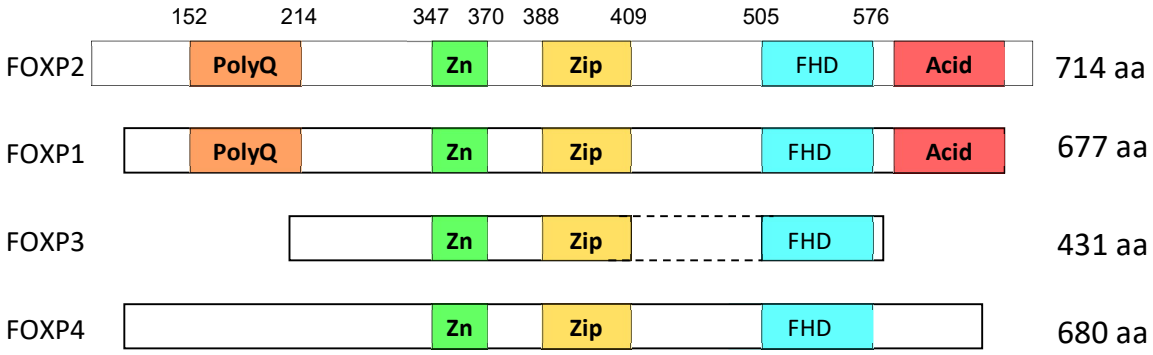


Figure 6: Domain topology schematic of the FOXP subfamily members. The numbers to the right indicate the length of the protein in number of amino acids. The glutamine-rich domain (PolyQ) is not found in FOXP3, however, all FOXP members have a zinc finger (Zn), leucine zipper (Zip) and conserved forkhead domain (FHD). Unique to FOXP1 and 2 is a C-terminal acid rich tail (Acid). The numbers above the domains indicate the start and end amino acid number of each domain in the full length FOXP2, respectively. FOXP3, being considerably shorter and lacking the polyQ domain, exhibits the greatest structural divergence from the other FOXP members.

For the rest of the FOX family members, the forkhead domain binds DNA as an obligate monomer with the exception of the FOXP subfamily members. The FOXP1, FOXP2 and FOXP3 forkhead domains have been shown to form homodimers *in vitro* using low resolution chromatography experiments (Bandukwala *et al*., 2011; Stroud *et al*., 2006;

Medina *et al*., 2016). Additionally, co-crystal structures of the FOXP2 and FOXP3 forkhead domain bound to DNA reveal domain-swapped forkhead domain dimers (Bandukwala *et al*., 2011; Chen *et al*., 2015; Stroud *et al*., 2006).

1.6.1 FOXP forkhead domain swapping

The FOXP2 and FOXP3 forkhead domains have the propensity to form domain-swapped dimers (Figure 7A) (Stroud *et al*., 2006; Bandukwala *et al*., 2011). Domain swapping is an uncommon oligomerisation event that involves the exchange of equivalent sudomain regions between associated monomeric subunits (Bennett *et al*., 1994). The term was first used to describe the unusual form of dimerisation observed in diphtheria toxin and has since been identified in a variety of protein families (Bennett *et al*., 1994; Liu and Eisenberg, 2002). Despite the identification of domain swapping in almost 40 proteins (Liu and Eisenberg, 2002), little is known about the mechanism (Kunu and Jernigan, 2004) or its role *in vivo* (Newcomer, 2002). Previous studies have suggested a role for domain swapping in HCC amyloid formation (Janowski *et al*., 2005), cell cycle control (for example, p13suc1) (Bourne *et al*., 1995) and the function of enzyme complexes (for example, T7 endonuclease I) (Hadden*et al*., 2001). Common to all domain swapped oligomers is the presence of a loop region that connects the exchanged and base subdomains (Newcomer *et al*., 2002). Furthermore, the loop region is often involved in the open interface that stabilises the protein oligomer (Newcomer *et al*., 2002).

Domain swapping of the FOXP forkhead domain is mediated by the extension of helix-2, facilitating an exchange of helix 1 and helix 2 with an adjacent monomer (Stroud *et al*., 2006; Bandukwala *et al*., 2011; Chen *et al*., 2015). The FOXP forkhead domain can be divided into two distinct subdomains: (i) the recognition helix and the wings that make up the DNA binding surface and (ii) helix 1 and helix 2 that make up the subdomain that is exchanged during domain swapping. The loop connecting the two subdomains (the hinge loop) mediates the domain swapping by folding to form part of

helix 2. The hinge loop is highly conserved across the FOX family members with the most notable exception being a single conserved proline residue (purple, Figure 4) that is replace by an alanine in the FOXP subfamily. Notably, exchange of the alanine with a proline prevents dimerisation of the FOXP forkhead domain, confirming the importance of the residue in the evolution of dimerisation in the FOXP subfamily (Chu *et al*., 2011; Bandukwala *et al*., 2011; Stroud *et al*., 2006). Furthermore, it has been shown that even subtle changes in the amino acid composition of the hinge greatly alters the propensity for dimerisation of the FOXP forkhead domain, for instance, substitution of Phe 540 with a Tyrosine in FOXP2 results in an approximately ten-fold increase in dimerisation (Perumal *et al*., 2015). Following domain swapping, the dimer is stabilised by hydrophobic interactions that consists of a number of aromatic residues, namely: Phe508, Tyr509, Trp533, Phe538, Ala539, Tyr540 and Phe541 (Stroud *et al*., 2006) (Figure 7).

The precise role of domain swap dimerisation in the FOXP subfamily is unclear, however, it is suggested to be crucial to the complete function of the FOXP TFs (Bandukwala *et al*., 2011). Current evidence suggests that domain swapping of the FOXP3 forkhead domain allows binding and association of two distant promoters (Figure 7B; Chen *et al*., 2015). Furthermore, at least two natural mutations that are known to prevent FOXP3 forkhead domain dimerisation, F371C and F373A, result in the immune disorder IPEX implying the necessity of the dimer (Bennett *et al*., 2001; Bandukwala *et al*., 2011). The phenomenon of gene kissing as a regulatory process is well documented, and it is plausible that the FOXP subfamily members are somehow involved in this. In this way distal response elements, possibly located on separate chromosomes, are brought together to bring about finely tuned gene regulation (Fanucchi *et* al., 2014).

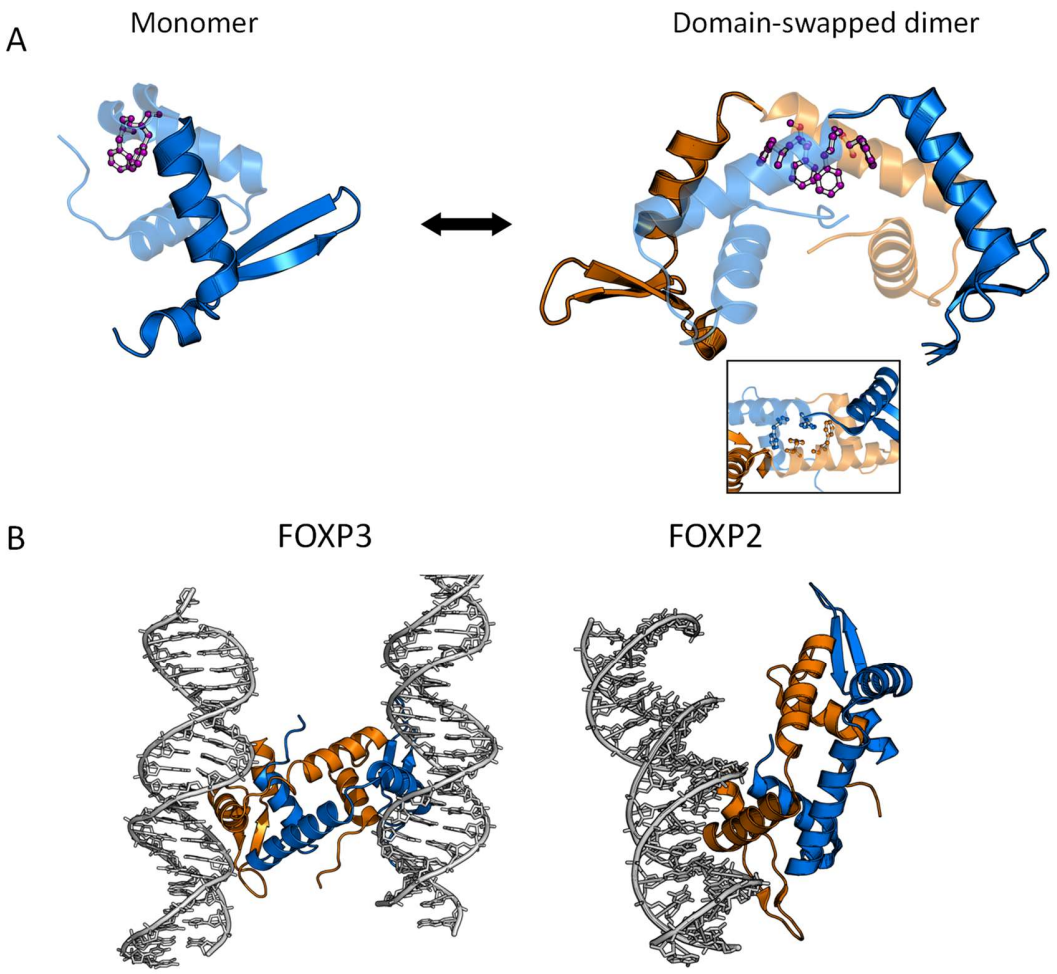


Figure 7: The FOXP forkhead domain exhibits the propensity to form domain-swapped dimers. (A)Two associated FOXP2 monomers exchange subdomain 2 (transparent) with one another to form a domain swapped dimer. The dimer interface (inset) is stabilised through van der Waals contacts between a number of hydrophobic residues. Models were rendered in Pymol™ v 0.99 (DeLano Scientific, 2006) using the PDB code: 2A07. (B) DNA binding of the FOXP3 and FOXP2 forkhead domain domain-swapped dimers. (Left) The FOXP3 forkhead domain domain-swapped dimer is capable of binding two separate DNA double helices (PDB ID: 3QRF). (Right)According to the crystal structure, the FOXP2 forkhead domain domain-

swapped dimer binds a single DNA double helix. Models were rendered in Pymol™ v 0.99 (DeLano Scientific, 2006) using the PDB codes: 3QRF and 2A07, respectively.

1.6.2 FOXP dimerisation regulates transcriptional activity

Formation of FOXP1/2/4 homo- and heterodimers regulate the differential expression of key role players during development of the central nervous and skeletal systems (Sin *et al*, 2014; Zhao *et al*., 2015). FOXP subfamily members 1, 2 and 4 are coexpressed in the developing human brain (Teramitsu *et al*., 2004; Takahashi *et al*., 2009; Konopka *et al*., 2009). Tissue-specific homo- and heterodimerisation of FOXP1/2/4 through a conserved leucine zipper domain modulate gene expression necessary for the development of a complete, functional nervous system (Li *et al*., 2004; Sin *et al.*, 2014). The various combinations of FOXP1/2/4 have a significant effect on the expression of genes involved in the Notch and WNT signalling pathways as well as genes crucial for the development of the central nervous system (Sin *et al*, 2014). Homodimers of FOXP1/2/4 act by either repressing or activating target genes. The formation of FOXP1/2/4 heterodimers significantly alters the method of regulation by either reducing the original functionality of the homodimer or by completely changing the method of gene regulation (i.e. a change from activation to repression and *vice versa*) (Sin *et al*, 2014). FOXP2 homodimers are known for their role in neurite outgrowth and synaptic plasticity. For instance, FOXP2 homodimers upregulate the NEUROD2 gene, whose exact function is not known, whereas FOXP2/4 heterodimers repress the expression of NEUROD2 (Vernes *et al*., 2011; Sin *et al.*, 2014). NEUROD2 is fundamental in central nervous system development and NEUROD2-null mice exhibit features common to FOXP2-null mice such as poor cerabellar development and motor deficits such as ataxia (Olsen *et al*., 2001).

Dimerisation is essential for DNA binding and transcriptional activity by the FOXP subfamily members (Li *et al*., 2004). Removal of the leucine zipper or deletion of a

highly conserved glutamic acid residue, E251, associated with the FOXP3 loss-of-function disease, IPEX, was shown to result in the complete loss of DNA binding and transcriptional activity *in vivo* (Li *et al*., 2004). In the same study, the DNA binding ability of the mutant FOXP members could be restored by the addition of an N-terminal Glutathione-S-transferase (GST) fusion tag. GST is an obligate dimer thereby facilitating the association of two GST fusion FOXP subunits; this suggests that two FOXP subunits are required to interact with DNA at the same time *in vivo*.

The FOXP2 forkhead domains are capable of binding DNA as both monomer and a dimer in the abscence of the leucine zipper (Stroud *et al*., 2006). However, the abscence of the leucine zipper results in unstable dimers that readily dissociate into monomers *in vitro* (Stroud *et al*., 2006). Thus, the problem of studying DNA binding by the FOXP2 forkhead domain is two-fold. An obligate FOXP2 forkhead domain dimer is necessary, to prevent the presence of mixed quaternary states during DNA binding studies, before the role dimerisation has in DNA binding can be ascertained. In contrast, the FOXP2 forkhead domain domain-swapped dimer is only shown to bind a single DNA strand (Figure 7b); again with little *in vitro* evidence to support otherwise (Stroud *et al*., 2006). According to the crystal structure, however, a distinct difference in the protein-DNA interactions is apparent between the FOXP2 forkhead domain monomer and FOXP2 forkhead domain domain-swapped dimer (Stroud *et al*., 2006). Taken together, It is of clear importance to understand the role that forkhead dimerisation plays in both the DNA binding propensity and transcriptional activity of the FOXP subfamily members.

1.7 FOXP2

FOXP2 is a transcriptional repressor involved in the development of the human brain and is the first gene product implicated in a hereditary human language disorder (Lai *et al*., 2001). A loss-of-function mutant, R553H FOXP2 was first discovered in two isolated cases, a British family (KE family) and an individual (CS), displaying severe deficits in vocalisation and the use of language, a condition referred to as verbal and

orofacial dyspraxia (Lai *et al*., 2001). Individuals with non-functional FOXP2 also displayed extra-linguistic deficits including below average IQ and abnormal brain structure (Watkins *et al.*, 2002). Furthermore, animal model studies of engineered *Mus musculus* expressing R552H FOXP2 (equivalent to the R553H mutation observed in *Homo sapiens*) revealed a role for FOXP2 in the development of the neural networks that govern fine motor coordination (Grozser *et al*., 2008; Enard *et al.,* 2009). The transgenic mice displayed abnormal cerebellar structure as well as impaired synaptic plasticity and motor learning (Grozser *et al*., 2008; Enard *et al.,* 2009). At a cellular level, neuronal cell culture (CTX and GE precursor cells) studies have revealed a role for FOXP2 in the regulation of differentiation of neuron precursors, particularily medium spiny neurons and interneurons of the embryonic forebrain (Chiu *et al*., 2014).

FOXP2 expression is altered in breast, B-cell lymphoma and osteosarcomas (Cuiffo *et al.*, 2014; Wong *et al*., 2016; Gascoyne *et al*., 2015). A role for FOXP2 in the regulation of metastasis in breast cancer cells has been reported (Cuiffo *et al*., 2014). Repression of FOXP2 expression, by MSC-deregulated miRNA, resulted in enhanced cancer stem cell morphology and ultimately poor prognosis in malignant breast cancer cohorts (Cuiffo *et al*., 2014). This is not the first case to report an aberrant expression of FOXP2 in cancer formation as an earlier study showed aberrant FOXP2 expression in neoplastic plasma cells (Campbell *et al*., 2010). These cases represent the potential for FOXP2 as an important factor in tumorigenesis and as such a novel therapeutic target for cancer therapy.

1.7.1 FOXP2 structure

FOXP2 contains a N-terminal glutamine rich domain (PolyQ), and multiple DNA binding domains in addition to the forkhead domain including a zinc finger domain (ZnF) and a leucine zipper domain (ZIP) (Vernes *et al*., 2006) (Figure 6). The PolyQ domain, also present in FOXP1, may be involved in regulatory protein-protein interactions (Lai *et* al., 2001). The ZnF and ZIP are closely associated toplogically and

appear to mediate homo- and heterotypic dimerisation among the FOXP members (Li *et al*., 2004; Bandukwala *et al*., 2011). It was initially suggested that the leucine zipper was necessary for dimerisation-dependent DNA binding (Li *et al*., 2004), however, it has since been shown that the isolated forkhead domain is able to bind DNA in both monomeric and dimeric form (Stroud *et al*., 2006).

1.7.2 DNA binding by the FOXP2 forkhead domain

Binding of the FOXP2 forkhead domain to DNA is dominated by non-specific interactions with few base specific contacts (Stroud *et al*., 2006). As with all wHTH TFs, including both the FOXP2 monomer and domain-swapped dimer, DNA binding is mediated through insertion of helix-3 into the DNA major groove (Stroud *et al*., 2006). Polar and apolar residues in helix-3 are responsible for sparse base-specific hydrogen bonding along the FOXP2 cognate sequence (5'-TGTTTAC-3') that governs sequence specificity. These residues include: Arg553, His554, Ser557 and Leu558 (Stroud *et al*., 2006; Nelson *et al*., 2014). Non-specific contacts between helix 1, the wing and the DNA backbone account for the majority of contacts and are responsible for stabilisation of the complex (Figure 8; Stroud *et al*., 2006).

The DNA binding is dominated by van der Waals contacts with few hydrogen bonds which suggests that the FOXP2 forkhead domain may recognise and bind a broad range of sequences (Stroud *et al*., 2006). Indeed, as with many other FOX TFs, FOXP2 recognises and binds several divergent secondary sequences (Zhu *et al*., 2009; Enard *et al*., 2009; Webb *et al*., 2017). FOXP2 binds the sequences with different rates and affinities indicating a concentration dependent regulatory mechanism (Webb *et al*., 2017). Alternatively, the different sites may promote site-dependent cooperative binding with other specific TFs at neighbouring response elements allowing several combinations for fine-tuning gene-specific regulation. The monomeric form of the FOXP2 forkhead domain exhibits a greater number of non-specific contacts than the domain-swapped dimeric form (Figure 8). Therefore, it may be that dimerisation of the

forkhead domain plays a physiological role in altering the DNA binding affinity. Unfortunately, the crystal structure lacks dynamics information that may be vital in determining a difference in DNA binding between the two quaternary states. It is clear then that a conclusion as to the role of dimerisation should be withheld until further solution and dynamics studies on the FOXP2 forkhead domain have been conducted.

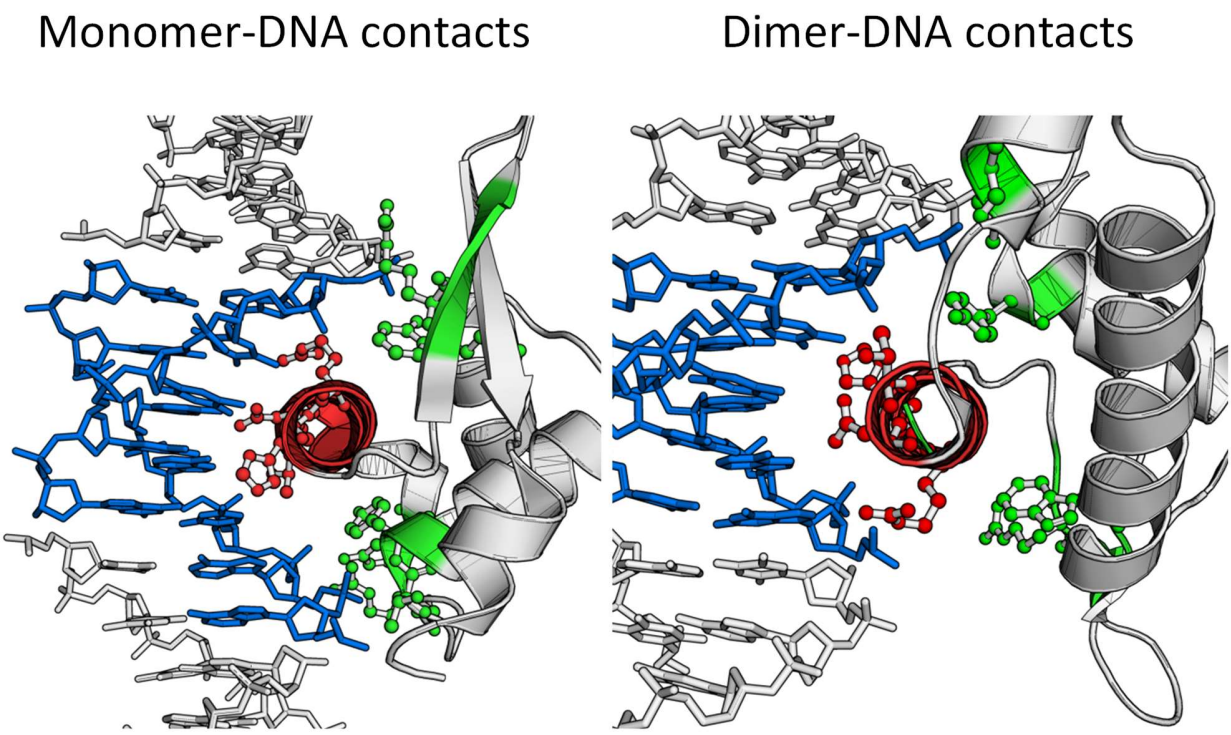


Figure 8: Comparison of DNA contacts made by the FOXP2 forkhead domain monomer (left) and the domain-swapped dimer (right). The residues involved in sequence specific contacts (red) remain the same, whereas those which make non-specific backbone contacts (green) differ considerably in number. Models were rendered in Pymol^{TM} v 0.99 (DeLano Scientific, 2006) using the PDB code: 2A07.

1.8 Problem identification

The FOXP2 forkhead domain is a promising model for elucidating the evolution of dimerisation in the FOX family of TF and may provide valuable insight into the mechanism of domain swapping and the evolution of TF dimerisation in general. The FOXP2 forkhead domain exists as a mixture of monomer and dimer *in vitro* unlike the obligate monomeric quaternary state of the rest of the FOX family members or the predominantly dimeric quaternary state of the FOXP3 forkhead domain (Stroud *et al*., 2006; Bandukwala *et al*., 2011). Understanding how dimerisation of the forkhead domain alters the mechanism of DNA binding will provide novel insight into the purpose of FOXP dimerisation as well as provide possible explanations for why the mutations that prevent FOXP3 forkhead dimerisation, resulting in IPEX syndrome. Furthermore, from an evolutionary perspective little is known about the role of the highly conserved hinge loop except that it plays a part in defining the sequence specificity of the forkhead domain (Marsden *et al*, 1998; Overdier *et al*., 1994). Given the importance of the hinge loop in both sequence specificity and dimerisation of the forkhead domain it is of high relevance to study the role of hinge loop mutations in the FOXP subfamily forkhead domain, the first FOX subfamily with a novel evolutionary mutation. Here we aim to determine the role that the evolutionary proline to alanine hinge loop mutation has on the DNA binding mechanism and transcriptional activity of FOXP2.

Although structural data for the FOXP2 forkhead domain-DNA interactions are available, no energetics or dynamics studies have been reported. This study will be the first insight into the thermodynamic signature and structural dynamics changes of the FOXP2 forkhead domain upon binding to the forkhead box consensus DNA sequence. Furthermore, the thermodynamic signature will be dissected into electrostatic and non-electrostatic components to provide a detailed picture of the types of interactions that drive FOXP2, and by extension FOX, specificity. To do this, an engineered form of a covalently-linked FOXP2 forkhead domain dimer (F541C mutation) and an obligate

FOXP2 forkhead domain monomer (A539P mutation) will be attempted. Comparative studies of the wild-type, F541C dimeric and A539P monomeric FOXP2 forkhead domains will be conducted.

This study can help to provide detailed insight into key amino acids responsible for the specificity divergence observed in the FOX family of transcription factors. In addition, differences in binding energetics coupled with *in vivo* studies will provide information on the role of dimerisation of the FOXP subfamily forkhead domain both in terms of defining DNA binding specificity as well as in regulational activity. Taken together, this thesis is a small but important step toward fully understanding the evolutionary driving forces that resulted in the segregation of the FOXP subfamily from the rest of the FOX TFs.

1.9 Objective and aims

1.9.1 Objective

The main aim of this thesis is to characterise the energetics of the obligate monomeric (A539P mutant), obligate dimeric (F541C mutant) and wild-type FOXP2 forkhead domains coupled with residue resolution backbone dynamics changes that occur during DNA binding to the consensus FOXP2 DNA sequence. Furthmore, the role of FOXP2 forkhead domain dimerisation *in vivo* will be probed using a luciferase reporter assay assessing both the wild-type and monomeric (A539P mutant) full length FOXP2 TFs.

1.9.2 Aims

To this end the aims of this research were to:

1. Incorporate the F541C and A539P mutations into the wild-type FOXP2 coding sequence and to subsequently recombinantly express and purify the F541C, A539P and wild-type FOXP2 forkhead domain.
2. Characterise the secondary, tertiary and quaternary structures of the A539P, F541C and wild-type FOXP2 forkhead domains to determine differences in the monomeric and dimeric FOXP2 forkhead domain.
3. Compare thermodynamic signatures for the binding of A539P and wild-type FOXP2 forkhead domains to an oligonucleotide containing a consensus binding site obtained by isothermal titration calorimetry. Confirm the stoichiometry of binding with a complementary binding assay (fluorescence anisotropy).
4. Determine the contribution of electrostatic and non-electrostatic interactions to DNA binding for both the A539P and wild-type FOXP2 forkhead domain by conducting several isothermal titration calorimetry binding experiments at various buffer ionic strengths.

5. Determine the change in heat capacity of DNA binding for the A539P and wild-type FOXP2 forkhead domain by conducting isothermal titration calorimetry experiments at varying temperatures.
6. Establish the changes in dynamics caused by the A539P hinge loop mutation for both the free and DNA bound forms of the FOXP2 forkhead domain using hydrogen-deuterium exchange mass spectrometry.
7. Assess the effect the A539P mutation has on the *in vivo* transcriptional activity of FOXP2 using a dual luciferase reporter assay.

Chapter 2

A key evolutionary mutation enhances DNA binding of the FOXP2 forkhead domain
Gavin Morris and Sylvia Fanucchi
Biochemistry 55, 1959-1967 (2016)

Motivation: Several co-crystal structures of the FOXP2 forkhead domain bound to DNA have been determined. Despite this, no *in vitro* DNA binding studies have been conducted with the FOXP2 forkhead domain.

Summary: This publication describes structural characterisation of the wild-type, monomeric A539P and dimeric F541C FOXP2 forkhead domains. The wild-type FOXP2 forkhead domain is shown to prefer binding to DNA as a monomer, however, the binding occurs through a different mechanism than the exclusively monomeric A539P mutant, despite having similar structural characteristics. This work provides the starting point for an engineered disulfide-linked FOXP2 forkhead domain dimer. Furthermore, the work provides the first thermodynamic signature for DNA binding by the FOXP2 forkhead domain and provides new evidence that the hinge loop of the forkhead domain is a key regulator of DNA binding specificity by FOX transcription factors.

Author contributions: Gavin Morris performed all experimental work, analysed data and wrote the manuscript. Sylvia Fanucchi was the principal investigator that supervised the project and assisted in data analysis and interpretation.

BIOCHEMISTRY
including biophysical chemistry & molecular biology

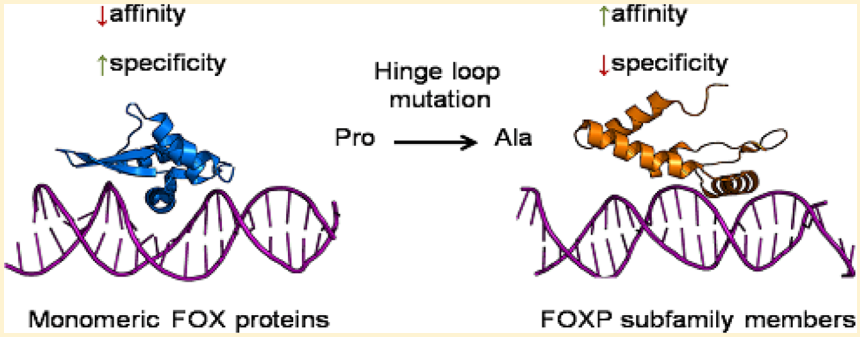
Article

pubs.acs.org/biochemistry

A Key Evolutionary Mutation Enhances DNA Binding of the FOXP2 Forkhead Domain

Gavin Morris and Sylvia Fanucchi*

Protein Structure−Function Research Unit, School of Molecular and Cell Biology, University of the Witwatersrand, 1 Jan Smuts Avenue, Braamfontein, 2050, Johannesburg, Gauteng, South Africa



ABSTRACT: Forkhead box (FOX) transcription factors share a conserved forkhead DNA binding domain (FHD) and are key role players in the development of many eukaryotic species. Their involvement in various congenital disorders and cancers makes them clinically relevant targets for novel therapeutic strategies. Among them, the FOXP subfamily of multidomain transcriptional repressors is unique in its ability to form DNA binding homo and heterodimers. The truncated FOXP2 FHD, in the absence of the leucine zipper, exists in equilibrium between monomeric and domain-swapped dimeric states in vitro. As a consequence, determining the DNA binding properties of the FOXP2 FHD becomes inherently difficult. In this work, two FOXP2 FHD hinge loop mutants have been generated to successfully prevent both the formation (A539P) and the dissociation (F541C) of the homodimers. This allows for the separation of the two species for downstream DNA binding studies. Comparison of DNA binding of the different species using electrophoretic mobility shift assay, fluorescence anisotropy and isothermal titration calorimetry indicates that the wild-type FOXP2 FHD binds DNA as a monomer. However, comparison of the DNA-binding energetics of the monomer and wild-type FHD, reveals that there is a difference in the mechanism of binding between the two species. We conclude that the naturally occurring reverse mutation (P539A) seen in the FOXP subfamily increases DNA binding affinity and may increase the potential for nonspecific binding compared to other FOX family members.

The control of gene expression by transcription factors (TF) is regulated by a number of processes including nonlinear interactions such as oligomerization, protein stability, and temporal delays involved in translation and protein processing.[1] TF oligomerization, through subunit combination-dependent DNA binding as well as through changes in protein complex stability, is a fundamental process used for the regulation of gene expression by many eukaryotic transcription factors.[2]

The forkhead box (FOX) proteins are a transcription factor family classified by a canonical winged helix DNA binding domain, the forkhead domain (FHD), first identified in the *Drosophila melanogaster Forkhead* gene product and mammalian HNF-3α TF (FOXA1).[3] Since the discovery of the FHD, the FOX family has been extended to encompass over 100 transcription factors classified into 19 subfamilies (FOXA-S) (for a concise review, see ref 4). The FOXP subfamily consists of four multidomain members (FOXP1−4). This subfamily is unique among the FOX proteins because the FOXP proteins are capable of forming homo- and heterodimers.[5] The FOXP subfamily members act as transcriptional activators and repressors in the differentiation of various cell-types. FOXP3 is involved in the differentiation of regulatory T cells (T_{reg}) that help to suppress the immune system, whereas FOXP1/2 and 4 share overlapping roles in the development of the lung, gut, and brain.[6−8] Dysfunction of these proteins leads to disease including, but not limited to, various cancers, an immune disorder, and a rare genetic speech disorder.[8−11]

The FOXP FHD has been shown in crystal structures to exist in two forms: a monomeric form which is comparable with the FHDs of other FOX proteins and a unique domain-swapped dimeric form.[13] A number of proteins have been shown to domain-swap, and although the exact mechanism of domain-swapping differs between proteins, it is clear that the presence of a flexible loop, termed the hinge loop, adjacent to the swapped region, is critical in all cases.[14] Changes in the composition of the hinge loop can drastically change the domain swapping propensity of these proteins.[15] The FOXP

Received: November 24, 2015
Revised: March 3, 2016
Published: March 7, 2016

ACS Publications © 2016 American Chemical Society

DOI: 10.1021/acs.biochem.5b01271
Biochemistry 2016, 55, 1959−1967

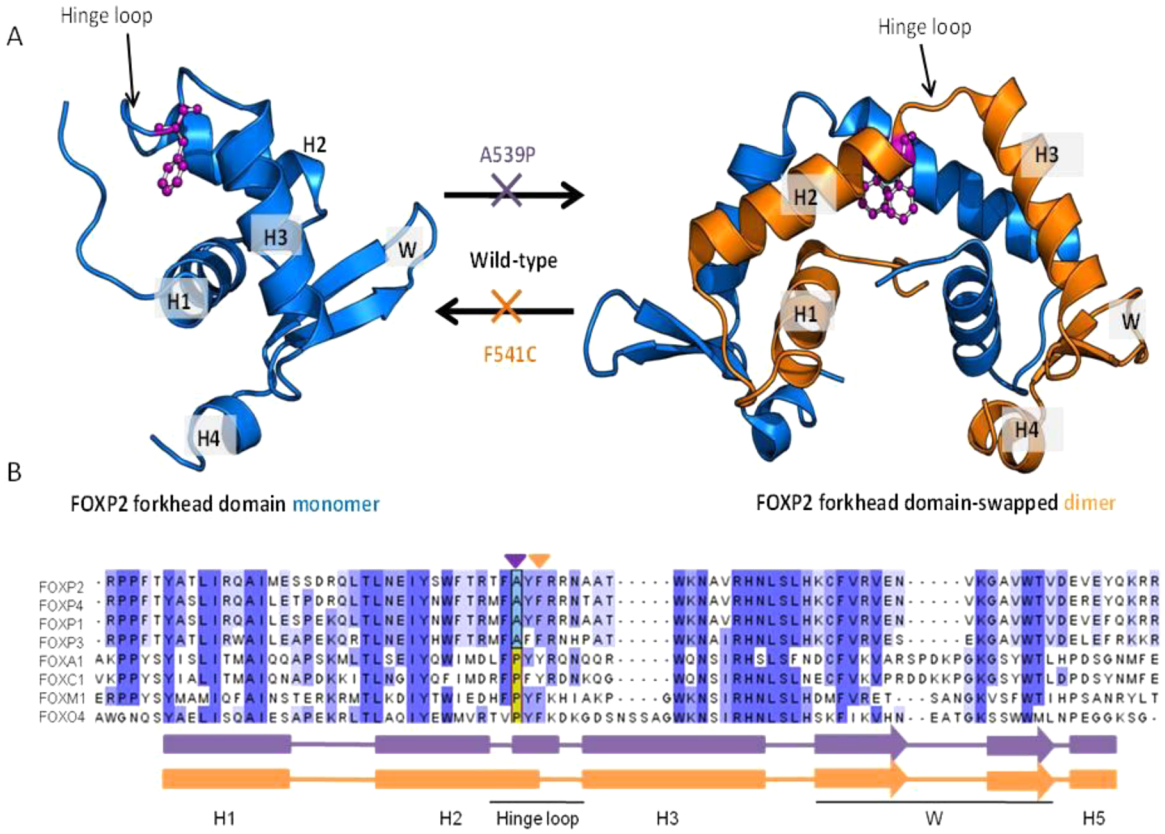


Figure 1. (A) Monomer−dimer equilibrium of the FOXP2 FHD. The α-helices and β-sheet wing of the FOXP2 FHD are labeled as H1−4 and W, respectively. The A539P and F541C mutations prevent dimerization and dissociation of the dimer complex, respectively. Phenylalanine 541 is shown in ball-and-stick in both monomeric and dimeric structures. (B) Multiple sequence alignment (performed using ClustalX v2.0)[18] of various FOX FHD sequences; the depth of blue indicates percentage conservation. The purple and orange arrows indicate the site of the A539P and F541C mutations, respectively.

subfamily are the only FOX proteins whose forkhead domains have been shown to form domain swapped dimers. The dimerization event involves the exchange of helices 3 and 4 as well as the β-sheet wing between associated monomers and occurs through the extension of helix 2 into a flexible hinge loop (Figure 1).[12,13] The hinge loop sequence, FPYF, is highly conserved in all FOX proteins with the exception of the FOXP subfamily members where the proline residue is replaced by an alanine (Figure 1). It has been shown that substitution of Ala539 with a proline residue in the FOXP2 FHD prohibits the extension of helix 2 and successfully prevents domain-swapping.[13] Thus, this unique incidence in the FOXP subfamily can be used to provide insight into the evolution of FOX domain-swapping and transcription factor dimerization. Evidence suggests that the FOXP2 FHD dimer is capable of binding two individual DNA strands, although the purpose of this is still unknown. Little else is known about the DNA binding properties of the dimer.[13,16] Given the increasing prevalence of FOXP2 (and the rest of the FOXP subfamily members) in a number of diseases, it is critical that the mechanism of dimerization and its role in DNA binding and hence transcriptional regulation are understood if novel therapeutic strategies are to be developed.

