School of Molecular and Cell Biology

University of the Witwatersrand

Johannesburg

Phosphorylation of the FOXP2 forkhead domain: the effect on structure and DNA binding using phosphomimetics

---

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

Ashleigh Anne Blane

940523

Supervisor: Dr S Fanucchi

05/06/2017

DECLARATION

I, **Ashleigh Anne Blane** (**940523**), am a student registered for the degree of MSc dissertation in the academic year 2015.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Thesis/Dissertation/Research Report **HAS NOT** been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

signature  _____ 05 day of June 2017

RESEARCH OUTPUTS

Conference poster presentation

Blane, A., and Fanucchi, S. Phosphorylation of the FOXP2 forkhead domain: the effects on structure and DNA binding using phosphomimetics. 25th SASBMB congress. East London, July 2016.

Publications

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

Webb, H., Steeb, O., Blane, A., Rotherman, L., Aron, S., Machanick, P., Dirr, H. & Fanucchi, S. (2017) The FOXP2 forkhead domain binds to a variety of DNA sequences with different rates and affinities. *The Journal of Biochemistry*, 1-10.
performed molecular docking studies

ABSTRACT

Transcription factors are proteins that are involved in the regulation of gene expression and are responsible for the tight control of transcription allowing a cell to react to changes in its environment. Transcription factors are thus highly regulated by a variety of mechanisms which include phosphorylation. Forkhead box P2 (FOXP2) is a transcription factor expressed in multiple tissues during embryonic development. FOXP2 like other FOX proteins contains a DNA binding domain known as the forkhead domain (FHD). The effect of phosphorylation of serine 557 in the FHD on the structure and DNA binding was done using a glutamate mutant (to mimic phosphorylation) and an alanine mutant (as a control). Structural characterisation was performed using size exclusion chromatography (SEC), intrinsic fluorescence and far-UV circular dichroism. The effect of phosphorylation on DNA binding was observed using electrophoretic mobility shift assay (EMSA) and isothermal titration calorimetry (ITC). Far-UV circular dichroism and intrinsic fluorescence of the mutants and wild type did not reveal any significant secondary or tertiary structural changes. SEC however revealed a decrease in dimerisation propensity in the Ser557 mutants when compared the wild type (WT). EMSA revealed that DNA binding of S557E is only observed at protein concentrations 40 times in excess of the DNA. DNA binding of the WT and S557A mutants is observed at 5 times and 20 times excess protein respectively. However, using ITC no DNA binding is observed for either S557E or S557A FOXP2 FHD. Thus, it is possible that phosphorylation of serine 557 in the FOXP2 FHD could be a mechanism for inactivation of FOXP2.

[257 words]

To my Mom and Dad, for all the coffee and cookie swirl Cs.

“You were a million years of work”, said God and his angels with needle and thread.

- Sleeping at last

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Sylvia Fanucchi for the endless guidance, encouragement, motivation and inspiring me with her passion for biochemistry. Without her none of this would have been possible.

Thank you to all my colleagues in the PSFRU, especially the FOXy group and my good friends (Alison Williams, Gary Robertson, Gavin Morris, Kim Wiid, Mia Steeb and Naadira Pahad), for all the encouragement, assistance and laughs.

I am enormously grateful to the National research foundation and the Wits postgraduate merit award for funding.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AIRS	Ambiguous interaction restraints
A260	Absorbance at 260 nm
A280	Absorbance at 280 nm
AU	Arbitrary units
CD	Circular dichroism
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphate
ds	Double stranded
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
FHD	Forkhead domain
FOX	Forkhead box
HADDOCK	High ambiguity driven docking
His-tag	Oligo-histidine tag
HTH	Helix-turn-helix
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl βD-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_d	Dissociation constant
LZ	Leucine zipper
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
Phos-WT	Phosphorylated-WT
PIC	Preinitiation complex
PKB	Protein kinase B
PMSF	Phenylmethylsulfonyl fluoride
PTM	Post translational modification
RNA	Ribonucleic acid
RPM	Revolutions per minute

SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SOC	Super optimal broth with catabolite repression
SUMO	Small ubiquitin like modifier
TBE	Tris borate EDTA
TBP	TATA box binding protein
TF	Transcription factor
Tris	Tris(hydroxymethyl)-aminomethane
UV	Ultra violet
Ve	Elution volume
Vo	Void volume
W-HTH	Winged-helix-turn-helix
WT	Wild type
ZF	Zinc finger
θMRE	Mean residue ellipticity

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

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1 Introduction

1.1. Transcription

Transcription factors (TF) are proteins that are involved in the regulation of gene expression. These proteins are responsible for the tight control of transcription allowing a cell to react to even slight changes in environment in an instant. This is achieved by allowing activation or repression of multiple genes (Huffman & Brennan, 2002).

In eukaryotes, RNA polymerase II transcribes DNA forming mRNA. RNA polymerase forms a complex with multiple general transcription factors on the DNA which is known as the basal transcription complex or the preinitiation complex (PIC) (Fig 1.1). Transcription is initiated when the PIC forms on specific regions of the DNA known as the promoter (Van Dyke *et al*, 1988; Roeder, 1991; Conaway & Conaway, 1993; Grünberg & Hahn, 2013). General transcription factors are responsible for the transcription of housekeeping genes which are expressed in order to maintain basic cellular function (Pedersen *et al*, 1999). In addition to general transcription factors there are other TFs that are responsible for transcription activation or repression of facultative genes. Facultative genes are differentially expressed depending on the tissue type and in response to stimuli (Burley, 1997).

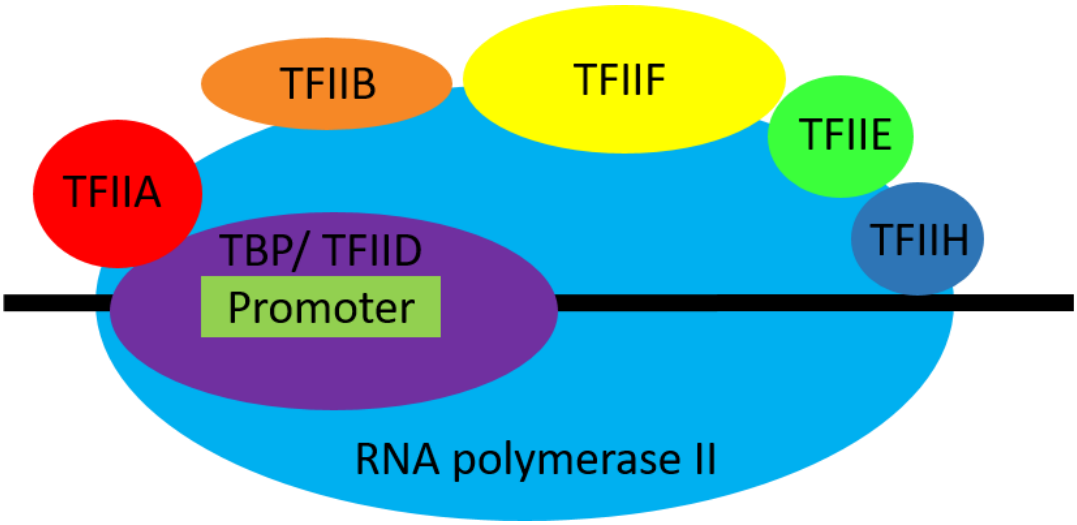

Fig 1.1. The preinitiation complex or basal transcription complex that forms on the promotor and initiates transcription. The TATA binding protein (TBP) binds the promotor followed by the other general transcription factors (TFIIA, TFIIB, TFIIF, TFIIE, TFIIE and TFIIH) and RNA polymerase II (Roeder, 1996).

1.2. Transcription factors

TFs tend to have multiple domains which generally facilitate DNA binding and protein-protein interactions. The DNA binding domains are involved in specifically binding regulatory sites on the DNA to regulate expression of specific genes. TFs either activate or repress transcription. This often occurs via protein-protein interactions with transcription machinery (Latchman, 1997). There are thousands of TFs classed into many families based on their DNA binding domains. Some examples of these domains include helix-turn-helix (HTH), zinc finger (ZF), leucine zipper (LZ), winged helix-turn-helix (W-HTH) and helix-loop-helix (Burley, 1997).

The classical HTH motif (Fig 1.2) has three alpha helices orientated in a triangular like shape. The turn between helix 2 and helix 3 is typically sharp and often well conserved. Helix 3 is inserted into the major groove during DNA binding. There are many variations of the basic HTH motif one of which is the winged-helix-turn-helix (W-HTH). The major difference between the HTH and W-HTH is the presence of a β-strand hairpin wing on the C-terminal end. The W-HTH, like the HTH motif, binds DNA by insertion of helix 3 into

the major groove. The wing however may make additional DNA contacts with the minor groove (in some cases) (Aravind *et al*, 2005).

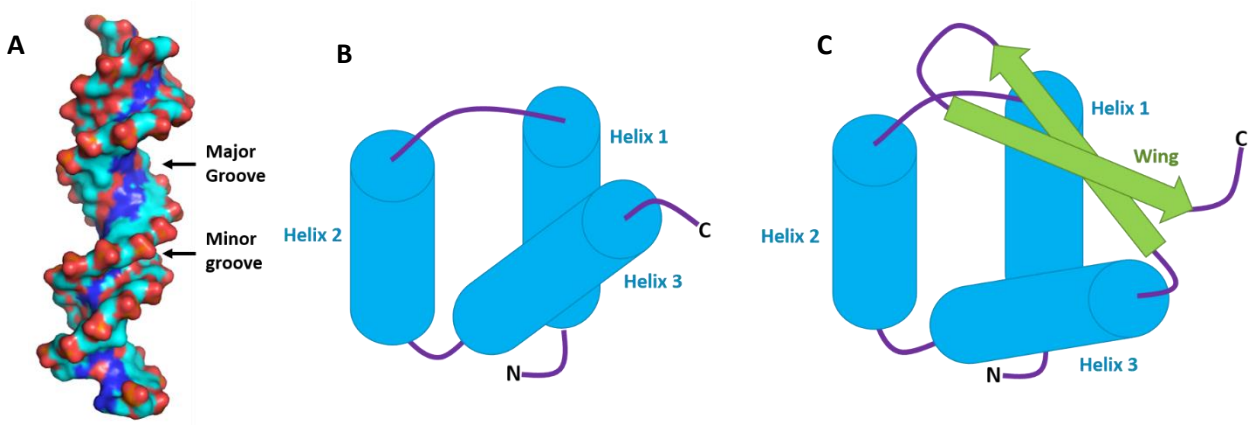

Fig 1.2. A: The major and minor groove of B-DNA. B: Schematic diagram of the helix-turn-helix. C: Schematic diagram of the winged-helix-turn-helix (W-HTH) motifs. The helices are displayed as blue cylinders and the β-strand hairpin in green. In the HTH and W-HTH helix 3 is inserted into the major groove of the DNA. Image adapted from (Aravind *et al*, 2005).

1.3. Regulation of transcription

Regulation of the transcription of genes is very important. It is this differential regulation of genes that allows different cell types to form in multi-cellular organisms. Regulation of transcription is achieved by regulatory DNA binding sites, transcription activators and repressors.

1.3.1. DNA binding

Transcription is initiated when the PIC forms on the promoter region of DNA (Fig 1.1). Common promotors in eukaryotes are the TATA box (found 25 to 30 bp upstream of the transcription start site), initiator sequence (which spans the transcription start site) and the downstream promotor element (which is located 25 bp downstream of the transcription start site) (Fig 1.3). First the TBP (TATA box binding protein) binds directly to promotor or proximal promotor elements after which several other components of the PIC assemble, including polymerase II, at the promotor and spans from position -30 to +30 (Sainsbury *et*

al, 2015). Eukaryotes contain regulatory regions of DNA known as enhancers or silencers to which specific TFs bind in order to regulate transcription. TFs which act as activators bind enhancer regions on the DNA whereas TFs which act as repressors bind the silencer regions on the DNA. Neither transcription activators nor repressors directly bind the promotor but they may interact with the PIC which forms at the promotor. These enhancer and silencer sites in the DNA can be found kilobases before and after the transcription start sites (Fig 1.3). These regulatory regions of DNA contain binding sites for multiple TFs and thus the control of transcription is also dependent on the types of TFs present in the cell at any given time (Pedersen *et al*, 1999).