In this work, we employ two hinge loop mutations to disrupt the equilibrium between the two quaternary states of the FOXP2 FHD in an attempt to determine the role of dimerization in DNA binding. An exclusively monomeric FOXP2 FHD mutant (A539P) has already been described.[13] To our knowledge, an exclusive FOXP2 FHD dimer has not been reported. In this work we used the bioinformatics tool, Disulfide-by-Design,[17] to identify a single residue (Phe541), situated within the dimer interface and on the axis of symmetry, as a candidate for substitution with cysteine. This F541C mutation is designed to allow the formation of an intersubunit disulfide bond between two associated FOXP2 FHD monomers, thereby effectively preventing their dissociation. Here we compare the DNA binding ability of the exclusively monomeric (A539P) and dimeric (F541C) FOXP2 FHD mutants with the wild-type FOXP2 FHD in order to determine the effect of dimerization of the FOXP2 FHD on DNA binding affinity. Furthermore, the energetics of this binding event is described for the first time. From this it is shown that the amino acid composition of the hinge loop is critical for dimerization and DNA binding affinity and may affect the sequence-specificity of the forkhead domain.

EXPERIMENTAL PROCEDURES

Mutation Site Identification. The bioinformatics tool Disulfide by Design was used to identify candidate residues for substitution with cysteine.[17] Phenylalanine at position 541 in the full length FOXP2 was chosen due to its location within the dimer interface and its position on the 2-fold symmetry axis of the FHD dimer. Therefore, only a single mutation is required to facilitate the formation of an interchain disulfide bond.

Site-Directed Mutagenesis. High fidelity overlapping primer PCR was used to introduce the A539P and F541C mutations at the corresponding codons of the wild-type FOXP2 FHD coding sequence.

DOI: 10.1021/acs.biochem.5b01271
Biochemistry 2016, 55, 1959−1967

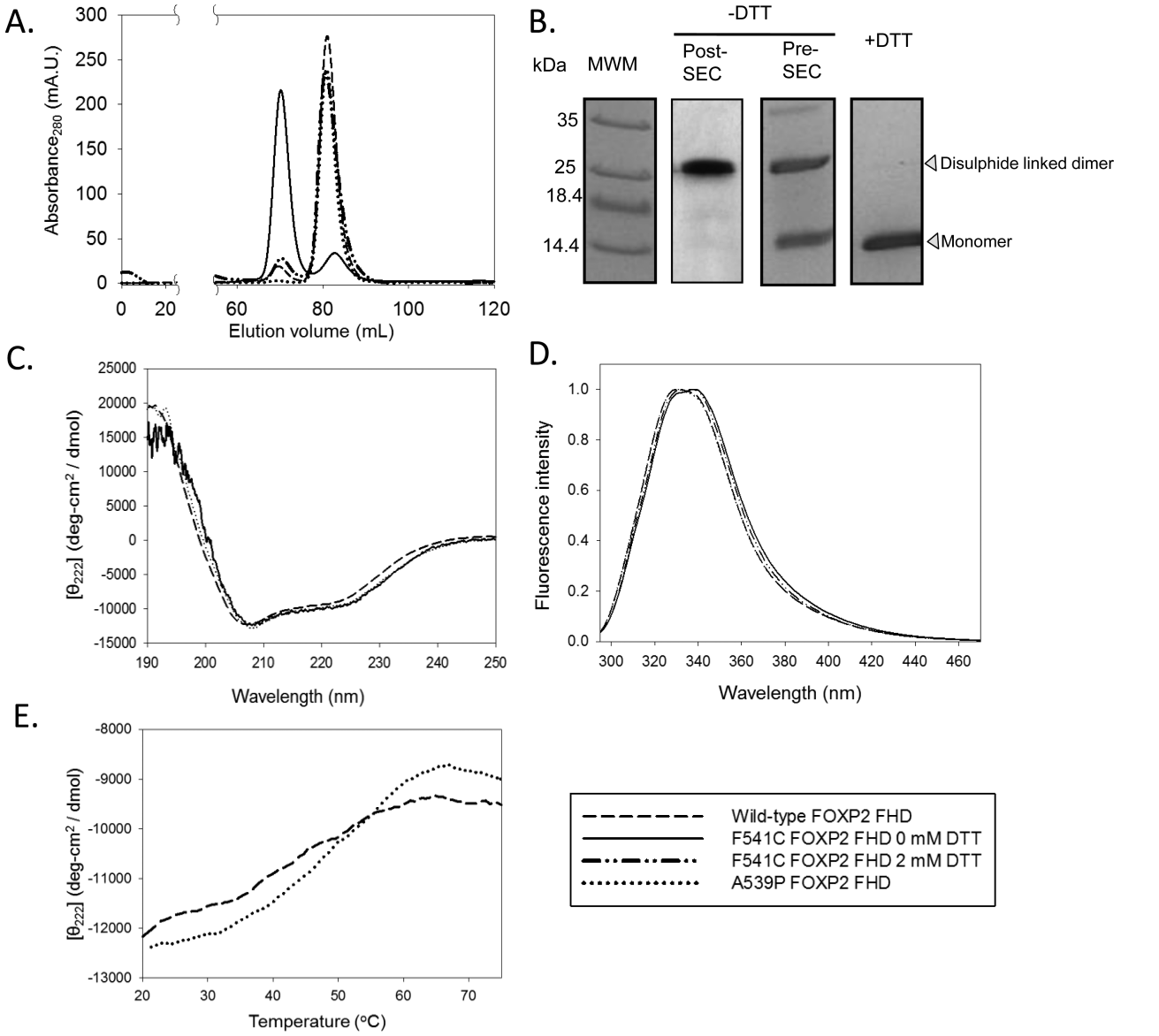


Figure 2. Secondary, tertiary, and quaternary structural characteristics of wild-type, monomeric A539P, and dimeric F541C FOXP2 FHDs. (A) Gel filtration elution profiles of wild-type, A539P, and F541C FOXP2 FHD under reducing and nonreducing conditions. The wild-type, A539P mutant, and F541C mutant under reducing conditions are predominantly monomeric (~13 kDa). Under nonreducing conditions F541C FOXP2 FHD is predominantly dimeric. (B) SDS-PAGE of F541C FHD under reducing (+DTT) and nonreducing (−DTT) conditions. The nonreduced sample prior to gel filtration (Pre-SEC) as well as the dimer fraction post gel filtration (Post-SEC) are shown. (C) Far-UV CD spectra for the wild-type, monomeric A539P, and dimeric F541C FOXP2 FHDs under nonreducing conditions. (D) Normalized intrinsic tryptophan fluorescence of monomeric A539P and dimeric F541C FOXP2 FHDs under reducing and nonreducing conditions. (E) Thermal melting curves of the wild-type and A539P FOXP2 FHD obtained by monitoring the far-UV CD absorbance at 222 nm while incrementally increasing temperature.

Protein Expression and Purification. The coding sequence of the FOXP2 FHD (amino acids 503−586) was codon optimized (Genscript, USA) and inserted into the multiple cloning site of a pET-11a plasmid (Novagen, Germany) under the control of a T7 inducible promoter. The wild-type and all mutant FOXP2 FHDs were expressed and purified as previously described.[19]

The disulfide linked F541C FOXP2 FHD dimers were further purified by size exclusion chromatography on a HiLoad 16/60 Superdex 75 prep grade size exclusion column (GE Healthcare, USA).

Structural Characterization. To determine the quaternary structure of the wild-type, monomeric A539P and dimeric F541C FOXP2 FHD, size exclusion chromatography was performed using ~300 μM protein on a HiLoad 16/60 Superdex 75 prep grade size exclusion column (GE Healthcare, USA). Following size exclusion, the dimeric F541C FOXP2 fraction was isolated, and free thiols were blocked by 1 h incubation (at 293 K in the dark) with 125 mM iodoacetamide before being resolved on a 16% nonreducing tricine SDS-PAGE gel.

Secondary structural characterization of the wild-type, monomeric A539P, and dimeric F541C FOXP2 FHD was performed using far-UV circular dichroism spectropolarimetry on a Jasco J1500 in the wavelength range of 190−250 nm using 10 μM protein in a 10 mm path length quartz cuvette. Raw data were converted to mean residue ellipticity ($[\theta]_{MRE}$) using the following formula:

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

Tertiary structural changes were analyzed using intrinsic tryptophan fluorescence on a Jasco FP-6300 fluorescence spectrophotometer with an excitation wavelength of 280 nm. The emission spectra were monitored over the wavelength range of 280−500 nm. Fluorescence

DOI: 10.1021/acs.biochem.5b01271
Biochemistry 2016, 55, 1959−1967

intensity of each spectrum was then normalized for comparison due to the quenching effect of the disulfide bond.

Oligonucleotides. Duplex cognate DNA containing a single binding site as determined by Nelson et al.[20] (bold), TTAGG-**TGTTTAC**TTTCATAG, was prepared to a stock concentration of 200 μM (Integrated DNA technology, South Africa).

Electrophoretic Mobility Shift Assay. Electrophoretic mobility shift assays were performed to determine the formation of monomer−DNA and dimer−DNA complexes. Samples of increasing protein/DNA ratios were prepared using 0−6 μM FOXP2 FHD with 1 μM cognate DNA. Binding reactions were conducted in EMSA binding buffer (10 mM HEPES pH 7.5, 100 mM KCl, 1 mM $MgCl_2$, 0.1 mg/mL BSA, and 20% glycerol) and incubated at 4 °C for 30 min. Each reaction sample was then resolved on a 10% polyacrylamide gel (containing 20% triethylene glycol) with a 1× TBE running buffer at 150 V for 2 h at 4 °C.

Fluorescence Anisotropy. Fluorescence anisotropy experiments were conducted on a PerkinElmer LS-50B fluorescence spectrophotometer fitted with an anisotropy filter. Briefly, 20 nM fluorescein- 5′-labeled DNA was incubated with increasing concentrations of wild-type, A539P and F541C FOXP2 FHD. Fluorescence anisotropy measurements were taken with an emission wavelength of 520 nm following excitation with a wavelength of 494 nm. Data were obtained in triplicate and averaged.

Isothermal Titration Calorimetry. The DNA binding thermodynamics of the wild-type and monomeric A539P FOXP2 FHD was monitored using isothermal titration calorimetry. Protein samples were prepared by extensive dialysis against fresh binding buffer (10 mM HEPES pH 7.5, 100 mM KCl). Protein concentration was determined by UV absorbance of 280 nm light, following correction for light scattering effects, using an extinction coefficient of 22 460 $M^{-1}\cdot cm^{-1}$. The DNA was extensively dialyzed against the same buffer used for the protein sample, and the concentration was assessed using UV absorbance at 260 nm with a theoretical extinction coefficient of 256 016 $M^{-1}\cdot cm^{-1}$ (Kibbe, 2007). Titration experiments were performed on a TA Instruments Nano-ITC by 20 10 μL injections of ∼50 μM cognate DNA into ∼5 μM FOXP2 FHD at 20 °C. The heats of saturation were averaged, and the value subtracted from all data points before data analysis. Data were fitted using a nonlinear least-squares method for independent site binding. Errors were determined as the standard deviation of three averaged independent titrations. Multiple site binding fits were attempted, but results did not make biochemical sense.

RESULTS

It has been shown that the isolated FOXP2 FHD is capable of existing as both a domain swapped dimer as well as a monomer in solution.[13] Because the full length protein has been shown to dimerize in vivo predominantly at the leucine zipper domain rather than the forkhead domain,[5] the question arises as to the relevance of the FHD dimer and its potential role in transcriptional regulation. To date, this question has not been answered. Work on the isolated FHD is complicated by the fact that the wild-type exists as a mixture of monomer and dimer species in solution. In order to address this, and to distinguish between the binding events of the monomer and dimer species, we have attempted here to create both an exclusively monomeric (A539P) and an exclusively dimeric (F541C) mutant. These species can then be studied, in comparison to the wild-type, so as to dissect the individual roles of the monomer and dimer in DNA binding and to establish the structural significance of the residues in the hinge loop region. In the work presented here we first confirm the successful creation of the monomer and dimer and whether there are any major structural differences between them and the wild-type. We then verify that all three species can bind DNA. Finally we compare the DNA binding affinities and energetics of the wild-type and monomeric FOXP2 FHD.

Structural Characterization of Wild-Type, Monomer and Dimer Mutants. Size exclusion chromatography was used so as to confirm the monomeric (∼13 kDa) or dimeric (∼26 kDa) status of the various FOXP2 FHD variants (Figure 2A). Because the dimeric mutant was engineered to be linked by a disulfide bond, all studies were conducted under both reducing (+2 mM DTT) and nonreducing conditions. The wild-type was found to be almost entirely monomeric at concentrations as high as 300 μM whether in the presence or absence of DTT. The A539P monomeric mutant was exclusively monomeric as was previously described.[13] We show here for the first time that the F541C mutation is capable of successfully capturing dimerization events of the FOXP2 FHD through the formation of a disulfide bond at the dimer interface that prevents the dissociation of the monomers. This can be seen by the size exclusion elution profile of F541C FOXP2 FHD, which closely matches that of the wild-type when under reducing conditions but displays a shift of the majority of the species to the dimeric state upon removal of the reducing agent. SDS-PAGE of the F541C mutant further confirms this result (Figure 2B). Under reducing conditions only a single band corresponding to the monomeric size can be seen, but under nonreducing conditions there are two bands suggesting that a high proportion of disulfide-linked dimers are present at equilibrium. Furthermore, nonreducing SDS-PAGE analysis of the dimeric peak obtained from the size exclusion column indicates that a pure dimeric form can indeed be isolated. The sizes of the monomeric and dimeric FOXP2 FHDs obtained (13 and 26 kDa, respectively) agree closely with the theoretical mass taken directly from the primary sequence suggesting that the proteins are globular in nature.

Far-UV circular dichroism and intrinsic tryptophan fluorescence spectroscopy were used to characterize the secondary and tertiary structure, respectively, of the mutants in comparison to the wild-type. All far-UV CD spectra are indicative of predominantly α-helical structure with a minor β-sheet component (Figure 2C) and suggest that the monomeric and dimeric mutants are almost identical in secondary structure to the wild-type. Furthermore, comparison of the monomeric and dimeric forms suggests no gross secondary structural changes to the forkhead domain following dimerization despite the extension of helix 2 observed in the wild-type dimer crystal structure.[13] Similarly, there is little difference in the fluorescence spectra between the wild-type, monomeric A539P, and monomeric F541C FOXP2 FHD species (under reducing conditions) (Figure 2D). The maximum emission wavelength of 330 nm is indicative of buried tryptophan residues.[21] The fluorescence spectrum of the F541C FHD under nonreducing conditions, however, shows a minor shift in maximum emission to a longer wavelength. This shift is recoverable following the addition of 2 mM DTT. This suggests that in the dimeric structure, the local environment of the tryptophan residues is slightly different than in the monomeric structure. In order to obtain some idea of the stability of the proteins in the absence of DNA, thermal unfolding of wild-type and monomeric mutant was monitored using CD as a probe (Figure 2E). The wild-type FOXP2 FHD unfolds in a linear manner starting at a temperature below 20 °C up to 60 °C similar to that seen in the homeodomain transcription factors.[26] In contrast the A539P FOXP2 FHD unfolds in a more

DOI: 10.1021/acs.biochem.5b01271
Biochemistry 2016, 55, 1959−1967

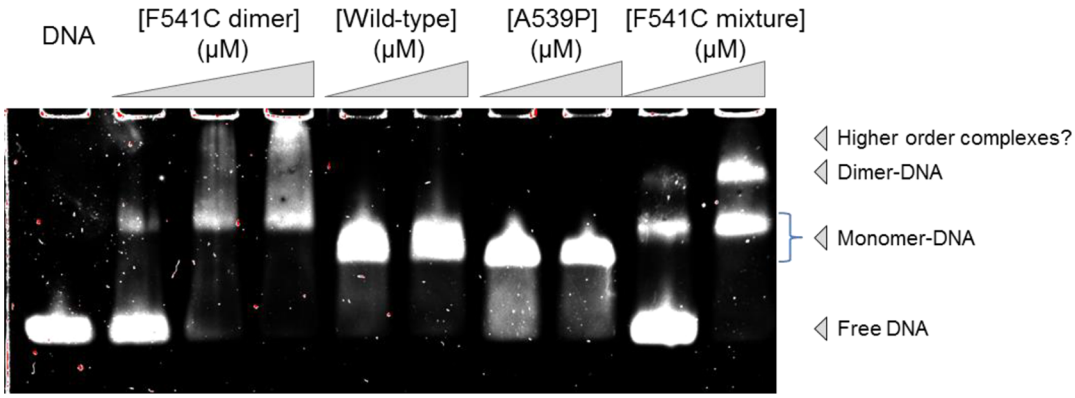


Figure 3. Protein–DNA complex formation of the wild-type and mutant FOXP2 FHDs determined by electrophoretic mobility shift assay. 300 nM cognate DNA was incubated with increasing concentrations of FOXP2 FHD (0–600 nM). The His-tag has been removed for A539P and wild type FOXP2 FHD but not for F541C FOXP2 FHD and this accounts for the slight shift in size of the complex.

sigmoidal fashion with a midpoint melting temperature of ~50 °C.

DNA Binding by Monomeric and Dimeric FOXP2 FHD. DNA binding was investigated using (i) electrophoretic mobility shift assays (EMSAs) to provide a qualitative perspective on the binding of the various species, (ii) fluorescence anisotropy (FA) to compare affinity of binding, and (iii) isothermal titration calorimetry (ITC) to provide energetic information.

The EMSA result shown in Figure 3 indicates that the wild-type FHD is indeed capable of binding DNA in the absence of the remainder of the protein. Furthermore, both the A539P and the F541C mutant FHDs are also capable of binding DNA. This can be seen by the lagging bands on the gel which are indicative of the larger protein–DNA complex and hence migrate at a slower rate than the free DNA. It is interesting to note that while there is only a single impeded band for the wild-type and A539P mutant, there are two bands for the F541C mutant mixture. This result suggests that the wild-type and A539P monomeric mutant only bind DNA as a monomer in the protein concentration range of 0–600 nM, but the F541C mutant is capable of binding to DNA in both monomeric and dimeric form under the same conditions. This result corresponds with the size exclusion result in Figure 2A, which indicates that the F541C mutant exists as a mixture of monomer and dimer species in solution. When the F541C dimer species is isolated, the EMSA results (shown in lanes 2, 3, and 4 on Figure 3) show a smear rather than single bands, which seems to suggest that higher order complexes are interacting with DNA. This may be a result of protein–DNA complex chaining, but further evidence is required to support this hypothesis.

DNA binding affinity was investigated using fluorescence anisotropy (FA) (Figure 4). This method was used because the DNA affinity of a number of other FOX FHDs has been reported using FA, and so it can be used as a means of comparison. Here we report the affinity of the wild-type and A539P mutant for cognate DNA. The binding isotherms of both wild-type and the monomeric mutant fit well to a single site binding model, and the proteins appear to have similar DNA binding affinities with K_D's of 892 ± 49.9 nM for the wild-type and 878 ± 66.2 nM for the monomeric mutant. This value is approximately an order of magnitude stronger than that reported for FOXM1 (~7 μM), 4-fold weaker than that reported for FOXO4 (~200 nM), and most similar to the ~1.5

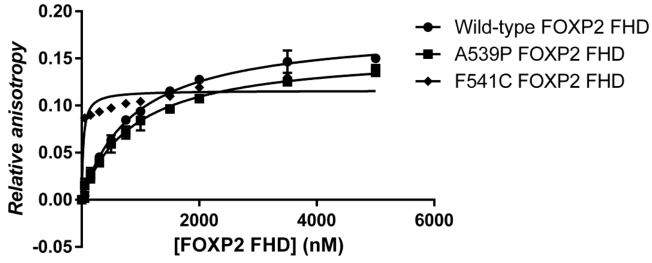
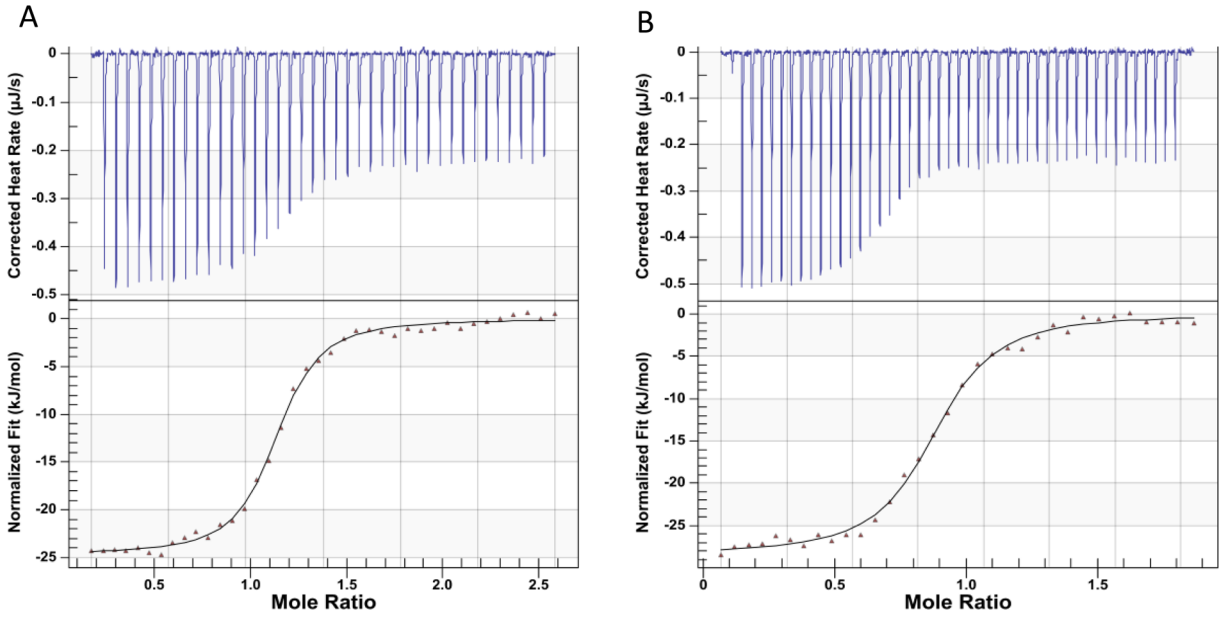


Figure 4. DNA binding isotherms of the wild-type A539P and F541C FOXP2 FHD as determined by fluorescence anisotropy sprectoscopy. 20 nM dsDNA containing a single FOXP2 binding site labeled at the 5′ end with fluorescein was incubated with increasing concentrations of FOXP2 FHD. The fluorescence anisotropy of each FHD concentration was measured and the data fitted with a single site binding model for wild-type and A539P FOXP2 FHD. A model could not be successfully fit to the F541C data. The K_D obtained for the wild-type and A539P FOXP2 FHD DNA binding are 892 ± 49.9 nM and 878 ± 66.2 nM, respectively. Errors indicate the standard deviation of three averaged replicates.

μM reported for FOXP1.[22−24] Unfortunately the purified dimer form of F541C FOXP2 FHD is unstable and prone to aggregation or higher order complex formation as suggested by the EMSA. As such its FA binding isotherm did not fit well to any binding model. However, it is interesting to speculate that, qualitatively, based on the greater anisotropy at lower protein concentrations, the dimer form has a greater affinity for DNA than the monomer form.

Thermodynamic parameters of binding were obtained for the wild-type and A539P monomeric mutant using isothermal titration calorimetry. This was done to compare the mode of binding and provide information as to whether the wild-type remains monomeric upon DNA binding. Double-stranded DNA oligonucleotides containing a single cognate binding site were injected into 5 uM wild-type or monomeric A539P FOXP2 FHD, and the heat generated or consumed by the binding reaction was measured (Figure 5, panels A and B respectively). A binding isotherm could unfortunately not be obtained for dimeric F541C FOXP2 FHD due to aggregation of the protein in the sample cell. A comparison of the DNA binding parameters for the wild-type and monomeric mutant is given in Figure 5C. While the free energy, enthalpy, and stoichiometry of the binding reaction is similar for both proteins, there are significant differences in the binding affinity and entropy. The large and exothermic enthalpy value obtained

DOI: 10.1021/acs.biochem.5b01271
Biochemistry 2016, 55, 1959−1967



C

	N	K_D (nM)	ΔG (kJ / mol)	ΔH (kJ / mol)	TΔS (kJ / mol / K)
Wild-type FOXp2 FHD	0.96 ± 0.15	66.94 ± 4.34	-40.26 ± 0.16	-24.46 ± 0.81	15.80 ± 0.70
A539P FOXP2 FHD	0.88 ± 0.01	102.30 ± 4.97	-39.23 ± 0.11	-27.60 ± 0.72	11.63 ± 0.84

Figure 5. Representative isotherms of (A) wild-type and (B) A539P FOXP2 FHD titrated into cognate sequence DNA. The isotherms were obtained by multiple small volume injections of 50−70 μM FOXP2 FHD into 5−7 μM DNA. (C) Summary of ITC thermodynamic parameters obtained for the wild-type and A539P FOXP2 FHDs. Errors are the standard deviation of three averaged independent titration experiments.

for both the wild-type and A539P FOXP2 FHD upon DNA binding suggests an enthalpically driven binding process as is commonly seen in major groove binding proteins.[25] Contrary to the FA results, the ITC results suggest that the A539P monomeric mutant has weaker affinity for the DNA than the wild-type. Because of the greater sensitivity of ITC as a method, we believe the difference in affinity between wild-type and monomeric mutant to be significant, indicating a difference in binding between the two proteins. This difference in the mode of binding is further supported by the difference in entropy of binding.

DISCUSSION

The FOXP subfamily of forkhead proteins is remarkable in its ability to dimerize. This is unlike any other FOX proteins studied to date. The dimer interface has been assigned to two distinct regions of the protein: the leucine zipper and the forkhead domain. The leucine zipper is believed to be the main contributor to this event and has been shown to be necessary for both dimerization and DNA binding of the full length protein.[5] Despite this, the isolated forkhead domain has been shown, in crystal structures, to form domain swapped dimers.[13,14,16] Although the role of the domain swapped dimer is unknown and indeed, its physiological relevance is still a matter of some contention, the crystal structures suggest that it can indeed bind DNA[13,16] and thus an intriguing possibility exists that differential binding between monomer and dimer species could provide an elegant means of transcriptional regulation. This would emphasize the significance of the evolution of domain swapping in the FOXP subfamily. In this paper we have investigated this possibility from a structural and thermodynamic perspective. We have compared the binding mechanism of the monomer and the dimer so as to gain insight into how the FHD interacts with DNA. Furthermore, we have compared the DNA binding mechanism of the wild-type with that of the monomer to provide information as to the quaternary structure of the wild-type upon interaction with DNA.

It is interesting that although the FOXP subfamily members share ∼75−92% identity in their FHD sequences, there are substantial differences in the propensity to form dimers between the FOXP subfamily members where the isolated FOXP3 FHD exists almost exclusively as a homodimer, while the FOXP1 and FOXP2 FHDs exist as a mixture of monomer and dimer under similar conditions.[13,16,22] Furthermore, the FOXP3 FHD dimer is known to bind two separate DNA strands at each binding site, but no such evidence has been reported for the FOXP1 and FOXP2 FHD dimer. This suggests that minor differences in sequence have a pronounced effect on the dimerization propensity of the FHD of these proteins which in turn could affect the manner in which they interact with DNA and regulate transcription. It is clear that the most critical sequence variation is located in the hinge loop connecting helices 2 and 3. A single residue substitution in the FOXP2 FHD, Y540F, for example, was shown to be sufficient to increase the propensity of the FOXP2 FHD for dimerization quite substantially.[27]

In this work, we have exploited the effect of the hinge loop on dimerization propensity by successfully creating two hinge

DOI: 10.1021/acs.biochem.5b01271
Biochemistry 2016, 55, 1959−1967

loop mutants that either inhibit or promote dimerization. Comparison of these mutants with the wild-type suggests that at concentrations under 300 μM, the wild-type FOXP2 FHD exists almost exclusively as a monomer (Figure 2). It is possible that dimerization of the FOXP2 FHD does occur but is difficult to detect due to the isolated dimer being unstable. Indeed this is confirmed by the disulfide-linked dimeric mutant reported here (Figure 2). The presence of disulfide linked FOXP2 FHD dimers provides evidence to suggest that transient dimerization events do occur in vitro in the absence of DNA even if dimers are not detected in the wild-type (Figure 2). The question that stems from this observation is whether the presence of DNA might stabilize the dimeric form of the FOXP2 FHD. This idea arises from the fact that the pH and ionic strength is different in the vicinity of DNA strands compared to in the cytoplasm. Indeed the FOXP2 FHD has been shown to adopt a less compact structure harboring exposed hydrophobic patches at the more acidic pH that would be expected surrounding DNA.[19] This could be necessary for crucial dimer interface contacts to occur. Studies in the presence of DNA, however, suggest that in addition to being monomeric in the absence of DNA, the wild-type FOXP2 FHD also binds to DNA as a monomer. This is implied by the single band in the EMSA (Figure 3), a FA binding isotherm identical to that of the monomeric mutant (Figure 4) and a stoichiometry of binding of 1 (Figure 5). These results are in support of the notion that the leucine zipper is required for the formation of FOXP2 FHD dimers capable of binding DNA.[5] A more in depth analysis of the DNA binding of wild-type in comparison to the monomeric mutant, however, reveals evidence to suggest that despite the apparent monomeric nature of the wild-type, the way in which the wild-type interacts with DNA is not identical to the way in which the monomer interacts with DNA.

Binding of DNA by proteins excludes water and associated counterions from the interface surface releasing them into bulk solvent, a process generally accompanied by a large increase in entropy of the system. The amount of water molecules excluded depends on whether the protein binds to the major or minor groove of the DNA, and this is readily determined by the thermodynamic profile obtained using isothermal titration calorimetry. Binding of a protein to the minor groove is generally driven by entropic gains produced by the release of a double water spine (located in the minor groove of DNA) upon binding that counteracts the enthalpic costs of the large structural changes that are required for residue insertion into the narrow space.[25] Conversely, DNA binding by major groove binding proteins is typically enthalpically driven by the formation of numerous base-residue hydrogen bonds in the larger major groove.

It can be seen that DNA binding of the wild-type FOXP2 FHD is largely enthalpically driven (ΔH = −24.46 ± 0.81 kJ/mol) as well as entropically favorable ($T\Delta S$ = 15.80 ± 0.70 kJ/mol/K) (Figure 5). It is thus clear that the FOXP2 FHD is a major groove sequence specific binding protein which agrees with the crystal structure of the monomer−DNA complex.[13] Entropy is generally regarded as the main driving factor behind the non-sequence specific component of DNA binding since it is associated with expulsion of counterions and water from the surface of the protein/DNA as well as conformational changes that occur upon complex formation. Interestingly there is a difference between the entropic parameter obtained for the wild-type and the A539P FOXP2 FHD, which suggests a difference in the mode of binding. Considering that the enthalpy of binding is similar, within error, it is likely that the increased entropy observed for the wild-type is the result of conformational changes that occur upon binding that are not permitted in the mutant due to the reduced flexibility in the A539P hinge loop. The increased flexibility of the wild-type compared to the monomer is also supported by the thermal unfolding curves (Figure 2E). The wild type unfolds in a less sigmoidal fashion than the monomeric mutant, indicative of a less cooperative transition to the unfolded state, implying that the wild-type native state is less folded and more flexible than that of the mutant. The increased range of motion inherent in the wild-type FOXP2 could allow for more nonspecific DNA backbone contacts to form, resulting in the observed increased affinity for the DNA (K_D = 66.94 ± 4.34 nM). This means that there is the potential for the wild-type FOXP2 FHD to find and bind sites different in sequence to the cognate recognition site (TGTTTAC). Future studies conducted with multiple sequences would be required to confirm this. Conversely, the presence of the helix disrupting proline in the hinge loop of the A539P FOXP2 FHD mutant appears to reduce DNA binding affinity (K_D = 102.3 ± 4.97 nM) as well as the entropy associated with binding ($T\Delta S$ = 11.63 ± 0.84 kJ/mol/K). The increased rigidity of the hinge loop stabilizes the monomeric form of the FOXP2 FHD reducing conformational changes induced by DNA binding including those that allow the wild-type to make stronger interactions with the DNA backbone.

Many eukaryotic transcription factors are intrinsically flexible and refold into a more stable form upon identifying and binding a particular DNA sequence.[28,29] This phenomenon, known as fly casting, allows TFs to scan large stretches of DNA sequence in a less defined structure and only form a stable complex at the sites that supply the necessary contacts for the protein to fold into a more stable structure. The high flexibility of the wild-type as understood by the energetics and thermal unfolding data presented here suggest that the wild-type FOXP2 FHD may be capable of fly casting. Another technique used by TFs in sequence identification is dynamic DNA readout whereby distant, intrinsically flexible protein motifs alter the dynamics that occur at the protein−DNA interface, facilitating the necessary DNA contacts for stable complex formation.[30] It is possible that the FOX FHD hinge loop sequence in part can control the dynamic DNA readout of the FHD allowing the highly similar FOX family members to identify and bind different DNA sequences. This mechanism helps to explain the difference in binding observed in the titration studies presented in this paper. The reduced entropy of binding observed in the A539P FOXP2 FHD may be due to restraint placed on flexibility of the hinge loop and hence the DNA contacting amino acids. Further protein dynamics studies in the absence and presence of DNA are necessary to support this.

Taken together, the hinge loop between helices 2 and 3 plays a key role in regulating the sequence specificity and DNA binding affinity of the FHD. The P539A mutation present in the FOXP subfamily could reduce the restraint on the FHD structure allowing it to present hydrophobic regions; a critical step in the dimerization of many domain swapping proteins. Further studies are required to prove this. The evidence provided here suggests that the composition of the hinge loop is a key regulator in both the DNA binding affinity and sequence-specificity of the forkhead domain.

DOI: 10.1021/acs.biochem.5b01271
Biochemistry 2016, 55, 1959−1967

CONCLUSION

The FOXP subfamily provides a unique system to study the evolution of dimerization in the FOX protein family and in transcription factors as a whole. In this paper we describe a covalent FOXP2 FHD dimer, F541C FOXP2 FHD, successfully engineered for use in comparative DNA binding studies. The F541C FOXP2 FHD is predominantly dimeric under nonreducing conditions, is correctly folded and is capable of binding DNA as both a monomer and dimer. While fluorescence anisotropy and electrophoretic mobility shift assays suggest that the wild-type FOXP2 FHD binds to DNA exclusively as a monomer; isothermal titration calorimetry the binding mechanism of the wild-type is not identical to that of the monomer. The greater affinity and entropy of binding of the wild-type compared to the monomer suggests that the wild-type has greater conformational freedom due to the presence of the flexible loop region. When in the presence of the leucine zipper in the full length protein, this increased conformational freedom in the FHD may promote dimerization upon DNA binding. Here we have shown that the evolutionarily important substitution of proline located in the FHD hinge loop with an alanine alters the DNA binding mechanism of the FOXP2 FHD, most likely by allowing it to adopt interactions with the DNA backbone not available to the structurally restrained monomeric FOX subfamily members.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Sylvia.fanucchi@wits.ac.za. Tel: +2711 717 6348. Fax: +2711 717 6351.

Funding

This work was supported by the University of the Witwatersrand, South African National Research Foundation Grant 80681 to S.F., Grant 68898 to H.W.D., the South African Research Chairs Initiative of the Department of Science and Technology Grant 64788 to H.W.D., and the Medical Research Council of South Africa.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank Mr. Mihai Silviu Tomescu for his assistance with setting up and optimization of the isothermal titration calorimetry experiments.

ABBREVIATIONS

TF, transcription factor; FHD, forkhead domain; FOX, forkhead box; CD, circular dichroism; DNA, DNA; DTT, dithiothreitol

REFERENCES

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

(2) 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.

(3) Weigel, D., and Jackle, H. (1990) The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell 63*, 455−456.

(4) 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.

(5) 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.

(6) 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.

(7) 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. 119* (Suppl 1), S197−S202.

(8) Lai, C. S., Gerrelli, D., Monaco, A. P., Fisher, S. E., and Copp, A. J. (2003) FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain 126*, 2455−2462.

(9) Bennett, C. L., and Ochs, H. D. (2001) IPEX is a unique X-linked syndrome characterized by immune dysfunction, polyendocrinopathy, enteropathy, and a variety of autoimmune phenomena. *Curr. Opin. Pediatr. 13*, 533−538.

(10) 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.

(11) 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.

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

(13) 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.

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

(15) Rousseau, F., Schymkowitz, J. W., Wilkinson, H. R., and Itzhaki, L. S. (2001) Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved prolines. *Proc. Natl. Acad. Sci. U. S. A. 98*, 5596−5601.

(16) Bandukwala, H. S., Wu, Y., Feuerer, M., Chen, Y., Barboza, 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.

(17) Craig, D. B., and Dombkowski, A. A. (2013) Disulfide by Design 2.0: a web based tool for disulfide engineering in proteins. *BMC Bioinf. 14*, 346.

(18) Larkin, M. A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics 23*, 2947−2948.

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

(20) 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.

(21) Lacowicz, J. R. (2002). *Topics in Fluorescence Spectroscopy*, 1st ed.; Kluwer Acedemic Publishers: New York.

(22) Chu, Y. P., Chang, C. H., Shiu, J. H., Chang, Y. T., Chen, C. Y., and Chuang, W. J. (2011) Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: insight into its domain swapping and DNA binding. *Protein Sci. 20*, 908−924.

(23) 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.

(24) Littler, D. R., Alvarez-Fernandez, M., Stein, A., Hibbert, R. G., Heidebrecht, T., Aloy, P., Medema, R. H., and Perrakis, A. (2010)

DOI: 10.1021/acs.biochem.5b01271
Biochemistry 2016, 55, 1959−1967

Structure of the FoxM1 DNA-recognition domain bound to a promoter sequence. *Nucleic Acids Res. 38*, 4527−4538.
(25) Privalov, P. L., Dragan, A. I., Crane-Robinson, C., Breslauer, K. J., Remeta, D. P., and Minetti, C. A. S. A. (2007) What drives proteins into the major or minor groove of DNA? *J. Mol. Biol. 365*, 1−9.
(26) Dragan, A. I., Read, C. M., Makeyeva, E. N., Milgotina, E. I., Churchill, M. E., Crane-Robinson, C., and Privalov, P. L. (2004) DNA binding and bending by HMG boxes: energetic determinants of specificity. *J. Mol. Biol. 343*, 371−393.
(27) 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.
(28) Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) Intrinsic disorder in transcription factors. *Biochemistry 45*, 6873−6888.
(29) Levy, Y., Onuchic, J. N., and Wolynes, P. G. (2007) Fly-Casting in protein-DNA binding: Frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc. 129*, 738−739.
(30) Fuxreiter, M., Simon, I., and Bondos, S. (2011) Dynamic protein-DNA recognition: beyond what can be seen. *Trends Biochem. Sci. 36*, 415−423.

DOI: 10.1021/acs.biochem.5b01271
Biochemistry 2016, 55, 1959−1967

A key evolutionary mutation enhances DNA binding of the FOXP2 forkhead domain

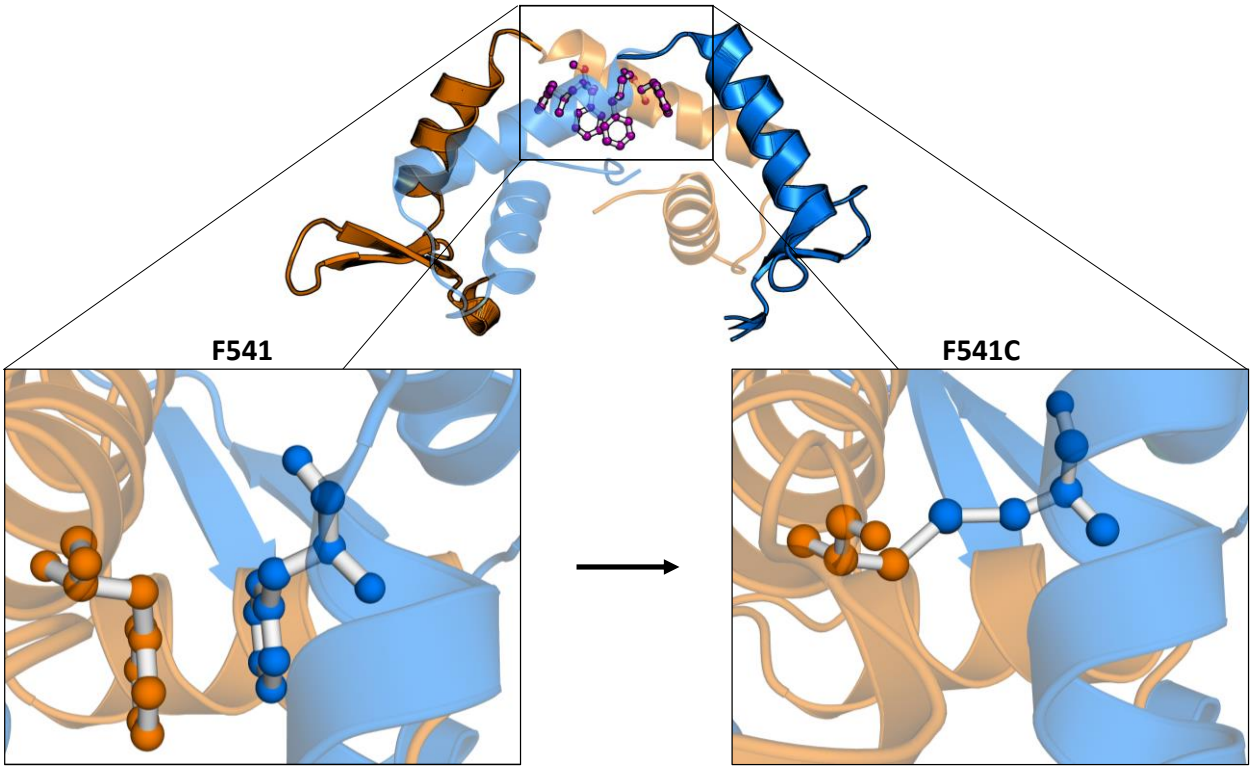
GAVIN MORRIS[1] AND SYLVIA FANUCCHI[1]

[1] Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg 2050. South Africa

Supplementary Material

-Table of Contents-

Item	Description
Supplementary Fig 1.	Representation of the proposed F541C mutation in the hinge loop of the FOXP2 forkhead domain.
Supplementary Fig 2.	Sequence data for the wild-type, A539P and F541C FOXP2 forkhead domain and the encoded protein product.
Supplementary Fig 3.	SDS-PAGE analysis of wild-type, A539P and F541C FOXP2 forkhead domain variant purity following chromatographic separation from bacterial milieu.



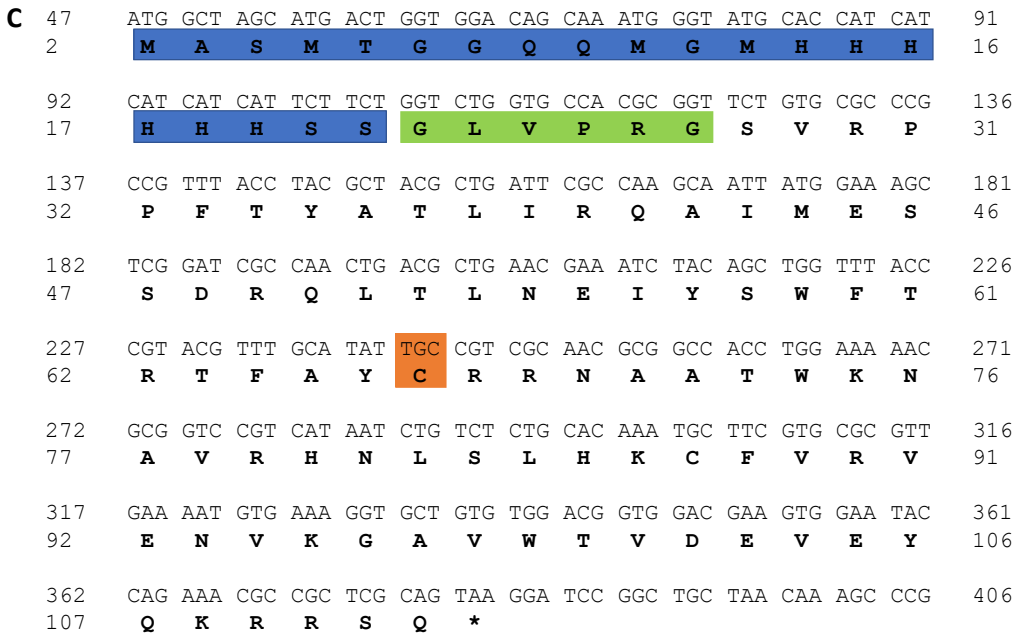
Supplementary Fig 1. Representation of the proposed F541C mutation in the hinge loop of the FOXP2 forkhead domain. A conserved Phenylalanine (F541) lies on the 2-fold axis of symmetry in the domain-swapped FOXP2 forkhead domain. In the domain-swapped dimer adjacent F541 residues form pi-stacking interactions in the dimer interface. Substitution of F541 with a cysteine residue has the potential to form a disulphide bond covalently linking the associated FOXP2 forkhead domain subunits preventing dissociation into the monomeric quaternary structure.

A

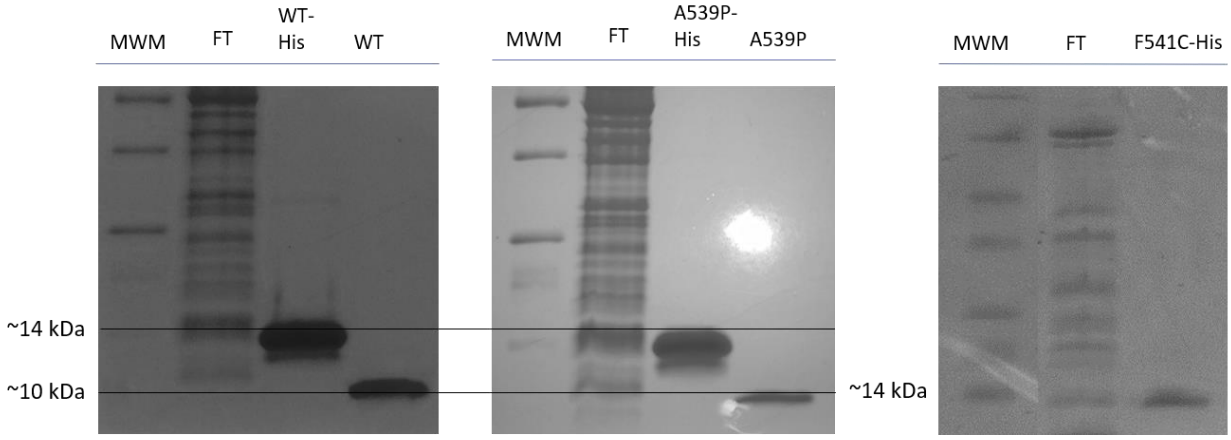
```
46    ATA CAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT ATG CAC   90
            M   A   S   M   T   G   G   Q   Q   M   G   M   H
91    CAT CAT CAT CAT CAT TCT TCT GGT CTG GTG CCA CGC GGT TCT GTG   135
15    H   H   H   H   H   S   S   G   L   V   P   R   G   S   V   29

136   CGC CCG CCG TTT ACC TAC GCT ACG CTG ATT CGC CAA GCA ATT ATG   180
30    R   P   P   F   T   Y   A   T   L   I   R   Q   A   I   M   44

181   GAA AGC TCG GAT CGC CAA CTG ACG CTG AAC GAA ATC TAC AGC TGG   225
45    E   S   S   D   R   Q   L   T   L   N   E   I   Y   S   W   59

226   TTT ACC CGT ACG TTT GCA TAT TTC CGT CGC AAC GCG GCC ACC TGG   270
60    F   T   R   T   F   A   Y   F   R   R   N   A   A   T   W   74

271   AAA AAC GCG GTC CGT CAT AAT CTG TCT CTG CAC AAA TGC TTC GTG   315
75    K   N   A   V   R   H   N   L   S   L   H   K   C   F   V   89

316   CGC GTT GAA AAT GTG AAA GGT GCT GTG TGG ACG GTG GAC GAA GTG   360
90    R   V   E   N   V   K   G   A   V   W   T   V   D   E   V   104

361   GAA TAC CAG AAA CGC CGC TCG CAG TAA GGA TCC GGC TGC TAA CAA   405
105   E   Y   Q   K   R   R   S   Q   *                           113
```

B

```
46    ATA CAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT ATG CAC   90
            M   A   S   M   T   G   G   Q   Q   M   G   M   H
91    CAT CAT CAT CAT CAT TCT TCT GGT CTG GTG CCA CGC GGT TCT GTG   135
15    H   H   H   H   H   S   S   G   L   V   P   R   G   S   V   29

136   CGC CCG CCG TTT ACC TAC GCT ACG CTG ATT CGC CAA GCA ATT ATG   180
30    R   P   P   F   T   Y   A   T   L   I   R   Q   A   I   M   44

181   GAA AGC TCG GAT CGC CAA CTG ACG CTG AAC GAA ATC TAC AGC TGG   225
45    E   S   S   D   R   Q   L   T   L   N   E   I   Y   S   W   59

226   TTT ACC CGT ACG TTT CCA TAT TTC CGT CGC AAC GCG GCC ACC TGG   270
60    F   T   R   T   F   P   Y   F   R   R   N   A   A   T   W   74

271   AAA AAC GCG GTC CGT CAT AAT CTG TCT CTG CAC AAA TGC TTC GTG   315
75    K   N   A   V   R   H   N   L   S   L   H   K   C   F   V   89

316   CGC GTT GAA AAT GTG AAA GGT GCT GTG TGG ACG GTG GAC GAA GTG   360
90    R   V   E   N   V   K   G   A   V   W   T   V   D   E   V   104

361   GAA TAC CAG AAA CGC CGC TCG CAG TAA GGA TCC GGC TGC TAA CAA   405
105   E   Y   Q   K   R   R   S   Q   *                           113
```



Supplementary Fig 2. Sequence data for the (A) wild-type, (B) A539P and (C) F541C FOXP2 forkhead domain and the encoded protein product (second line). All protein were expressed as a fusion protein with an N-terminal hexa-histidine tag (blue) and a thrombin cleavage site (green). The mutations were successfully incorporated using site-directed mutagenesis (purple and orange).



Supplementary Fig 3. SDS-PAGE analysis of wild-type, A539P and F541C FOXP2 forkhead domain variant purity following chromatographic separation from bacterial milieu. Cleavage of the hexa-histidine tag was performed by incubating the purified protein for 4 hours at 20 ^{o}C in 10 mM Tris pH 8.0, 100 mM NaCl with 1 U bovine thrombin. Thrombin and uncleaved fusion protein were removed by sequential immobilised-nickel affinity and benzamidine affinity chromatography.

Chapter 3

The hinge loop plays a pivotal role in DNA binding and transcriptional activity of FOXP2

Gavin Morris, Stoyen Stochev, Heini W. Dirr and Sylvia Fanucchi

(Biological Chemistry, in press)

Motivation: The data obtained in chapter 2 revealed a significant difference in the thermodynamic signature of DNA binding between the wild-type and A539P FOXP2 forkhead domain, however, the inherent low-resolution nature of thermodynamic data prompted further investigation to determine the factors that contribute to the difference. Furthermore, the effect that the A539P mutation has on the *in vivo* function of FOXP2 has not been explored.

Summary: The thermodynamic components of the signatures obtained, from chapter 2, were dissected by determining the change in heat capacity and salt concentration dependence of the protein-DNA interaction for both the wild-type and A539P FOXP2 forkhead domain. In addition, changes in the backbone dynamics and solvent accessibility of the proteins were assessed to determine structural changes in the two proteins during DNA binding. Finally, the transcriptional activity of FOXP2 was assessed using an *in vivo* luciferase reporter assay to determine the effect the A539P mutation has on FOXP2 in the cellular environment.

A significant difference in the heat capacity changes and salt dependence was observed between the wild-type and A539P FOXP2 forkhead-DNA interactions suggesting that the mutation alters the forkhead domains capacity to form hydrophobic, hydrogen bonding and coulombic interactions with a consensus DNA sequence. In addition, the structural dynamics studies suggest a significant conformational adjustment occurs in the wild-type FOXP2 forkhead domain that does not occur in the A539P mutant. Finally, the mutation significantly altered the transcriptional activity of FOXP2. Together these data suggest the hinge loop not only controls the dimerisation of the forkhead domain but also alters the mechanism by which the proteins recognise and interact with the consensus binding site. Further, the oligomeric state of the FOXP2 forkhead domain regulates transcriptional activity in the cellular environment.

Author contributions: Gavin Morris performed all experimental work, analysed data and wrote the manuscript. Stoyen Stoychev assisted with the setup and data analysis of the hydrogen-deuterium exchange mass spectrometry experiments. Heini W. Dirr assisted with experimental design. Sylvia Fanucchi was the PI that supervised the project and assisted in data analysis and interpretation.

DE GRUYTER Biological Chemistry

The forkhead domain hinge-loop plays a pivotal role in DNA binding and transcriptional activity of FOXP2

Journal:	*Biological Chemistry*
Manuscript ID	BIOLCHEM-2018-0185.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	26-Apr-2018
Complete List of Authors:	Morris, Gavin; University of the Witwatersrand, School of molecular and cell biology Stoychev, Stoyan; Council for Scientific and Industrial Research Naicker, Previn; Council for Scientific and Industrial Research Dirr, Heini; University of the Witwatersrand, School of molecular and cell biology Fanucchi, Sylvia; University of the Witwatersrand, School of molecular and cell biology; University of the witwatersrand
Section/Category:	Protein Structure and Function
Keywords:	Forkhead box, FOXP, DNA-binding, hinge loop, backbone dynamics

SCHOLARONE™
Manuscripts

Reply to Reviewer's comments for the manuscript entitled "The forkhead domain hinge-loop plays a pivotal role in DNA binding and transcriptional activity of FOXP2"

We thank both reviewers for their insightful and thoughtful analysis of our work.

The reviewer's comments are given below and our response to each point is indicated in blue underneath each of the reviewers' comments. In the manuscript we have highlighted changes to the text to accommodate the reviewers' suggestions in yellow.

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author
The FOXP members of FOX transcription factors are unusual in that they can form domain-swapped dimers via their Forkhead domains (FHDs) in addition to dimerization via the leucine zipper regions. In the present manuscript, Morris et al. analyzed molecular consequences of a FOXP-specific residue exchange in the FHD hinge loop, P539A, that seems to correlate with the ability to form domain-swapped FHD dimers, using ITC, H/D exchange MS and fluorescence quenching analyses. ITC analyses were performed at different temperatures and at different salt concentrations to analyze the heat capacity change and the importance of electrostatics upon DNA binding. The authors also employed a dual luciferase assay to assess the consequences of the FHD residue exchange on transcriptional activity. The authors find that DNA binding energetics, dynamics and recognition features are changed due to the FOXP-specific hinge loop sequence.

Based on their analyses, the authors conclude that wt FOXP binds DNA in a less sequence-specific manner and depends more on electrostatic interactions compared to the A539P mutant. H/D exchange experiments indicated that the A539P mutant even in isolation resembled the DNA-bound, less dynamic and solvent exposed form of the wt FHD. These altered dynamics were associated with a significant reduction in the ability of a full-length version of the A539P mutant to stimulate transcription in a luciferase reporter assay.

The results are novel and interesting and are generally well presented. The work appears to be technically sound.

Specific comments:

1. Dimerization of FOX transcription factors depends primarily on the leucine zipper, which is present in both wt and A539P mutant FOXP. Moreover, the wt as well as the A539P mutant FHD appear to be largely monomeric under the chosen experimental conditions (p. 11). Thus, the main difference between the FHDs may be in the altered dynamics rather than in the presence or absence of a FHD dimer in isolation. While the domain swap may occur when the A539P mutant FHD binds DNA, the authors should avoid the confusing term "monomeric mutant".

We have replaced the term monomeric mutant with the term A539P FOXP2 FHD throughout the paper

2. The authors conclude that binding of wt FHD to DNA is less sequence specific than for the A539P mutant. It would be gratifying to see this suggestion tested experimentally. The authors may use their

For Review Only

dual luciferase reporter assay to this effect. Alternatively, they may resort to and discuss existing data concerning the sequence specificity of FOXP and other sub-family members.
Unfortunately we do not have the resources to test sequence specificity experimentally at this stage although, indeed, it would be interesting and satisfying to do so. Instead, we have done as the reviewer suggested and included a paragraph in the discussion (highlighted in yellow) on why this mutation can lead to sequence specificity differences based on our existing data as well as on other data reported in the literature.

Reviewer: 2

Comments to the Author
In the manuscript (BIOLCHEM-2018-0185) titled "The forkhead domain hinge-loop plays a pivotal role in DNA binding and transcriptional activity of FOXP2", Morris et al have studied an evolutionary mutation (P539A) that differentiates FOXP subfamily of transcription factors from the rest of the FOX proteins. They conducted ITC and HDXMS experiments and supplemented these studies with some in vivo data using luciferase reporter assays. The manuscript is well written. Here are some minor comments:

1. Introductory paragraphs describe the domain organization and location of the mutation studied. It may be a good idea to include a figure/schematic showing the domain organization (leucine zipper domain, forkhead domain), domain swapping, helix 2, hinge loop, the location of mutation etc to make it easier for the readers to understand better the background of this work.
   This is now included as figure 1 in the introduction.
2. For ITC data, I suggest that the original isotherms be provided in the supplementary material.
   We have included these in the supplementary material as requested (Figures 2-5)
3. Is the protein-DNA complex shown in the figures a theoretical model or crystal structure? Mention the PDB ID in the figure legend if it is a crystal structure or NMR structure.
   It is a crystal structure and it has now been made clear in the figure legend of figure 4
4. Specify the strain of bacteria in the materials and methods section (Page 18, line 32).
   This has been included on line 1 of the protein expression and purification section of the materials and methods
5. Specify the buffer for protein purification in the materials and methods section.
   This has been included
   Provide a PAGE image of purified protein along with molecular weight markers in the same gel to be included in the supplementary material.
   This is included in the supplementary material as figure 1
6. Page 19, line 6, is the symbol used for degree Celsius correct?
   This has been corrected
7. Specify the buffer used to dissolve the oligonucleotides.
   This has been done
8. On page 20, line 20, explain the various terms used in the equation.
   This has been explained in subsequent paragraph
9. On page 20, line 25, "performed at in binding buffer". Is something missing between at and in?
   The temperature was missing and has now been added

For Review Only

For Review Only

The forkhead domain hinge-loop plays a pivotal role in DNA binding and transcriptional activity of FOXP2

Gavin Morris[1], Stoyan Stoychev[2], Previn Naicker[2], Heini W. Dirr[1] and Sylvia Fanucchi[1,*]

[1]Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of the Witwatersrand, 1 Jan Smuts Ave, Braamfontein, 2050 Johannesburg, Gauteng, South Africa

[2]CSIR Biosciences, CSIR, Meiring Naude Road, Brummeria, 0001 Pretoria, Gauteng, South Africa

*Corresponding author
e-mail: sylvia.fanucchi@wits.ac.za

<Running title: The hinge-loop regulates DNA binding and dynamics of FOXP2>

For Review Only

Abstract

FOX proteins are a ubiquitously expressed family of transcription factors that regulate development and differentiation of a wide range of tissues in animals. The FOXP subfamily members are the only known FOX proteins capable of forming domain-swapped forkhead domain dimers. This is proposed to be due to an evolutionary mutation (P539A) that lies in the forkhead domain hinge loop, a key region thought to fine-tune DNA sequence specificity in the FOX transcription factors. Considering the importance of the hinge loop in both the dimerisation mechanism of the FOXP forkhead domain and its role in tuning DNA binding, a detailed investigation into the implications of mutations within this region could provide important insight into the evolution of the FOX family. ITC and HDXMS were used to study the thermodynamic binding signature and changes in backbone dynamics of FOXP2 forkhead domain DNA binding. Dual luciferase reporter assays were performed to study the effect that the hinge-loop mutation has on FOXP2 transcriptional activity *in vivo*. We demonstrate that the change in dynamics of the hinge-loop region of FOXP2 alters the energetics and mechanism of

DNA binding highlighting the critical role of hinge loop mutations in regulating DNA binding characteristics of the FOX proteins.

Keywords: backbone dynamics; DNA-binding; electrostatics; forkhead; FOXP; hinge-loop.

Introduction

The forkhead box (FOX) family is a prominent transcription factor family responsible for the differentiation and growth regulation of a large variety of cell types across multiple vertebrate taxa. FOX transcription factors play a role in numerous processes including development and differentiation, cell cycle regulation and homeostasis (Hannenhalli and Kaestner, 2009). The FOX family of transcription factors is identified by a highly conserved DNA binding domain, the forkhead domain (FHD) which is a variant of a helix-turn-helix motif termed a winged helix (Clark et al., 1993; Gajiwala and Burley, 2000). FOX family members are subdivided into 19 (A-S) subfamilies based on sequence divergence from the forkhead domain and homology of the flanking sequences (Lai et al., 1993; Kaestner et al., 2000). All the FOX proteins are multidomain proteins and show substantial deviation in the number and type of domains flanking the forkhead domain. The FOXP subfamily (FOXP1-4) is characterised by a leucine zipper and zinc finger domain located N-terminal to the forkhead domain. FOXP1, 2 and 4 also possess a long N-terminal polyglutamine tract of unknown function (Lai et al., 2001).

FOXP2, the first gene product to be associated with language acquisition and use, plays a vital part in the development of brain regions necessary for complex language comprehension (Lai et al., 2001). Loss-of-function mutations in the FOXP2 FHD cause a congenital mental retardation that presents as a difficulty in fine motor control of the orofacial muscles required for complex language production (orofacial dyspraxia) as well as reduced language cognition (Fujita et al., 2008). FOXP2 also plays a significant role in non-communicable disease. Aberrant expression of FOXP2 has been observed in breast cancer, diffuse large B-cell lymphoma and multiple myeloma (Cuiffo et al., 2014; Wong et al ., 2016). FOXP2 has also been implicated in schizophrenia, autism and most recently Huntington's disease (Gong et al., 2004; Spaniel et al., 2011; Hachigan et al., 2017).

For Review Only

Three of the four FOXP subfamily members (FOXP1/2/4) share overlapping regions of expression in the developing and adult vertebrate brain (Teramistu, 2004; Takahashi et al., 2009) As a result of this overlapping expression pattern as well as the presence of a conserved leucine zipper domain, the FOXP subfamily is unique among FOX transcription factors in its ability to form both homo- and heterotypic dimers via the leucine zipper domain (Li et al., 2004). Dimerisation of the FOXP subfamily members has been shown to be essential *in vivo* and is responsible for the differential expression of a large array of target genes (Li et al., 2004; Sin et al., 2016). In addition to dimerisation through the leucine zipper, crystal structures reveal that the isolated FOXP FHD is capable of forming 3D-domain swapped dimers (Stroud et al., 2006). Domain swapping is an uncommon form of oligomerisation whereby a portion of the associating domains is exchanged with the equivalent region in the oligomerisation partner. This is usually facilitated through a "hinge-loop" - a random coil between the exchanged and fixed regions of the monomer (Rousseau et al., 2006). The FOXP FHD has a far greater propensity to form a domain swapped dimer than the forkhead domain of other members of the FOX family. This is attributed to a single proline to alanine substitution in the hinge-loop between helix 2 and 3 that promotes extension of helix 2 into the hinge-loop (Stroud et al., 2006). Indeed, reversal of this hinge-loop mutation successfully prevents dimerisation of the FOXP FHD (Stroud et al., 2006; Chu et al., 2011).

All FOX FHD structures reveal that the forkhead domain interacts with DNA via helix 3 – the recognition helix – which inserts into the major groove (Clark et al., 1993; Stroud et al., 2006). Despite almost complete sequence conservation within the recognition helix of the forkhead domain as well as high conservation of the recognition DNA sequence (TA/GTTT/GG/AA/GT/C), the FOX family members display a large deviation in target gene activity (Pierrou et al., 1994) Furthermore the FOX FHD has been shown to bind to more than one DNA sequence with varying rates and affinity and sequence specificity is believed to be a form of transcriptional regulation (Webb et al., 2017) The region of the FOX FHD that is considered most likely to contribute to sequence specificity is the region immediately prior to the recognition helix consisting of helix 2 and the less highly conserved hinge-loop region (Figure 1) (Pierrou et al., 1994; Overdier et al., 1994; Marsden et al., 1998).

It is interesting, from both an evolutionary and functional perspective that of all the FOX proteins, (of which there are over 100) only the four members of the FOXP subfamily have been discovered to have developed the ability to form forkhead domain dimers. This unique

https://mc.manuscriptcentral.com/bc
For Review Only

ability is entirely due to the amino acid sequence within the hinge-loop region and has lead us to question the significance of this region in FOX protein function. Indeed, mutations in the hinge-loop of FOXP3 that have been shown to reduce dimer propensity have also been associated with disease (Bandukwala et al., 2011). Does this mean that the formation of forkhead domain dimers specifically (over and above dimers formed via the leucine zipper) is significant for FOXP function? In this work we aim to determine the role that the FOXP proline to alanine hinge-loop mutation has on the DNA binding mechanism of the FOXP2 FHD. This study will provide insight into the complexities of transcriptional regulation via FOXP proteins and will distinguish between proteins with an exclusively monomeric forkhead domain and those with the potential to form domain swapped dimers, hopefully providing a clue as to whether the FOXP subfamily has evolved to develop a unique mode of action compared to the other members of the family.

Results

In order to compare the DNA binding mechanism of a FOX FHD that has the potential to form a dimer to one that does not, we have worked with both the wild-type FOXP2 FHD and the exclusively monomeric A539P hinge-loop mutant. The thermodynamics of DNA binding has been investigated in this study at different temperatures and at different salt concentrations to obtain information on the heat capacity and the electrostatics of binding respectively. The dynamics and flexibility of wild-type and A539P mutant upon DNA binding has been studied so as to assess whether restrictions on backbone flexibility affect DNA binding. Finally a luciferase reporter assay was also performed so as to understand the effect of the monomeric mutation on FOXP2 transcriptional ability and hence to provide physiological relevance to the *in vitro* work.

Heat capacity of DNA binding

The heat capacity of biochemical interactions provides insight into the arrangements and types of bonds involved in the interaction. When DNA binds to a protein, a number of factors will contribute to the change in heat capacity (ΔC_P) upon DNA binding. These factors include the structural changes that may accompany DNA binding, changes in ionic shielding as well as changes in the solvation of the protein and DNA. Usually, sequence-specific DNA binding is

https://mc.manuscriptcentral.com/bc

For Review Only

accompanied by a large negative heat capacity change, while non sequence specific binding does not necessarily yield a large negative value.

Temperature dependant DNA binding studies of the wild-type FOXP2 FHD and the A539P mutant show significant differences (Figure 2). The wild-type FOXP2 FHD has an unusually small ΔC_p upon binding DNA of -0.88 ± 0.05 kJ/mol/K. This is nearly half of that obtained for the A539P FOXP2 FHD mutant where ΔC_p = -1.56 ± 0.04 kJ/mol/K. This suggests that the binding of the wild-type is less sequence specific than the A539P mutant and may involve more backbone rather than base interactions. The data for both the wild-type and A539P FOXP2 FHD agree well with the linear model fitted for determining heat capacity indicating that no additional temperature-dependent equilibria with a significant enthalpy change occur across the temperature range of 10-30 ^{o}C.

Salt studies of binding

The polyelectrolytic nature of DNA results in the formation of steep counter-ion concentration gradients around the phosphate backbone, a process known as counter-ion condensation (Manning, 1978; Record et al., 1978; Privalov et al., 2011). Exclusion of counter-ions from the DNA backbone during protein binding to the DNA results in a measurable cratic entropy of mixing that can be used to dissect the total entropic term into salt-dependent (electrostatic) and salt-independent (non-electrostatic) components. This analysis can be performed by fitting a linear regression to the double log plot of the observed association constant ($\log K_a$) as a function of salt concentration (log[S]) (Figure 3A) (Privalov et al., 2011). The affinity constant at 1 M salt (K_{nel}) is presumed to be a measure of the salt independent (non-electrostatic) component of the free energy of binding (Privalov et al., 2011). The total entropy change upon binding can then be dissected into electrostatic (ΔS_{el}) and non-electrostatic (ΔS_{nel}) components. The number of counter-ions (N) excluded from the DNA backbone during formation of the protein-DNA interface can be obtained from the gradient of the linear regression by considering the DNA phosphate counter-ion occupancy of NaCl (ψ = 0.64) (Olmsted et al., 1995; Privalov et al., 2011).

As with the heat capacities, there is a difference in the DNA binding electrostatics of the wild-type FOXP2 FHD compared to the A539P mutant. The wild-type displays a significant dependence on electrostatic interactions with most of the entropy originating from the electrostatic component ($T\Delta S_{el}$ = 14.64 ± 0.38 kJ/mol) while the non-electrostatic contribution to the entropy, ($T\Delta S_{nel}$) is -1 ± 0.38 kJ/mol (Figure 3B). The A539P mutant FOXP2 FHD has a smaller $T\Delta S_{el}$ component than the wild-type with a value of 7.53 ± 0.16 kJ/mol and a lower

For Review Only

$T\Delta S_{nel}$of -4.11 ± 0.16 kJ/mol (Figure 3B). As would be expected following these results, the wild-type FOXP2 FHD makes a higher number of ionic contacts with the DNA backbone than the does the A539P mutant. The wild-type forms approximately 4 (N = 2.58 ± 0.27) ionic contacts with the DNA while the mutant forms approximately 2-3 (N = *1*.67 ± 0.028).

Dynamics of DNA binding

Since differences in the thermodynamic DNA binding signatures were observed between the wild-type FOXP2 FHD and the A539P mutant, and particularly since a difference in the conformational entropy component was found, hydrogen-deuterium exchange experiments were performed in the absence and presence of DNA so as to identify the regions of the forkhead domain responsible for the observed differences (Figure 4).

The wild-type FOXP2 FHD displays a decrease in backbone solvent accessibility upon binding to DNA in regions presumed to shielded by the protein-DNA interface (helix-3 and wing-1) as well as, somewhat surprisingly, in regions distal from the interface (helix-1 and helix-2). Interestingly, the A539P FOXP2 FHD resembles the less dynamic, less solvent exposed DNA-bound form of the wild-type whether it is in the presence or absence of DNA suggesting that the flexibility of the hinge-loop plays a significant role in the global dynamics of the forkhead domain.

Since the wild-type showed a change in backbone dynamics upon DNA binding, fluorescence quenching studies were performed in the presence of a collisional quencher, in order to assess whether the changes in backbone dynamics are as a result of significant refolding events. Stern-Volmer constants were determined for both the wild-type and A539P FOXP2 FHD in the absence and presence of DNA (Figure 5). Despite only the wild-type showing a change in backbone dynamics upon binding, both proteins showed a significant reduction in the Stern-Volmer constants upon binding to DNA implying a change in the local environment of the tryptophan residues. The FOXP2 FHD contains three tryptophan residues; two flanking the hinge loop (Trp533 and Trp548) and a third within the loop connecting the recognition helix and the β-sheet wing (Trp573). The largest contributor to the observed decrease in the Stern-Volmer constant upon binding is likely Trp573 which becomes entrenched in the protein-DNA interface. The higher solvent accessibility of the tryptophan residues in the DNA bound form of the wild-type FOXP2 FHD compared to the mutant corroborates the findings of the deuterium exchange experiments suggesting that the wild-type binds in a more open conformation.

https://mc.manuscriptcentral.com/bc

For Review Only

Luciferase reporter assay

The FOXP subfamily members are the only FOX proteins known to form homo and heterotypic dimers. FOXP dimerisation can occur at two interfaces i.e. via the leucine zipper domain and also via the forkhead domain. Dimerisation through the leucine zipper domain has been shown to be necessary for correct FOXP2 function *in vivo* (Li et al., 2004). However, no studies have attempted to understand the role of only the FOXP2 FHD dimerisation *in vivo*. Indeed, since the domain swapped dimer has only been identified in crystal structures of the isolated forkhead domain, there is question as to whether dimerisation of the forkhead domain is physiologically significant or whether it is simply an artefact of the crystallisation conditions.

Luciferase reporter assays were performed with the full length proteins that therefore contain both the forkhead domain and the leucine zipper domain. The assays were performed with both wild-type and A539P mutant FOXP2 to assess whether the dimerisation of the forkhead domain has any detectable role in transcriptional activity (Figure 6). Overexpression of wild-type and A539P mutant FOXP2 resulted in an approximately 7-fold and 5-fold induction of luciferase activity, respectively. Thus disruption of forkhead domain dimerisation via the A539P mutation significantly lowers transcriptional activity of FOXP2 but does not inactivate it (**$p<0.012$).

Discussion

The FOXP subfamily is unique among FOX proteins in its ability to form dimers. Particularly it has the capacity to form dimers at two interfaces: the leucine zipper and the forkhead domain. The focus of this work has been on the forkhead domain dimer. FOXP proteins are capable of forming forkhead domain dimers due to a proline to alanine substitution in the hinge-loop region of the forkhead domain that allows domain swapping to occur. Here we study the effect that this hinge-loop mutation has on the DNA binding mechanism of the protein in an attempt to obtain evidence that the FOXP subfamily has evolved a novel mechanism for DNA binding, recognition, specificity and hence transcriptional regulation. We have accomplished this by isolating the wild-type FOXP2 FHD and the A539P FOXP2 FHD mutant and comparing their interaction with DNA.

For Review Only

Changes induced upon DNA binding

In our previous work, we examined the structure and DNA-binding thermodynamics of the wild-type FOXP2 FHD and the A539P mutant (Prabhu and Sharp, 2005). We found that there is little structural difference between the wild-type and the A539P mutant. In fact, we only detected a low proportion of wild-type dimer, concluding that the apo wild-type existed almost entirely as a monomer under the conditions we used for the study (Prabhu and Sharp, 2005). Despite this, we found that there was a distinct difference between the wild-type and the A539P mutant in their thermodynamic DNA binding signatures (Prabhu and Sharp, 2005). In this work, we expand on this discovery and clearly establish that the source of this difference lies in the role of the hinge-loop in regulating DNA binding dynamics.