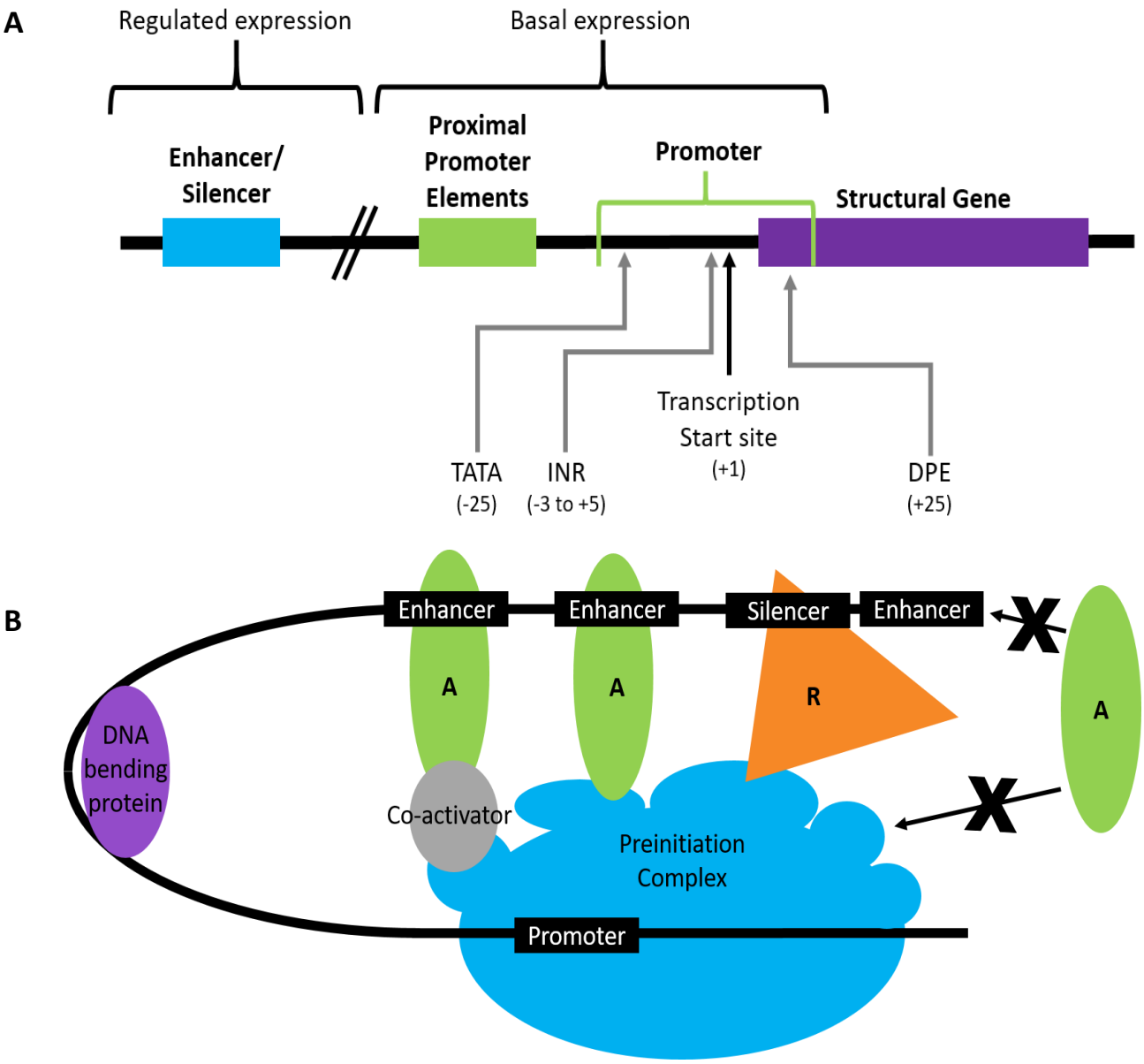

Fig 1.3. A: Structure of RNA polymerase II promoter. The structural gene (purple) is transcribed when the transcription initiation complex forms at the promoter region (green) Common eukaryotic promoters are indicated in grey. Enhancer/ silencer sites (blue) are bound by TFs which either activate or repress transcription of specific genes. **B: Transcription factors bound to the preinitiation complex and enhancer/ silencers.** The preinitiation complex (PIC, blue) is bound to the promotor. The DNA is looped by a DNA bending protein (purple). Activators (A, green) bound to enhancers either interact with the PIC directly or indirectly through co-activators (grey). Repressors (R, orange) bind silencers and either prevent activators from binding enhancer sites or prevent them from binding the PIC.

1.3.2. Transcription activators and repressors

Transcription activation is defined as the increase in the rate of gene expression whereas transcription repression is defined as the decrease in the rate of gene expression. The manner in which TFs either act as repressors or activators varies. TFs function as transcription activators by interacting with either the PIC or other co-activators (Fig 1.3). They activate transcription by either increasing the rate of the basal transcription complex formation or enhancing its level of activity (Roberts, 2000). Many TFs act as transcriptional repressors. The first mechanism of transcriptional repression is competitive binding to the transcriptional activator's DNA binding site (Fig 1.4). Transcription repression is sometimes achieved by binding to the activation domain on a transcription activator or by interaction (either directly or indirectly) with members of the basal transcription complex. Transcription repressors can also bind to their own specific regulatory DNA sequences (Cowell, 1994; Latchman, 1997). Some TFs can even act as both activators and repressors (Sharrocks, 2001).

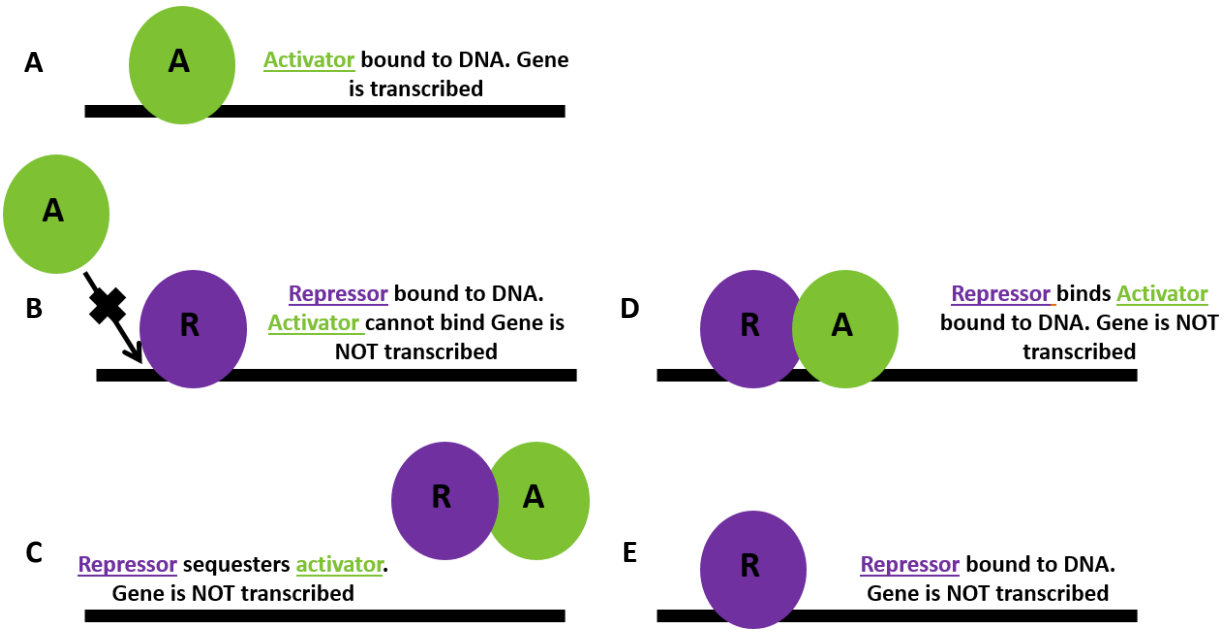

Fig 1.4. Mechanisms of repression of gene expression by transcription factors. A: Transcription of the gene is active when the activator (A) is bound to enhancer region of the DNA. **B**: Transcription is repressed when the repressor (R) competitively binds the activator's (A) DNA binding site. **C**: The repressor (R) binds the activator (A) preventing the activator from binding the DNA, therefore repressing gene expression. **D**: The repressor (R) directly binds the activator (A) or the preinitiation complex thereby inhibiting activation of gene expression. **E**: The repressor (R) binds to a specific inhibitory DNA sequence (silencer) preventing gene expression.

1.3.3. Regulation of transcription factors

In order for cells to respond to stimuli and changes in environment, differential expression of specific genes is required. This is generally achieved through signal transduction pathways, which through multiple proteins, relay information from the cell's surface to the nucleus. TFs are generally the end targets of these pathways as they are responsible for regulation of gene expression (Tootle & Rebay, 2005). Since TFs are responsible for the tight regulation of gene expression they are, themselves highly regulated. Regulation of TFs occurs by a variety of mechanisms of which binding of ligands, interactions with other proteins or post translational modifications are examples (Latchman, 1997).

1.3.4. Post translational modifications

Post translational modification (PTM) of TFs is one of the ways in which TFs with highly similar binding sequences can regulate very distinct genes. TFs have been shown to undergo many types of PTMs. These include phosphorylation, acetylation, ubiquintation, glycosylation and sumoylation (Benayoun & Veitia, 2009). PTMs of TFs, either individually or in combination, have been shown to influence protein-protein interactions, DNA-interactions, cellular localisation, activity and stability of TFs (Tootle & Rebay, 2005). PTMs have been shown to occur on multiple sites on a protein (multisite modification). This is important as multiple pathways can converge on the same target TF via different modifications. These different modification patterns present on the target TF at any given time can cause different structural and functional changes (Benayoun & Veitia, 2009; Tootle & Rebay, 2005)

Glycosylation is the addition of O-linked β-N-acetylglucosamine to the hydroxyl group of serine and threonine residues. Serine and threonine residues can also be phosphorylated post translationally. Phosphorylation is the addition of a phosphate group to these residues. Phosphorylation and glycosylation therefore acts on the same residues and in fact may act competitively (Vosseller *et al*, 2001). Lysine residues however are the target for acetylation, sumoylation and ubiquitination. Acetyltransferase mediates the transfer of an acetyl group to specific lysine residues (Bannister *et al*, 2000). Sumoylation is the attachment of a small protein, known as Small ubiquitin like modifier (SUMO), to specific lysine residues. Ubiquitination like sumoylation is the addition of ubiquitin to specific amino acid residues and is catalysed in a multiple step process (Geiss-Friedlander & Melchior, 2007). Sumoylation, ubiquitination, acetylation, glycosylation and phosphorylation of specific residues affect the activity, location, structure and stability of their targets (Gill, 2004; Verger *et al*, 2003)

1.3.5. Phosphorylation

One of the most studied PTMs is phosphorylation as it is a regulatory mechanism for a wide variety of proteins including TFs. Phosphorylation is the esterfication of a phosphate group to the hydroxyl group of either tyrosine, serine or threonine residues, which is facilitated by kinases (Tarrant & Cole, 2009). Kinases generally phosphorylate residues by transferring the γ- phosphate from high energy donor molecules such as adenosine triphosphate and guanosine triphosphate. Kinases fall into two very broad classes (S/T kinases and Y kinases) (Tootle & Rebay, 2005).

Phosphorylation generally acts as a regulatory mechanism for proteins, including TFs, in two ways. Firstly, the addition of the charged phosphate group can induce local and long range conformational changes in proteins. This is often due to alteration of electrostatic interactions and hydrophobicity of specific regions on proteins. The change in conformation is a regulatory mechanism as this may change the nature of the binding sites therefore affecting protein-protein and protein-DNA interactions that occur (Johnson & Barford, 1993). The second way in which phosphorylation regulates protein activity is by steric and electrostatic repulsion, due the addition of the negative charge, which blocks the proteins' active sites or binding sites without any conformational changes occurring (Johnson & Barford, 1993). Many TFs such as members of the Ets (E26 transformation-specific) family (Wasylyk *et al*, 1997), NF-IL6 and STAT3 (Akira, 1997) have been shown to be regulated by phosphorylation in this way.

1.4. FOX proteins

FOX (forkhead box) proteins are a family of TFs which contain a highly conserved winged helix domain (called the forkhead domain), which is typically involved in DNA binding. Despite the conserved forkhead domain (FHD), FOX proteins are functionally diverse. FOX proteins have been shown to be involved in embryonic development in various tissues (Shu *et al*, 2007), the immune system (Koh *et al*, 2009), metabolism (Friedman & Kaestner, 2006) and have been implicated in cancer (Sharma *et al*, 2005).

Fig 1.5. displays the first solved crystal structure of a FOX FHD: that of the FOXA3 forkhead domain (FHD) bound to DNA (Clark *et al*, 1993). This can be considered a generic model for all FOX FHDs. The FOX forkhead domain is a W-HTH motif and is between 90 and 100 residues in length. The FHD comprises of three α-helices (helix 1, helix 2 and helix 3) and two flanking loops which form wing like structures (wing 1 and wing 2) (Friedman & Kaestner, 2006; Clark *et al*, 1993). These winged regions are the least conserved regions of the FHD. The FHD binds DNA by insertion of helix 3 into the major groove. Helix 3 is highly conserved in all FOX proteins (Friedman & Kaestner, 2006; Clark *et al*, 1993).