The hydrogen-deuterium exchange experiments (Figure 4) suggest that the wild-type FOXP2 FHD has a more flexible and solvent exposed backbone in the absence of DNA than the A539P mutant. There is a distinct change in wild-type backbone solvent accessibility upon binding DNA where the backbone becomes less solvent exposed. The A539P mutant, on the other hand, shows little change in solvent accessibility whether in the presence or absence of DNA and, in fact, in both instances it resembles the more rigid conformation of the DNA-bound wild-type. This is an interesting result because it implies that a reduction in the flexibility of the hinge-loop as introduced by the mutation causes a global reduction in the dynamics of the entire protein. It makes sense that a weak interaction network exists between the two subdomains that are exchanged during domain swapping of the FOXP2 FHD. The hinge-loop appears to control this network. Thus the reduction in backbone flexibility caused by the A539P mutation in the hinge-loop may prevent rearrangement of secondary and tertiary structural interactions that occur freely in the wild-type FOXP2 FHD. When the more dynamic wild-type binds to DNA, its backbone becomes more restricted and resembles that of the A539P mutant. However, the fluorescence quenching results indicate that the wild-type tryptophan residues remain more exposed to the solvent when in the presence of DNA than the mutants tryptophan residues (Figure 4). Considering that the location of Trp537 indicates that it is likely to be a direct reporter of events at the protein-DNA interface, the quenching results imply that the wild-type forkhead domain does not insert as deeply into the DNA major groove as the mutant. This could be because the inherent flexibility in the wild-type forkhead domain might reduce the amount of time it remains associated with the DNA, preventing it from inserting as deeply into the major groove as the more conformationally restricted A539P mutant does. Furthermore, Trp548 (which lies proximal to the hinge loop) is also likely to be a significant contributor to the difference in signal observed in the dynamic fluorescence

For Review Only

quenching study of the wild-type and mutant FOXP2 FHD (Figure 5). The hinge loop of the forkhead domain has been shown to exist in either a random coil or α-helical conformation depending on the subfamily under question (Marsden et al., 1998; Chu et al., 2011). The different conformations of the hinge loop are thought to govern the orientation of the recognition helix in the major groove of DNA, fine tuning specific DNA contacts between the recognition helix and the DNA bases (Marsden et al., 1998). The exchange of an alanine residue with proline in the hinge loop of FOXP2 could increase the solvent accessibility of Trp548 by restricting the dynamics of the hinge loop to a more open conformation. Considering the marked difference in solvent accessibility between the wild-type and mutant FOXP2 FHD observed in this study, it appears that the residues in the C-terminal portion of the hinge loop, particularly Ala539, have a significant role in determining the structural conformation of this important region and in doing so fine tune the DNA contacts made during sequence specific binding. Such an observation is corroborated by the large change in thermodynamic binding signatures seen in this study (Figures 1 and 2). The much larger reduction in tryptophan solvent accessibility observed in the A539P mutant, compared to the wild-type upon DNA binding, suggests that a conformational switch to a more concealed conformation occurs in the mutant hinge loop not seen in the wild-type. Such a conformational switch undoubtedly contributes to the observed difference in the ΔC_P of binding and provides some evidence that this residue is a key component in defining the DNA binding of the forkhead domain, perhaps acting as a small energetic checkpoint for sequence recognition and specificity.

Interactions formed with the DNA

It is largely accepted that the heat capacity derived from macromolecular interactions is due to hydration/dehydration of polar and non-polar residues. This has been shown to be true for protein-DNA interactions, although it must be noted that other contributing factors have also been suggested including binding coupled refolding of disordered regions (Murphy et al., 1992; Spolar and Record, 1994). The negative heat capacity values upon DNA binding of both wild-type and mutant (Figure 2) indicate solvation of polar groups and burial of hydrophobic groups at the interface. However, the larger value of ΔC_P obtained for the A539P mutant suggests that its DNA-protein interface could be larger possibly owing to a different orientation of the recognition helix in the DNA major groove (Liu et al., 2008). The less negative value for the wild-type suggests there are more polar residues (and hence more water molecules) at the interface of the wild-type with DNA which results in less freedom of movement and a lower

For Review Only

capacity to retain heat (Liu et al., 2008). The idea that the wild-type-DNA interface is more polar is corroborated by the salt study (Figure 3) which suggests that the wild-type forms a greater number of electrostatic interactions at the DNA interface than the mutant. The thermodynamics of binding also confirms this result. While the enthalpy of binding is more thermodynamically favourable for the mutant, the electrostatic entropy component is more favourable for the wild-type. This implies that although the A539P mutant has a larger binding interface and forms more sequence specific contacts with the DNA bases, the wild-type forms more ionic contacts with the DNA phosphate backbone, which are likely the cause of the higher DNA affinity reported for the wild-type (Morris and Fanucchi, 2016).

It is clear that both specific interactions with the DNA bases and non-specific interactions with the phosphate backbone occur upon DNA binding. Although we show here that the A539P mutant may have a higher specificity of binding than the wild-type (due to the decreased number of ionic contacts and greater heat capacity of binding), both wild-type and mutant do have some degree of specificity for their targets as they do not interact with a random sequence (Webb et al., 2017). Specific site recognition and formation of a stable protein-DNA complex relies on two requirements. One is hydrogen bonds formed between the residues of the protein recognition motif and the nitrogenous bases which contributes to the enthalpic binding term and change in heat capacity (direct readout). The other is shape complementarity and conformational adjustments to maximise electrostatic interactions between the DNA backbone and the surrounding protein structure through shape and charge complementarity (indirect readout) (Steffen et al., 2002). In agreement with the heat capacity studies (Figure 2), the entropic component associated with conformational changes is the minor contributing factor to the entropy of binding (Figure 3). Therefore in order to form site specific interactions with the DNA, both the wild-type and the A539P FOXP2 FHD depend on electrostatic and hydrogen bonds with the backbone of the DNA target sequence and flanking regions (direct readout) rather than large conformational changes. The NMR structures of the forkhead domain of FOXP1 and FOXO4 agree with this observation, having a stable folded structure in solution in the absence of DNA (Weigelt et al., 2001; Chu et al., 2011).

Chimeric protein and NMR studies have shown that the region spanning helix-2 and the hinge loop is known to be a significant contributor to the specificity of the forkhead domain (Pierrou et al., 1994; Overdier et al., 1994; Marsden et al., 1998). The data obtained in this study corroborates those findings by showing that the protein-DNA interactions that occur between the forkhead domain and the consensus binding site differ significantly due to a point mutation in the hinge loop region. The difference in the number and types of interactions has the

For Review Only

potential to alter the specificity of the forkhead domain by altering the depth of groove insertion (and so the number of base specific hydrogen bonds) as well as facilitating novel backbone ionic contacts that would otherwise be restricted by a rigid hinge loop.

It is interesting, considering the highly conserved nature of the recognition helix, that the various members of the FOX family are able to recognise different specific sequences. According to our results, it is possible that specificity is linked to backbone flexibility. One of the regions that appears to have a remarkable change in backbone flexibility upon binding of the wild-type to DNA is the wing region, located directly C-terminal to the recognition helix (Figure 4). Of all the regions in the FOX FHD, the β-sheet wing is one of the least conserved. Interestingly, the majority of known forkhead domain structures reveal that there are contacts made between the wing and the DNA largely between positively charged residues and the phosphate backbone (Stroud et al., 2006; Tsai et al., 2007; Litter et al., 2010). Considering both the flexibility and the variability of this region, these contacts may be key to distinguishing binding site selectivity between FOX family members. Perhaps the wing acts to regulate activity of the FOX proteins through protein-protein interactions with other transcription factors as observed in the RHR domain of NFAT and the FOXP2 FHD co-crystal structure (Wu et al., 2006). This could provide an additional mechanism to alter gene regulation by the various FOX transcription factors despite the highly conserved recognition sequence of the FOX family. Recently, several studies have aimed to disrupt the DNA binding of FOXM1 with specific small molecule compounds due to its role in the progression and survival of several cancers (Gormally et al., 2014; Marisco and Gormally, 2015). The flexibility and variability of the wing region among FOX family members may prove an ideal target for further drug studies, both to provide specificity towards FOX proteins and possibly even to control the coordination of specific gene expression regimes by disrupting necessary protein-protein interactions (Fontaine et al., 2015).

The role of the forkhead domain hinge loop *in vivo*

Regulation of gene expression depends on the cellular context and many transcription factors can act to either increase or reduce appropriate gene targets by coalescing multiple cellular signals through macromolecular complex formation and post-translation modification. Indeed, FOXP2 acts as both a transcriptional repressor and activator depending on the gene target (Spiteri et al., 2007). The reporter assay described here shows the transcriptional activation capability of FOXP2 for a cognate promoter sequence. The study was performed on full length FOXP2 wild-type and full length FOXP2 A539P mutant. Although the A539P mutation has

https://mc.manuscriptcentral.com/bc
For Review Only

been shown to abolish dimerisation of the isolated forkhead domain *in vitro* (Stroud et al., 2006; Chu et al., 2010), it cannot be said for certain whether the mutation alters the quaternary structure of full length FOXP2 *in vivo*. It is very likely that full length A539P FOXP2 is still capable of forming dimers if only via the leucine zipper interface. It is probable, though, given the evidence from the isolated forkhead domain, that the A539P mutation will prevent dimerisation of the full length protein at the forkhead domain dimer interface *in vivo*.

The results from the study show that FOXP2 wild-type is more transcriptionally active than the A539P mutant; however the hinge-loop mutation does not completely inactivate FOXP2 activity (Figure 6). Therefore, the ability of the forkhead domain to form dimers may not be absolutely essential for transcriptional activity of FOXP2 although it may be required for optimal activity. This hints at the idea of dimerisation of the FOXP2 FHD acting as a regulatory mechanism. From the results of this study, we have shown that wild-type has a more flexible backbone than the mutant and it has the ability to associate with the DNA with a more dynamic open structure. The fact that wild-type is more transcriptionally active than the mutant implies that backbone flexibility and loose association with the DNA may be an important consequence of the evolutionary hinge loop mutation unique to the FOXP subfamily members by improving DNA affinity (increasing occupancy time at the promoter) as well as by providing a point for domain-swapping to occur at the appropriate promoter site.It is certainly possible that the wild-type assumes the monomeric form when locating and binding the consensus site in a more open conformation after which it can form hetero- or homo-typic domain-swapped dimers with other FOXP subfamily member forkhead domains in a context and sequence dependent manner. Precisely what effect forkhead domain dimerisation could have on multi-component transcriptional complex formation, aggregation of proximal promoter sites or how it is controlled remains unclear. The quaternary state of the FOXP2 FHD may alter the recruitment of transcription factors necessary for appropriate target gene regulation. For example, NFAT has been shown to interact directly with the monomeric FOXP2 FHD (Wu et al., 2006). The role of FOXP2 FHD dimerisation in protein-protein interactions with NFAT and other transcription factors is an important avenue for future studies.

Conclusions

The members of the FOX family of transcription factors have become clinically relevant targets in recent years owing to their overarching involvement in the progression and survival of a cancerous cell state in several tissues. Understanding the structural biology of specific protein-DNA interactions is essential to the design of novel therapeutics that target and control

https://mc.manuscriptcentral.com/bc

For Review Only

the transcriptional master regulators of a diseased cell state. Here we present data that shows a distinct difference in the DNA binding capacity between the wild-type and the A539P mutant FOXP2 FHD. We demonstrate that the inherent change in dynamics of the hinge-loop region (the region spanning helix-2 and helix-3) in the FOXP2 FHD regulates DNA binding both *in vitro* and *in vivo* by altering the global dynamics of the FOXP2 FHD, highlighting the critical role of the hinge loop in protein dynamics and in altering DNA binding characteristics of closely related transcription factor family members. We show that the wild-type has a more flexible backbone structure which allows it to make more electrostatic contacts with the DNA than the A539P mutant however the majority of these interactions are non-specific and furthermore it inserts into the DNA more shallowly than the mutant. This loose binding appears to be necessary for efficient transcription. Overall these results indicate that the hinge-loop connecting helix-2 and helix-3 plays a significant role in regulating the mechanism of DNA binding of the forkhead domain. This is an important distinguishing factor between FOXPs and other FOX proteins that do not have the flexible hinge-loop.

Materials and methods

Protein expression and purification

T7 *Escherichia coli* bacteria were transformed with a pET-11a plasmid housing the coding sequence for human FOXP2 FHD (residues 501-584) fused with an N-terminal hexahistidine tag under the control of an IPTG inducible promoter. Site-directed mutagenesis was performed to generate the A539P mutation using the Quickchange site-directed mutagenesis kit (Agilent, USA).Cultures were grown to an OD_{600} of 0.6-0.8 at 37°C in an aerobic shaker rotating at ~200 rpm. Once the appropriate optical density was reached the cultures were cold shocked at 4°C for 10 minutes before 0.5 mM IPTG was added to induce heterologous protein expression. The culture was then incubated at 20°C for 20 hours in an aerobic shaking incubator at 200 rpm to allow for sufficient soluble heterologous protein expression.

Immobilised metal ion affinity chromatography (IMAC) was utilised for coarse purification of the FOXP2 FHD-fusion protein from the bacterial cell milieu (Supplementary Figure 1). The IMAC column was equilibrated with ten column volumes of equilibration buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 25 mM imidazole). The cell lysate was then loaded onto the column at a flow rate of 2 ml/min. The column was washed with ten column volumes of high

For Review Only

salt buffer (20 mM Tris-HCl pH 7.5, 1.5 M NaCl, 25 mM imidazole) before the bound protein was eluted in a single step with elution buffer (20 mM Tris-HCl, 500 mM NaCl, 300 mM imidazole). The N-terminal His-tag was removed by incubation of the protein with thrombin for 4 hours at 20°C. The protein was further purified using successive rounds of IMAC, benzamidine affinity chromatography and size-exclusion chromatography as described elsewhere (Blane and Fannuchi, 2015; Morris and Fannuchi, 2016).

Oligonucleotides

The following duplex cognate DNA sequence TTAGG**TGTTTAC**TTTCATAG containing a single binding site (in bold) has been shown to have a strong affinity for the FOXP2 FHD (Nelson et al., 2013; Morris and Fannuchi, 2016; Webb et al., 2017) and was used exclusively in this study. Duplex DNA was synthesised at Integrated DNA technology, South Africa and was prepared to a stock concentration of 200 μM in TBE buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH ~ 8.3). The determination of DNA concentration was performed in triplicate using UV-absorbance and the average was taken as the final concentration for downstream experiments.

Heat capacity of DNA binding

The heat capacity of binding for the wild-type and A539P FOXP2 FHD was determined using isothermal titration calorimetry on a NanoITC instrument (TA instruments, New Jersey, USA; Supplementary Table 1). The heat capacity value was obtained from the gradient of a linear regression fitted to the enthalpies of a series of five titrations performed at five temperatures in the range of 10 to 30°C. A typical titration consisted of 5 μl injections of 70-100 μM protein into 7-10 μM DNA. Both the protein and DNA were dialysed thoroughly against the same binding buffer (10 mM HEPES pH 7.5, 100 mM NaCl and 0.02% NaN_3). Blank titrations of protein into buffer and buffer into DNA were performed to account for heats of dilution. Enthalpies were determined from a nonlinear independent sites regression using the NITPIC software package (Keller et al., 2012). Each titration experiment was performed at least in duplicate and the replicates were averaged. Errors are the standard deviation of the averaged replicates.

Salt effects on DNA binding

https://mc.manuscriptcentral.com/bc

For Review Only

The electrostatic contributions to DNA binding of both the wild-type and the A539P mutant were calculated by performing DNA binding experiments at increasing salt concentrations and applying the following equation (Privalov et al., 2011; Supplementary Table 1):

$$log(K_a) = -N \cdot log[Salt] + log(K_{ne})$$

The number of electrostatic contacts formed between the protein and the DNA (N) and the non-electrostatic association constant (K_{nel}) were determined by fitting a linear regression to the double log plot of salt concentration and the corresponding experimental K_a. The association constants (K_a) for DNA binding were obtained from non-linear regression of an independent sites model fitted to isotherms determined by isothermal titration calorimetry at increasing salt (NaCl) concentrations. K_{nel} was calculated at a 1 M salt concentration. The entropy was dissected into contributions from the conformational changes and counterion exclusion using the Gibbs-Helmholtz equation. The entropy at 293 K was chosen as the standard to compare the wild-type FOXP2 FHD to the A539P FOXP2 FHD. The number of ionic contacts with the DNA backbone were estimated from N by considering the DNA phosphate counter-ion occupancy of NaCl, ψ=0.64 (Olmsted et al., 1995; Privalov et al., 2011).
Isothermal (293 K) titrations of 70-100 µM wild-type or 70-100 µM A539P FOXP2 FHD into 7-10 µM DNA were performed at 293.15 K in binding buffer (10 mM HEPES pH 7.5 and 0.02% NaN_3) adjusted to one of five salt concentrations in the range of 50-150 mM NaCl. Each titration consisted of forty 5 µl injections of protein into 950 µl cognate DNA using a NanoITC instrument (TA Instruments, New Jersey,USA). Each titration was performed at least in duplicate and the replicate values were averaged. Errors are the standard deviation of the averaged values. Each binding isotherm was fit with a non-linear independent sites model minimising the chi-squared between the model and experimental data points.

Dynamic quenching studies

The Stern-Volmer coefficient ($K_{SV,}$) describes the solvent accessibility of the tryptophan residues based on the degree to which they are quenched by increasing concentrations of quencher (Q) in solution (Eftink and Ghiron, 1976):

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

The Stern-Volmer coefficient (K_{SV}) was determined for the apo and DNA bound form of the wild-type and A539P FOXP2 FHD (Supplementary Figure 9). Fluorescence measurements were performed on a Jasco FP-6300 fluorimeter with an excitation wavelength of 295 nm.

For Review Only

Samples consisted of 2 µM FOXP2 FHD in binding buffer (10 mM HEPES pH 7.5, 100 mM NaCl and 0.02 % NaN_3) with increasing acrylamide (quencher) concentrations from 0-250 mM. For DNA bound studies 2 µM FOXP2 FHD was incubated with an equivalent of cognate DNA oligonucleotide for 30 minutes at 20°C prior to performing the experiments. Buffer only and DNA-buffer blanks were subtracted before data analysis. All experiments were performed in triplicate and averaged following analysis. Errors are the standard deviation of the averaged data.

Hydrogen-deuterium exchange mass spectrometry

The *in vitro* structural dynamics of the apo and DNA bound forms of wild-type and A539P FOXP2 FHD were studied by hydrogen-deuterium exchange mass spectrometry (Supplementary Figures 7 and 8). Labelling, quenching and proteolytic cleavage experiments were performed on a PAL HDX system (LEAP Technologies, USA) coupled to an Agilent 1100 HPLC system (Agilent, USA). Mass spectrometry was performed on an AB Sciex 6600 TripleTOF mass spectrometer (AB Sciex, USA). Protein labelling consisted of a 10-3600 s incubation of 20-30 µg of wild-type or A539P FOXP2 FHD, with and without an equivalent mass of cognate DNA oligonucleotide, in binding buffer (10 mM HEPES pH 7.5, 100 mM NaCl) made up with 90% D_2O at 293 K. Samples were then quenched by a 2-fold dilution in quench buffer (20 mM phosphate pH 4.5, 100 mM TCEP and 3.4 M guanadine-HCl) held at 0°C. The protein was then fragmented by incubation on an inline Porozyme immobilised pepsin chromatography column (Life Technologies) at 4°C for 30 s before being desalted on an Acclaim PepMap trap column (0.3×5 mm) and subsequently loaded onto a Kinetex C_{18} reverse phase chromatography column (Phenomenex, USA). Peptides were separated onto the mass spectrometer at a flow rate of 200 µl/min with an elution gradient of 5-40 % buffer B (80% ACN/0.1% FA).

Peptide mass analysis was performed on an AB Sciex 6600 TripleTOF in Data Dependent Acquisition (DDA) mode. In DDA mode, precursor scans were acquired from *m/z* 360-1500 using an accumulation time of 250 ms followed by 30 product scans, acquired from *m/z* 100-1800 at 100 ms each, for a total scan time of 3.3 s. Charge ions, falling between 1^+ - 5^+, were automatically fragmented in the Q2 collision cell using nitrogen gas. The collision energy was chosen automatically based on the *m/z* and the charge. Peptide identification was performed in PEAKS 6 (Bioinformatics Solutions Inc., USA). The degree of deuteron incorporation was determined with the HDXaminer software package (Sierra Analytics, USA).

For Review Only

Luciferase reporter assays

HEK293 cell cultures were plated at a density of 1×10^4 cells per well in a 96 well plate and grown to confluency in antibiotic free DMEM medium at 37°C with 5% CO_2. Each well was transfected with 1.6 µg of pcDNA4 vector containing the full length wild-type FOXP2 coding sequence or the full length A539P FOXP2 coding sequence obtained by site-directed mutagenesis of the wild-type coding sequence using the Quikchange Lightening site-directed mutagenesis kit followed as per manufacturer's instructions (Agilent, USA; Supplementary Figure 11). Negative controls were performed by the addition of transfection reagent, transfection with 1.6 µg pGL4 vector under the control of a 6X FOXP2 consensus sequence synthesised by Integrated DNA Technology (Cape Town, South Africa; Supplementary Figure 10) and transfection with 1.6 µg pRL-TK vector encoding *Renilla* firefly luciferase under control of a cognate tyrosine kinase promoter. Transfections were performed using Fugene HD transfection reagent as per manufacturer's instructions (Promega). Transfected cells were incubated for a further 24 hours. Luciferase assays were performed using the Dual-Glo luciferase assay kit followed as per manufacturer's instructions (Promega). Luminescence readings were taken on a GLoMax 96 Microplate Luminometer (Promega). Transfection efficiency was normalised by co-transfection with a pRL-TK vector encoding *Renilla* luciferase under control of a tyrosine kinase promoter and by taking the ratio of the firefly luciferase activity to *Renilla* luciferase activity. Replicates of at least three were performed for each FOXP2 sequence and control.

For Review Or

Acknowledgements

We would like to acknowledge Kerry Hanmer and Dr Demetra Mavri-Damelin from the school of Molecular and Cell Biology at the University of the Witwatersrand for their assistance in culturing the HEK293 cells used in the dual luciferase reporter assays. We also extend our gratitude to Dr Sonja Vernes (Max Planck Institute for Psycholinguistics, Netherlands) for the generous gift of the pcDNA4-FOXP2 vector used in the dual luciferase reporter assays.
Funding: This work was supported by the University of the Witwatersrand; South African National Research Foundation [Grant 80681 to S.F, 68898 to H.W.D.]; the South African Research Chairs Initiative of the Department of Science and Technology [Grant 64788 to H.W.D] and the Medical Research Council of South Africa.

References

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

Blane, A. and Fanucchi, S. (2015) Effect of pH on the Structure and DNA Binding of the FOXP2 Forkhead Domain. Biochemistry *54*, 4001-4007

Clark, K.L., Halay, E.D., Lai, E. and Burley, S.K. (1993) Co-crystal structure of the HNF-3/forkhead DNA-recognition motif resembles histone H5. Nature *36*, 412-420.

Chu, Y. P., Chang, C. H., Shiu, J. H., Chang, Y. T., Chen, C. Y. and Chuang, W. J. (2011) Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: Insight into its domain swapping and DNA binding. Protein Sci. *20*, 908–924.

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.

deHaseth, P. L., Gross, C. A., Burgess, R. R. and Record Jr., M. T. (1977) Measurement of binding constants for protein-DNA interactions by DNA- cellulose chromatography. Biochemistry *16*, 4777–4783.

Eftink, M.R. and Ghiron, C.A. (1976) Exposure of tryptophanyl residues in proteins. Quantitative determination by fluorescence quenching studies. Biochemistry *15*, 672-680

Fontaine, F., Overman, J. and François, M. (2015) Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and obstacles. Cell Regen. (London) *4*, 2.

Fujita, E., Tanabe, Y., Shiota, A., Ueda, M., Suwa, K., Momoi, M. Y. and Momoi, T. (2008) Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. Proc. Natl. Acad. Sci. USA *105*, 3117–22.

Gajiwala, K.S. and Burley, S. K. (2000) Winged helix proteins. Curr Opin Struct. Biol. *10*, 110-116

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

For Review Only

Gormally, M. V., Dexheimer, T. S., Marsico, G., Sanders, D. A., Lowe, C., Matak-Vinkoviä, D., Michael, S., Jadhav, A., Rai, G., Maloney, D. J., et al. (2014) Suppression of the FOXM1 transcriptional programme via novel small molecule inhibition. Nat. Commun. *5*.

Hachigian, L.J., Carmona, V., Fenster, R.J., Kulicke, R., Heilbut,A., Sittler, A., Pereira de Almeida, L., Mesirov, J.P., Gao, F., Kolaczyk, E.D. and Heiman, M. (2017) Control of Huntington's disease-associated phenotypes by the striatum enriched transcription factor Foxp2. Cell Rep *21*, 2688

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

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

Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A. and Schuck, P. (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. Anal. Chem. *84*, 5066–5073.

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. USA *90*, 10421–10423.

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

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*, 809–22.

Littler, D. R., Alvarez-Fernández, 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, C.-C., Richard, A. J., Datta, K. and LiCata, V. J. (2008) Prevalence of temperature-dependent heat capacity changes in protein-DNA interactions. Biophys. J. *94*, 3258–3265.

Manning, G. S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. Q. Rev. Biophys. *11*, 179.

For Review Only

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.

Marsico, G. and Gormally, M. V. (2015) Small molecule inhibition of FOXM1: How to bring a novel compound into genomic context. Genomics Data *3*, 19–23.

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

Murphy, K. P. and Freire, E. (1992) Thermodynamics of structural stability and cooperative folding behavious in proteins. Adv. Protein Chem. *43*, 313–361.

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

Olmsted, M. C., Bond, J. P., Anderson, C. F. and Record, M. T. (1995) Grand canonical Monte Carlo molecular and thermodynamic predictions of ion effects on binding of an oligocation (L8+) to the center of DNA oligomers. Biophys. J. *68*, 634–647.

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*, 2755–66.

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*, 5002–12.

Prabhu, N.V. and Sharp, K.A. (2005) Heat capacity in proteins. Annu. Rev. Phys. Chem. *56*, 521-548

Privalov, P. L., Dragan, A. I. and Crane-Robinson, C. (2011) Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. Nucleic Acids Res. *39*, 2483-2491

Record, J. M. T., Anderson, C. F., Lohman, T. M., (1978) Thermodynamic analysis of ion effects on binding and conformational equilibria of proteins and nucleic-acids - roles of ion association or release, screening, and ion effects on water activity. Q. Rev. Biophys. *11*, 103–178.

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

For Review Only

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

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

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., Vernes, S. C., Fisher, S. E., Ren, B. and Geschwind, D. H. (2007) Identification of the Transcriptional Targets of FOXP2, a Gene Linked to Speech and Language, in Developing Human Brain. Am. J. Hum. Genet. *81*, 1144–1157.

Spolar, R. and Record, M. (1994) Coupling of local folding to site-specific binding of proteins to DNA. Science *263*, 777–784.

Steffen, N. R., Murphy, S. D., Tolleri, G. W. and Lathrop, R. H. (2002) DNA sequence and structure: direct and indirect recognition in protein-DNA binding. Bioinformatics *18*, S22-30.

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.

Takahashi, H., Takahashi, K. and Liu, F.-C. (2009) FOXP Genes, Neural Development, Speech and Language Disorders. In: Advances in experimental medicine and biology, pp 117–129.

Teramitsu, I. (2004) Parallel FoxP1 and FoxP2 Expression in Songbird and Human Brain Predicts Functional Interaction. J. Neurosci. *24*, 3152–3163.

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.

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*, 1-10

Weigelt, J., Climent, I., Dahlman-Wright, K. and Wikstrom, M. (2001) Solution structure of the DNA binding domain of the human forkhead transcription factor AFX (FOXO4). Biochemistry *40*, 5861–5869.

Wong, K. K., Gascoyne, D. M., Soilleux, E. J., Lyne, L., Spearman, H., Roncador, G., Pedersen, L. M., Møller, M. B., Green, T. M. and Banham, A. H. (2016) FOXP2-

For Review Only

positive diffuse large B-cell lymphomas exhibit a poor response to R-CHOP therapy and distinct biological signatures. Oncotarget *7*, 52940–52956.

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.

Figure legends

Figure 1 Anatomy of the FOXP2 forkhead domain.
FOXP2 has two dimerisation domains: the leucine zipper (ZIP) and the forkhead domain (FHD). The FOXP2 FHD can form a domain-swapped dimer (inset) whereby associated FOXP2 forkhead domain monomers exchange the domain-swapped subdomain (helix-1 and –2; blue). The exchange (transparent) involves the extension of helix-2 into the hinge loop that connects the two subdomains of the FOXP2 forkhead domain. This is thought to be possible because of an evolutionary hinge loop mutation (P539A) exclusive to the FOXP subfamily (Ala539 shown in purple). The DNA binding subdomain (helix-3 and the wing; orange) are responsible for the formation of protein-DNA interface. The figure was produced using the crystal structure of the FOXP2 FHD bound to DNA (PDB ID: 2A07).

Figure 2 Temperature dependence of DNA binding of the wild-type (triangles and solid line) and A539P (circles and dashed line) FOXP2 FHD.
The heat capacity values are ΔC_p = -0.88 ± 0.048 kJ/mol/K for wild-type and ΔC_p = -1.56 ± 0.04 kJ/mol/K for the A539P FOXP2 FHD. Error bars are the sample standard deviation of the average of at least duplicate or triplicate titrations.

Figure 3 Salt dependent studies of DNA binding by the wild-type and A539P FOXP2 FHD.
(A) The number of counter-ions displaced (N) during binding can be determined from the gradient of the linear regression which can then be related to the number of ionic contacts made. The steeper gradient for the wild-type (solid line) implies a greater number of backbone interactions with the DNA compared to the mutant (dashed line). Error bars are the sample standard deviation of the average of at least duplicate or triplicate titrations. The ΔH displays a minor dependence on the salt concentrations used in this study (Supplementary Figure 6). (B) The entropy of binding at 100 mM NaCl was dissected into electrostatic ($T\Delta S_{el}$) and non-

https://mc.manuscriptcentral.com/bc

For Review Only

electrostatic ($T\Delta S_{nel}$) components using the enthalpy of binding (ΔH) averages from all salt concentrations. The wild-type (grey) displays a greater entropy of binding but the magnitude of the enthalpy and non-electrostatic free energy of binding is smaller compared to the mutant (black).

Figure 4 The change in backbone dynamics of the wild-type FOXP2 FHD induced by DNA binding as observed with hydrogen-deuterium exchange mass spectrometry.
The change in deuterium uptake is calculated by subtracting deuterium uptake for the DNA bound wild-type FOXP2 FHD from the deuterium uptake from the unbound form at the incubation time of 3600 s. The difference in deuteron uptake is mapped to the crystal structure of the monomeric wild-type FOXP2 FHD bound to the consensus DNA sequence (PDB ID: 2A07). The site of Ala539 in the hinge-loop is shown in purple. Insets: Deuterium uptake plots for helix-2 (H2; Ser522-Thr535), the recognition helix (H3; Trp548-His559) and the β-sheet wing (W1; Cys561-Glu577) of the wild-type (blue) and A539P (orange) FOXP2 FHD in the presence (dashed line) and absence (solid line) of DNA.

Figure 5 (A) Stern-Volmer constants (K_{SV}) for the free and DNA bound forms of the wild-type (grey) and A539P (black) FOXP2 FHD. There is a substantial reduction in tryptophan fluorescence quenching upon DNA binding of both wild-type and mutant however the wild-type displays greater quenching than the mutant. Hence wild-type tryptophan residues are more exposed upon DNA binding than the mutant. Errors represent the sample standard deviation of three independent replicates. (B) The positions of the three tryptophan residues in the FOXP2 FHD.

Figure 6 Dual luciferase reporter assay for FOXP2 in HEK293 cells.
Reporter assays were performed with a vector encoding firefly luciferase under the control of a promoter containing 6 sequential cognate binding sites (TTAGG**TGTTTAC**TTTCA). Cells were transfected with an empty pcDNA4 vector (grey) or pcDNA4 vectors designed to transiently over express either wild-type FOXP2 (blue) or A539P mutant (orange). Error bars represent the standard deviation of at least three independent replicates.

For Review Only

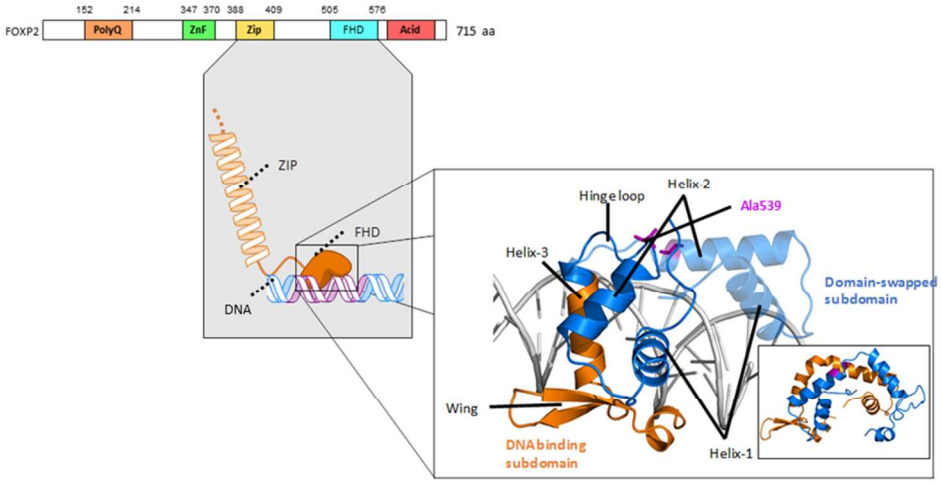


Figure 1 Anatomy of the FOXP2 forkhead domain. FOXP2 has two dimerisation domains: the leucine zipper (ZIP) and the forkhead domain (FHD). The FOXP2 FHD can form a domain-swapped dimer (inset) whereby associated FOXP2 forkhead domain monomers exchange the domain-swapped subdomain (helix-1 and –2; blue). The exchange (transparent) involves the extension of helix-2 into the hinge loop that connects the two subdomains of the FOXP2 forkhead domain. This is thought to be possible because of an evolutionary hinge loop mutation (P539A) exclusive to the FOXP subfamily (Ala539 shown in purple). The DNA binding subdomain (helix-3 and the wing; orange) are responsible for the formation of protein-DNA interface. The figure was produced using the crystal structure of the FOXP2 FHD bound to DNA (PDB ID: 2A07).

254x190mm (96 x 96 DPI)

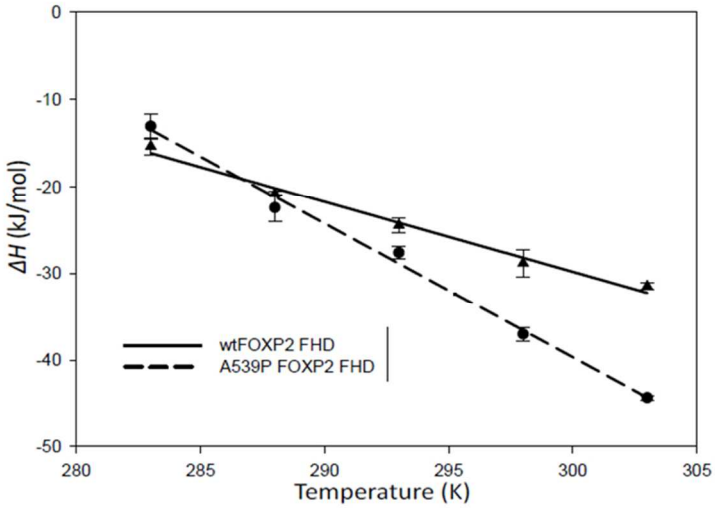


Figure 2 Temperature dependence of DNA binding of the wild-type (triangles and solid line) and A539P (circles and dashed line) FOXP2 FHD. The heat capacity values are ΔCp = -0.88 ± 0.048 kJ/mol/K for wild-type and ΔCp = -1.56 ± 0.04 kJ/mol/K for the A539P FOXP2 FHD. Error bars are the sample standard deviation of the average of at least duplicate or triplicate titrations.

254x190mm (96 x 96 DPI)

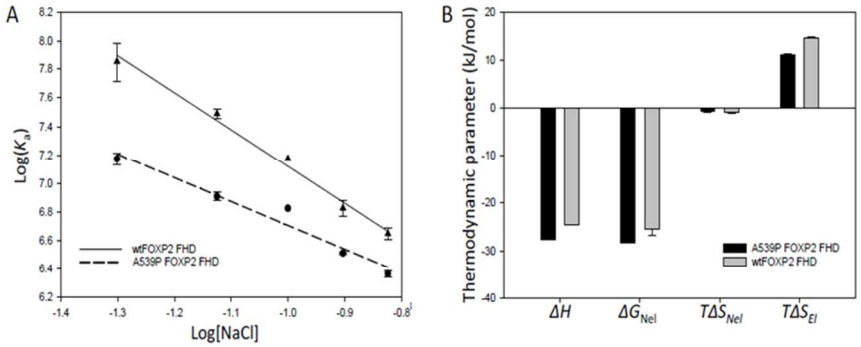


Figure 3 Salt dependent studies of DNA binding by the wild-type and A539P FOXP2 FHD. (A) The number of counter-ions displaced (N) during binding can be determined from the gradient of the linear regression which can then be related to the number of ionic contacts made. The steeper gradient for the wild-type (solid line) implies a greater number of backbone interactions with the DNA compared to the mutant (dashed line). Error bars are the sample standard deviation of the average of at least duplicate or triplicate titrations. The ΔH displays a minor dependence on the salt concentrations used in this study (Supplementary Fig 6). (B) The entropy of binding at 100 mM NaCl was dissected into electrostatic (TΔSel) and non-electrostatic (TΔSnel) components using the enthalpy of binding (ΔH) averages from all salt concentrations. The wild-type (grey) displays a greater entropy of binding but the magnitude of the enthalpy and non-electrostatic free energy of binding is smaller compared to the mutant (black).