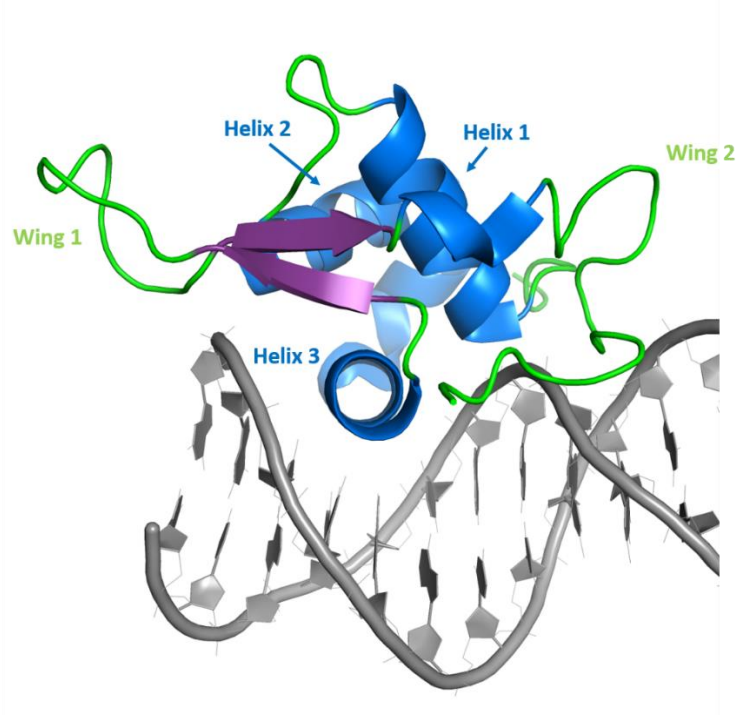

Fig 1.5. FOXA3 forkhead domain bound to DNA. Helix 3 is wedged into the major groove of the DNA (grey). PDB ID: 1VTN (Clark et al. 1993). Image rendered with the PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

1.4.1. FOXP subfamily

FOX proteins are arranged into various subfamilies (A-S) based on different structural characteristics. The four members of the FOXP subfamily are distinguished from other FOX proteins as the FHD is found at the C-terminal rather than the N-terminal end (Chu *et al*, 2011). In addition to the FHD, the FOXP proteins also share additional similar structural motifs which include a glutamine-rich region, a zinc finger and a leucine zipper which are unique to the FOXP subfamily (Fig 1.6) (Wang *et al*, 2003). The glutamine rich region has been shown in other TFs to function as a transcriptional activation/ repression domain (Shu *et al*, 2001; Wang *et al*, 2003). The leucine zipper and zinc finger motifs have been implicated in homo and heterodimer formation of the full length FOXP proteins (Wang *et al*, 2003).

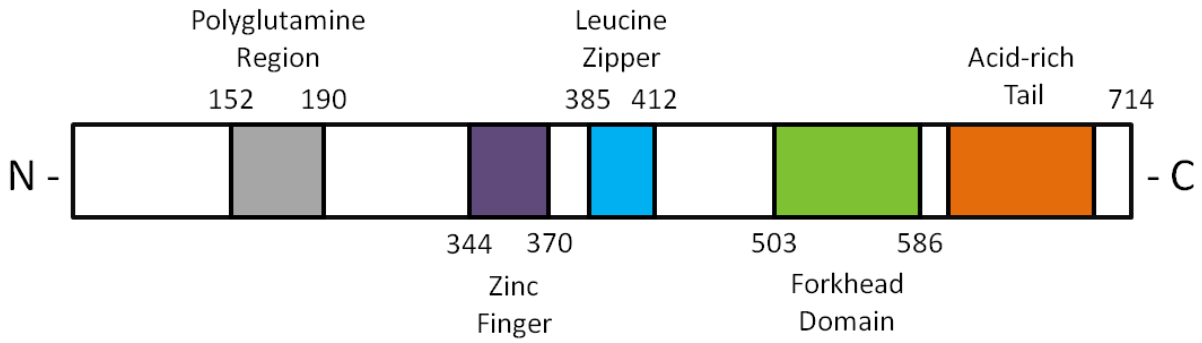

Fig 1.6. Schematic diagram depicting the various domains of FOXP proteins. The polyglutamine region (grey), zinc finger (purple), leucine zipper (blue), the forkhead domain (green) and acid rich tail (orange). FOXP2 numbering is used.

1.2.1.1. FOXP function

The four members of the FOXP subfamily display diverse functions. FOXP1 has been shown to be involved in tumour suppression (Banham *et al*, 2001). FOXP3, one of the more studied members has been implicated in regulation of the immune system (Koh *et al*, 2009). FOXP4, the least studied member, has been found to be expressed in multiple tissues within the heart, lungs, kidneys, testis, brain and liver (Teufel *et al*, 2003).

FOXP2 is expressed in intestinal, neural and cardiovascular tissues during embryonic development (Shu *et al*, 2001). The importance of FOXP2 was first recognised when the R553H mutation in its forkhead domain was discovered. The mutation was discovered in a family, known as the KE family, suffering from a speech disorder classified as verbal dyspraxia. Verbal dyspraxia is defined as the inability to sequence mouth movements resulting in difficulty with pronunciation. In addition to verbal dyspraxia, members of the KE family, despite displaying close to normal nonverbal IQ, also display defects of written language and receptive linguistic skills (Fisher & Scharff, 2009).

1.2.1.2. FOXP structure

Members of the FOXP subfamily, like other FOX proteins, bind DNA by insertion of helix 3 into the DNA major groove. FOXPs however, unlike other FOX proteins, have the ability to form a domain swapped dimer with the FHD. In addition to the formation of a domain swapped dimer of the FHD, the full length FOXP proteins may also form homo and heterodimers (Wang *et al*, 2003; Li *et al*, 2004). This is mediated by the leucine zipper/ zinc finger region.

In domain swapping of FOXP2, FOXP3 and FOXP1, helix 3, strand 2 and strand 3 are swapped by two FOXP2 monomers. Domain swapping of the FHD is made possible due to the extension of the turn connecting helix 2 and helix 3 (Bandukwala *et al*, 2011; Chu *et al*, 2011; Stroud *et al*, 2006). There are multiple residues that interact along the hydrophobic dimer interface of the FOXP2 domain swapped dimer which include Tyr509, Tyr531, Tyr540, Trp533, Trp548, Phe507, Phe534, Phe538 and Phe541 from both monomers (Stroud *et al*, 2006). It is possible that the FOXP2 domain swapped dimer may simultaneously bind two separate DNA strands, due to the proximity of the two DNA binding surfaces of the domain swapped dimer. It has been postulated that the FOXP2 domain swapped dimer may loop DNA or even mediate interchromosomal contacts in this way (Stroud *et al*, 2006). In Fig 1.7 the domain swapped dimer and monomer of the FOXP2 FHD are displayed (PDB ID: 2A07). The main difference between the dimer and the monomer is that in the dimer helix 2 is extended to form the hinge region.

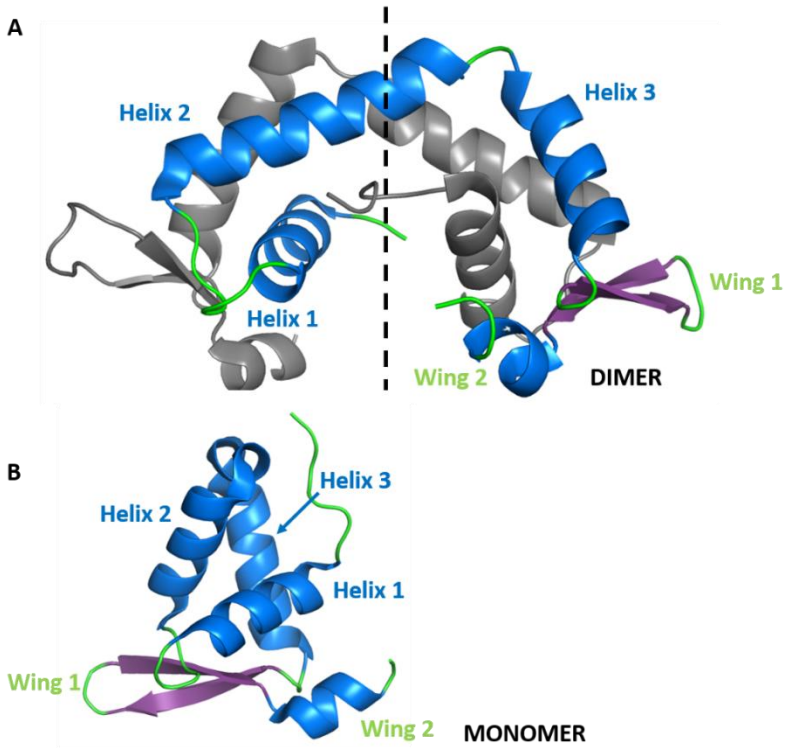

Fig 1.7. The FOXP2 FHD domain swapped dimer (A) and monomer (B). One monomer of the domain swapped dimer is displayed in grey. The dimer is formed by the interchange of helix 3, strand 1 and strand 2 (PDB ID: 2A07)(Stroud *et al*, 2006). Image rendered with PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

1.2.1.3. FOXP2 DNA binding and sequence specificity

The FOXP2 FHD has been shown to bind DNA in the dimeric and monomeric form. As with other FOX proteins the monomeric FHD of FOXP2 interacts with DNA by inserting helix 3 into the major groove. The main DNA interactions formed by the monomeric FHD, as determined from the crystal structure (PDB ID: 2A07) occur through Asn550, Arg553 and His554. Asn550 and His554 from direct hydrogen bonds with the DNA while Arg553 forms a water mediated hydrogen bond with the DNA bases. Hydrogen bonds to the DNA phosphate backbone are formed by Ser557, Thr547, Arg504, Trp573, Tyr509, Arg583 and Arg584. In addition to these hydrogen bonds various residues also form extensive van der Waals interactions with the DNA. Certain aromatic and hydrophobic residues from helix 1 and helix 3 interact with the DNA stabilising the DNA protein interaction (Stroud *et al*, 2006).

As the forkhead domain of FOX proteins is so highly conserved, it is predicted that FOX proteins have similar DNA sequences to which they will bind. The general FOX consensus sequence is 5' [A/G][C/T][A/C]AA[C/T]A 3' (Carlsson & Mahlapuu, 2002). Regions in the FHD such as wing 1 and wing 2 as well as helix 3 confer DNA binding specificity. FOX proteins including FOXA3 utilise the wing 1 and wing 2 regions to extensively bind to the DNA minor groove and backbone. As the FOXP2 forkhead domain has shorter wing 1 and wing 2 regions, there are limited DNA contacts. It has been postulated that this could result in a decrease in binding affinity of FOXP2 compared to some other FOX proteins (Stroud *et al*, 2006). Per the crystal structure, FOXP2 utilises many van der Waals forces and fewer hydrogen bonds to make contacts in the major groove compared to other FOX proteins, which could allow the FOXP2 FHD more flexibility in terms of DNA binding sequences (Stroud *et al*, 2006). The core binding sequence specifically for FOXP2 was defined as 5' A[C/T]AAATA 3' (Wang *et al*, 2003). More recently however, a cognate sequence was determined by Nelson *et al*, (2013), which is 5'-TG**TTTA**C-3'.

1.5. Phosphorylation of FOX proteins

Phosphorylation has been studied in multiple members of the FOX protein family, and most extensively in the FOXO subfamily. FOXO proteins have been shown to be phosphorylated by ATK/protein kinase B (PKB) at specific sites, in multiple domains including the FHD, *in vivo*. When these ATK/PKB sites are phosphorylated, FOXOs become susceptible to further phosphorylation by casein kinase-1 (Biggs *et al*, 1999). Phosphorylation, *in vivo*, of a residue in the wing 2 region of the FOXO FHD, results in a decrease in DNA binding affinity due to the addition of the negative charge in this basic region (Zhang *et al*, 2002). In addition, phosphorylation of residues in the nuclear localisation sequence, found adjacent to the FHD, prevents the import of FOXO into the nucleus. This occurs because the phosphorylated FOXO can bind the protein 14-3-3 which shields the NLS sequence from nuclear import proteins (Brunet *et al*, 1999). FOXOs have also been shown to be phosphorylated by additional kinases and this phosphorylation alters DNA binding (Obsil & Obsilova, 2008). Other members of the FOX family have been shown to be subject to phosphorylation *in vivo* including FOXP3 (Nie *et al*, 2013), FOXC2 (Berry *et al*, 2005) and FoxM1 (Chen *et al*, 2009).

1.6. Possible phosphorylation of FOXP2

As mentioned previously the FHD binds DNA by inserting helix 3 into the major groove of the DNA. The interaction is mediated by multiple hydrogen bonds and Van der Waals interactions. There is a highly-conserved serine in most FOX FHDs (Fig 1.8) found on the C-terminal end of helix 3, in FOXP2 this is Ser557. This serine residue has been shown to be phosphorylated in some FOX proteins which include FOXA3 (Shiromizu *et al*, 2013), FOXO3 (Lehtinen *et al*, 2006) and FoxO1(Brent *et al*, 2008). In FOXA3 Ser172 was identified as a phosphorylation site using mass spectrometry in colorectal cancer cells (Shiromizu *et al*, 2013). In addition to this, in FOXO3, Ser215 (Lehtinen *et al*, 2006) and in FoxO1, Ser218 (Brent *et al*, 2008) have been shown to be phosphorylated *in vivo*. The role of phosphorylation of this serine residue in FOXA3 and FOXO3 however, has not been determined. In FoxO1 mammalian Ste20-like kinase 1 phosphorylates four serine residues in the FHD, one of which is Ser215 which corresponds to Ser557 in FOXP2. The phosphorylation of these residues completely disrupts DNA binding of FoxO1 *in vitro* (Brent *et al*, 2008). In FOXP2 it is unknown if Ser557 is a phosphorylation site. However, given its conservation across all members of the FOX family and the fact that the serine in this position in other family members has been shown to be phosphorylated, Ser 557 can be considered as a likely phosphorylation site. In the crystal structure solved by Stroud *et al*, (2006) (PDB ID: 2A07) Ser557 makes a hydrogen bond to the DNA backbone (Fig 1.10A). If Ser557 in FOXP2 is phosphorylated, it is currently unknown which kinase may be responsible for its phosphorylation, and furthermore its role when phosphorylated is unknown.

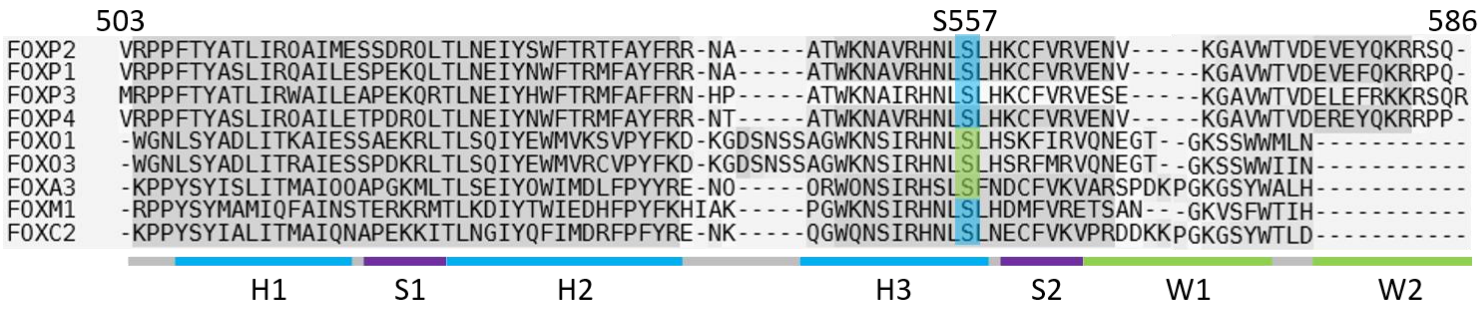

Fig 1.8. Multiple sequence alignment of the FHD of multiple FOX proteins. Ser557 is highlighted in blue in FOX proteins where it has not been shown to be phosphorylated or in green in proteins where it is known to be phosphorylated. The secondary structure is represented by the coloured bar below and the helices (H), strands (S) and wings (W) are indicated. Multiple sequence alignment done using T-Coffee (Notredame *et al*, 2000; Di Tommaso *et al*, 2011).

Phosphorylation of Ser557 would result in three additional oxygen atoms. As the p*K*a of the phosphate group is 5.7 (Śmiechowski, 2010) two of the oxygen atoms will be negatively charged at physiological pH. The effect of phosphorylation could therefore result in hydrogen bonds forming at the new hydrogen bond acceptor sites (Fig 1.9). Alternatively, the additional negative charge after phosphorylation could hinder DNA binding because of electrostatic repulsion and steric hindrance.

To determine what effect phosphorylated Ser557 will have on the structure and function of the FOXP2 FHD, a phosphomimetic was used in this study. A phopshomimetic is the mutation of a potentially phosphorylated amino acid (serine, threonine or tyrosine) to a charged amino acid such as aspartate or glutamate in order to mimic the phosphorylated residue. Phosphomimetics have been used successfully to mimic the phosphorylation of serine, threonine and even tyrosine (Lowy, 1996 and Potter and Hunter, 1998). The use of aspartate and glutamate to mimic phosphorylation is not without challenges. For instance, there is a size difference between the phosphate group and the carboxylate group of aspartate and glutamate as well as two rather than three additional oxygens. This means that in many instances using aspartate and glutamate to mimic phosphorylated residues may not work due to differences in bulk and charge.

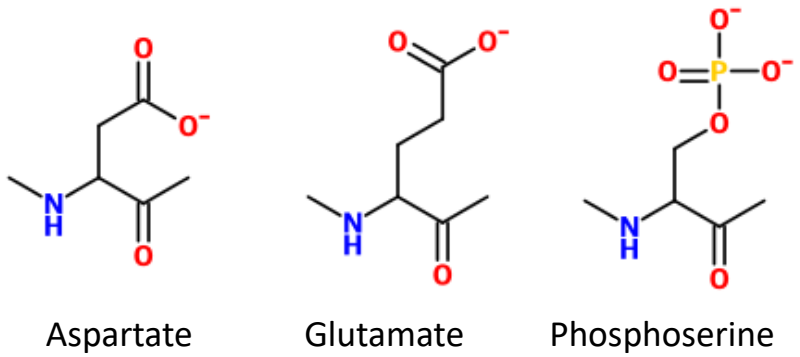

Fig 1.9. The structure of aspartate, glutamate and phosphorylated serine (www.emolecules.com).

In this study for these reasons a S557E (Fig 1.10C) mutant was used to mimic phosphorylation of Ser557. Glutamate was chosen rather than aspartate as it is structurally like a phosphoserine as it contains an extra carbon making it a similar length to phosphoserine. Glutamate has a p*K*a of approximately 4, therefore it will be negatively charged like phosphoserine at pH 7.5 (physiological). A S557A mutant was also made to act as a control (Fig 1.10B). The purpose of the alanine control is to ensure that any changes observed between the WT (Fig 1.10A) and phosphomimetic are due to changes in charge and bulk and not because of the disruption of protein structure or function due to the mutation (Tarrant & Cole, 2009). A phosphomimetic was used rather than phosphorylating Ser557 because the kinase that may phosphorylate this serine residue is unknown and it would be challenging to ensure that only Ser557 becomes phosphorylated.

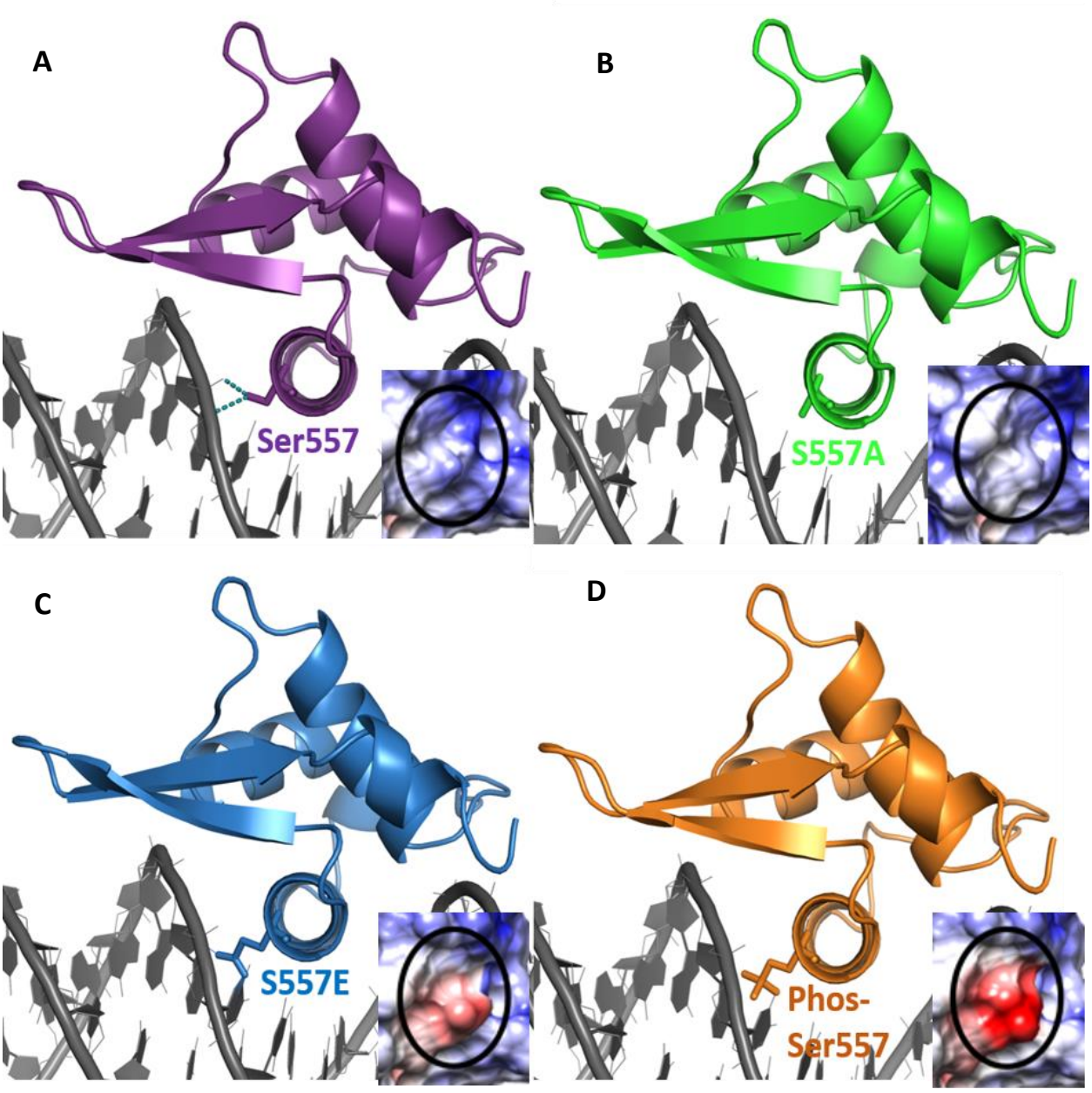

Fig 1.10. The structure of the FOXP2 FHD (PDB ID: A207) bound to DNA showing the WT, phosphorylated serine and the two serine mutants. The surface charge (red is negative and blue is positive) at position 557 is indicated in the panel at the bottom right of each structure and circled in black. **A**: Ser557 is shown and the hydrogen bonds are displayed as blue dashes. **B**: S557A mutant bound to DNA. **C**: S557E mutant bound to DNA. **D**: WT bound to DNA with Ser557 phosphorylated. The increase in negative charge and bulk in the phosphorylated WT is clear. The S557E mutant is similar to the phosphorylated WT however the smaller size of the residue and decreased charge is notable. The mutants were predicted using the wild type crystal structure (2A07) (Stroud *et al*, 2006). Cartoon Image rendered with PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC. Surface charge Image rendered with Chimera 1.7 (Pettersen *et al*, 2004)

1.7. Main aim

The aim is to use phosphomimetics to determine the effect of phosphorylation of Ser557 on the structure and DNA binding of the FOXP2 forkhead domain.

1.8. Objectives

To produce the phosphomimetic (S557E) and the control (S557A) FOXP2 FHD using site directed mutagenesis.

To express the wild type and various mutant FOXP2 FHDs using the pET-11a plasmid and to purify the respective proteins using His-tag affinity chromatography.

To investigate the effect of the phosphomimetic of Ser557 on the equilibrium between the monomer and dimer of the FOXP2 FHD using size exclusion chromatography.

To investigate the effect of the phosphomimetic of Ser557 on the secondary structure of the FOXP2 FHD using far-UV circular dichroism.

To investigate the effect of the phosphomimetic of Ser557 on the tertiary structure of the FOXP2 FHD using intrinsic tryptophan fluorescence.

To investigate the effect of the phosphomimetic of Ser557 on DNA binding using electrophoretic mobility shift assay and isothermal titration calorimetry.

CHAPTER 2 Materials and Methods

2.1. Materials

Table 2.1. The manufacturer of the non-standard reagents and materials used in each experiment. All other reagents were of analytical grade.

Method	Material	Manufacturer
Mutagenesis, sequencing and transformation	pET-lla plasmid	GenScript, USA
	XL10 Gold ultra-competent cells	New England Biolab, USA
	Site directed mutagenesis kit	Strategene, USA
	Ampicillin	Melford, UK
	T7 *E. coli* competent cells	New England Biolab, USA
	GeneJet mini prep kit	Thermo Fisher scientific, EU
	Primers for mutagenesis	Iquaba biotech, RSA
	Sybr Gold nucleic acid gel stain	Invitrogen, USA
Expression	Isopropyl βD-1-thiogalactopyranoside	Melford, UK
	Lysozyme	Sigma-aldrich, USA
	DNase	Merck, Germany
	Phenylmethylsulfonyl fluoride	Roche, Germany
Purification and cleavage	HisTrap column	GE healthcare, Sweden
	Benzamidine column	GE healthcare, Sweden
Size exclusion chromatography	Superdex 16/60 75 column	GE healthcare, Sweden
	Low range gel filtration calibration kit	GE healthcare, Sweden
EMSA and ITC	Duplex DNA	Integrated DNA technologies-WhiteSci scientific, RSA
Tricine SDS-PAGE	Page Ruler low range unstained protein ladder 100-3.4 kDa	Thermo Fisher scientific, EU
	Unstained protein molecular weight marker 116-14.4 kDa	Thermo Fisher scientific, EU
	Precision plus unstained protein ladder	Bio-Rad, USA

2.2. Methods

2.2.1. Mutagenesis, sequencing and transformation

The codon optimised gene sequence of FOXP2 FHD (residues 504-594) were incorporated into the pET-11a vector (GenScript, USA). In addition to the FOXP2 gene, an N-terminal insert containing a six histidine tag (to aid in purification of the protein) and a thrombin cleavage site (in order to cleave the histidine tag) was incorporated. The pET-11a vector allows for selection with ampicillin and expression control via a T7 polymerase promotor (Fig. 2.1).