254x190mm (96 x 96 DPI)

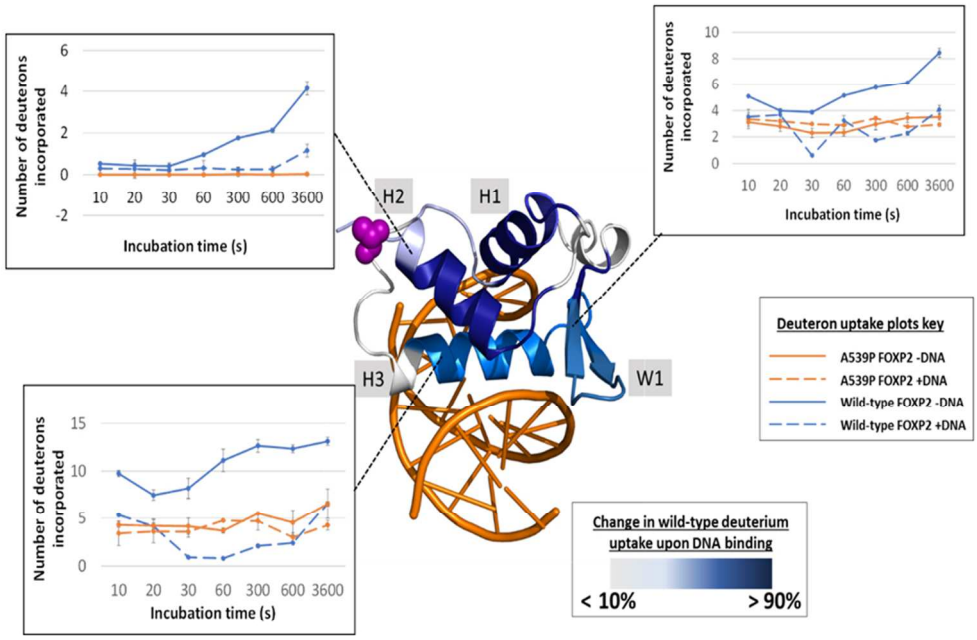


Figure 4 The change in backbone dynamics of the wild-type FOXP2 FHD induced by DNA binding as observed with hydrogen-deuterium exchange mass spectrometry. The change in deuterium uptake is calculated by subtracting deuterium uptake for the DNA bound wild-type FOXP2 FHD from the deuterium uptake from the unbound form at the incubation time of 3600 s. The difference in deuteron uptake is mapped to the crystal structure of the monomeric wild-type FOXP2 FHD bound to the consensus DNA sequence (PDB ID: 2A07). The site of Ala539 in the hinge-loop is shown in purple. (Insets) Deuterium uptake plots for helix-2 (H2; Ser522-Thr535), the recognition helix (H3; Trp548-His559) and the β-sheet wing (W1; Cys561-Glu577) of the wild-type (blue) and A539P (orange) FOXP2 FHD in the presence (dashed line) and absence (solid line) of DNA.

254x190mm (96 x 96 DPI)

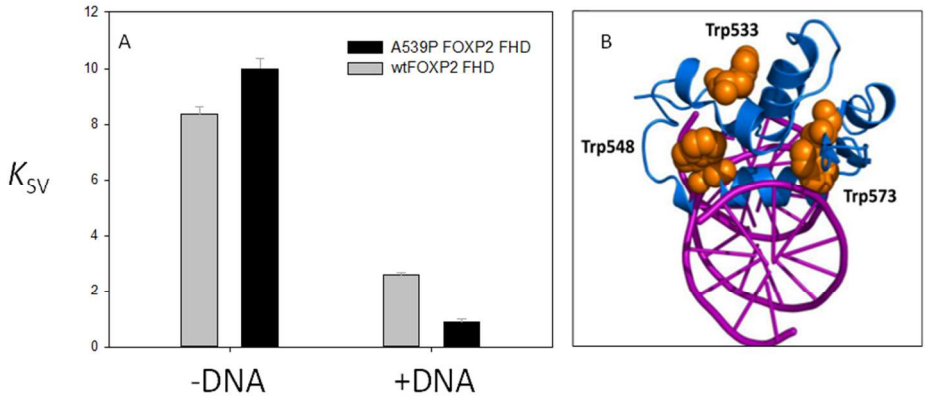


Figure 5 (A) Stern-Volmer constants (KSV) for the free and DNA bound forms of the wild-type (grey) and A539P (black) FOXP2 FHD. There is a substantial reduction in tryptophan fluorescence quenching upon DNA binding of both wild-type and mutant however the wild-type displays greater quenching than the mutant. Hence wild-type tryptophan residues are more exposed upon DNA binding than the mutant. Errors represent the sample standard deviation of three independent replicates. (B) The positions of the three tryptophan residues in the FOXP2 FHD.

254x190mm (96 x 96 DPI)

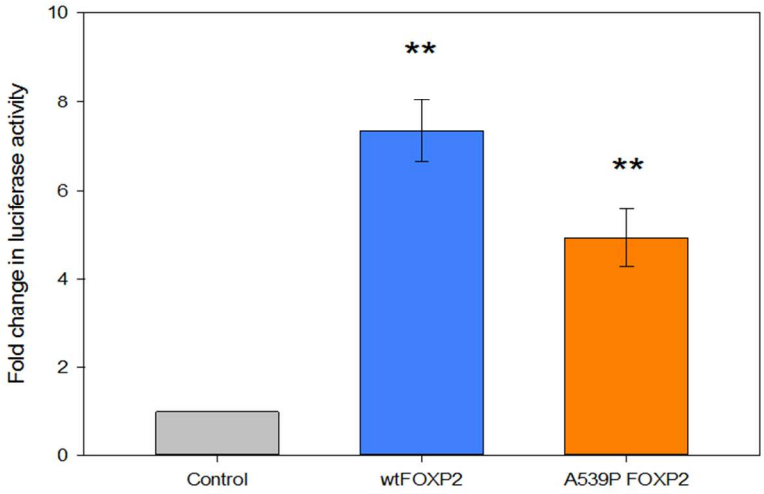


Figure 6 Dual luciferase reporter assay for FOXP2 in HEK293 cells. Reporter assays were performed with a vector encoding firefly luciferase under the control of a promoter containing 6 sequential cognate binding sites (TTAGGTGTTTACTTTCA).31 Cells were transfected with an empty pcDNA4 vector (grey) or pcDNA4 vectors designed to transiently over express either wild-type FOXP2 (blue) or A539P mutant (orange). Error bars represent the standard deviation of at least three independent replicates.

254x190mm (96 x 96 DPI)

The hinge loop plays a pivotal role in DNA binding and transcriptional activity of FOXP2

GAVIN MORRIS[1], STOYAN STOCHEV[2], PREVIN NAIKER[2],HEINRICH W. DIRR[1] AND SYLVIA FANUCCHI[1]

[1] Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg 2050. South Africa

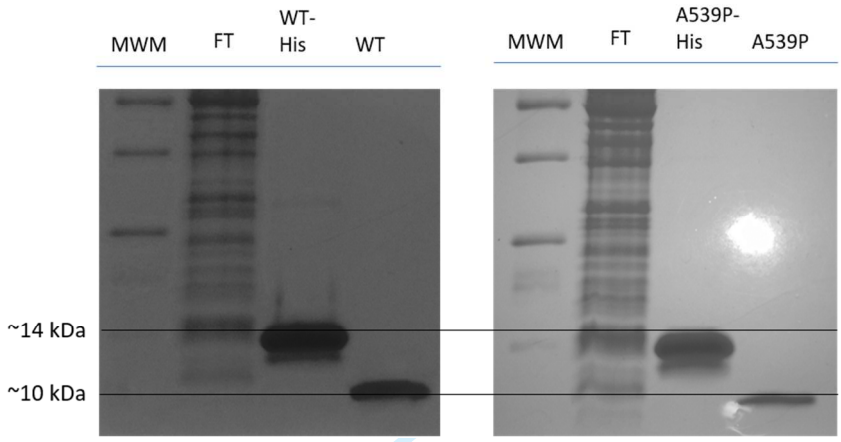
[2]CSIR Biosciences, CSIR, Meiring Naude Road, Brummeria,0001, Pretoria, Gauteng, South Africa

Supplementary Material

-Table of Contents-

Item	Description
Supplementary Fig 1:	SDS-PAGE analysis of wild-type and A539P FOXP2 forkhead domain variant purity following chromatographic separation from bacterial milieu.
Supplementary Table 1.	Thermodynamic signatures obtained for DNA binding of wild-type and A539P FOXP2 FHD at various temperatures and under different NaCl concentrations
Supplementary Fig 2:	Representative isotherms for the determination of the change in heat capacity of binding for the wild-type FOXP2 forkhead domain binding to consensus DNA.
Supplementary Fig 3:	Representative isotherms for the determination of the change in heat capacity of binding for the A539P FOXP2 forkhead domain binding to consensus DNA.
Supplementary Fig 4:	Representative isotherms for the determination of the effect of salt concentration for the wild-type FOXP2 forkhead domain binding to consensus DNA.
Supplementary Fig 5:	Representative isotherms for the determination of the effect of salt concentration for the A539P FOXP2 forkhead domain binding to consensus DNA.
Supplementary Fig 6.	Dependence of the enthalpy of binding on the concentration of salt.
Supplementary Fig 7.	Peptide coverage map for the wild-type and A539P FOXP2 forkhead domain.
Supplementary Fig 8.	Deuteron uptake plots for all peptides all peptides identified for apo and DNA bound forms of the wild-type and A539P FOXP2 forkhead domain.
Supplementary Fig 9.	Dynamic fluorescence quenching of free and DNA bound wild-type or A539P FOXP2 forkhead domain.
Supplementary Fig 10.	Confirmation of the inserted 6xNel promoter in front of the *firefly* luciferase coding sequence in the pGL4.10-LUC.
Supplementary Fig 11.	Sequence data obtained for the forkhead domain of the full length wild-type and A539P FOXP2.

For Review Only



Supplementary Fig 1. SDS-PAGE analysis of wild-type and A539P FOXP2 forkhead domain variant purity following chromatographic separation from bacterial milieu. Cleavage of the hexa-histidine tag was performed by incubating the purified protein for 4 hours at 20 ^{o}C in 10 mM Tris pH 8.0, 100 mM NaCl with 1 U bovine thrombin. Thrombin and uncleaved fusion protein were removed by sequential immobilised-nickel affinity and benzamidine affinity chromatography.

For Review Only

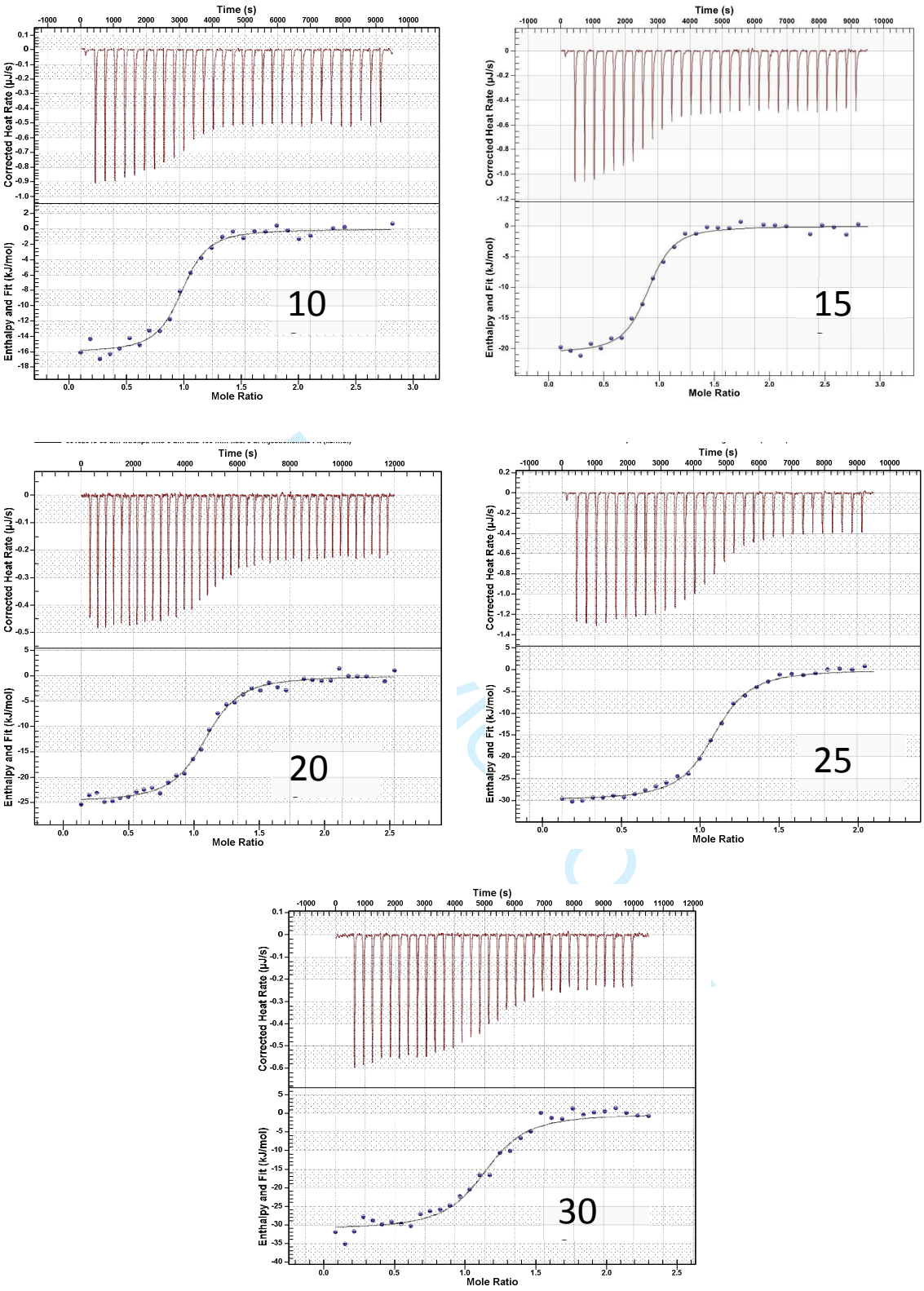
https://mc.manuscriptcentral.com/bc

Supplementary Table 1. Thermodynamic signatures obtained for DNA binding of wild-type and A539P FOXP2 FHD at various temperatures and under different NaCl concentrations

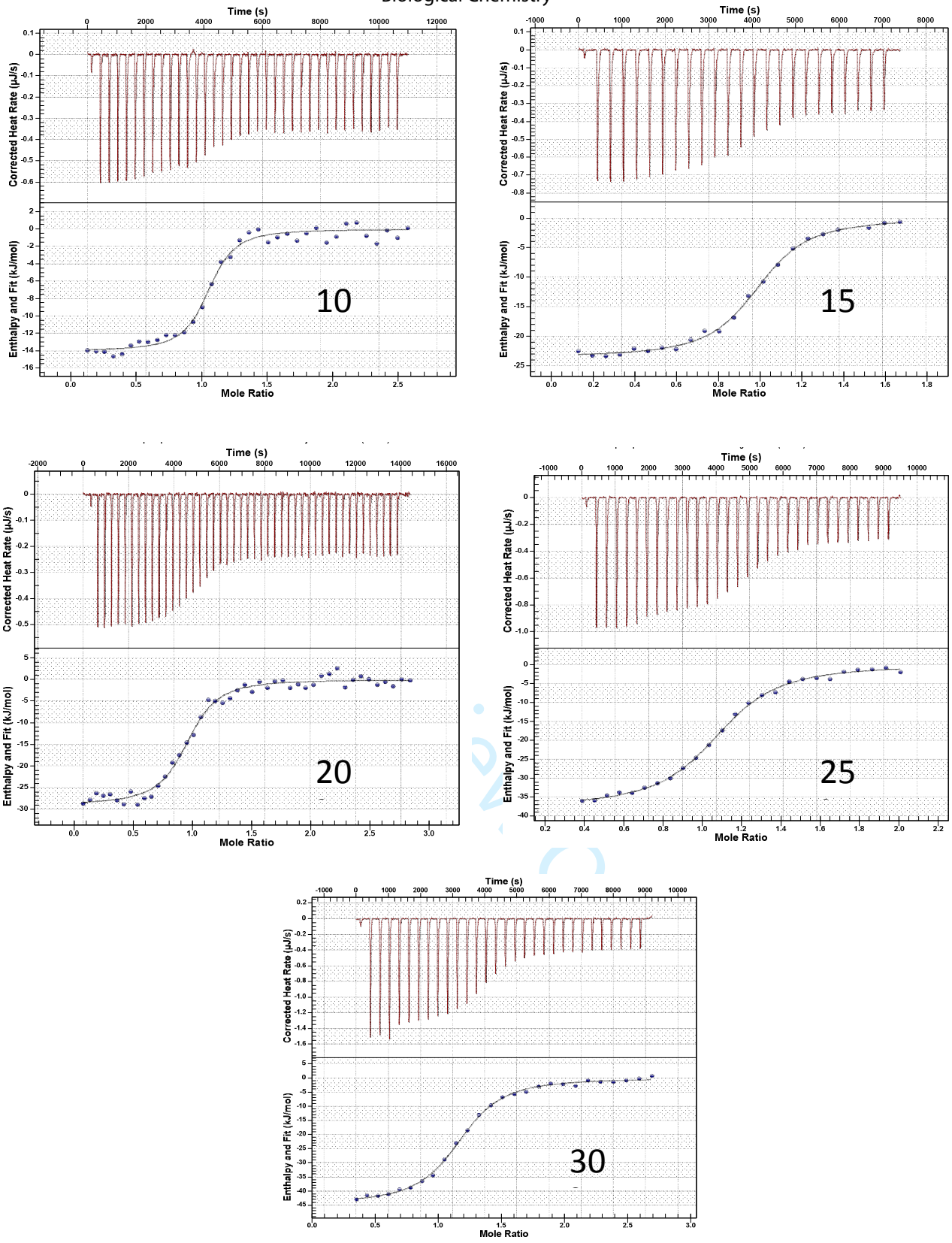
System	Temperature (K)	Salt concentration (M)	*N*	*ΔG* (kJ/mol)	$Log_{10}(K_a)$ (M^{-1})	*ΔH* (kJ/mol)	*TΔS* (kJ/mol)
Wild-type	283	0.1	0.89 ± 0.10	-38.89 ± 1.98	7.02 ± 0.34	-15.35 ± 0.99	23.54 ± 2.97
	288	0.1	0.92 ± 0.07	-38.94 ± 1.47	7.00 ± 0.26	-20.80 ± 0.26	18.14 ± 1.20
	293	0.1	0.96 ± 0.15[a]	-40.26 ± 0.16[a]	7.17 ± 0.05[a]	-24.46 ± 0.81[a]	15.80 ± 0.70[a]
	298	0.1	1.02 ± 0.07	-39.16 ± 1.53	6.89 ± 0.28	-28.82 ± 1.56	10.34 ± 3.09
	303	0.1	1.17 ± 0.06	-37.78 ± 3.04	6.55 ± 0.54	-31.47 ± 0.38	6.31 ± 3.42
	293	0.05	0.87 ± 0.028	-44.06 ± 0.76	7.85 ± 0.14	-26.04 ± 3.95	19.52 ± 6.18
	293	0.075	1.15 ± 0.22	-42.02 ± 0.17	7.49 ± 0.03	-26.32 ± 2.67	15.66 ± 2.98
	293	0.125	0.99 ± 0.19	-38.30 ± 0.30	6.83 ± 0.05	-20.39 ± 0.45	17.91 ± 0.16
	293	0.15	1.19	-37.46	6.67	-16.90	17.47
A539P	283	0.1	1.08 ± 0.07	-39.04 ± 0.34	7.04 ± 0.05	-13.04 ± 1.41	26 ± 1.75
	288	0.1	0.99 ± 0.03	-38.72 ± 1.32	6.97 ± 0.23	-22.43 ± 1.58	16.30 ± 2.89
	293	0.1	0.88 ± 0.01[a]	-39.23 ± 0.11[a]	6.82 ± 0.05[a]	-27.6 ± 0.72[a]	11.63 ± 0.84[a]
	298	0.1	1.13 ± 0.08	-39.24 ± 2.02	6.88 ± 0.36	-37.03 ± 0.77	2.21 ± 2.79
	303	0.1	1.05 ± 0.15	-39.51 ± 1.51	6.78 ± 0.27	-44.31 ± 0.26	-4.80 ± 1.77
	293	0.05	1.07 ± 0.09	-40.25 ± 0.23	7.17 ± 0.04	-23.26 ± 1.99	16.99 ± 1.76
	293	0.075	1.08 ± 0.08	-38.76 ± 0.17	6.90 ± 0.03	-25.65 ± 1.39	14.816 ± 1.31
	293	0.125	1.02 ± 0.09	-36.54 ± 0.05	6.51 ± 0.01	-21.08 ± 2.37	12.67 ± 1.63
	293	0.15	0.98 ± 0.07	-35.71 ± 0.13	6.36 ± 0.02	-23.15 ± 2.58	12.57 ± 2.71

[a] Data obtained from Morris and Fanucchi, 2016

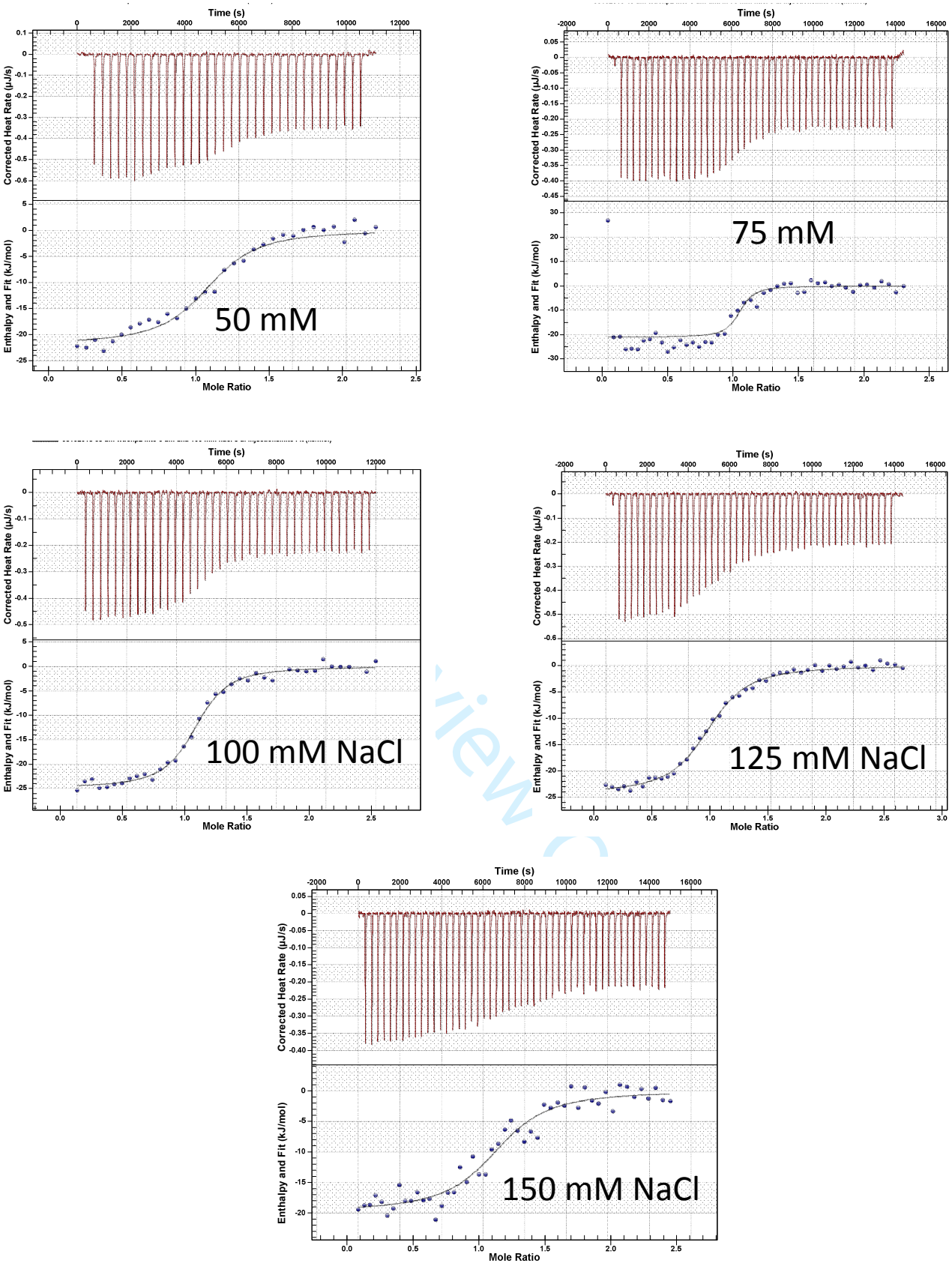
Errors represent the standard deviation of at least 2 replicates.



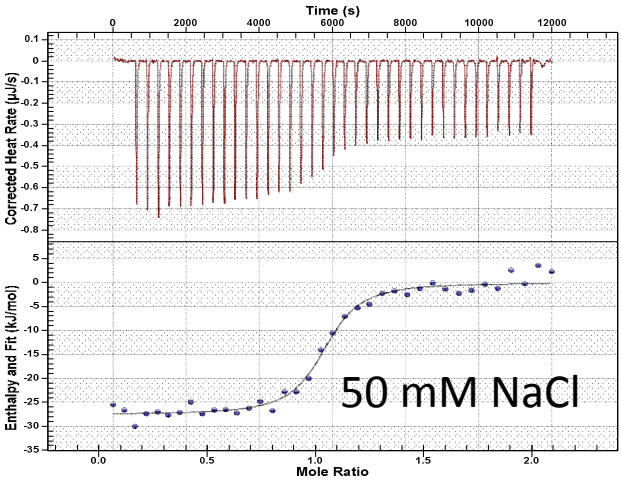
Supplementary Fig 2: Representative isotherms for the determination of the change in heat capacity of binding for the wild-type FOXP2 forkhead domain binding to consensus DNA. The temperature of each titration is shown on the respective trace.

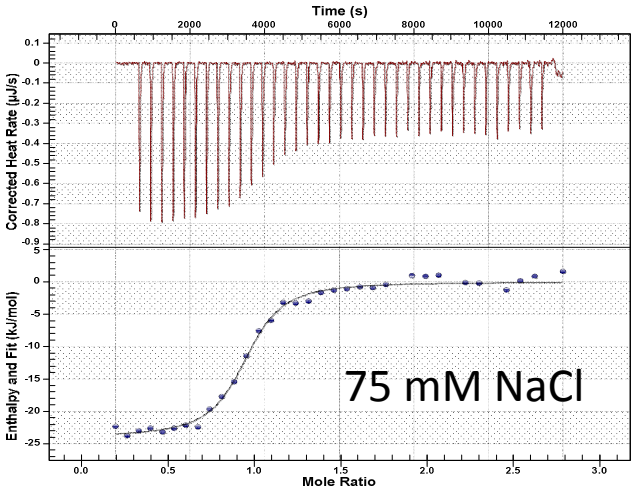


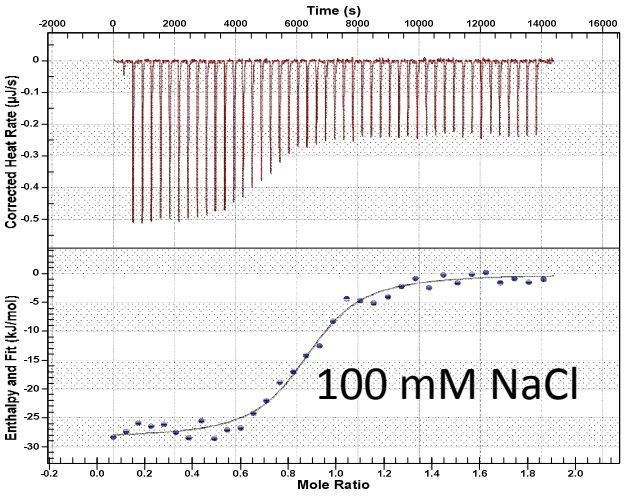
Supplementary Fig 3: Representative isotherms for the determination of the change in heat capacity of binding for the A539P FOXP2 forkhead domain binding to consensus DNA. The temperature of each titration is shown on the respective trace.

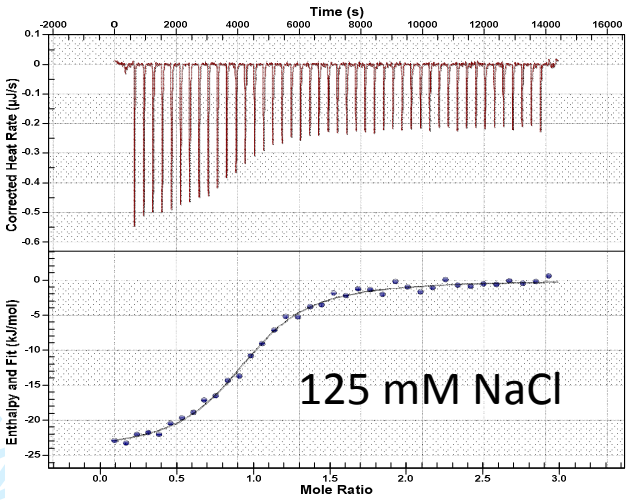


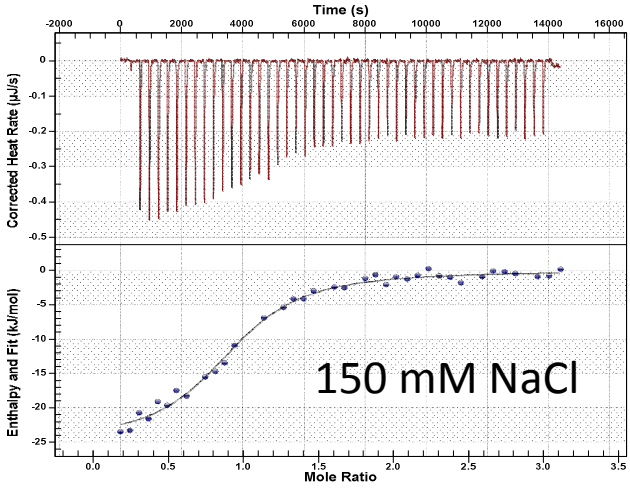
Supplementary Fig 4: Representative isotherms for the determination of the effect of salt concentration for the wild-type FOXP2 forkhead domain binding to consensus DNA. The salt concentration of each titration is shown on the respective trace.



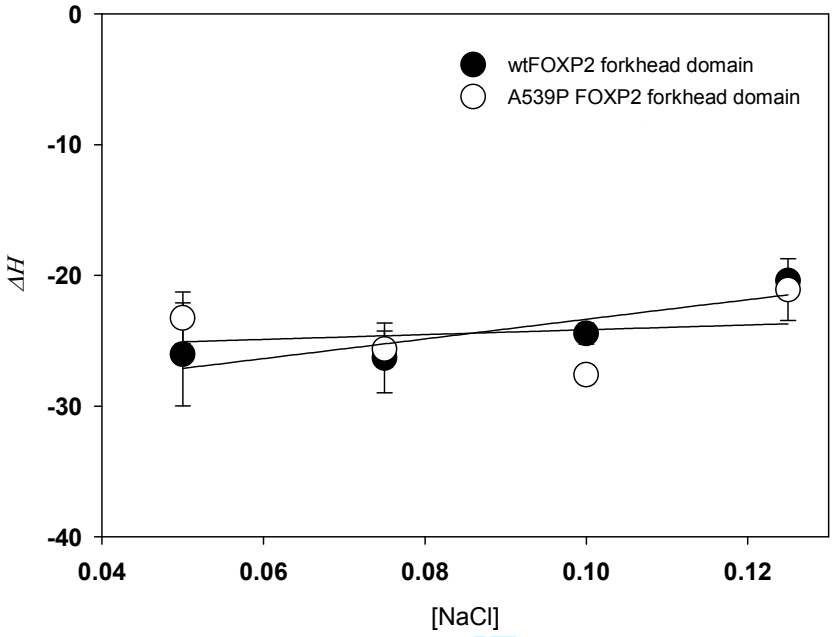






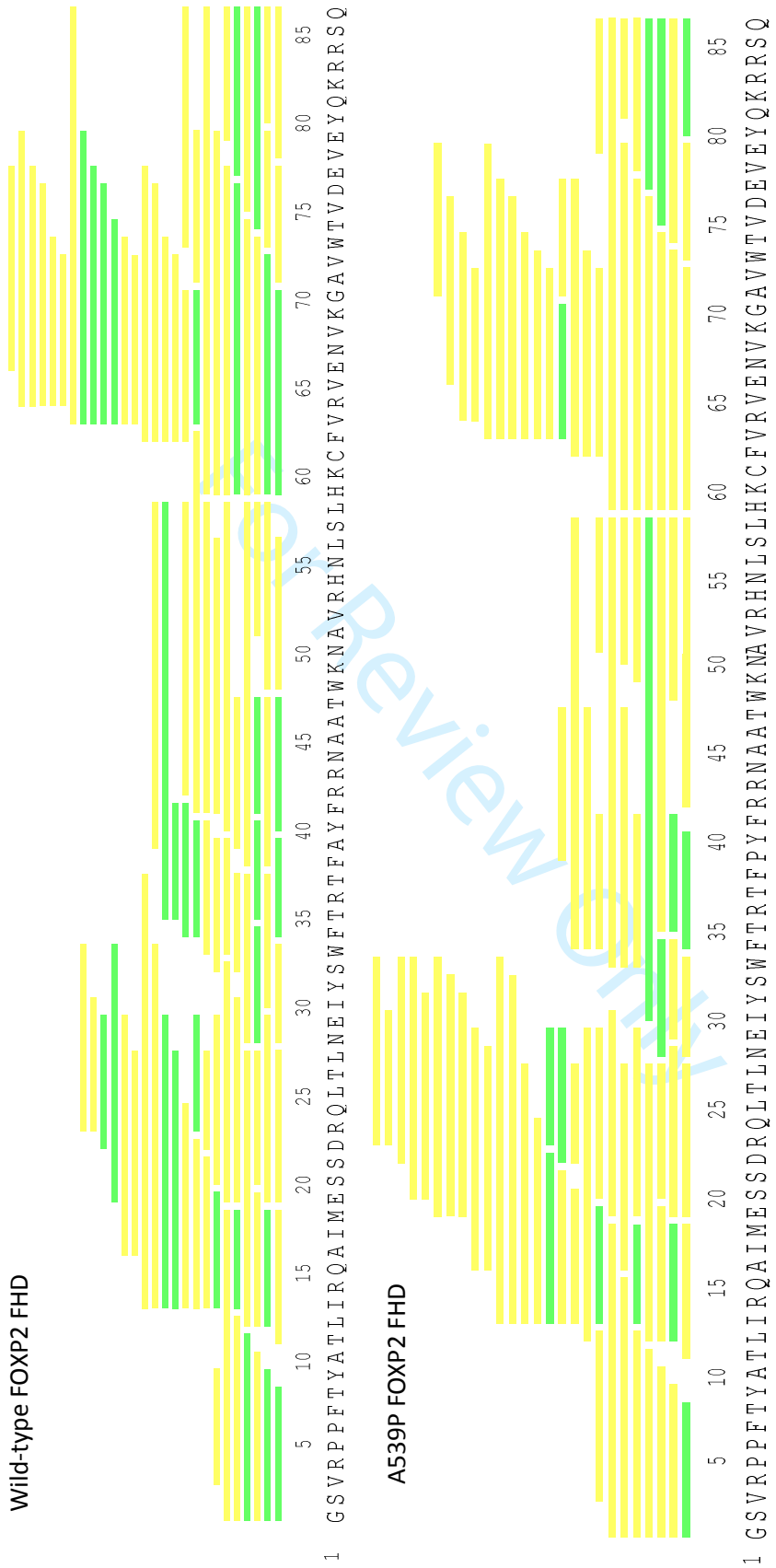


Supplementary Fig 5: Representative isotherms for the determination of the effect of salt concentration for the A539P FOXP2 forkhead domain binding to consensus DNA. The salt concentration of each titration is shown on the respective trace.



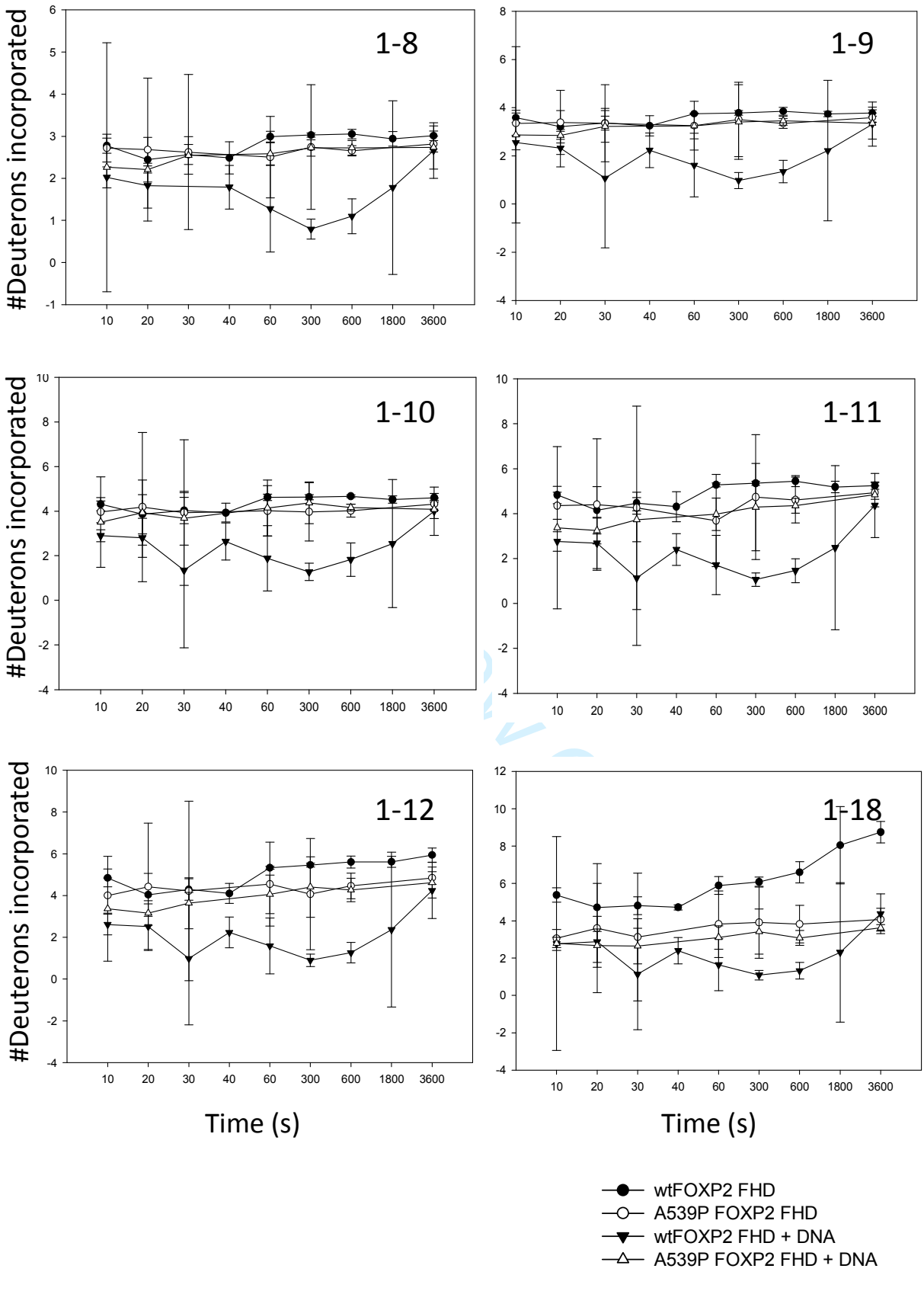
Supplementary Fig 6. Dependence of the enthalpy of binding on the concentration of salt. The enthalpy of binding displays very little correlation to the concentration of salt having an $R^2 << 0.95$ for both the wild-type and A539P FOXP2 forkhead domain. The enthalpies across the salt concentrations were averaged and used to dissect the entropic term into electrostatic and non-electrostatic terms.

For Peer Review Only

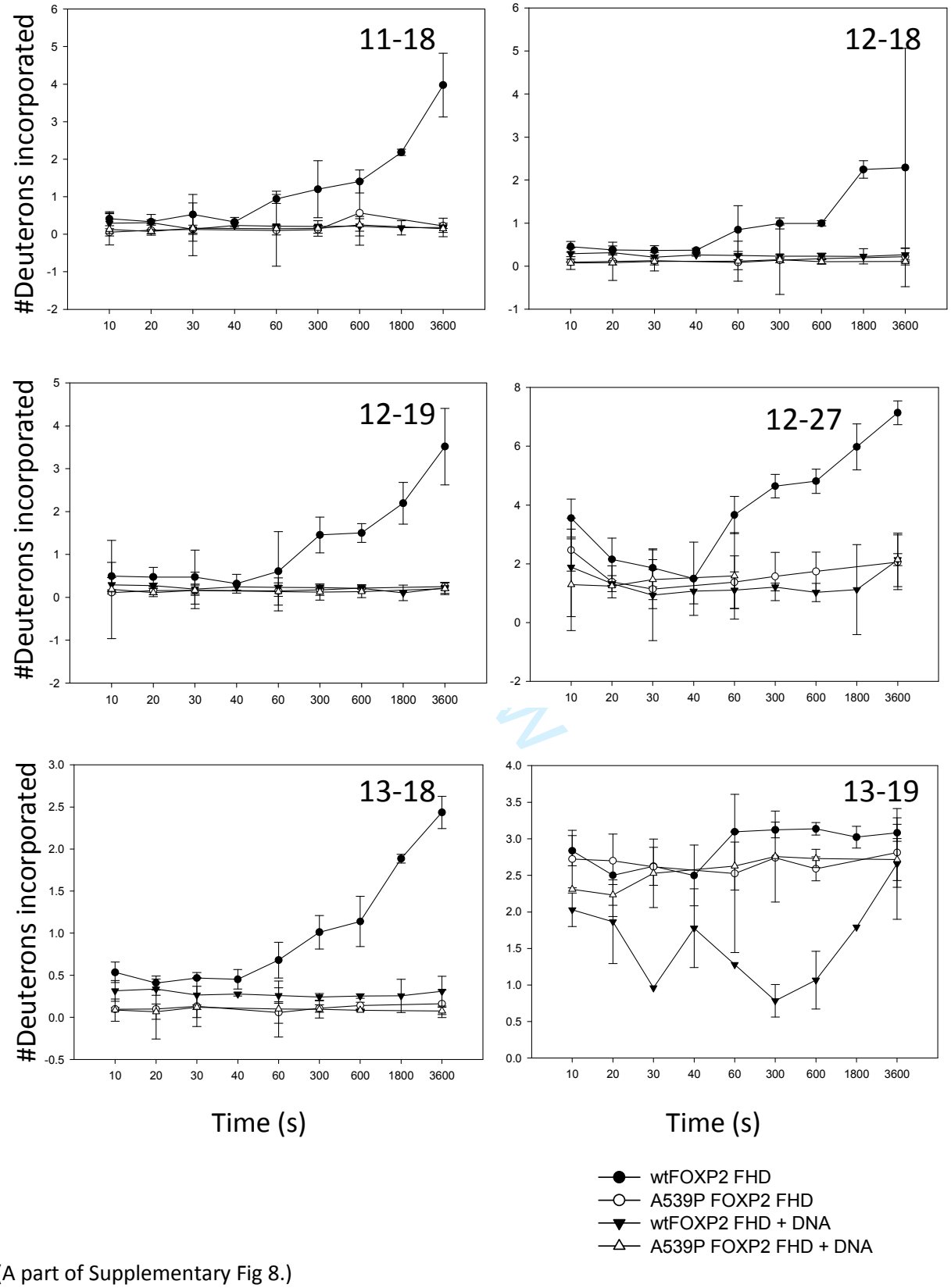


Supplementary Fig 7. Peptide coverage map for the wild-type and A539P FOXP2 forkhead domain. Full coverage of both proteins was obtained for the free and DNA bound forms. Only peptides with a confidence score >80 were kept.

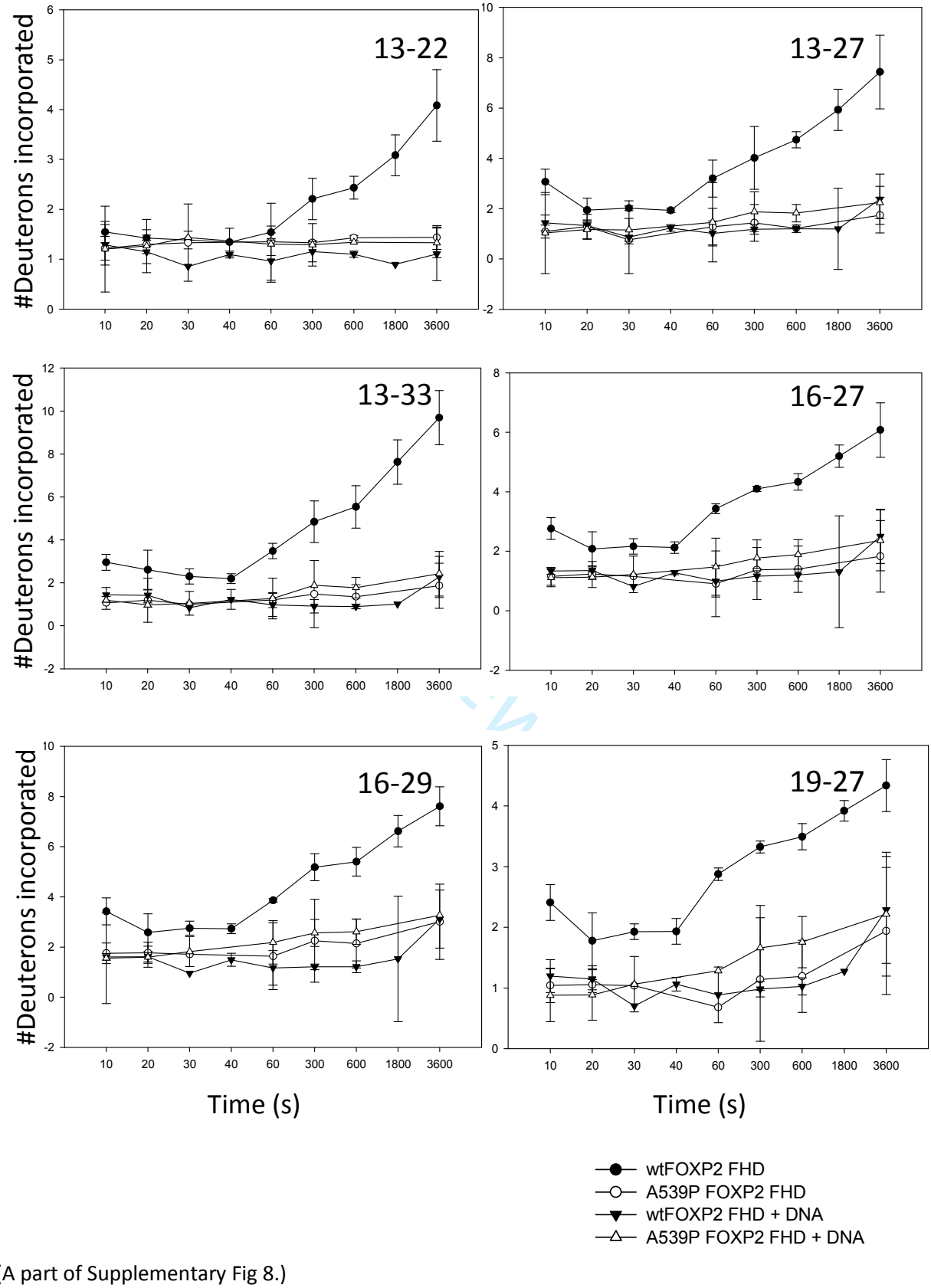
For Review Only



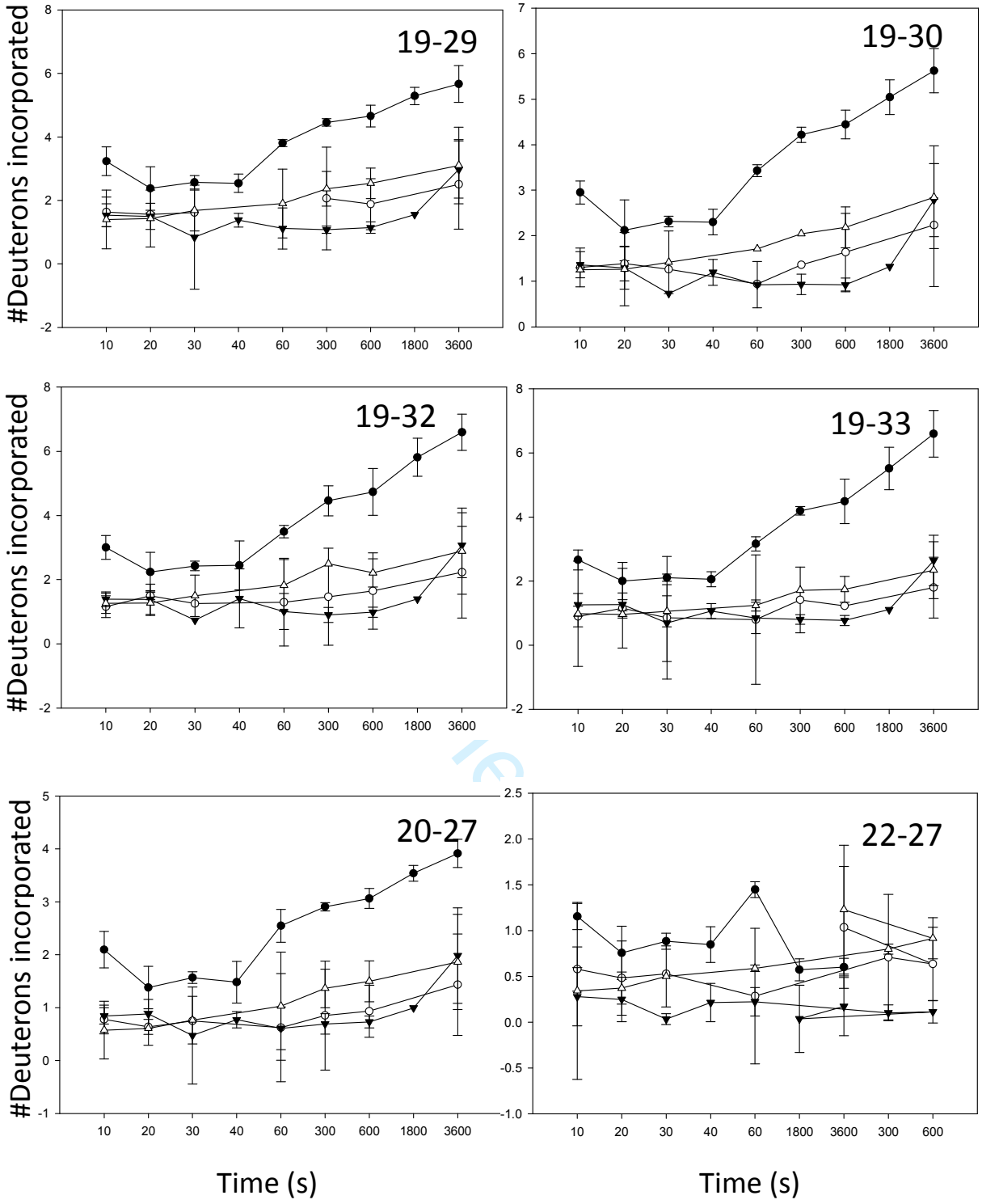
(A part of Supplementary Fig 8.)



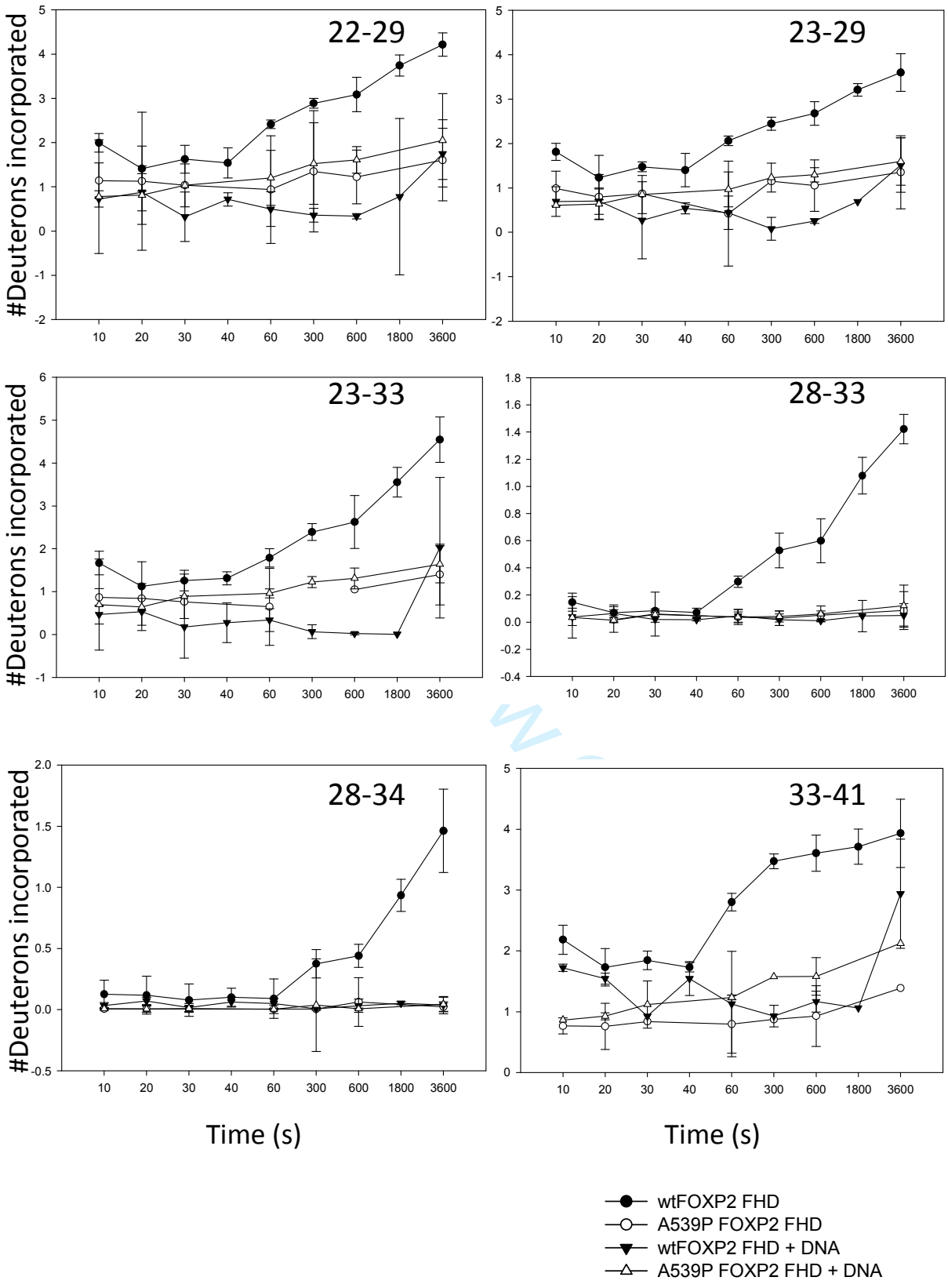
(A part of Supplementary Fig 8.)



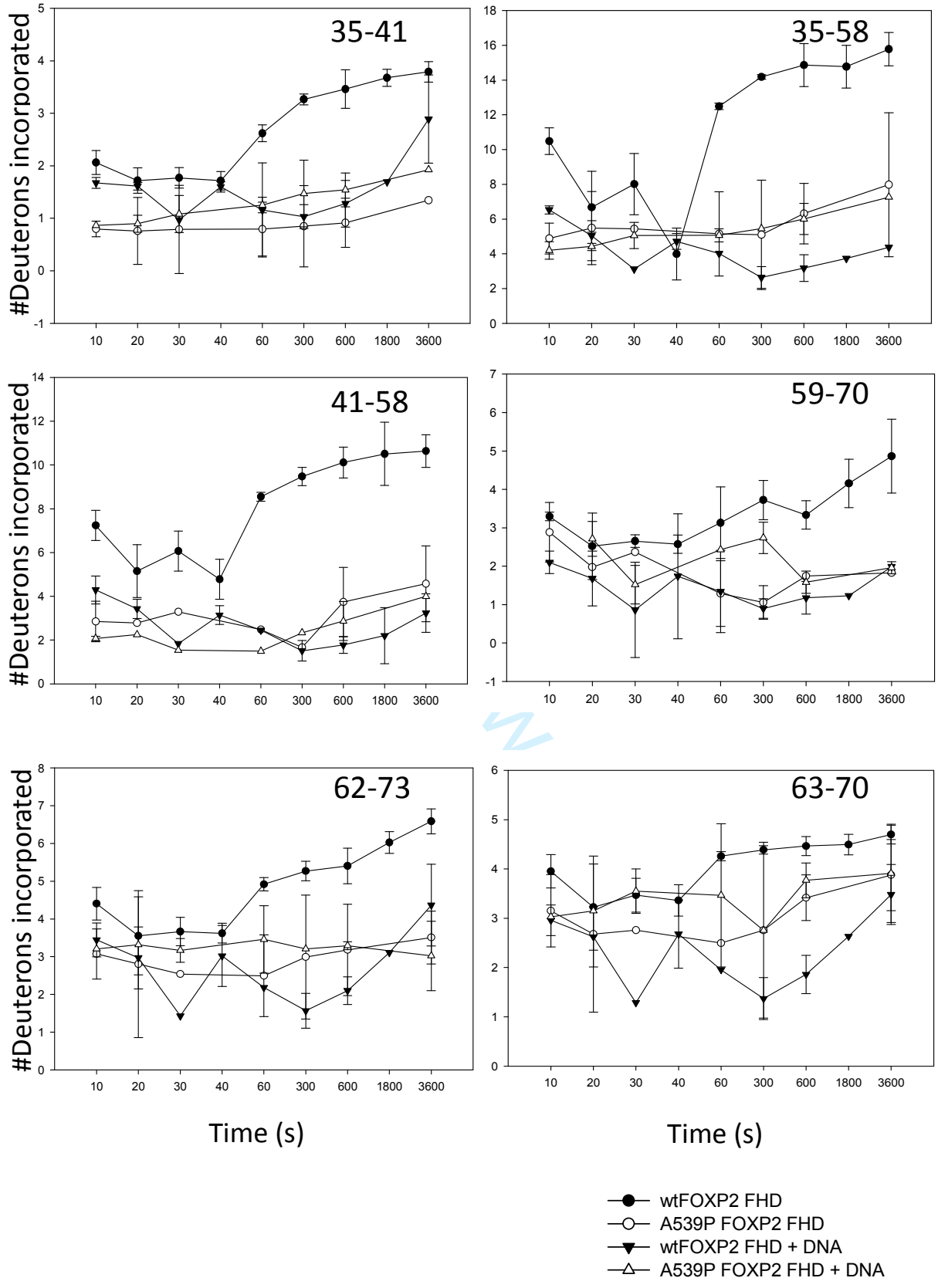
(A part of Supplementary Fig 8.)



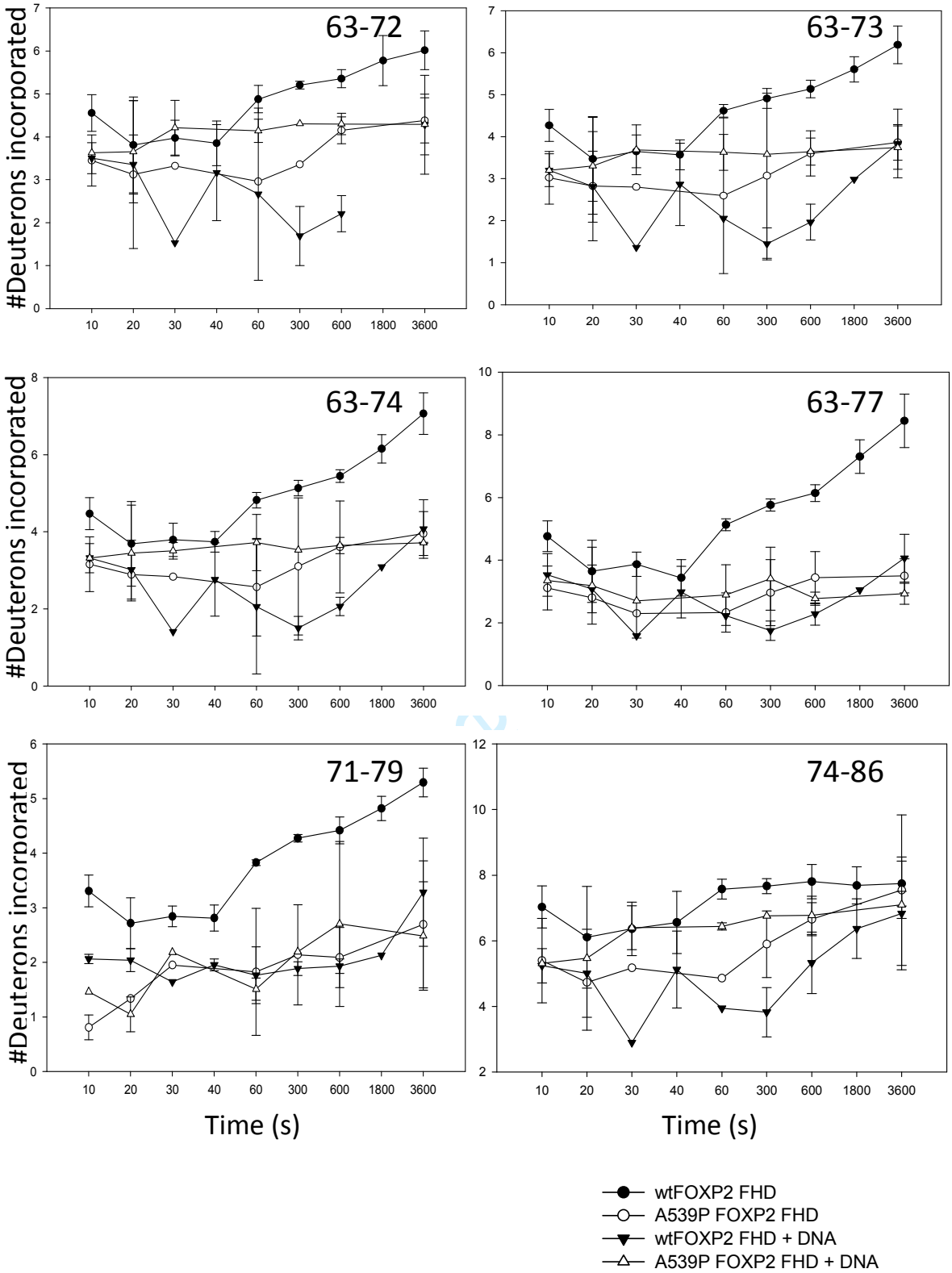
(A part of Supplementary Fig 8.)



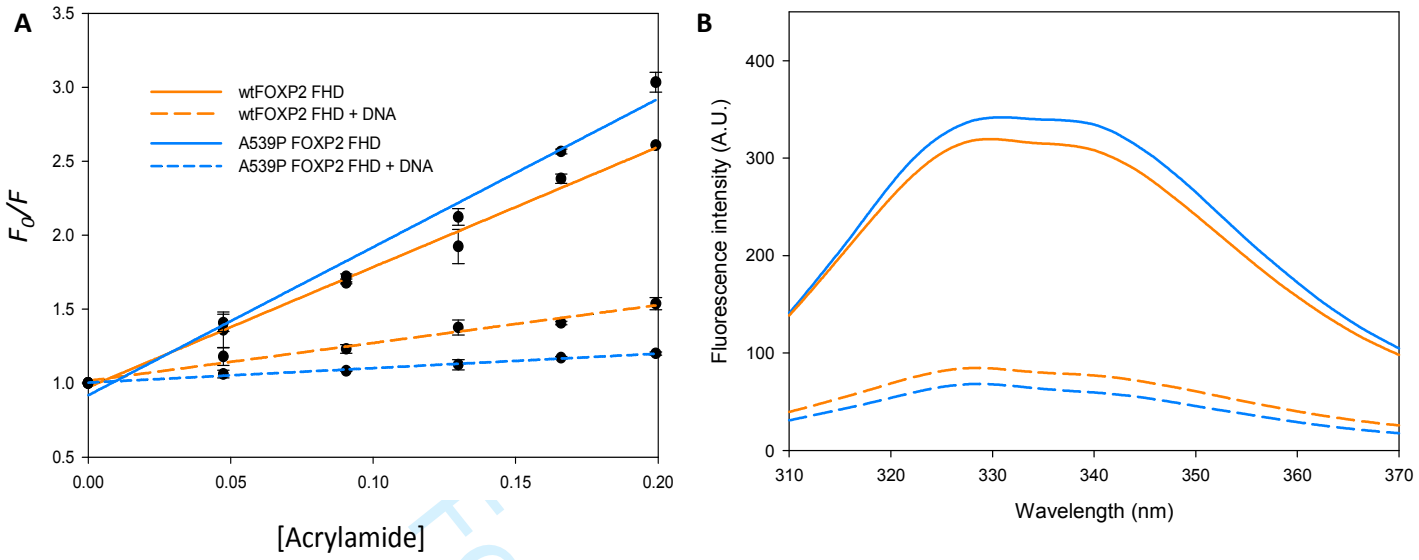
(A part of Supplementary Fig 8.)



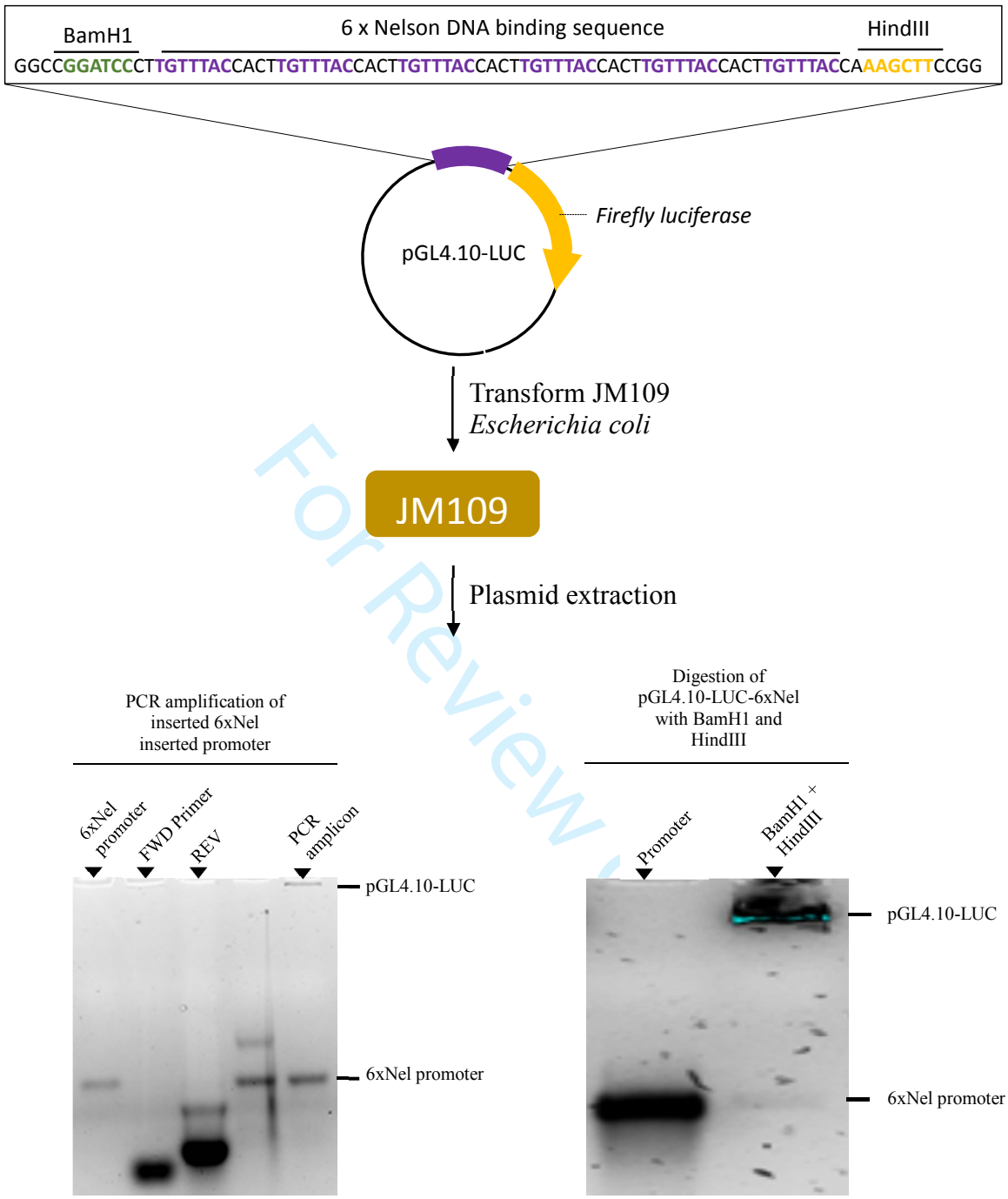
(A part of Supplementary Fig 8.)



Supplementary Fig 8 (Page 7 of 7). Deuteron uptake plots for all peptides all peptides identified for apo and DNA bound forms of the wild-type and A539P FOXP2 forkhead domain. Peptides are labelled by their residue numbers (top right corner of each plot). Lines are guides for the eyes only and not a fitted model. Error bars represent the standard deviation of three independent experiments.



Supplementary Fig 9. Dynamic fluorescence quenching of free and DNA bound wild-type or A539P FOXP2 forkhead domain. (A) Stern-Volmer plots for the wild-type and A539P FOXP2 forkhead domain with and without DNA. The emission intensity at 330 nm of the unquenched sample (F_0) was divided by the emission intensity at 330 nm (F) in the presence of increasing quencher concentration (50 – 200 mM acrylamide) following excitation at 295 nm. The data points were fitted with a linear regression and the gradient taken as the Stern-Volmer coefficient (K_a), all regressions displayed an $R^2 > 0.95$. Error bars represent the standard deviation of three independent replicates. (B) Fluorescence emission spectra of the wild-type (orange) and A539P (blue) FOXP2 forkhead domain in the absence (solid) and presence (dashed) of DNA. Significant quenching can be observed upon DNA binding likely due to the direct interaction of Trp573 with the DNA backbone stablising the transition moment of the excited Trp.



Supplementary Fig 10. Confirmation of the inserted 6xNel promoter in front of the *firefly* luciferase coding sequence in the pGL4.10-LUC. PCR amplification of the inserted promoter and digestion of the pGL4.10-LUC-6xNel plasmid confirmed successful insertion of the bespoke promoter sequence. Samples were resolved on 2% agarose gels with 0.5 X TBE buffer.

For Review Only

A

FOXP2 protein [Homo sapiens]

Sequence ID: AAI43868.1 Length: 714 Number of Matches: 1

Range 1: 486 to 714 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
489 bits(1258)	1e-170	Compositional matrix adjust.	228/229(99%)	228/229(99%)	0/229(0%)

```
Query  1    MSSEIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNAA  60
            MSSEIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNAA
Sbjct  486  MSSEIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNAA  545

Query  61   TWKNAVRHNLSLHKCFVRVENVKGAVWTVDEVEYQKRRSQKITGSPTLVKNIPTSLGYGA  120
            TWKNAVRHNLSLHKCFVRVENVKGAVWTVDEVEYQKRRSQKITGSPTLVKNIPTSLGYGA
Sbjct  546  TWKNAVRHNLSLHKCFVRVENVKGAVWTVDEVEYQKRRSQKITGSPTLVKNIPTSLGYGA  605

Query  121  ALNASLQAALAESSLPLLSNPGLINNASSGLLQAVHEDLNGSLDHIDSNGNSSPGCSPQP  180
            ALNASLQAALAESSLPLLSNPGLINNASSGLLQAVHEDLNGSLDHIDSNGNSSPGCSPQP
Sbjct  606  ALNASLQAALAESSLPLLSNPGLINNASSGLLQAVHEDLNGSLDHIDSNGNSSPGCSPQP  665

Query  181  HIHSIHVKEEPVIAEDEDCPMSLVTTANHSPELEXDREIEEEPLSEDLE  229
            HIHSIHVKEEPVIAEDEDCPMSLVTTANHSPELE DREIEEEPLSEDLE
Sbjct  666  HIHSIHVKEEPVIAEDEDCPMSLVTTANHSPELEDDREIEEEPLSEDLE  714
```

B

FOXP2 protein [Homo sapiens]

Sequence ID: AAI43868.1 Length: 714 Number of Matches: 1

Range 1: 384 to 714 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
594 bits(1531)	0.0	Compositional matrix adjust.	318/331(96%)	318/331(96%)	0/331(0%)

```
Query  1    MQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNLVSSVTMSKNMLETSPQSLP  60
            MQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNLVSSVTMSKNMLETSPQSLP
Sbjct  384  MQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNLVSSVTMSKNMLETSPQSLP  443

Query  61   QTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNIPMSSEIAPNYEFYKNADVR  120
            QTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNIPMSSEIAPNYEFYKNADVR
Sbjct  444  QTPTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNIPMSSEIAPNYEFYKNADVR  503

Query  121  PPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFPYFRRNAATWKNAVRHNLSLHKCFVR  180
            PPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTF YFRRNAATWKNAVRHNLSLHKCFVR
Sbjct  504  PPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNAATWKNAVRHNLSLHKCFVR  563

Query  181  VENVKGAVWTVDEVEYQKRRSQKITGSPTLVKNIPTSLGYGAALNASLQAALAESSLPLL  240
            VENVKGAVWTVDEVEYQKRRSQKITGSPTLVKNIPTSLGYGAALNASLQAALAESSLPLL
Sbjct  564  VENVKGAVWTVDEVEYQKRRSQKITGSPTLVKNIPTSLGYGAALNASLQAALAESSLPLL  623

Query  241  SNPGLINNASSGLLQAVHEDLNGSLDHIDSNGNXSPGXSPQPHXHSIHXKEXPVIAEXEV  300
            SNPGLINNASSGLLQAVHEDLNGSLDHIDSNGN SPG SPQPH HSIH KE PVIAE E
Sbjct  624  SNPGLINNASSGLLQAVHEDLNGSLDHIDSNGNSSPGCSPQPHIHSIHVKEEPVIAEDED  683

Query  301  CPMSLVTTANHSPXLEXDREXEEEPXSEDXE  331
            CPMSLVTTANHSP LE DRE EEEP SED E
Sbjct  684  CPMSLVTTANHSPELEDDREIEEEPLSEDLE  714
```

Supplementary Fig 11. Sequence data obtained for the forkhead domain of the full length (A) wild-type and (B) A539P FOXP2. The introduction of the A539P mutation can clearly be seen, red box. Mismatches aside from the intended mutation are caused by poor signal quality at the end of the sequence and presents as an X.