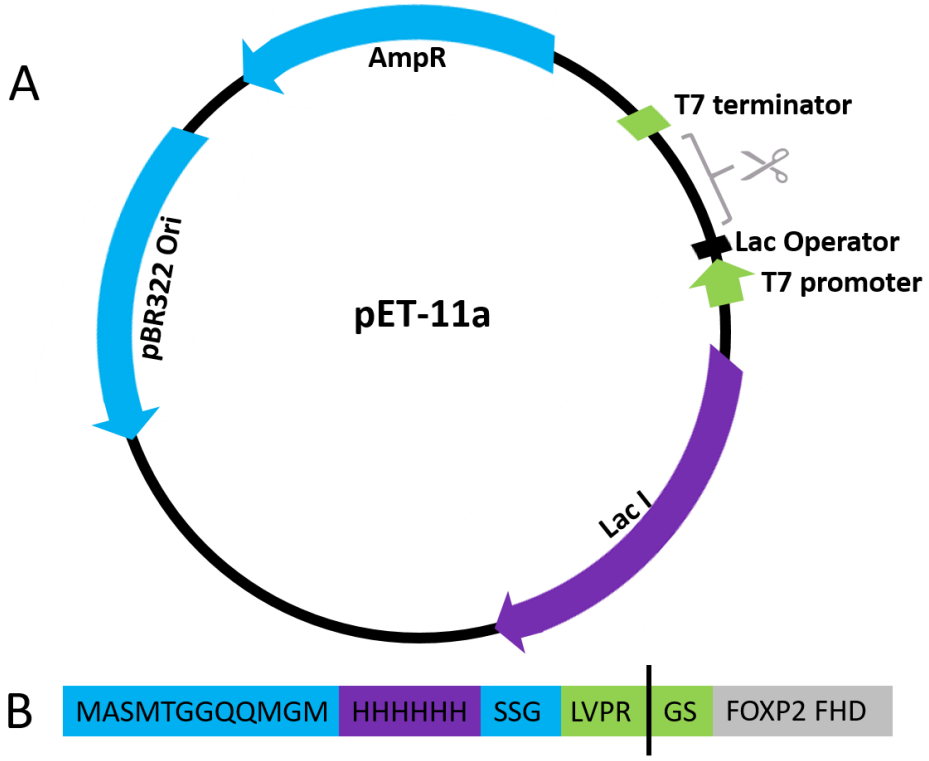

Fig. 2.1. **A: Schematic diagram of the pET-11a plasmid**. The FHD gene sequence was inserted in between the T7 promotor and terminator indicated by the scissors **B**: **The exact N-Terminal insert on the FOXP2 FHD**. The histidine tag is displayed in purple; the thrombin cleavage site is displayed in green and the line shows where thrombin cleaves.

XL10 gold ultra-competent cells were transformed with the pET-11a plasmid, containing the FOXP2 FHD insert. These cells were chosen as they are ultra-competent (the

transformation will likely work even if there is only a small portion of plasmid, which is likely when using a PCR product) and they contain a high copy number of plasmid. The transformation was performed by heat shock at 42 °C for 45 seconds. The transformed cells were cultured at 37 °C for 1 hour while shaking in super optimal broth with catabolite repression (SOC) media. SOC consists of 2% (w/v) tryptone, 0.5% (w/v) yeast, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 20 mM glucose. The cells were subsequently grown overnight at 37 °C on agar plates containing 0.1 mg/ml of ampicillin for selection. Colonies were selected and were cultured in 2X YT media (1.6% (w/v) tryptone, 1% (w/v) yeast extract and 0.5% (w/v) NaCl) containing 0.1 mg/ml ampicillin at 37 °C for 16 hours with vigorous shaking. The culture was used to make up 1:1 glycerol stocks (cell culture: 80% glycerol) and a mini prep was done to extract the plasmid (GeneJet plasmid mini prep kit). The isolated plasmid was sent for sequencing (Inqaba Biotec, RSA) to ensure the gene sequence was correct.

The WT isolated plasmid was subsequently used as a template for site-directed mutagenesis in order to generate the S557A and S557E mutants. Site-directed mutagenesis is a PCR reaction designed to produce specific point mutations in a plasmid. This is achieved by using long primers which encode the desired mutation meaning that in each cycle of PCR more mutant plasmids are produced. The WT template is digested using Dpn1 after the PCR is finished resulting in a product containing only intact mutant plasmid (Papworth *et al*, 1996). Primers (Table 2.2) encoding each mutant were designed using PrimerX (http://www.bioinformatics.org). Complementary primers with the mutation in the centre were designed with melting temperatures between 75 and 85 °C and with a GC content between 40 and 60 %. The primers were terminated with either a C or G. The primers were synthesised at Inqaba Biotec (RSA) and provided as a lyophilised pellet. The primers were resuspended in milli Q water at an appropriate concentration.

Table 2.2. Primers designed for mutagenesis using primer X. The CG content, melting temperature (TM), length and number of bases changed are displayed.

	Primer Sequence	CG content (%)	TM (°C)	Length	Number of base changes
S557A Forward Primer	5' CGCGGTCCGTCATAATCTG**G**CTCTGCA CAAATGCTTCG 3'	55.26	83.3	38	1
S557A Reverse Primer	5' CGAAGCATTTGTGCAGAG**C**CAGATT ATGACGGACCGCG 3'				
S557E Forward Primer	5' CGCGGTCCGTCATAATCTG**GAA**CTGC ACAAATGCTTCGTGC 3'	53.66	80.2	41	3
S557E Reverse Primer	5' GCACGAAGCATTTGTGCAG**TTC**CAGA TTATGACGGACCGCG 3'				

Mutagenesis was performed using the QuikChange site directed mutagenesis kit (Stratagene, USA) according to the manufacturer's protocol. The components and cycling conditions used are displayed in Table 2.3 and Fig 2.2. The PCR products obtained were run on a 1% agarose gel and stained with Sybr Gold DNA stain (Invitrogen, Life technologies). The PCR products were used to transform XL10 gold ultracompetent cells, cultured and sequenced in the same manner as the WT.

Table 2.3. Setup of site- directed mutagenesis reactions.

Reagent	sample
Reaction buffer (10x)	5 µl
Ds DNA template (10-100 ng)	50 ng
Forward primer	125 ng
Reverse primer	125 ng
dNTP mix	1 µl
QuickSolution reagent	1.5 µl
Milli Q water	Volume made up to 50 µl
PfuUltra HF DNA polymerase (2.5 U/µl)	1 µl

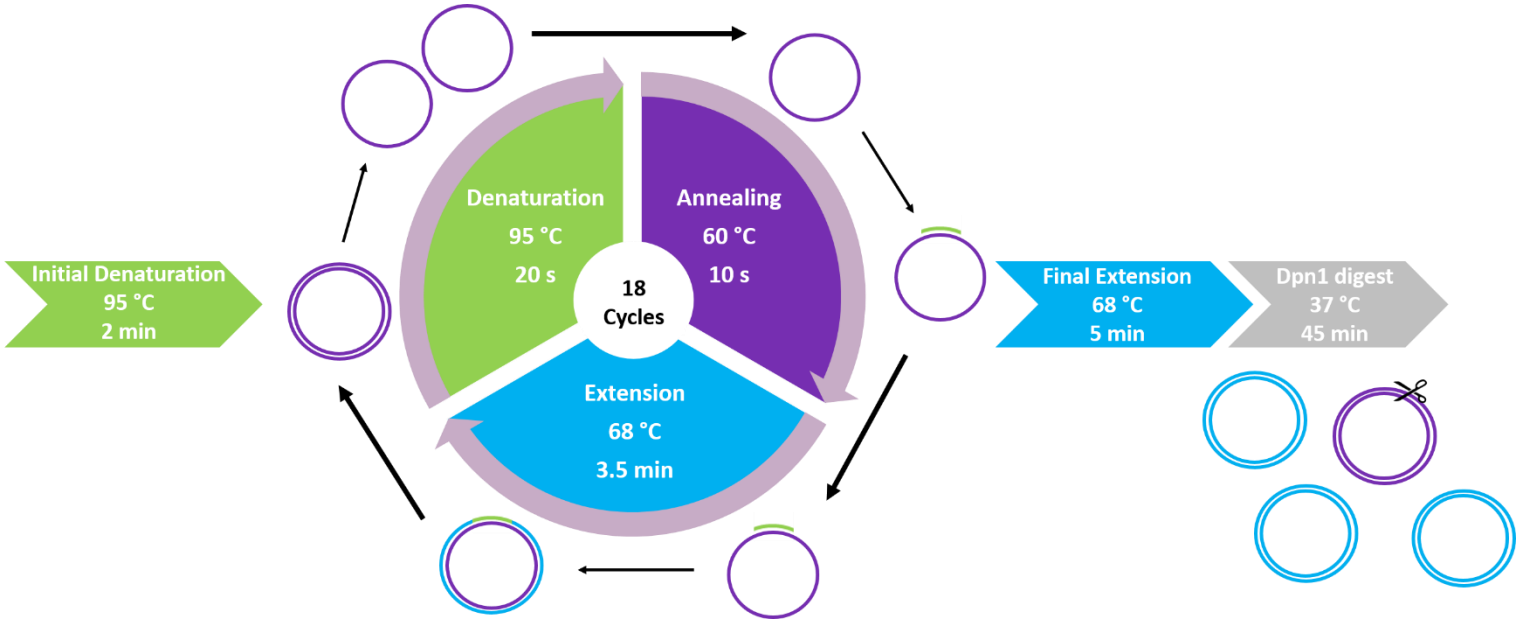

Fig. 2.2. **Schematic diagram indicating the PCR cycles used for all mutagenesis reactions**. Methylated WT template plasmid (purple) was denatured. Primers (green), encoding the mutation, bound to denatured template. The rest of the plasmid was built from the primer by polymerase, which resulted in the synthetic strand containing the mutation (blue). These new plasmids (not methylated) containing the mutation also acted as templates in subsequent PCR cycles. After 18 cycles of PCR Dpn1 was used to digest methylated plasmid (initial WT template) leaving a non-methylated product which contained the mutation.

The sequenced WT and mutant plasmids were used to transform competent T7 *E. coli* cells using the same transformation protocol. Multiple colonies from each plate were selected and grown separately in 2X YT media at 37 °C for 16 hours with vigorous shaking. The cell culture was used to make glycerol stocks that were used for expression.

2.2.2. Expression trials

Expression trials were performed to determine the optimal expression conditions of the mutants. Glycerol stocks of the S557E and S557A mutants were used to inoculate separate 100 ml flasks of 2X YT media containing 0.1 mg/ml of ampicillin. The cells were grown overnight at 37 °C with vigorous shaking. The cultures were used to inoculate flasks of 100 ml 2X YT media and 0.1 mg/ml ampicillin at a 1 in 50 dilution. The flasks were incubated at 37 °C until an O.D $_{600}$ of 0.6 was reached. The flasks were cooled to 20 °C and isopropyl βD-1-thiogalactopyranoside (IPTG) (0 mM, 0.1 mM, 0.3 mM, 0.5 mM and 0.7 mM) was added. The cultures were incubated at 20 °C and 1 ml samples were taken at various times (0 hours, 4 hours, 6 hours and 16 hours). The samples were centrifuged for 5 minutes at 5000 xg and the supernatant was discarded. The pellet was resuspended in 100 µl equilibration buffer (20 mM Tris, 30 mM imidazole and 500 mM NaCl, pH 7.5). The samples were sonicated to lyse the cells and were then centrifuged at 13 000 xg for 10 minutes. The supernatant was separated from the pellet and the pellet was resuspended in equilibration buffer. Reducing sample buffer (12 % SDS, 30% glycerol, 0.05% Coomassie blue G-250 and 250 mM Tris/HCl at pH 7) was added to both the supernatant and the pellet. All samples were resolved on a 16 % tricine gel according to the protocol by Schägger, (2006).

2.2.3. SDS-PAGE

Tricine SDS-PAGE was performed per the protocol by Schägger, (2006). SDS-PAGE is a well-used technique in which molecules are separated through a gel matrix via an electric charge. The protein samples are denatured, reduced and exposed to SDS (Sodium dodecyl sulfate), a denaturing anionic surfactant, which binds the protein and not only gives it a

uniform charge but also forces it into a rod like shape. This ensures that the samples are fully denatured and are thus separated by size and not by charge or shape as they are drawn through a polyacrylamide matrix by an electric current. There are multiple methods of SDS-PAGE, however the glycine (Laemmli, 1970) and tricine (Schägger, 2006) are two of the most commonly used. These two methods differ in the buffering system (glycine-tris and tricine-tris) and subsequently the trailing ions. The p*K*a of the trailing ion effects its mobility and therefore the rate at which it in the gel. Tricine is ideal for separations of smaller proteins and peptides (Schägger & Gebhard, 1987). SDS-PAGE is a versatile technique which is frequently used to estimate sizes of proteins and observe if there are other protein contaminants in a sample.