Chapter 4
General Discussion

4.1 FOXP2 forkhead domain dimers

Domain-swap dimerisation of the FOXP2 forkhead domain has been observed in DNA bound co-crystal structures and the forkhead domains purported to dimerise *in vitro* in the abscence of DNA (Stroud *et al*., 2006). In this study, several attempts were made to purify wild-type FOXP2 forkhead domain dimer to study but were not successful, in agreement with previous efforts (Blane and Fanucchi, 2015; Perumal *et al*., 2015). The dimer dissociation equilibrium constant of the wild-type FOXP2 forkhead domain is approximately 2.4 mM, (Perumal *et al*., 2015). This was readily apparent in size-exclusion studies where concentrations up to 300 μM displayed minimal formation of wild-type FOXP2 forkhead domain dimer (Chapter 2, Figure 2A). To obtain sufficient FOXP2 forkhead domain dimer for downstream studies, an obligate dimer mutant was designed and structurally characterised. Substitution of a single residue (Phe541) within the dimer hydrophobic core with a cysteine was sufficient to result in a covalently linked FOXP2 forkhead domain dimer (Chapter 2). The F541C mutation dramatically improved the proportion of FOXP2 forkhead domain dimer that could be purified (Chapter 2, Figure 2A). Furthermore, structural and electrophoretic mobility shift assay DNA binding studies showed that the disulfide-linked F541C FOXP2 forkhead domain dimer was fully folded and capable of binding the FOXP2 DNA consensus site. However, the mutant dimer aggregated rapidly at room temperature preventing detailed DNA binding studies. Attempts at determining binding constants for the F541C FOXP2 forkhead domain dimer, with fluorescence anisotropy and isothermal titration calorimetry studies, at room temperature resulted in spurious and irreproducible isotherms (data not published). Nevertheless, the F541C FOXP2 forkhead domain

provided useful insight into the formation of FOXP2 forkhead domain dimers. Firstly, the presence of a cysteine in the dimer interface acted as a covalent snare to trap associated FOXP2 forkhead domain monomers preventing their dissociation suggesting that the wild-type FOXP2 forkhead domain may be forming short lived unstable dimers that were not readily detectable. Secondly, the F541C FOXP2 forkhead domain dimer could be used in the low resolution electrophoretic mobility shift DNA binding assays as a marker to compare with the wild-type and monomeric A539P FOXP2 forkhead domains (Chapter 2, Figure 3).

To assess the possibility that dimerisation may be promoted by DNA binding electrophoretic mobility shift assays, fluorescence polarisation assays and isothermal titration calorimetry were performed with the wild-type and monomeric A539P FOXP2 forkhead domains. Stoichiometries determined from the DNA binding analyses reported that the wild-type FOXP2 forkhead domain preferentially bound as a monomer with no detectable dimer-DNA interaction (Chapter 2 Figures 3, 4 and 5). Together, the studies performed on the F371C FOXP3 mutant and those presented in this work suggest that the FOXP2 TF binds to DNA as monomer, translocates to the site of action and upon encountering another FOXP TF dimerises to form an active complex (Bandukwala *et al*., 2011).

4.2 The hinge loop affects the specificity and affinity of the FOXP2 forkhead domain

The hinge loop flexibility regulates the specificity of the FOXP2 forkhead domain. Like many TF DBDs, the FOXP2 forkhead domain binds several divergent response element sequences with different rates and affinities (Webb *et al*., 2017). To date, two regions have been identified as clear regulators of forkhead domain specificity, namely, the region spanning helix 2 to the hinge loop and the extended loop wings (which show the greatest sequence divergence across the FOX family) (Overdier *et al*., 1994; Obsil and Obsilova, 2008; Cirillo and Zaret, 2007; Murphy *et al*., 2004). The hinge loop displays

a high degree of sequence divergence except for 4 highly conserved amino acids, F(A/P)YF, where only the FOXP subfamily have an alanine at the second position (Chapter 1, Figure 4). The alanine residue is of clear importance in dimerisation of the FOXP2 forkhead domain as reversing the evolutionary proline to alanine mutation abolishes dimerisation of the FOXP subfamily forkhead domain (Chapter 2, Figure 2A; Chu *et al*., 2011; Bandukwala *et al*., 2011; Stroud *et al*., 2006). Therefore, it appears that the role of the hinge loop is two-fold: the hinge loop controls the propensity of the forkhead domain to form domain swapped dimers, and it alters the specificity of the forkhead domain for DNA.

The work presented in this thesis clearly demonstrates that the hinge loop plays a fundamental role in regulating the specificity and affinity of the FOXP2 forkhead domain for the consensus FOXP2 sequence. The A539P mutation resulted in a significant change in the thermodynamic signature of DNA binding to FOX consensus sequence (Chapter 2, Figure 5). The wild-type FOXP2 forkhead domain bound the FOX DNA consensus sequence with a greater affinity, lower enthalpic and greater entropic term (Chapter 2, Figure 5). To determine the origin of the difference in the binding energetics the change in heat capacity and the salt dependence of the binding was determined for both the wild-type and A539P FOXP2 forkhead domains. A significant decrease in the heat capacity of binding (ΔC_p) for the wild-type (-0.88 kJ/mol/K) was observed compared to that obtained for the A539P (-1.56 kJ/mol/K) FOXP mutant (Chapter 3, Figure 1) The change in heat capacity of DNA binding stems from burial of apolar (and to a lesser degree polar surfaces) in the protein-DNA interface, conformational changes in both the protein and DNA and vibrational relaxation of protein and DNA residues (Privalov *et al*., 2007; Prabhu and Sharp, 2005; Dragan *et al*., 2004; Ha *et al*., 1989; Spolar and Record; 1994; Jen-Jacobson *et al*., 2000b). Considering that the same oligonucleotide was used in both systems, that the hinge loop does not directly interact with the DNA, and that electrostatic contacts do not contribute significantly to the enthalpy of protein-DNA interactions, the increase in ΔC_p observed for the wild-type FOXP2 forkhead domain must stem from either a decrease in the

apolar surface area buried in the interface or conformational changes that occur in the wild-type that do not occur in the A539P mutant, or some combination of thereof (Privalov *et al*.,2011).

The DNA binding affinity of the FOXP2 forkhead domain is highly dependent on the ionic strength of the environment where approximately half the free energy of DNA binding originates from the formation of ionic contacts with the consensus site DNA (Chapter 3, Figure 2). The salt-dependence studies of DNA binding show that the wild-type FOXP2 forkhead domain makes an additional two ionic contacts with the consensus DNA sequence compared to the mutant (Chapter 3, Figure 2). The origin of these additional ionic contacts stems from the exclusion of an extra counter-ion from the DNA phosphate-sugar backbone either because of newly allowed ionic interactions or an increased footprint on the DNA backbone. This data suggests that the evolutionary mutation that is present in the FOXP subfamily forkhead domain reduces DNA binding specificity in exchange for improved DNA binding affinity. Considering the proposed role of the FOXP subfamily in binding two separate DNA strands, the additional ionic contacts may be necessary to provide sufficient attractive forces to stabilise the large ternary complex at the cost of shallower groove insertion (as observed in the crystal structures), although, further studies are required to prove this (Bandukwala *et al*., 2011; Stroud *et al*., 2006).

The entropic component of DNA binding for both the wild-type and A539P FOXP2 forkhead domain is dominated by counter ion exclusion ca. two ionic contacts from the protein-DNA interface with very little contribution from DNA conformational changes (Chapter 3, Figure 2). The conformational entropy is largely dominated by structural rearrangement of the DNA double helix during protein binding (Privalov *et al*., 2009). This observation agrees with the data obtained for several major groove binding DBDs where only minor readjustments of the DNA structure are required for insertion of the recognition motif into the groove (Privalov *et al*., 2009; Privalov *et al*., 2011). The large DNA structural rearrangements required for insertion of the recognition motif into the

much narrower minor groove incurs a significant enthalpic cost that is compensated for by a much larger entropy both in terms of conformational changes and counter ion exclusion (Privalov *et al*., 2009; Privalov *et al*., 2011). To assess the possibility of conformational changes contributing to the ΔC_p, hydrogen-deuterium exchange mass spectrometry was performed on the DNA bound and free forms of both the wild-type and A539P FOXP2 forkhead domains.

4.3 Conformational switching in the FOXP2 forkhead domain

The FOXP2 forkhead domain is comprised of two distinct subdomains: the recognition helix and the wing region (subdomain 1) make up the FOXP2 forkhead domain-DNA interface and the region spanning helix 1 and helix 2 makes up the domain swapped subdomain (subdomain 2) (Figure 9). A significant reduction in deuteron uptake was observed for both subdomains upon DNA binding (Chapter 3, Figure 3). Interestingly, significant stabilisation of subdomain 2 only occurs for the wild-type FOXP2 forkhead domain and not for the A539P mutant. The first and second helix are distal from the DNA and therefore the reduction in deuteron uptake must be the result of stabilisation of the helical secondary structure coupled with the formation of an interface between the two subdomains. This suggests that the wild-type FOXP2 forkhead domain adopts a conformationally flexible structure in solution, a requirement for domain swapping to occur (Newcomer *et al*., 2002). It is surprising then that very little wild-type FOXP2 forkhead domain dimer was observed by gel filtration studies perhaps due to a relatively small protein-protein interface (Stroud *et al*., 2006). The fact that a significant stabilisation of subdomain 2 is only observed in the wild-type (and not in the A539P mutant) FOXP2 forkhead domain corroborates the notion that DNA binding stabilises the monomeric form of the DBD.

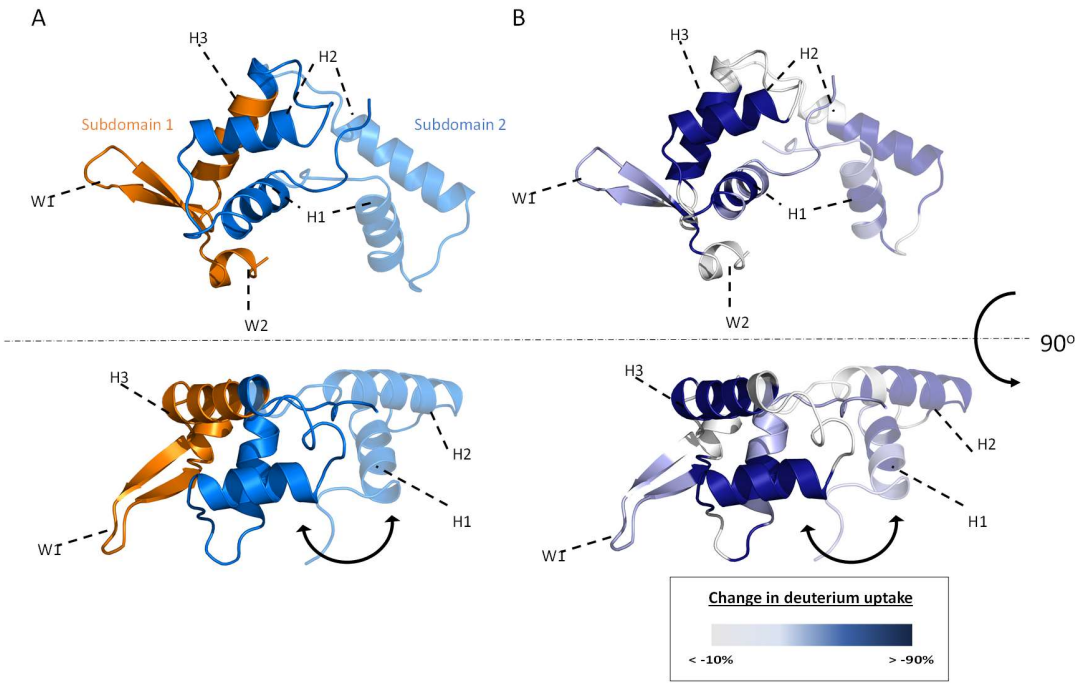


Figure 9: Cartoon representation of the FOXP2 forkhead domain divided into the two subdomains. (A) Subdomain 2 (blue) is exchanged between associated FOXP2 forkhead domain monomers (transparent). Subdomain 1 (orange) is responsible for the formation of the protein-DNA interface. (B) Difference heat map of the solution and DNA bound FOXP2 forkhead domain, determined by HDXMS in this study, accurately maps to the monomeric form of the forkhead domain. It is clear that subdomain 2 shows a significant change in conformational dynamics despite not directly interacting with the DNA.

The large degree of conformational change, in the wild-type but not the A539P FOXP2 forkhead domain, caused by DNA binding can account for the significant difference observed in the ΔC_p of binding. Little change in deuteron uptake occurs between the free and DNA bound forms of the A539P mutant likely due to restriction of the structural flexibility placed on the forkhead domain by the structural rigidity of the proline side chain. Together the reduced backbone dynamics and tryptophan solvent accessibility of the A539P mutation suggest formation of a larger solvent-inaccessible protein-DNA interface than that for the wild-type promoting the notion of deeper major

groove insertion (Chapter 3, Figures 3 and 4). The higher degree of salt-dependent contacts, the change in heat capacity of binding and the increased solvent accessibility of the tryptophan residues of the wild-type FOXP2 forkhead domain reveal that the dynamic hinge loop promotes non-specific backbone contacts over deep major groove insertion.

The FOXP2 forkhead domain displays features of the search-recognition mode switching mechanism employed by other eukaryotic transcription factors (Slutksky and Mirny, 2004; Iwahara and Levy., 2013; Zandarashvili *et al*., 2012). According to the solution structural studies, the FOXP2 forkhead domain resembles the DNA bound form, however, the energetic and dynamics studies reveal that significant conformational and dynamics stabilisation occurs during binding. Transcription factors, such as Egr-1, have been shown through NMR studies to switch between a scanning mode and site recognition mode (Iwahara and Levy, 2013). The search mode and recognition mode populations have distinct local structural and dynamic differences despite the global protein structure between them being similar. Electrostatic and hydrogen bonding contacts are made between the N-terminal residues of helix-1, C-terminal residues of helix-2 and the sugar-phosphate backbone of the consensus binding site (Chapter 1, Figure 8). These three regions are significantly stabilised by DNA binding and may act as a switch between the search and recognition modes of FOXP2 forkhead domain. The large discrepancy in energetics and dynamics changes in binding between the wild-type and the A539P FOXP2 forkhead domain promotes the notion that the composition of the hinge loop plays a fundamental role in defining the search and recognition mode populations. The search mode of the FOXP2 forkhead could be characterised by a more open conformation with fewer interactions between subdomain 1 and subdomain 2 (Figure 10) whereas the closed monomeric conformation (the recognition mode) is only formed when the necessary contacts with DNA backbone contacts are made by helices 1 and 2 stabilising the contacts between the two subdomains (Figure 10 recognition mode). Considering the large sequence divergence within the N-terminal residues of helix-1, the C-terminal residues of helix-2 and the

hinge loop regions of the FOX family forkhead domain (Chapter 1, figure 4) and the surprisingly high conservation of the recognised DNA binding sequence it is possible that evolutionary changes in these regions do not only serve to alter the propensity for dimerisation (hinge loop) but also the energetics of switching between the search and recognition modes of each family member. Such a mechanism would allow for fine temporal and context-specific tuning of the appropriate gene response in tissues that co-express several FOX family members.

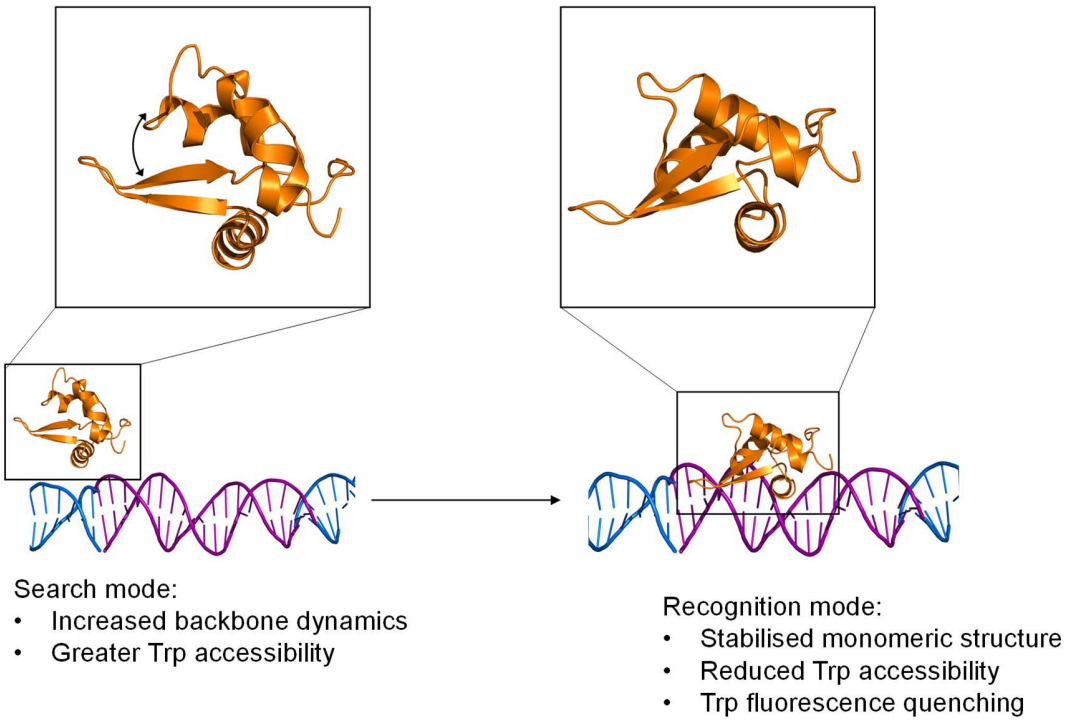


Figure 10: Switching between search and recognition modes in the forkhead domain may be controlled by the composition and inherent flexibility of the hinge loop region. The large degree of stabilisation of the FOXP2 forkhead domain (orange) upon binding to the consensus recognition site (purple) observed in HDXMS experiments suggests that DNA binding results in conformational restrictions that promote contacts between the two the subdomains of the protein.

4.4 In vivo mechanism

The activity of FOXP1/2/4 is regulated by the formation of homo- and heterodimers *in vivo* (Li *et al*., 2004). For cross dimerisation to occur, the proteins must exist as stable monomers, as with many other dimeric transcription factors, they can interact with DNA as monomers as well. Two well-studied studies systems, the bZip transcription factors and arc repressor, are known to prefer the so-called monomeric pathway whereby monomers non-specifically bind to DNA and co-localise at the site of action where they form functional dimers before transcriptional activation/repression occurs (Markovitz and Levy, 2009; Cranz *et al*., 2004; Kohler *et al*., 1999). Co-localisation along the DNA duplex both reduces both the search time of finding a suitable binding site as well as increases the probability of encountering a binding partner due to the reduced dimensionality of facilitated diffusion along the DNA backbone (Cranz *et al*., 2004; Kohler *et al*., 1999). It is possible that the FOXP subfamily members employ this tactic as well allowing formation of homo/heterotypic partners, depending on the degree of each partner's expression, at the site of activity. In this way the regulation of FOXP1/2/4 target genes can be controlled by the expression patterns of the respective transcription factors in a tissue context manner (Sin *et al*., 2015). If this is the case then it is likely that the forkhead domain of FOXP1/2/4 remains in a transient monomeric state capable of recognising and interacting with the consensus enhancer element, as observed in the studies presented in this thesis. In addition, there is crystallographic evidence that the FOXP2 forkhead domain interacts with binding partners (NFAT) in the monomeric form adding an additional layer of control in gene regulation (Bandukwala *et al*.,2011).

In vivo analysis, using luciferase reporter assays, of the wild-type and A539P mutation, purported to prevent forkhead domain dimerisation, shows that FOXP2 is capable of transcriptional activity with and without presumed dimerisation of the forkhead domain (Chapter 3, Figure 5). It is evident based on several studies that dimerisation of FOXP2, through the leucine zipper domain, is essential for transcriptional activity. However, here we have shown that dimerisation of FOXP2 forkhead domain is not essential to

activity. Perhaps, forkhead domain dimerisation is another step in regulating FOXP2 activity in the cell such that the population of the monomeric and dimeric forms of the forkhead domain is controlled by several environmental factors including the nature of binding partners and transcription complex formation or through post-translational modifications. Alternatively, dimerisation may occur as a critical step in a sequence of events leading to transcriptional complex formation and full activity. Therefore, in this work, a possible mechanism by which FOXP2 enacts transcriptional control is proposed (Figure 11). First, it makes sense that FOXP2 complies with the monomeric pathway identifying and binding target sites as monomers. Second, through random encounter or promoted by external factors, the FOXP2 monomers dimerise through the first dimerisation interface, the leucine zipper. Finally, the FOXP2 forkhead domains, once brought into close proximity of one another through internal motion of the protein itself or through recruitment of host factors, are able to undergo domain swapping as this process only requires exchange of the subdomain not directly associated with the DNA. Such a mechanism would agree with the purported role of the FOXP subfamily members being capable of binding and collating two independent sites. This mechanism would also provide an explanation for heterodimerisation as a regulatory mechanism with each FOXP subfamily member being capable of identifying different enhancer elements or by recruiting differential host factors.

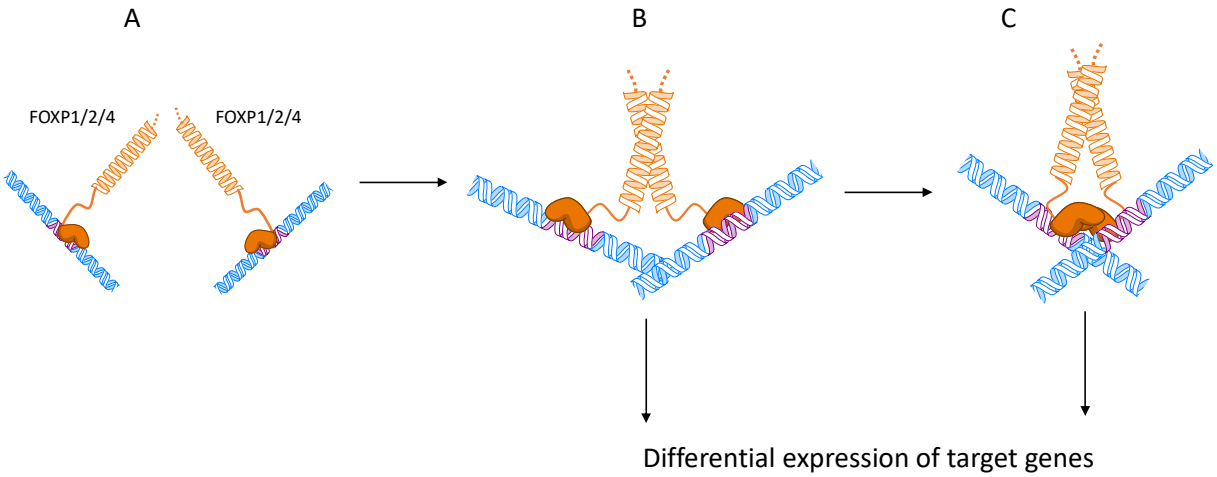


Figure 11: Proposed mechanism for differential gene expression controlled by the oligomeric state of the FOXP2 forkhead domain. (A) FOXP2 binds non-specifically and locates the target binding site (purple) as a monomer. (B) Two FOXP2 monomers bound at independent distal enhancer elements and dimerise through the leucine zipper possibly with the assistance of host factors. (C) The monomeric forkhead domains bound to the target site are brought closer together either through internal motion of the protein or with the assistance of additional host factors and dimerise to form the final active complex.

4.6 Limitations of the mutations used in the experiments

It must be noted that, although the mutations used throughout this study have shown minimal impact on the secondary and tertiary structures of the FOXP2 forkhead domain as determined by CD spectropolarimetry and instrinsic tryptophan fluorescence (Chapter 2, Figure 2) it cannot be said for certain that the mutant FOXP2 forkhead domains (both the A539P and the F541C mutants) do not affect subtle underlying interaction netoworks within the protein. Because of this, care should be taken during interpretation of the results presented herein. The mutations served to provide insight into the basis of the evolution of the FOXP subfamily members and not as an absolute

representation of FOX evolution. Further studies should be conducted on the unaltered FOXP2 forkhead domain structure using high fidelity methods such as NMR to further substantiate the claims made in this study.

4.7 Conclusion

The FOX proteins are a diverse and large family of transcription factors responsible for regulating critical processes in development and differentiation of almost every tissue in humans. FOX proteins all share a highly conserved DNA binding domain, the forkhead box, responsible for recognising a conserved DNA motif (TGTTTAC). Apart from their natural roles in differentiation and development nearly all of the FOX transcription factors are integral to the progression of congenital disorders as well as several cancers marking them as clinically relevant transcription factor family. Of the 15 FOX subfamilies, only the FOXP subfamily proteins require dimerisation to be transcriptionally active. Dimerisation of the FOXP subfamily members occurs at two distinct interfaces, one through a conserved leucine zipper domain and the other through domain-swapping of the forkhead domain. Domain-swapping of the FOXP forkhead domain is purported to occur due to mutation of a highly conserved alanine residue located in the so-called hinge loop connecting the two exchanged subdomains, one consisting of the recognition helix and the two wings characteristic of the forkhead domain motif and the other comprising of helices 1 and 2. The precise role of the hinge loop has not been determined, however, it clearly plays a role in both the propensity of the forkhead domain to form domain swapped dimers (FOXP subfamily members) and in controlling the specificity of forkhead domain. Here we provide evidence to suggest that the residue composition of the hinge loop is crucial in controlling the DNA binding mechanism of the forkhead domain despite not being directly involved in the protein-DNA interface. The energetics of binding for the wild-type and a monomeric (A539P) mutant FOXP2 forkhead domain differ substantially. The wild-type FOXP2 forkhead domain has a higher affinity for the cognate binding site but has a shallower major groove insertion and relies on the formation on additional sugar-phosphate backbone

contacts when compared to the A539P monomeric mutant. The results of the *in vivo* study also indicate that dimerisation of the FOXP2 forkhead domain is not essential for transcriptional activity, revealing a regulatory role for the oligomeric state of the FOXP2 forkhead domain. Additionally, forkhead domain dynamics are controlled by the flexibility of the hinge loop indicating that the forkhead domain regulates the switch between the non-specific solution structure (proposed to be the search mode) and the stable recognition mode upon specific site binding in accordance with the speed-stability paradox proposed for other eukaryotic transcription factors. Taken together this work offers insight into the role of the hinge loop in dimerisation, DNA binding and transcriptional activity of the FOXP2 transcription factor and provides a step in understanding the complex nature of transcriptional control by the FOXP subfamily of transcription factors.

Appendix A: Methodology

Plasmids

A pET-11a plasmid (Novagene, USA) containing a codon optomised FOXP2 forkhead domain (FOXP2FHD) with an N-terminal hexahistidine tag coding sequence and thrombin cleavage site for post purification removal of the hexahistidine tag (pET-11a-FOXP2FHD) was purchased form Genscript, USA. The plasmid contains an ampicillin resistance gene necessary for selection of transformed bacterial cells, a T7 promoter for IPTG induction of heterologous protein expression and a bacterial origin of replication for high copy number.

Transformation

A high plasmid concentration is required for downstream applications including PCR, sequencing and transformation of mammalian cells. To facilitate the demands for a high amount of each plasmid JM109 *Escherichia coli* were transformed with the respective plasmid. Briefly, super competent JM109 cells were incubated with the ~10 ng of the plasmid for 30 minutes at 4 ^{o}C. The cells were then heat shocked at 42 ^{o}C for 45 seconds and then immediately placed at 4 ^{o}C for 2 minutes. Fresh sterile SOC medium (2 % (m/v) tryptone, 0.5 % (m/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 MM glucose) was added to the cells and the cultures were incubated at 37 ^{o}C for 1 hour. Spread plates of the cultures were performed on sterile LB agar plates (1 % (m/v) tryptone, 0.5 % (m/v) yeast extract, 85 mM NaCl and 1 % (m/v) agar) containing 100 µg/mL ampicillin to select for transformed cells. Spread plates of empty SOC media and untransformed cells as well as spread plates of LB agar plates without ampicillin were performed as controls. The plates were incubated at 37

°C for 16 hours in a sterile incubator. Fresh sterile LB broth (1 % (m/v) tryptone, 0.5 % (m/v) yeast extract, 85 mM NaCl) was inoculated with colonies selected from the LB agar ampicillin-resistance selection plates and the cultures incubated at 37 °C until an OD_{600} of 0.6 was obtained. Small volume stocks of the culture were then stored in a 50 % glycerol solution at -80 °C until further use.

To ensure coding sequence integrity the plasmids were extracted from the stock culture using a GeneJet plasmid Miniprep kit (ThermoFischer Scientific, USA) and then sequenced by Inqaba biotechnology (South Africa).

Generation of FOXP2 forkhead domain mutants

The mutations were introduced using site-directed mutagenesis with the Quikchange lightening site-directed mutagenesis kit (Agilent, USA). Two mutations were introduced: F541C (for the dimerization studies) and A539P (for the monomer studies). Primers were designed to introduce the minimum number of base changes and to maintain codon optimisation for bacterial expression. The quality of the primers was assessed with Gene Runner software v3.04 (Hastings Software Inc., NY, USA). Following PCR, the plasmids were sequenced (Inqaba biotechnology, Pretoria, South Africa) to confirm the successful introduction of the mutation. JM109 *E.coli* were transformed with the mutant FOXP2 plasmids as previously described for plasmid storage and production when needed.

Heterlogous protein expression

The heterologous expression system was constructed by transforming T7 express *E.coli* with a pET-11a plasmid encoding for the wild-type, F541C or A539P FOXP2 forkhead domain following the same protocol as mentioned previously (see Transformation). 1 mL Stocks of transformed cells were stored at -80 °C in a 50 % glycerol solution. When needed frozen stocks of the transformed cells were thawed and used to inoculate a small volume starter culture before large scale expression. The starter culture consisted of a

1000-fold dilution of the transformed cell stock in fresh sterile 2 x YT media (1.6 % (m/v) tryptone, 1 % (m/v) yeast extract and 85 mM NaCl). The starter culture was incubated at 37 ^{O}C for 16 hours with vigorous shaking (230 rpm). The starter culture was then diluted 50-fold in fresh sterile 2 x YT media for large scale expression, typically 2 L of media, of wild-type, F541C or A539P FOXP2 forkhead domain. Following dilution of the starter culture the large-scale expression culture was grown to an OD_{600} of 0.6-0.8 at 37 ^{o}C before being cold-shocked at 4 ^{o}C for 10 minutes. Expression of the FOXP2 forkhead domain was controlled by an IPTG (isoproylthiogalactopyranoside) inducible T7 promoter system. For optimum soluble protein expression, the cultures were incubated at 20 ^{o}C for 20 hours after induction of expression by the addition of 0.6 mM IPTG. Expression trials were conducted prior to large scale expression to optimise expression temperature and the concentration of IPTG required to induce soluble expression. Cells were harvested by centrifugation at 5000 xg for 20 minutes, resuspended in immobilised metal-ion affinity chromatography (IMAC) equilibration buffer (see Protein purification) with 1 mg/mL Chicken egg lysozyme and stored at -20 ^{o}C for no more than 2 weeks.

Protein purification

Complete lysis by chemical, mechanical and enzymatic methods ensured total protein extraction from the stored expressed protein cell stocks. Expressed protein cell stocks were thawed completely at 20 ^{o}C after which 1 mM PMSF (phenylmethane sulfonylfluoride) was added to the stock to prevent heterologous protein degradation. The cells were then further lysed by 5 rounds of 10 s pulse sonication. 0.01 mg/mL DNase I was added to the cell lysate to reduce viscosity and assist in removal of DNA during subsequent purification steps. The insoluble fraction was removed from solution by centrifugation at 24 000 x g for 20 minutes.

Immobilised metal ion affinity chromatography (IMAC) was utilised for coarse purification of the FOXP2 forkhead domain-fusion from the bacterial cell milieu. The

soluble fraction of the cell lysate was loaded onto an Nickel-NTA affinity column equilibrated with 10 column volumes of IMAC equilibration buffer (10 mM tris-HCl pH 7.5, 30 mM imidazole and 500 mM NaCl). The column was then washed with five column volumes of a high salt concentration buffer (10 mM tris-HCl pH 7.5, 30 mM imidazole and 1.5 M NaCl) to remove bound DNA from the immobilised FOXP2 forkhead domain. The FOXP2 forkhead domain-fusion was eluted by a single step elution with IMAC elution buffer (10 mM tris-HCl pH 7.5, 300 mM imidazole, 500 mM NaCl). Following the first round of IMAC purification the protein was adialysed into thrombin cleavage buffer (100 mM Tris pH 8.0, 2 mM $CaCl_2$ and 100 mM NaCl). The N-terminal histag was removed by incubation of the protein with Thrombin for 4 hours at 20 ^{O}C. The undigested fusion protein, isolated histag and thrombin were then removed by a second round of IMAC followed by Benzamidine affinity chromatography. Fine purification of the FOXP2 forkhead domain was then performed by size-exclusion chromatography with a sephadex S75 16/60 column.

Assessment of protein purity

Samples of the purified protein were resolved on a 16 % discontinuous polyacrylamide gel with a tricine buffer system and stained with Coomassie brilliant blue R-250 to assess the degree of purity. Only protein samples with a purity of at least 95%, determined by densitometry, were pooled and used for subsequent experiments. DNA contamination levels were assessed using the ratio of 280 nm light absorbance to 260 nm light absorbance (A_{280}/A_{260}) by the protein sample. Absorbance measurements were performed on a Jasco V-630 Uv/Vis Absorbance spectrophotometer in scanning mode. Sample concentrations were adjusted to give an absorbance of 280 nm wavelength light in the range of 0.5 to 1. All samples used in experiments had an A_{280}/A_{260} of at least 1.7.

Protein concentration determination

Protein concentration was determined by absorbance of 280 nm wavelength light using the beer-Lambert law with an extinction coefficient of 16 900 $M^{-1}cm^{-1}$. Measurements were taken on a Jasco V630 UV/Vis absorbance spectrophotometer in single wavelength mode. Before all experiments protein samples were centrifuged for 10 minutes at 12 000 x g to remove aggregates. Scattering was accounted for by log extrapolation using absorbance values measured at 320 and 340 nm (Birdsall *et al*., 1983). The absorbance of the undiluted solution was determined by extrapolation of a linear regression fitted to the absorbance values of five dilutions. The determination of concentration was repeated in triplicate and the average taken as the final concentration of the sample for further experiments.

Secondary structure characterisation

Secondary structure of the wild-type, F541C and A539P FOXP2 forkhead domain was characterised by far-UV circular dichroism. Measurements were taken on a Jasco J-1800 circular dichroism spectropolarimeter scanning a wavelength range of 190-250 nm with a 2 mm pathlength. Samples were dialysed into a CD appropriate buffer (10 mM HEPES pH 7.5, 100 mM $NaSO_4$ and 0.02% NaN_3) and diluted to 10 μM for the measurements. The ellipticity, measured in millidegrees, can be normalised for the size of the protein using the following equation:

$$\theta_{MRE} = \frac{100 \times \theta}{C \times n \times l}$$

Where θ is the ellipticity measure in millidegrees, *C* is the concentration of the protein, *n* is the number of amino acids in the protein and *l* is the path length of the sample. Data with a high tension voltage greater than 600 were discarded before analysis.

Thermal unfolding of the wild-type and A539P FOXP2 forkhead domain were monitored with far-UV circular dichroism. Measurements were taken on a Jasco J-1800 circular dichroism spectropolarimeter following the 222 nm signal, a wavelength that gives an indication of the α-helical content of a protein, over a temperature range of 20-80 ^{o}C with a gradient of 1 ^{o}C/min.

Tertiary structure characterisation

Tertiary structural changes were analyzed using intrinsic tryptophan fluorescence on a Jasco FP-6300 fluorescence spectrophotometer with an excitation wavelength of 280 nm. The emission spectra were monitored over the wavelength range of 280−500 nm. Fluorescence intensity of each spectrum was then normalised to the maximum fluorescence intensity value for comparison due to the quenching effect of the disulfide bond.

Quaternary structure characterisation

To determine the quaternary structure of the wild-type, monomeric A539P and dimeric F541C FOXP2 FHD, size exclusion chromatography was performed using ~300 μM protein on a HiLoad 16/60 Superdex 75 prep grade size exclusion column (GE Healthcare, USA) equilibrated with SEC buffer (10 mM HEPES pH 7.5, 500 mM NaCl and 0.02% NaN_3). The eluted protein was detected by 280 nm light absorbance. More accurate measurement of the size of the protein may be determined by multiple angle light scattering. Following size exclusion, the dimeric F541C FOXP2 fraction was isolated, and free thiols were blocked by 1 h incubation (at 293 K in the dark) with 125 mM iodoacetamide before being resolved on a 16% nonreducing tricine SDS-PAGE gel to confirm disulfide-linked dimers.