Samples of 50 µl were mixed with 16 µl of reducing sample buffer (12% (w/v) SDS, 6% (w/v) β-mercaptoethanol, 30% (w/v) glycerol, 0.005% Coomassie blue G-250 and 150 mM Tris-HCl, pH 7.5). Samples were subsequently boiled for 5 minutes and sonicated 3 times for 1 second to ensure the samples were fully denatured. The samples were vortexed and centrifuged to ensure the entire sample was mixed and all at the bottom of the tube. Samples (10 µl) were run on a 10% discontinuous gel consisting of acrylamide/bis-acrylamide stock (49.5% T and 3% C) and gel buffer (3 M Tris, 1 M HCl and 0.3% SDS at pH 8.45) cast and run using the BioRad mini protean electrophoresis system. The anode and cathode buffers were 1 M Tris at pH 8.9 (pH adjusted with HCl) and 1 M Tris, 1 M Tricine and 1% SDS at~pH 8.25 (pH not adjusted). The gels were run at 45 V until the samples entered the separating gel at which point the voltage was increased to 160 V.

2.2.4. Expression

The wild type and both mutants were expressed in the same manner as previously described (Blane & Fanucchi, 2015). An overnight culture consisting of 50 ml 2X YT medium with 0.1 mg/ml of ampicillin was inoculated with 1 ml of glycerol stock and was grown up overnight at 37 °C. Larger volumes of 2X YT media with 0.1 mg/ml ampicillin were inoculated with a 1 in 50 dilution of overnight culture. Cells were grown at 37 °C with shaking until an O.D $_{600}$ of 0.6 was reached. The cultures were then cooled to 20 °C before

0.5 mM IPTG was added. The cultures were incubated at 20 °C with shaking for 20 – 22 hrs. The cells were pelleted by centrifugation at 5000 xg for 25 minutes at 4 °C and the medium was discarded. The pellet was resuspended in 50 ml of equilibration buffer per litre of culture and frozen at – 20 °C. Cells were thawed at 20 °C, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 0.1 mg/ml of lysozyme was added. The lysate was then sonicated (power 18) X 5 for 30 seconds on ice. The lysate was left to incubate at 20 °C for 30 minutes after the addition of 0.01 mg/ml DNase 1 and 2 mM Mg Cl_2. The lysate was centrifuged at 23 000 xg for 30 minutes at 4 °C to separate the soluble and insoluble fractions. The supernatant was used in the subsequent purification.

2.2.5. Purification and Cleavage

Immobilised metal affinity chromatography (IMAC) was the first purification step used. In this technique, the recombinantly expressed protein, with a histidine tag, specifically binds a column with either nickel or cobalt embedded. This is because the imidazole side chain of histidine has a high affinity to chelated metals. The rest of the cell lysate should not interact with the column as the His-tagged protein does and should there elute from the column. At this point multiple washing steps, such as salt or detergent, may be done to remove any DNA, proteins or lipids bound to the protein immobilised on the column. The pure protein is then eluted from the column with imidazole as imidazole competes with the His-tag to be the nickel or cobalt column (Arnau *et al*, 2006).

A 5 ml (1.6 X 2.5 cm) HisTrap column (GE healthcare, USA) charged with Ni^{2+} was equilibrated with 10 column volumes of equilibration buffer (20 mM Tris, 500 mM NaCl and 30 mM imidazole, pH 7.5). The supernatant was loaded on to the column and the flow through was collected. After the flow through passed through the column, the column was allowed to re-equilibrate with equilibration buffer. Salt wash buffer (20 mM Tris, 1.5 M NaCl and 30 mM imidazole, pH 7.5) was used to wash the column for at least 5 column volumes to remove any DNA bound to the protein. After re-equilibration the His-tagged FHD was eluted using elution buffer (20 mM Tris pH 7.5, 500 mM NaCl and 500 mM imidazole). Fractions of the contamination peak and the main protein peak were collected.

Samples of the supernatant, pellet, flow through and all protein fractions were resolved on a 16 % tricine gel.

As the His-tag made interfere with the function of the protein the His-tag was removed before subsequent studies. The His-tag is cleaved using thrombin as a thrombin cleavage site was incorporated into the recombinant protein (Fig 2.1). Once cleavage has occurred the thrombin, cleaved of His-tag and any protein with the His-tag still attached must be removed from the cleaved protein. The solution is applied to a benzamidine and IMAC columns connected in series. The thrombin binds to the benzamidine column while the His-tag binds to the IMAC thus the cleaved protein elutes immediately (Young *et al*, 2012).

Fractions containing the FHD were pooled and dialysed against thrombin cleavage buffer (20 mM Tris, 2 mM KCl and 100 mM NaCl, pH 8). Protein was centrifuged at 10 000 xg for 15 minutes to remove aggregates. Approximately 1 unit of thrombin was added per mg of protein and the solution was left to incubate at 20 °C for 3 to 4 hours. After addition of thrombin, a sample was taken every half an hour and analysed on a 16% tricine SDS-PAGE gel to test for cleavage. The solution was filtered through a 0.2 μM filter and was loaded on a pre-equilibrated 1 ml (0.7 X2.5 cm) benzamidine (GE healthcare, USA) and IMAC column connected in series. The cleaved FHD was collected in fractions as it flowed through the columns. Any uncleaved protein was eluted using elution buffer. Samples of all fractions collected were resolved on a 16 % tricine gel.

2.2.6. Protein purity and concentration

To ensure the protein was sufficiently pure, fractions of protein obtained from the purification were resolved on a 16% tricine gel and stained using Coomassie blue as described in section 2.2.3. Fractions containing the protein of interest without protein contamination were pooled and dialysed against a buffer of 10 mM HEPES, pH 7.5 and 100 mM NaCl. The protein was then centrifuged at 10 000 xg for 10 minutes to remove any aggregates. An absorbance spectrum from 250 to 340 nm was obtained using a JASCO

V-630 spectrophotometer. The A280/A260 ratio was used to detect the presence of any nucleic acid contamination.

The protein solution was centrifuged at 10 000 xg for 10 minutes and the buffer was the filtered through a 22 nM filter. In order to determine the concentration of the protein, the absorbance readings at 280, 340 and 260 nm were obtained for five dilutions of the protein. This was repeated in triplicate. The 340 nm reading was subtracted from the 280 nm reading to account for aggregation. The Beer-Lambert law (Eq. 1) was used to calculate the protein concentration. The theoretical extinction coefficient obtained from Expasy ProtParam (Gasteiger *et al*, 2005) used was 22460 $M^{-1}.cm^{-1}$. The extinction coefficient is calculated using equation 2 based of the primary structure using extinction coefficients for tyrosine, tryptophan and cysteine in water (Gill & von Hippel, 1989).

Eq 1. $$A = \varepsilon.c.l$$

Where A is the absorbance at 280 nm (A.U), ε is the extinction coefficient of the FHD at 280 nm ($M^{-1}.cm^{-1}$), c is the concentration of protein (M) and l is the path length (cm).

Eq 2. $$\varepsilon\,(prot) = n(Tyr)\times\varepsilon\,(Tyr) + n(Trp)\times\varepsilon\,(Trp) + n(Cys)\times\varepsilon\,(Cys)$$

Where n refers to the number of amino acid residues, ε is the molar extinction coefficient of the amino acid and $\varepsilon\,(prot)$ refers to the predicted molar extinction coefficient of the protein.

2.2.7. Structural characterisation

2.2.7.1. Size exclusion chromatography

Gel filtration or size exclusion chromatography is a form of column chromatography where molecules are passed through a column containing a porous resin in order to separate them based on hydrodynamic volume/ Stokes radius. This technique was first described by (Ruthven *et al*, 1954). The Stokes radius is determined from a molecule's molecular mass, shape and size. The smaller the molecule, or rather the smaller the hydrodynamic volume the more time the molecule spends in the pores. This means that molecules within the fractionation range of the column are separated based on hydrodynamic volume with larger molecules eluting first. There are now a variety of resins commonly used for size exclusion chromatography which include dextran (Sephadex/Sephacryl), agarose (Sepharose) and polyacrylamide polymers. Superdex resin is made up of a matrix consisting of agarose with covalently linked dextran chains. It is often used as it offers high resolution and is generally non-reactive with biological samples (Blurnenfeld & Gardner, 1985).

Size exclusion chromatography was used to determine concentrations at which the WT, S557A and S557E FOXP2 FHDs were monomeric. A HiLoadTM 16/60 SuperdexTM 75 prep grade column (GE healthcare, USA) was equilibrated with equilibration/ running buffer (10 mM Tris pH 7,5, 500 mM NaCl and 30 mM imidazole). Imidazole was included in the running buffer as it tends to help prevent the FOXP2 FHD from interacting with the column resin resulting in sharper peaks. One ml of protein at concentrations ranging between 25 and 200 μM of WT and mutant was loaded on to the column. Protein standards (GE healthcare low range gel filtration calibration kit) were run according to the manufacturer's instructions in equilibration/ running buffer. The standards consisted of conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa). A standard curve was used to approximate the size of the WT and mutant FHDs. It was constructed by plotting the log Mw of each standard against the ratio of their respective elution volumes (Ve) and void volume (Vo). A linear

regression was fitted to the curve and the subsequent equation was used to calculate the sizes of the FHD and mutants based on their elution volumes.

To determine dissociation constant of the WT FHD, four concentrations (25, 50, 100 and 200 μM) were passed through the size exclusion column. The area under the monomer and dimer peaks were calculated using AKTA primeview 4 software. This was used to calculate the portion of protein in each peak. This was used to determine the protein concentration of each peak which was subsequently used to determine the K_d using equation 3 as described by Perumal *et al*, (2015).

Eq 3. $$K_d = [Monomer]^2/[Dimer]$$

where K_d refers to the dissociation constant and the monomer and dimer are given in μM.

2.2.7.2.Far-UV circular dichroism

Circular dichroism is a spectroscopic technique in which the differential absorption of right and left handed circularly polarised light by chiral molecules is measured. The average molar absorption coefficient (Δε) is the difference between the left and right absorption coefficients. The spectra obtained in the far-UV range indicate the secondary structure of protein backbone. This is because of the chirality of the α-carbon displayed by all the amino acids with the exception of glycine (Greenfield, 2006). The shape of the far-UV spectrum (260 -190 nm) of a protein that is predominantly α- helical shows troughs at 220 and 218 nm with a peak at 190 nm. Data at lower wavelengths is often noisy due to a lack of buffer transparency (Brahms & Brahms, 1980). Far-UV circular dichroism was used to probe significant secondary structural changes between the WT and mutant FHDs.

The WT and mutant FHDs were dialysed into CD buffer (10 mM HEPES, pH 7.5 and 50 mM $NaSO_4$). The far-UV CD spectra of 10 μM samples were measured on a J-1500 CD spectrometer in triplicate. The spectra were measured from 250 to 180 nm. The scan speed was 100 nm/min and the bandwidth was 1 nm. Ten accumulations were taken. The spectra

were converted from θ (m.deg) to θ_{MRE} (deg.cm^2.dmol^{-1}) using equation 4 below where θ is the ellipticity and n is the number of peptide bonds.

Eq 4. $$\theta_{MRE} = \frac{(\theta\ (m.deg).10^6)}{(l\ (mm).\ c(\mu M).\ n)}$$

Where θ refers to ellipticity (m.deg), l refers to the path length (mm), c refers to the protein concentration (μM) and n refers to the number of peptide bonds.

2.2.7.3. Intrinsic fluorescence

Fluorescence spectroscopy makes use of the fluorescence phenomenon where a molecule (the fluorophore) absorbs a specific wavelength of light. This light excites electrons to a higher energy level and light is then emitted when the electrons move back to their ground state (Fig. 2.3). The emitted light is of a lower frequency/ longer wavelength than the light used to excite the fluorophore (Lakowicz, 2006).