Oligonucleotides

Duplex cognate DNA containing a single binding site as determined by Nelson et al. (bold), TTAGG**TGTTTAC**TTTCATAG, was prepared to a stock concentration of 200 μM (Integrated DNA technology, South Africa). Concentration of the oligonucleotide was determined by UV-absorption. The A_{260} values for a dilution series of five dilutions was fitted with a linear regression and extrapolated to give the absorbance value of the undiluted sample. The Beer-Lambert law was used to determine the concentration from the extrapolated absorbance value using an extinction coefficient of 256,016 $M^{-1}cm^{-1}$. Absorbance values were corrected for with a buffer blank prior to analysis. The determination of DNA concentration was performed in triplicate and the average was taken as the final concentration for downstream experiments.

Electrophoretic mobility shift assay

Electrophoretic mobility shift assays were performed to determine the formation of monomer−DNA and dimer−DNA complexes. Samples of increasing protein/DNA ratios were prepared using 0−6 μM FOXP2 forkhead domain with 1 μM cognate DNA. Binding reactions were conducted in EMSA binding buffer (10 mM HEPES pH 7.5, 100 mM KCl, 1 mM MgCl2, 0.1 mg/mL BSA, and 20% glycerol) and incubated at 4 °C for 30 min. Each reaction sample was then resolved on a 10% polyacrylamide gel (containing 20% triethylene glycol) with a 1× TBE running buffer at 150 V for 2 h at 4 °C.

Fluorescence anisotropy

Fluorescence anisotropy experiments were conducted on a PerkinElmer LS-50B fluorescence spectrophotometer fitted with an anisotropy filter. Briefly, 20 nM fluorescein- 5’ - labeled DNA was incubated with increasing concentrations of wildtype, A539P and F541C FOXP2 forkhead domain in binding buffer (10 mM

HEPES pH 7.5, 100 mM NaCl and 0.02% NaN_3). Fluorescence anisotropy measurements were taken with an emission wavelength of 520 nm following excitation with a wavelength of 494 nm. The best fit for the wild-type and A539P FOXP2 forkhead domain data was found to be an 1:1 protein to DNA binding model. The F541C FOXP2 forkhead isotherm could not be fitted with any of the traditional models, including a 1:1 protein to DNA binding model and a 2:1 binding model. Data were obtained in triplicate and averaged. Errors are the standard deviation of the averaged replicates.

Isothermal titration calorimetry

Preliminary DNA binding studies of the wild-type and A539P FOXP2 forkhead domain were conducted using isothermal titration calorimetry at a temperature of 20 ^{o}C. A typical titration consisted of 40 5 µL injections of 70-100 µM into 7-10 µM cognate DNA oligonucleotide (see Oligonucleotides). Protein and DNA were thoroughly dialysed against the same binding buffer (10 mM HEPES pH 7.5, 100 mM NaCl and 0.02% NaN_3) prior to the titration experiment. Protein into buffer and buffer into DNA blanks were performed to account for heats of dilution (mentioned as heats of saturation in the case where the last ten heats were average and substracted from the isotherm in protein-DNA binding experiments) and subtracted from the experimental titrations prior to data analysis. Data was fitted with an independent sites nonlinear regression using the NanoAnalyze software included with the instrument. Titrations were performed in triplicate with protein purified from independent cultures. Errors are the standard deviation of the averaged thermodynamic parameters.

Heat capacity studies

The heat capacity of binding for the wild-type and A539P FOXP2 forkhead domain were determined using isothermal titration calorimetry on a TA instruments NanoITC instrument (TA instruments,New Jersey, USA). The heat capacity value was obtained from the gradient of a linear regression fitted to the enthalpies of a series of five titrations

performed at the following temperatures: 10,15,20,25 and 30 ºC. A typical titration consisted of forty 5 µL injections of 70-100 µM protein into 7-10 µM cognate DNA (see Oligonucleotides). Both the protein and DNA were dialysed thoroughly against the same binding buffer (10 mM HEPES pH 7.5, 100 mM NaCl and 0.02% NaN_3). Blank titrations of protein into buffer and buffer into DNA were performed to account for heats of dilution. Finally, the heats of the last 10 injections were averaged and the value subtracted from all injections to account for residual heats of dilution before data analysis. Data was fitted by a nonlinear independent sites regression using the NITPIC software package (ref). Each titration experiment was performed at least in duplicate and the replicates were averaged. Errors are the standard deviation of the averaged replicates.

Salt effects on DNA binding

The entropy associated with DNA binding can be dissected into conformational changes of the protein or DNA and the exclusion of counter ions from the polyanionic backbone of DNA by protein residue sidechains (Privalov *et al*., 2011). This gives insight into the specificity of the contacts made between the protein and DNA as well as the degree of protein refolding that occurs during the binding process. The entropy of counter ion exclusion is dependent on the concentration of counters ions as well as the number of charge-charge contacts made between the protein and the DNA backbone. There is a growing body of evidence to support the following equation:

$$\log(K_a) = -N \cdot \log[Salt] + \log(K_{NEL})$$

By DNA binding experiments at increasing salt concentrations it is possible to determine the number of electrostatic contacts (*N*) made between the protein and the cognate DNA sequence. The *N* value is then used to alculate the number of ionic contacts by considering the occupancy of phosphate associated Na ions, ψ = 0.64 (Olmsted *et al*., 1995). Once the $K_{NEL,}$ the K_a at 1 M salt concentration, has been

determined it is possible to determine the weight of the contribution of counter ion exclusion to the entropy of binding.

Titrations of either 70-100 μM wild-type or 70-100 μM A539P FOXP2 forkhead domain into 7-10 μM cognate DNA (see Oligonucleotides) were performed at five salt concentrations in the range of 50-150 mM NaCl. Each titration consisted of ~40 5 μL injections of protein into 950 μL cognate DNA using a TA instruments NanoITC instrument (TA Instruments, New Jersey,USA). The number of electrostatic contacts formed between the protein and the DNA (*N*) and the non-electrostatic association constant (K_{NEL}) were determined by fitting a linear regression to the double log plot of salt concentration (independent variable) and the corresponding experimental association constant (dependent variable). Each titration was performed at least in duplicate and the replicate values were averaged. Errors are the standard deviation of the averaged values.

The entropy was dissected into contributions from the conformational changes and counterion exclusion using the Gibbs-Helmholtz equation. The entropy at 293 K was chosen as the standard to compare the wild-type FOXP2 forkhead domain to the A539P FOXP2 forkhead domain.

$$(1)\ \Delta G_{NEL} = -RTlnK_{NEL}$$
$$(\mathbf{2})\Delta G_{NEL} = \Delta H - T\Delta S_{NEL}$$
$$(3)\Delta S_{293K} = \Delta S_{NEL} + \Delta S_{EL}$$

The association constant at 1 M salt concentration, K_{NEL}, can be used to determine the ΔG_{NEL}, the non-electrostatic component of the free energy of DNA binding using the classical relationship between free energy and association constant (1). The enthalpic term is negligibly dependent on the concentration of salt allowing for the dissection of the electrostatic and non-electrostatic components of the entropic term using the Gibbs-Helmholtz free energy equation (2 and 3) (Privalov *et al*., 2007)

Hydrogen-deuterium exchange mass spectrometry

The *in vitro* structural dynamics of the apo and DNA bound forms of wild-type and A539P FOXP2 forkhead domain were studied by hydrogen-deuterium exchange mass spectrometry. Labelling, quenching and proteolytic cleavage experiments were performed on a LEAP Technologies PAL HDX system (LEAP Technologies, USA) coupled to an Agilent 1100 HPLC system (Agilent, USA). Mass spectrometry was performed on a AB Sciex 6600 TripleTOF mass spectrometer (AB Sciex, USA). Protein labelling consisted of a 10-3600s incubation of 20-30 µg of wild-type or A539P FOXP2 forkhead domain, with and without an equivalent mass of cognate DNA oligonucleotide, in 90 % D_2O at 20 ^oC. Samples were then quenched by a 2-fold dilution in quench buffer (20 mM Phosphate pH 4.5, 100 mM TCEP and 3.4 M guanadine-HCl) held at 0 ^oC. The protein was then fragmented by incubation on an inline Porozyme immobilised pepsin chromatography column (Life Technologies) at 4 ^oC for 30s before being desalted on a Acclaim PepMap trap column (0.3 x 5 mm) and subsequently loaded onto a Phenomenex Kinetex C_{18} reverse phase chromatography column (Phenomenex, USA). Peptides were separated onto the mass spectrometer at a flow rate of 200 µL/min with an elution gradient of 5-40 % buffer B (80% ACN/0.1%FA).

Peptide mass analysis was performed on a AB Sciex 6600 TripleTOF in Data Dependent Acquisition (DDA) mode. In DDA mode precursor scans were acquired from *m/z* 360-1500 using an accumulation time of 250 ms followed by 30 product scans, acquired from *m/z* 100-1800 at 100 ms each, for a total scan time of 3.3 s. Charge ions, falling between 1^+ - 5^+, were automatically fragmented in the Q2 collision cell using nitrogen gas. The collision energy was chosen automatically based on the *m/z* and the charge. Peptide identification was performed in PEAKS 6 (Bioinformatics Solutions Inc., USA). The following settings were used during the peptide search: no enzymatic cleavage specificity, precursor mass of 5 ppm and product error of 0.1 Da. The degree of deuteron incorporation was determined with the HDXaminer software package (Sierra Analytics, USA).

Dynamic fluorescence quenching

Dynamic fluorescence quenching can provide insight into the local structural changes of the wild-type and A539P FOXP2 forkhead domain. There are three tryptophans in the forkhead domain covering almost the entire structure. A single tryptophan is intimately involved in DNA recognition and binding (Trp573), therefore, changes in the fluorescence and solvent accessibility of the remaining two tryptophans as probed by dynamic fluorescence quenching provides more information on gross structural changes during binding. The Stern-Volmer coefficient ($K_{SV,}$ equation x) describes the solvent accessibility of the tryptophan based on the degree to which it is quenched by increasing concentrations of quencher in solution.

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

The Stern-Volmer coefficient (K_{SV}) was determined for the apo and DNA bound form of the wild-type and A539P FOXP2 forkhead domain. Fluorescence measurements were performed on a Jasco FP-6300 in emission wavelength scanning mode with an excitation wavelength of 295 nm. Samples consisted of 2 μM FOXP2 forkhead domain in binding buffer (10 mM HEPES pH 7.5, 100 mM NaCl and 0.02 % NaN_3) with increasing acrylamide (quencher) concentrations from 0-250 mM in increments of 50 mM. For DNA bound studies 2 μM FOXP2 forkhead domain was incubated with an equivalent of cognate DNA oligonucleotide for 30 minutes at 20 ^{o}C. Buffer only and DNA-buffer blanks were subtracted before data analysis. All experiments were performed in triplicate and averaged following analysis. Errors are the standard deviation of the averaged data.

Dual luciferase reporter assay

HEK293 cell cultures were plated at a density of 1 x 10^4 cells per well in a 96 well plate and grown to confluency in antibiotic free DMEM medium at 37 ^oC with 5% CO_2. Each well was transfected with 1.6 µg of pcDNA4 vector containing the full length wild-type FOXP2 coding sequence or the full length A539P FOXP2 coding sequence generated by site-directed mutagenesis of the wild-type coding sequence using the Quikchange Lightening site-directed mutagenesis kit followed as per manufacturer's instructions (Agilent, USA). Negative controls were performed by the addition of transfection reagent, transfection with 1.6 µg pGL4 vector under the control of a 6X FOXP2 consensus sequence synthesised by Integrated DNA Technology (Cape Town, South Africa) and transfection with 1.6 µg pRL-TK vector encoding *Renilla firefly* luciferase under control of a cognate tyrosine kinase promoter. Transfections were performed using Fugene HD transfection reagent as per manufacturer's instructions (Promega). Transfected cells were incubated for a further 24 hours. Luciferase assays were performed using the Dual-Glo luciferase assay kit followed as per manufacturer's instructions (Promega). Luminescence readings were taken on a GLoMax 96 Microplate Luminometer (Promega). Transfection efficiency was normalised by co-transfection with a pRL-TK vector encoding *Renilla* luciferase under control of a tyrosine kinase promoter and by taking the ratio of the *firefly* luciferase activity to *Renilla* luciferase activity. Replicates of at least three were performed for each FOXP2 sequence and control.

References

Alber, T. (1992) Structure of the leucine zipper. Curr. Opin. Genet. Dev. **2**, 205–210.

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.

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

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.

Bandukwala, H. S., Wu, Y., Feuerer, M., Chen, Y., Barboza, 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.

Benayoun, B. A., Caburet, S. and Veitia, R. A. (2011) Forkhead transcription factors: Key players in health and disease. Trends Genet.

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. **91**, 3127–3131.

Birdsall, B., King, R. W., Wheeler, M. R., Lewis, C. A., Goode, S. R., Dunlap, R. B., and Roberts, G. C. K. (1983) Correction for light absorption in fluorescence studies of protein-ligand interactions. *Anal. Biochem. 132*, 353–361.

Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. and Xie, X. S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. Proc. Natl. Acad. Sci. U. S. A. **103**, 5752–7.

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.

Bourne, Y., Arvai, a S., Bernstein, S. L., Watson, M. H., Reed, S. I., Endicott, J. E., Noble, M. E., Johnson, L. N. and Tainer, J. a. (1995) Crystal structure of the cell cycle-regulatory protein suc1 reveals a beta-hinge conformational switch. Proc. Natl. Acad. Sci. **92**, 10232–10236.

Bowers, J. M. and Konopka, G. (2012) The role of the FOXP family of transcription factors in ASD. Dis. Markers.

Brennan, R. G. (1993) The winged-helix DNA-binding motif: Another helix-turn-helix takeoff. Cell.

Campbell, A. J., Lyne, L., Brown, P. J., Launchbury, R. J., Bignone, P., Chi, J., Roncador, G., Lawrie, C. H., Gatter, K. C., Kusec, R., et al. (2010) Aberrant expression of the neuronal transcription factor FOXP2 in neoplastic plasma cells. Br. J. Haematol. **149**, 221–230.

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**.

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

Chiu, Y. C., Li, M. Y., Liu, Y. H., Ding, J. Y., Yu, J. Y. and Wang, T. W. (2014) Foxp2 regulates neuronal differentiation and neuronal subtype specification. Dev. Neurobiol. **74**, 723–738.

Chu, Y. P., Chang, C. H., Shiu, J. H., Chang, Y. T., Chen, C. Y. and Chuang, W. J. (2011) Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: Insight into its domain swapping and DNA binding. Protein Sci. **20**, 908–924.

Cirillo, L. A. and Zaret, K. S. (2007) Specific Interactions of the Wing Domains of FOXA1 Transcription Factor with DNA. J. Mol. Biol. **366**, 720–724.

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.

Coulocheri, S. A., Pigis, D. G., Papavassiliou, K. A. and Papavassiliou, A. G. (2007) Hydrogen bonds in protein-DNA complexes: Where geometry meets plasticity. Biochimie.

Cranz, S., Berger, C., Baici, A., Jelesarov, I. and Bosshard, H. R. (2004) Monomeric and Dimeric bZIP Transcription Factor GCN4 Bind at the Same Rate to Their Target DNA Site. Biochemistry **43**, 718–727.

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.

deHaseth, P. L., Gross, C. A., Burgess, R. R. and Record Jr., M. T. (1977) Measurement of binding constants for protein-DNA interactions by DNA- cellulose chromatography. Biochemistry **16**, 4777–4783.

Dragan, A. I., Read, C. M., Makeyeva, E. N., Milgotina, E. I., Churchill, M. E. A., Crane-Robinson, C. and Privalov, P. L. (2004) DNA binding and bending by HMG boxes: Energetic determinants of specificity. J. Mol. Biol. **343**, 371–393.

Dyson, H. J. and Wright, P. E. (2016) Role of intrinsic protein disorder in the function and interactions of the transcriptional coactivators CREB-binding Protein (CBP) and p300. J. Biol. Chem.

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., Brückner, M. K., Schreiweis, C., Winter, C., Sohr, R., et al. (2009) A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice. Cell **137**, 961–971.

Fanucchi, S., Shibayama, Y. and Mhlanga, M. M. (2014) Are genes switched on when they kiss? Nucleus **5**, 103–112.

Fontaine, F., Overman, J. and François, M. (2015) Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and obstacles. Cell Regen. (London, England) **4**, 2.

Gao, S., Wang, Y., Wang, M., Li, Z., Zhao, Z., Wang, R. X., Wu, R., Yuan, Z., Cui, R., Jiao, K., et al. (2017) MicroRNA-155, induced by FOXP3 through transcriptional repression of <i>BRCA1</i>, is associated with tumor initiation in human breast cancer. Oncotarget **8**, 41451–41464.

Gascoyne, D. M., Spearman, H., Lyne, L., Puliyadi, R., Perez-Alcantara, M., Coulton, L., Fisher, S. E., Croucher, P. I. and Banham, A. H. (2015) The forkhead transcription factor FOXP2 is required for regulation of p21waf1/cip1 in 143B osteosarcoma cell growth arrest. PLoS One **10**.

Givaty, O. and Levy, Y. (2009) Protein Sliding along DNA: Dynamics and Structural Characterization. J. Mol. Biol. **385**, 1087–1097.

Gormally, M. V., Dexheimer, T. S., Marsico, G., Sanders, D. A., Lowe, C., Matak-Vinkoviä, D., Michael, S., Jadhav, A., Rai, G., Maloney, D. J., et al. (2014) Suppression of the FOXM1 transcriptional programme via novel small molecule inhibition. Nat. Commun. **5**.

Gorman, J. and Greene, E. C. (2008) Visualizing one-dimensional diffusion of proteins along DNA. Nat. Struct. Mol. Biol. **15**, 768–774.

Gray, P. A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D. I., Tsung, E. F., Cai, Z., et al. (2004) Mouse brain organization revealed through direct genome-scale TF expression analysis. Science (80-.). **306**, 2255–2257.

Groszer, M., Keays, D. A., Deacon, R. M. J., de Bono, J. P., Prasad-Mulcare, S., Gaub, S., Baum, M. G., French, C. A., Nicod, J., Coventry, J. A., et al. (2008) Impaired Synaptic Plasticity and Motor Learning in Mice with a Point Mutation Implicated in Human Speech Deficits. Curr. Biol. **18**, 354–362.

Groszer, M., Keays, D. A., Deacon, R. M. J., de Bono, J. P., Prasad-Mulcare, S., Gaub, S., Baum, M. G., French, C. A., Nicod, J., Coventry, J. A., et al. (2008) Impaired Synaptic Plasticity and Motor Learning in Mice with a Point Mutation Implicated in Human Speech Deficits. Curr. Biol. **18**, 354–362.

Ha, J. H., Spolar, R. S. and Record, M. T. (1989) Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. J. Mol. Biol. **209**, 801–816.

Hadden, J. M., Convery, M. A., Déclais, A. C., Lilley, D. M. J. and Phillips, S. E. V. (2001) Crystal structure of the Holliday junction resolving enzyme T7 endonuclease I. Nat. Struct. Biol. **8**, 62–67.

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

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.

Harrington, R. E. (1992) DNA curving and bending in protein–DNA recognition. Mol. Microbiol.

Hegde, N. S., Sanders, D. A., Rodriguez, R. and Balasubramanian, S. (2011) The transcription factor FOXM1 is a cellular target of the natural product thiostrepton. Nat. Chem. **3**, 725–731.

Horn, D., Kapeller, J., Rivera-Brugués, N., Moog, U., Lorenz-Depiereux, B., Eck, S., Hempel, M., Wagenstaller, J., Gawthrope, A., Monaco, A. P., et al. (2010) Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. Hum. Mutat.

Iwahara, J. and Clore, G. M. (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. Nature **440**, 1227–1230.

Iwahara, J. and Levy, Y. (2013) Speed-stability paradox in DNA-scanning by zinc-finger proteins. Transcription **4**, 58–61.

Iwahara, J., Zweckstetter, M. and Clore, G. M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. Proc. Natl. Acad. Sci. U. S. A. **103**, 15062–15067.

Janowski, R., Kozak, M., Abrahamson, M., Grubb, A. and Jaskolski, M. (2005) 3D domain-swapped human cystatin C with amyloidlike intermolecular β-sheets. Proteins Struct. Funct. Genet. **61**, 570–578.

Jen-Jacobson, L. (1997) Protein-DNA recognition complexes: Conservation of structure and binding energy in the transition state. Biopolym. - Nucleic Acid Sci. Sect.

Jen-Jacobson, L., Engler, L. E., Ames, J. T., Kurpiewski, M. R. and Grigorescu, A. (2000b) Thermodynamic Parameters of Specific and Nonspecific Protein-DNA Binding. Supramol. Chem. **12**, 143–160.

Jen-Jacobson, L., Engler, L. E. and Jacobson, L. A. (2000a) Structural and thermodynamic strategies for site-specific DNA binding proteins. Structure **8**, 1015–1023.

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

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

Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A. and Schuck, P. (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. Anal. Chem. **84**, 5066–5073.

Kibbe, W. A. (2007) OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Res. 35*.

Klug, A. (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. Q. Rev. Biophys. **43**, 1–21.

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. **96**, 11735–11739.

Kolomeisky, A. B. (2011) Physics of protein–DNA interactions: mechanisms of facilitated target search. Phys. Chem. Chem. Phys. **13**, 2088–2095.

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.

Koo, C. Y., Muir, K. W. and Lam, E. W. F. (2012) FOXM1: From cancer initiation to progression and treatment. Biochim. Biophys. Acta - Gene Regul. Mech.

Kundu, S. and Jernigan, R. L. (2004) Molecular mechanism of domain swapping in proteins: An analysis of slower motions. Biophys. J. **86**, 3846–3854.

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

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

Laoukili, J., Stahl, M. and Medema, R. H. (2007) FoxM1: At the crossroads of ageing and cancer. Biochim. Biophys. Acta - Rev. Cancer.

Levine, M. (2010) Transcriptional enhancers in animal development and evolution. Curr. Biol.

Levy, Y., Onuchic, J. N. and Wolynes, P. G. (2007) Fly-casting in protein-DNA binding: Frustration between protein folding and electrostatics facilitates target recognition. J. Am. Chem. Soc. **129**, 738–739.

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**, 809–22.

Littler, D. R., Alvarez-Fernández, 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, C.-C., Richard, A. J., Datta, K. and LiCata, V. J. (2008) Prevalence of Temperature-Dependent Heat Capacity Changes in Protein-DNA Interactions. Biophys. J. **94**, 3258–3265.

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

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

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. **109**, E553–E562.

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

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

Luse, D. S. (2013) The RNA polymerase II preinitiation complex: Through what pathway is the complex assembled? Transcription.

Manning, G. S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. Q. Rev. Biophys. **11**, 179.

Manning, G. S. (1978) Limiting laws and counterion condensation in polyelectrolyte solutions. V. Further development of the chemical model. Biophys. Chem. **9**, 65–70.

Marcovitz, A. and Levy, Y. (2013) Obstacles may facilitate and direct DNA search by proteins. Biophys. J. **104**, 2042–2050.

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.

Marsico, G. and Gormally, M. V. (2015) Small molecule inhibition of FOXM1: How to bring a novel compound into genomic context. Genomics Data **3**, 19–23.

Massari, M. E. and Murre, C. (2000) Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. Mol. Cell. Biol. **20**, 429–440.

Medina, E., Córdova, C., Villalobos, P., Reyes, J., Komives, E. A., Ramírez-Sarmiento, C. A. and Babul, J. (2016) Three-Dimensional Domain Swapping Changes the Folding Mechanism of the Forkhead Domain of FoxP1. Biophys. J. **110**, 2349–2360.

Meerschaut, I., Rochefort, D., Revençu, N., Pètre, J., Corsello, C., Rouleau, G. A., Hamdan, F. F., Michaud, J. L., Morton, J., Radley, J., et al. (2017) FOXP1-related intellectual disability syndrome: A recognisable entity. J. Med. Genet. **54**, 613–623.

Miller, J., McLachan, a D. and Klug, a. (1985) Repetitive zinc-binding doains in the protein transcription factor IIA from Xenopus oocytes. EMBO J. **4**, 1609–1614.

Murphy, K. P. and Freire, E. (1992) Thermodynamics of structural stability and cooperative folding behavious in proteins. Adv. Protein Chem. **43**, 313–361.

Murphy, T. C., Saleem, R. A., Footz, T., Ritch, R., McGillivray, B. and Walter, M. A. (2004) The wing 2 region of the FOXC1 forkhead domain is necessary for normal DNA-binding and transactivation functions. Investig. Ophthalmol. Vis. Sci. **45**, 2531–2538.

Myatt, S. S. and Lam, E. W. F. (2008) Targeting FOXM1. Nat. Rev. Cancer.

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. **110**, 12349–12354.

Newcomer, M. E. (2002) Protein folding and three-dimensional domain swapping: A strained relationship? Curr. Opin. Struct. Biol.

Noy, A., Sutthibutpong, T. and A. Harris, S. (2016) Protein/DNA interactions in complex DNA topologies: expect the unexpected. Biophys. Rev.

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

Olmsted, M. C., Bond, J. P., Anderson, C. F. and Record, M. T. (1995) Grand canonical Monte Carlo molecular and thermodynamic predictions of ion effects on binding of an oligocation (L8+) to the center of DNA oligomers. Biophys. J. **68**, 634–647.

Olson, J. M., Asakura, A., Snider, L., Hawkes, R., Strand, A., Stoeck, J., Hallahan, A., Pritchard, J. and Tapscott, S. J. (2001) NeuroD2 Is Necessary for Development and Survival of Central Nervous System Neurons. Dev. Biol. **234**, 174–187.

Olson, J. M., Asakura, A., Snider, L., Hawkes, R., Strand, A., Stoeck, J., Hallahan, A., Pritchard, J. and Tapscott, S. J. (2001) NeuroD2 is necessary for development and survival of central nervous system neurons. Dev. Biol. **234**, 174–187.

Ostrow, A. Z., Kalhor, R., Gan, Y., Villwock, S. K., Linke, C., Barberis, M., Chen, L. and Aparicio, O. M. (2017) Conserved forkhead dimerization motif controls DNA replication timing and spatial organization of chromosomes in *S. cerevisiae*. Proc. Natl. Acad. Sci. **114**, E2411–E2419.

Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. and Sigler, P. B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution. Nature.

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**, 2755–66.

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

Pan, Y., Tsai, C. J., Ma, B. and Nussinov, R. (2010) Mechanisms of transcription factor selectivity. Trends Genet.

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.

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**, 5002–12.

Pilarsky, C., Wenzig, M., Specht, T., Saeger, H. D. and Grützmann, R. (2004) Identification and Validation of Commonly Overexpressed Genes in Solid Tumors by Comparison of Microarray Data. Neoplasia **6**, 744–750.

Prabhu, N. V. and Sharp, K. A. (2005) HEAT CAPACITY IN PROTEINS. Annu. Rev. Phys. Chem. **56**, 521–548.

Privalov, P. L., Dragan, A. I. and Crane-Robinson, C. (2011) Interpreting protein/DNA interactions: Distinguishing specific from non-specific and electrostatic from non-electrostatic components. Nucleic Acids Res.

Privalov, P. L., Dragan, A. I., Crane-Robinson, C., Breslauer, K. J., Remeta, D. P. and Minetti, C. A. S. A. (2007) What Drives Proteins into the Major or Minor Grooves of DNA? J. Mol. Biol.

Record, J. ~M. ~T., Anderson, C. ~F. C. F., Lohman, T. M. T. ~M., Record, M. T., Anderson, C. ~F. C. F., Lohman, T. M. T. ~M., Record, J. ~M. ~T., Anderson, C. ~F. C. F., Lohman, T. M. T. ~M., Record, M. T., et al. (1978) Thermodynamic Analysis of Ion Effects on Binding and Conformational Equilibria of Proteins and Nucleic-Acids - Roles of Ion Association or Release, Screening, and Ion Effects on Water Activity. Q.~Rev.~Biophys. **11**, 103–178.

Rentzeperis, D., Jonsson, T. and Sauer, R. T. (1999) Acceleration of the refolding of Arc repressor by nucleic acids and other polyanions. Nat. Struct. Biol. **6**, 569–73.

Ribeiro, J., Melo, F. and Schüller, A. (2015) PDIviz: Analysis and visualization of protein-DNA binding interfaces. Bioinformatics **31**, 2751–2753.

Riggs, A. D., Bourgeois, S. and Cohn, M. (1970) The lac represser-operator interaction. III. Kinetic studies. J. Mol. Biol. **53**, 401–417.

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B. and Mann, R. S. (2010) Origins of Specificity in Protein-DNA Recognition. Annu. Rev. Biochem. **79**, 233–269.

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

Rousseau, F., Schymkowitz, J. W., Wilkinson, H. R. and Itzhaki, L. S. (2001) Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. Proc. Natl. Acad. Sci. U. S. A. **98**, 5596–5601.

Sakaguchi, S., Miyara, M., Costantino, C. M. and Hafler, D. A. (2010) FOXP3 + regulatory T cells in the human immune system. Nat. Rev. Immunol.

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. **94**, 3842–3847.

Seeman, N. C., Rosenberg, J. M. and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. **73**, 804–808.

Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F. and Rabinovich, D. (1989) The conformation of the DNA double helix in the crystal is dependent on its environment. Nature **342**, 456–460.

Shigekawa, T., Ijichi, N., Ikeda, K., Horie-Inoue, K., Shimizu, C., Saji, S., Aogi, K., Tsuda, H., Osaki, A., Saeki, T., et al. (2011) FOXP1, an Estrogen-Inducible Transcription Factor, Modulates Cell Proliferation in Breast Cancer Cells and 5-Year Recurrence-Free Survival of Patients with Tamoxifen-Treated Breast Cancer. Horm. Cancer **2**, 286–297.

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

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. **97**, 8868–8873.

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

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.

Simicevic, J., Schmid, A. W., Gilardoni, P. A., Zoller, B., Raghav, S. K., Krier, I., Gubelmann, C., Lisacek, F., Naef, F., Moniatte, M., et al. (2013) Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics. Nat. Methods **10**, 570–576.

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

Slutsky, M. and Mirny, L. A. (2004) Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential. Biophys. J. **87**, 4021–4035.

Sollis, E., Graham, S. A., Vino, A., Froehlich, H., Vreeburg, M., Dimitropoulou, D., Gilissen, C., Pfundt, R., Rappold, G. A., Brunner, H. G., et al. (2016) Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. Hum. Mol. Genet. **25**, 546–557.

Spolar, R. and Record, M. (1994) Coupling of local folding to site-specific binding of proteins to DNA. Science (80-.). **263**, 777–784.

Spolar, R. and Record, M. (1994) Coupling of local folding to site-specific binding of proteins to DNA. Science (80-.). **263**, 777–784.

Srinivasan, A. R., Sauers, R. R., Fenley, M. O., Boschitsch, A. H., Matsumoto, A., Colasanti, A. V. and Olson, W. K. (2009) Properties of the nucleic-acid bases in free and Watson-Crick hydrogen-bonded states: Computational insights into the sequence-dependent features of double-helical DNA. Biophys. Rev.

Steffen, N. R., Murphy, S. D., Tolleri, G. W. and Lathrop, R. H. (2002) DNA sequence and structure: direct and indirect recognition in protein-DNA binding. Bioinformatics **18**, S22-30.

Strauch, M. A. (2001) Protein-DNA Complexes: Specific. Encycl. Life Sci. 1–7.

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.

Svingen, T. and Tonissen, K. F. (2006) Hox transcription factors and their elusive mammalian gene targets. Heredity (Edinb).

Takahashi, H., Takahashi, K. and Liu, F.-C. (2009) FOXP Genes, Neural Development, Speech and Language Disorders. In Advances in experimental medicine and biology, pp 117–129.

Takayama, K., Horie-Inoue, K., Ikeda, K., Urano, T., Murakami, K., Hayashizaki, Y., Ouchi, Y. and Inoue, S. (2008) FOXP1 is an androgen-responsive transcription factor that negatively regulates androgen receptor signaling in prostate cancer cells. Biochem. Biophys. Res. Commun. **374**, 388–393.

Takayama, K., Suzuki, T., Tsutsumi, S., Fujimura, T., Takahashi, S., Homma, Y., Urano, T., Aburatani, H. and Inoue, S. (2014) Integrative Analysis of FOXP1 Function Reveals a Tumor-Suppressive Effect in Prostate Cancer. Mol. Endocrinol. **28**, 2012–2024.

Takayama, Y. and Clore, G. M. (2011) Intra- and intermolecular translocation of the bi-domain transcription factor Oct1 characterized by liquid crystal and paramagnetic NMR. Proc. Natl. Acad. Sci. **108**, E169–E176.

Takayama, Y. and Clore, G. M. (2012) Interplay between minor and major groove-binding transcription factors Sox2 and Oct1 in translocation on DNA studied by paramagnetic and diamagnetic NMR. J. Biol. Chem. **287**, 14349–14363.

Teramitsu, I. (2004) Parallel FoxP1 and FoxP2 Expression in Songbird and Human Brain Predicts Functional Interaction. J. Neurosci. **24**, 3152–3163.

Todeschini, A. L., Georges, A. and Veitia, R. A. (2014) Transcription factors: Specific DNA binding and specific gene regulation. Trends Genet.

Tolosa, A., Sanjuán, J., Dagnall, A. M., Moltó, M. D., Herrero, N. and de Frutos, R. (2010) FOXP2 gene and language impairment in schizophrenia: Association and epigenetic studies. BMC Med. Genet. **11**.

Triulzi, T., Tagliabue, E., Balsari, A. and Casalini, P. (2013) FOXP3 expression in tumor cells and implications for cancer progression. J. Cell. Physiol.

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.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. and Luscombe, N. M. (2009) A census of human transcription factors: Function, expression and evolution. Nat. Rev. Genet.

Vernes, S. C., Nicod, J., Elahi, F. M., Coventry, J. A., Kenny, N., Coupe, A. M., Bird, L. E., Davies, K. E. and Fisher, S. E. (2006) Functional genetic analysis of mutations implicated in a human speech and language disorder. Hum. Mol. Genet. **15**, 3154–3167.

Vernes, S. C., Oliver, P. L., Spiteri, E., Lockstone, H. E., Puliyadi, R., Taylor, J. M., Ho, J., Mombereau, C., Brewer, A., Lowy, E., et al. (2011) FOXP2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet. **7**.

Viswamitra, M. A. (1982) Structural diversity in DNA: From monomer structures to oligonucleotides. Cold Spring Harb. Symp. Quant. Biol. **47**, 25–31.

Von Hippel, P. H. and Berg, O. G. (1989) Facilitated target location in biological systems. J. Biol. Chem.

Vuzman, D. and Levy, Y. (2012) Intrinsically disordered regions as affinity tuners in protein–DNA interactions. Mol. BioSyst. **8**, 47–57.

Wang, I.-C., Chen, Y.-J., Hughes, D., Petrovic, V., Major, M. L., Park, H. J., Tan, Y., Ackerson, T. and Costa, R. H. (2005) Forkhead box M1 regulates the transcriptional network of genes essential for mitotic progression and genes encoding the SCF (Skp2-Cks1) ubiquitin ligase. Mol. Cell. Biol. **25**, 10875–94.

Watkins, K. E., Dronkers, N. F. and Vargha-Khadem, F. (2002) Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. Brain **125**, 452–464.

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. and Jackle, H. (1990) The fork head domain: A novel DNA binding motif of eukaryotic transcription factors. Cell.

Weigelt, J., Climent, I., Dahlman-Wright, K. and Wikstrom, M. (2001) Solution structure of the DNA binding domain of the human forkhead transcription factor AFX (FOXO4). Biochemistry **40**, 5861–5869.

Wong, K. K., Gascoyne, D. M., Soilleux, E. J., Lyne, L., Spearman, H., Roncador, G., Pedersen, L. M., Møller, M. B., Green, T. M. and Banham, A. H. (2016) FOXP2-positive diffuse large B-cell lymphomas exhibit a poor response to R-CHOP therapy and distinct biological signatures. Oncotarget **7**, 52940–52956.

Wright, P. E. and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J. Mol. Biol. **293**, 321–331.

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.

Wunderlich, Z. and Mirny, L. A. (2009) Different gene regulation strategies revealed by analysis of binding motifs. Trends Genet.

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. **109**, E1724–E1732.

Zandarashvili, L., Esadze, A., Vuzman, D., Kemme, C. A., Levy, Y., and Iwahara, J. (2015) Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proc. Natl. Acad. Sci. 112*, E5142–E5149.

Zhao, H., Zhou, W., Yao, Z., Wan, Y., Cao, J., Zhang, L., Zhao, J., Li, H., Zhou, R., Li, B., et al. (2015) Foxp1/2/4 regulate endochondral ossification as a suppresser complex. Dev. Biol. **398**, 242–254.

Zhou, H.-X. (2011) Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. Proc. Natl. Acad. Sci. **108**, 8651–8656.

Zhou, J., Wang, Y., Wang, Y., Yin, X., He, Y., Chen, L., Wang, W., Liu, T. and Di, W. (2014) FOXM1 modulates cisplatin sensitivity by regulating EXO1 in ovarian cancer. PLoS One **9**.

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.