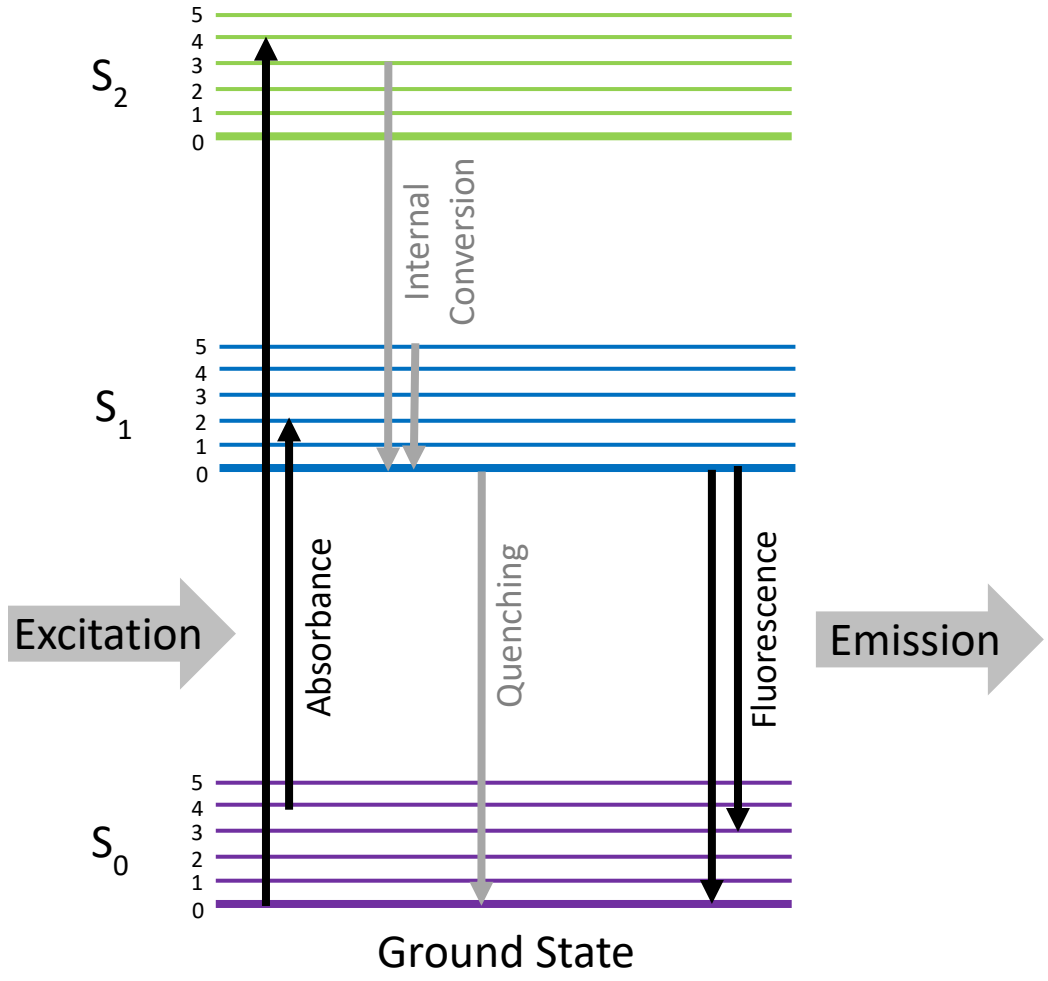

Fig. 2.3. Jablonski diagram showing the fluorescence phenomenon. S_0, S_1 and S_2 are energy levels and the lines indicate different vibrational energy levels. Light of a specific wavelength is used to excite electrons and the light is absorbed. Electrons fall back to the lowest vibrational energy level with energy loss due to heat and vibration (internal conversions). As the electron drops back down to the ground state the energy can either be emitted via light (fluorescence) or lost due to quenching (Lakowicz, 2006).

Proteins may contain amino acids which act as fluorophores (tryptophan, tyrosine, phenylalanine and cysteine). Cysteine and phenylalanine exhibit extremely weak signals and are thus not used as intrinsic fluorophores. Tryptophan is most often used as it has a good signal and quenches the quantum yield of nearby tyrosine. Both tyrosine and tryptophan are excited at 280 nm and tryptophan can be exclusively excited at 295 nm. These intrinsic fluorophores can be used to probe tertiary structural changes in the protein around them. This is due to the sensitivity of the fluorophore to the polarity of its

environment which alters the emissions spectrum. A hyposochromic or blue shift indicates a decrease in the polarity which often indicates burying of the fluorophore within the protein and thus reducing its exposure to the polar solvent. A bathochromic or red shift indicates an increase in polarity which occurs when the fluorophore becomes more exposed to the polar solvent (Lakowicz, 2006).

Intrinsic fluorescence of the WT, S557A and S557E mutant FHDs was used to determine if there are any tertiary structural changes in the FHD caused by mutating Ser557 to an alanine or glutamate. Fluorescence spectra of 2 μM protein in 10 mM HEPES, pH 7.5 and 100 mM NaCl were determined in triplicate using a JASCO FP-6300 spectrofluorometer. Both 295 and 280 nm were used to excite the fluorophores (tryptophan at 295 nm and tryptophan and tyrosine at 280 nm) and the emission spectra were recorded from 280 to 450 nm. The excitation and emission band widths were set at 5 and 2.5 nm respectively. The scanning speed was 500 nm/min and three accumulations were recorded.

2.2.8. DNA binding studies

The FOX family of proteins has a defined DNA consensus sequence which is 5' [A/G][C/T][A/C]AA[C/T]A 3' (Carlsson and Mahlapuu, 2002). The core binding sequence specifically for FOXP2 was defined as 5' A[C/T]AAATA 3' (Wang *et al*, 2003). More recently Nelson *et al,* (2013) found the sequence (5'-T**GTTTA**C-3') to which FOXP2 binds with a high affinity. The FOXP2 2A07 crystal structure (Stroud *et al,* 2006) contains the Wang sequence 5'-AACTATGAAA**CAAATT**TTCCT-3'. Cognate DNA containing the Nelson sequence binding site (5'-AGGT**GTTTA**CTTTCATAG-3') was used for this work.

2.2.8.1. Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) is an electrophoresis based technique that is used to observed protein-DNA interactions. After allowing time for the protein to bind DNA, the solution is run on a 10% polyacrylamide gel. Protein-DNA complexes are larger than the free DNA and thus are retarded by the gel. This should result in the formation of two distinct bands visible on the gel after staining with a nucleic acid stain (Garner & Revzin, 1981). EMSAs, however, are often of low resolution because of the smearing of bands. These smears are a result of dissociation and re-association of the complex as the gel is run. The technique however is suitable to indicate if the complex associates and may give an indication of the dissociation constant for the binding interaction.

The WT and mutant FHDs were dialysed against a buffer consisting of 10 mM HEPES, pH 7.5 and 100 mM NaCl. Reactions (50 µl) were set up containing increasing protein concentration from 0 to 20 µM (0 to 40X in excess of DNA), 0.5 µM DNA in EMSA binding buffer (10 mM HEPES, pH 7.5, 100 mM KCl, 1 mM EDTA, 1 mM $MgCl_2$, 0.1 mg/ml bovine serum albumin and 10 % (v/v) glycerol) and MilliQ water. The samples were left to incubate on ice for 1 hour. A 10% polyacrylamide gel was prepared with TBE buffer and was left to equilibrate to 4 °C along with running buffer (TBE buffer) for 1 hour. The gel was run at 4 °C for 1.5 hours immediately after loading the samples. The gel was stained with a 1:10 000 dilution of SYBR gold in TBE buffer for 5 minutes. The gels were visualised under UV light using the BioRad gel doc system.

2.2.8.2. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is based on measuring the heat absorbed or released as one of the reactants is titrated into the other. The isotherm obtained can be used to determine thermodynamic properties such as the binding affinity (K_a), enthalpy, entropy and free energy. The K_a and enthalpy (ΔH) can be derived directly from the titration. This can be used to calculate the Gibbs free energy (ΔG) and entropy (ΔS) (Ladbury & Chowdhry, 1996) using equations 5 and 6.

Eq 5. $\Delta G = RTln\,(K_a)$

Where ΔG refers to Gibbs free energy (kJ/mol), R refers to the gas constant, T refers to temperature (K) and K_a refers to the dissociation constant.

Eq 6. $\Delta G = \Delta H - T\Delta S$

Where ΔG refers to Gibbs free energy (kJ/mol), ΔH refers to the enthalpy (kJ/mol), T refers to the temperature (K) and ΔS(refers to the entropy (kJ/mol).

ITC was used to study binding interactions between WT and mutant proteins and DNA. Titrations were done on a Nano isothermal titration calorimeter (TA instruments, USA) as previously described (Morris & Fanucchi, 2016). The WT, S557A, S557E FHDs and DNA were dialysed against 10 mM HEPES, pH 7.5 and 100 mM NaCl. A sample of the buffer was then filtered through a 0.2 µM filter to remove any particles. The protein and DNA samples were centrifuged at 10 000 xg for 10 minutes. The protein (100 µM) was loaded into the sample syringe and DNA (5 µM) was loaded into the sample cell. The concentration of DNA was determined using the same method as the protein using the experimental extinction coefficient of 256 016 $M^{-1}\cdot cm^{-1}$ (Kibbe, 2007). The ITC was allowed to equilibrate while stirring at 250 rmp. Once equilibrated, the titration was started. Injection volumes of 5 µl were used. Approximately 40 injections were done.

2.2.9. Molecular Docking

Molecular docking was done in an attempt to computationally observe differences in binding between the WT and mutants in order to elucidate reasons for the differences. Nelson DNA was modelled using the build nucleic acid function in Accelrys Discovery studio 4.1. The monomeric FHD structure was obtained from chain K of the 2A07 crystal structure (Stroud *et al*, 2006). Serine 557 was mutated to either an alanine or a glutamate using Chimera 1.7 (Pettersen *et al*, 2004). A model containing phosphorylated serine 557 was made in the PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC using the PyTMs plugin. The DNA and protein models were energy minimised using Chimera 1.7 (Pettersen *et al*, 2004). All the protein models were docked to the DNA model using

the easy interface HADDOCK (Dominguez *et al*, 2003). High ambiguity driven docking approach (HADDOCK) makes use of ambiguous interaction restraints (AIRS) which is biochemical/biophysical information to aid in docking. The docked structures are subsequently ranked based on their intermolecular energy which considers electrostatic and Van der Waals energy along with the AIRs. The AIRs in the easy interface which was used were automatically generated based on user defined active residues (residues involved directly in the interaction) and passive residues (the surface residues around the binding site). Two residues were defined as active residues in the protein (Arg553 and Asn550). Arg553 was used because when it is mutated, DNA binding is abolished (Vernes *et al*, 2006) and thus makes an important DNA contact. Asn550 was chosen as it is central in helix 3 and makes hydrogen bonds to the DNA backbone in the crystal structure. His554, Lys549, Arg553 and His559 were defined as passive residues as they are the surface exposed residues of helix 3 surrounding Arg553 and Asn550 (Dominguez *et al*, 2003). The docked models were grouped into clusters based on their similarity. The best scored structure of the cluster with the protein bound to the binding site was analysed.

CHAPTER 3 Results

The effect of phosphorylation of Ser557 in the FOXP2 FHD on structure and DNA binding is unknown. In order to determine the effect of phosphorylation of serine 557 may have, serine 557 was mutated to both a glutamate to act as a phosphomimic, and an alanine to act as a control. Structural characterisations (size exclusion chromatography, far-UV circular dichroism and intrinsic tryptophan fluorescence) and DNA binding studies (EMSA, ITC and molecular docking) were conducted with the WT FOXP2 FHD and, S557A and S557E mutants to observe any changes caused by phosphorylation of serine 557.

3.1. Mutagenesis and sequencing

Primers for mutagenesis (Table 2.2) were designed using primer X. The pET-11a plasmid incorporated with the WT FOXP2 FHD gene was used as a template for mutagenesis. The PCR products obtained from the mutagenesis reactions were resolved on a 1 % agarose gel (Fig. 3.1). A large band was present in each lane which is likely the intact mutant plasmids. The digested plasmid is present as a band at the bottom of the gel and as a feint smear along the lanes. The PCR products were used to transform T7 *E. coli* cells. The cells were grown up and the plasmids were extracted using a mini prep kit. The extracted plasmids were sequenced and the correct mutations were found to have occurred (Fig. 3.2.).

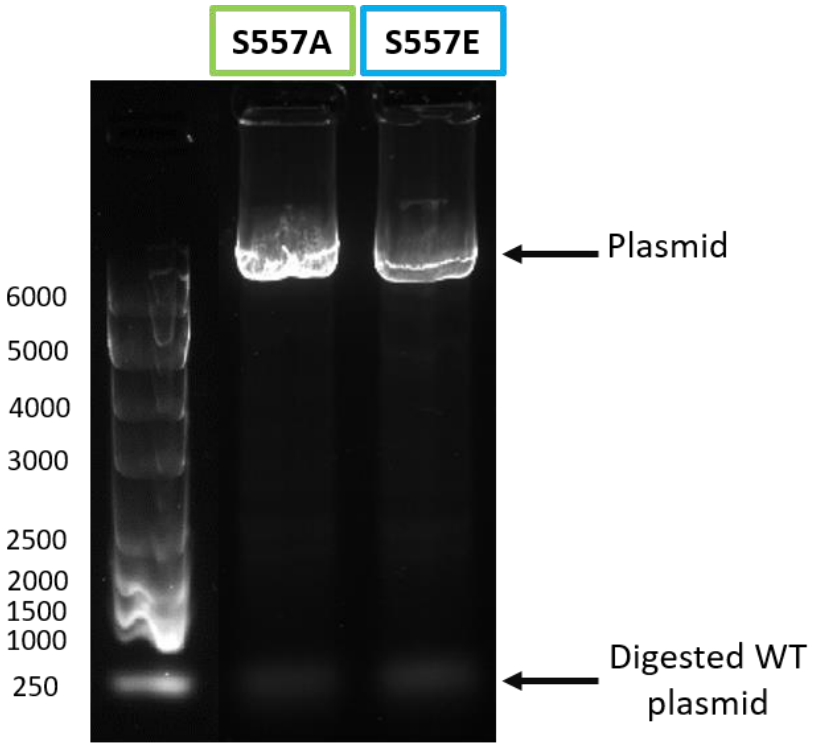

Fig. 3.1. PCR product of S557A and S557E site- directed mutagenesis. PCR product obtained from PCR reaction resolved on a 1% agarose gel stained with SYBR gold nucleic acid stain. The intact plasmid is visible as a clear band at the top of the gel. Lane 1 contains the O'GeneRuler 1 kb DNA Ladder (Thermo fisher scientific, EU).

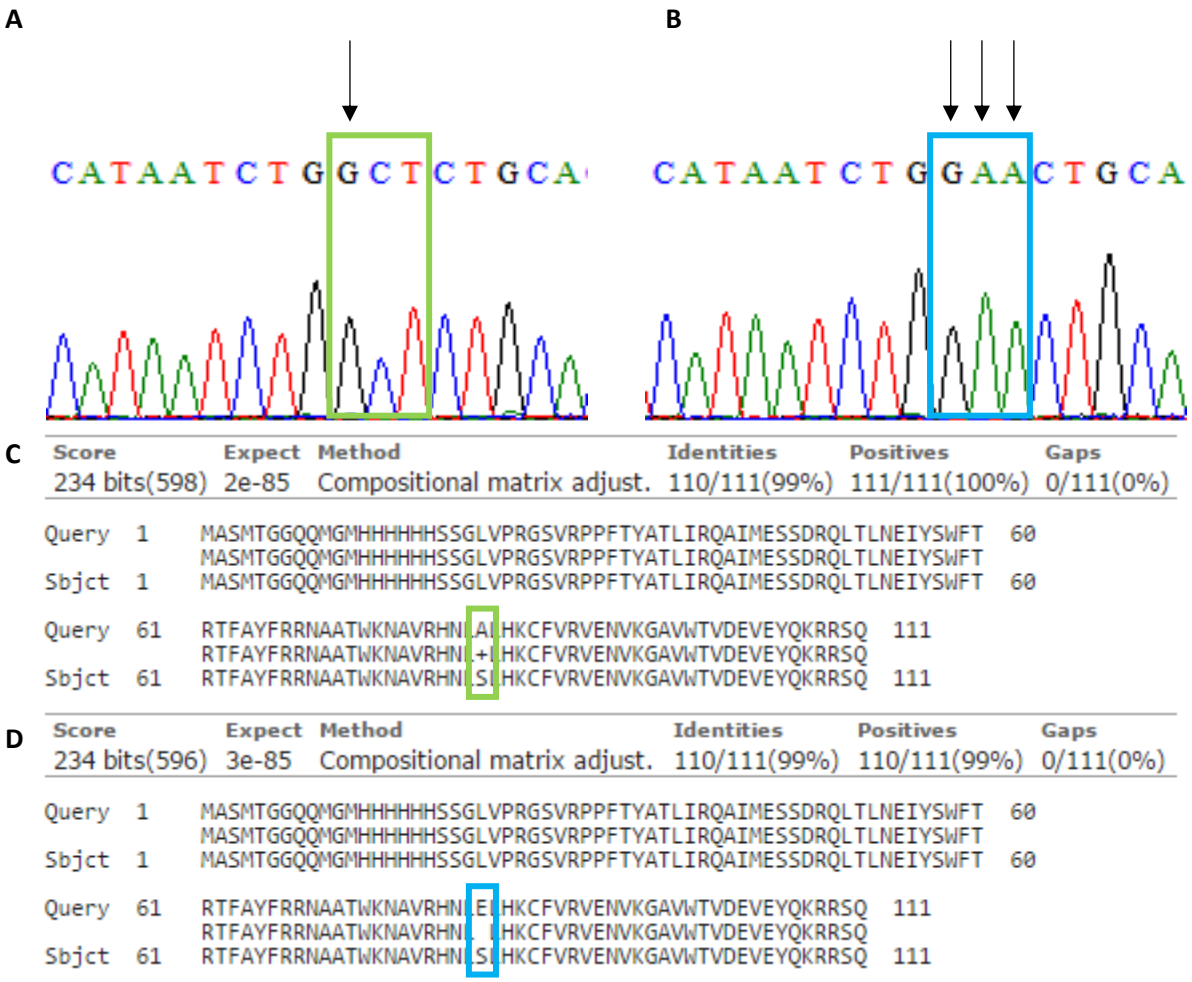

Fig. 3.2. Sequencing results of S557A (A) and S557E (B) mutant plasmids and alignment with the WT FOXP2 FHD plasmid (C and D). The green and blue blocks in A and B indicate the alanine and glutamate codons respectively. The arrows above indicate nucleotides which were changed. The DNA sequencing of the mutant plasmids (Query) were translated to protein using Expasy translate (Gasteiger *et al*, 2005) and aligned with the WT protein (Sbjct) sequence using NCBI Blast (blast.ncbi.nlm.nih.gov). The correct mutation has occurred in the S557A (**C**) and S557E (**D**) plasmids.

3.2. Expression trials

Expression trials were performed to determine the optimal expression conditions of the mutant FHDs only, as expression conditions for the WT have been previously determined (Blane & Fanucchi, 2015). The conditions chosen were based on the optimal expression conditions for the WT. Expression trials were conducted at 20 °C with varying concentrations

of Isopropyl βD-1-thiogalactopyranoside (IPTG) to determine the best conditions to express the S557A and S557E FHDs. In figure 3.3 the 16% tricine SDS-PAGE gels of the S557A expression trial are shown. The best soluble expression conditions appear to be with 0.5 mM IPTG for 22 hours at 20 °C.

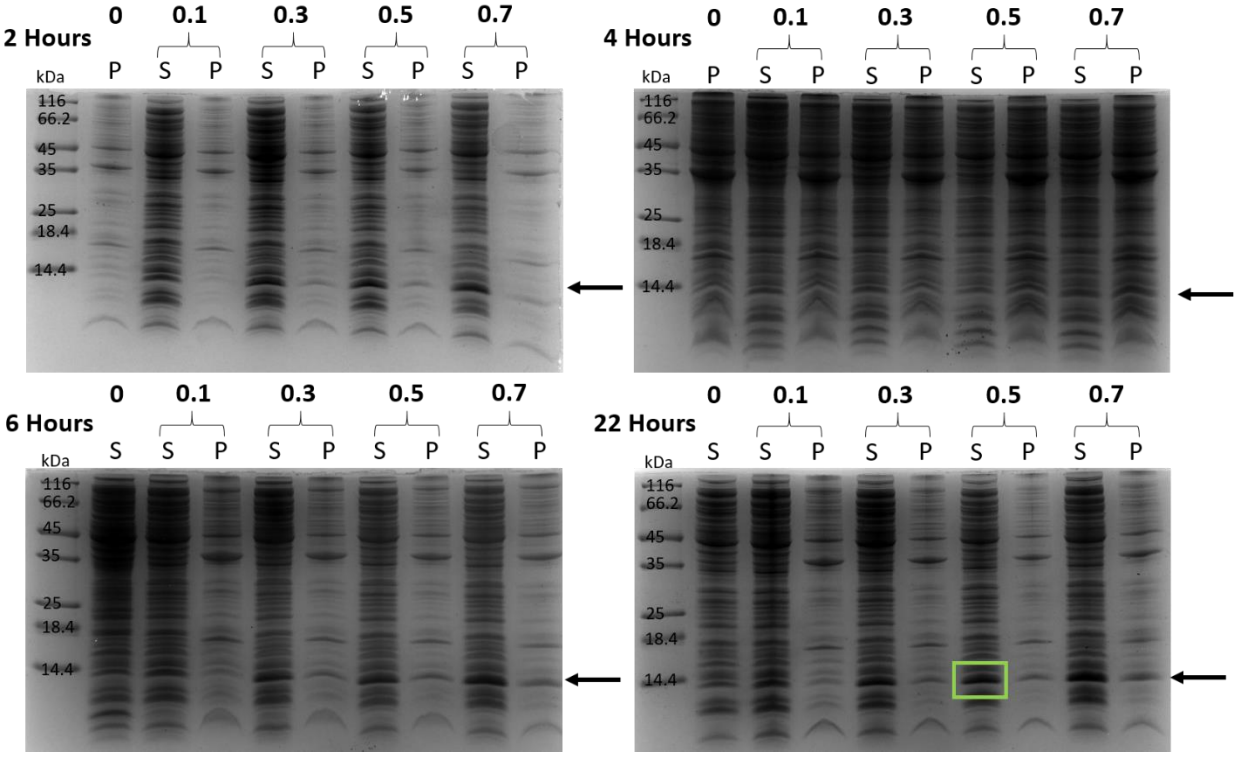

Fig. 3.3. S557A expression trials (2, 4, 6 and 22 hours). The number in bold indicates the IPTG concentration in mM. S and P stand for supernatant and pellet respectively. The arrow indicates the expected position of the FHD band (~14 kDa). The best apparent expression conditions appear to be using 0.5 mM IPTG and incubating at 20 °C for 22 hours (green block). The sizes of the bands in the Thermo scientific unstained protein molecular weight marker (166-14.4 kDa) are indicated on each band in kDa.

In figure 3.4 the 16% tricine SDS-PAGE gels of the S557E expression trial is shown. The best soluble expression conditions appear to be with 0.5 mM IPTG for 22 hours at 20 °C.

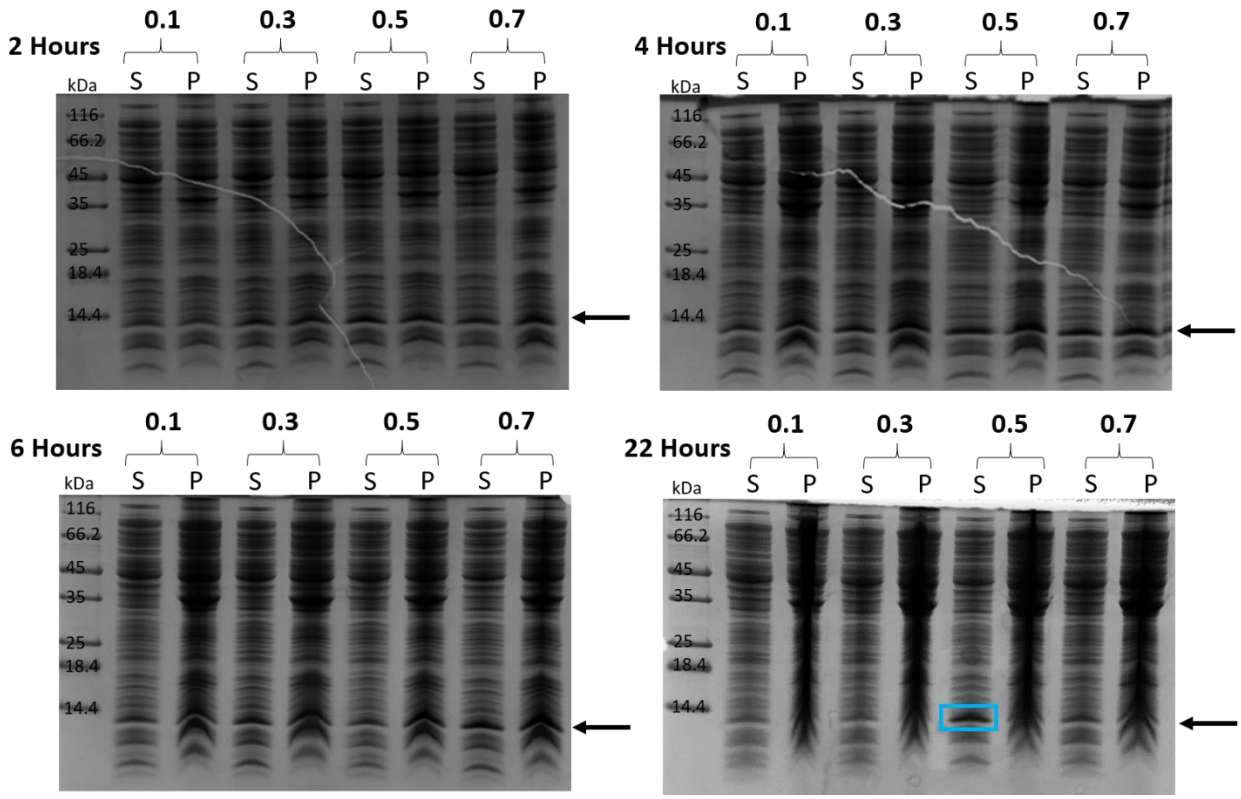

Fig. 3.4. S557E expression trials (2, 4, 6 and 22 hours). The number in bold indicates the IPTG concentration in mM. S and P stand for supernatant and pellet respectively. The arrows indicate the expected position of the FHD band (~14 kDa). The best apparent expression conditions appear to be using 0.5 mM IPTG and incubating at 20 °C for 22 hours (blue box). The sizes of the bands in the Thermo scientific unstained protein molecular weight marker (166-14.4 kDa) are indicated on each band in kDa.

3.3. Expression and purification

In order to obtain sufficient quantities of the WT, S557E and S557A FOXP2 FHDs to perform structural characterisations and binding studies, the FHDs were expressed using the pET-11a expression vector in T7 *E.coli* cells. The proteins were subsequently purified using IMAC and the His-tag was subsequently cleaved. The following sections show an example of how the WT FOXP2 FHD was purified (the S557A and S557E mutants were purified the same way).

3.3.1. Immobilised metal affinity chromatography

The WT FOXP2 FHD was expressed for 22 hours at 20 °C after induction with 0.5 mM IPTG. The soluble fraction of the cell lysate was loaded on to the IMAC column and fractions were

collected of the eluted protein (F1-F3 in Fig. 3.5.B). Fractions 2 and 3 in Fig 3.5B contained large quantities of FOXP2 FHD and were thus pooled and the His-tag was cleaved using thrombin. There were still large quantities of contaminants in the pooled fractions but these containments aggregated during dialyses against thrombin cleavage buffer and were removed by centrifugation at 10 000 xg for 10 minutes.

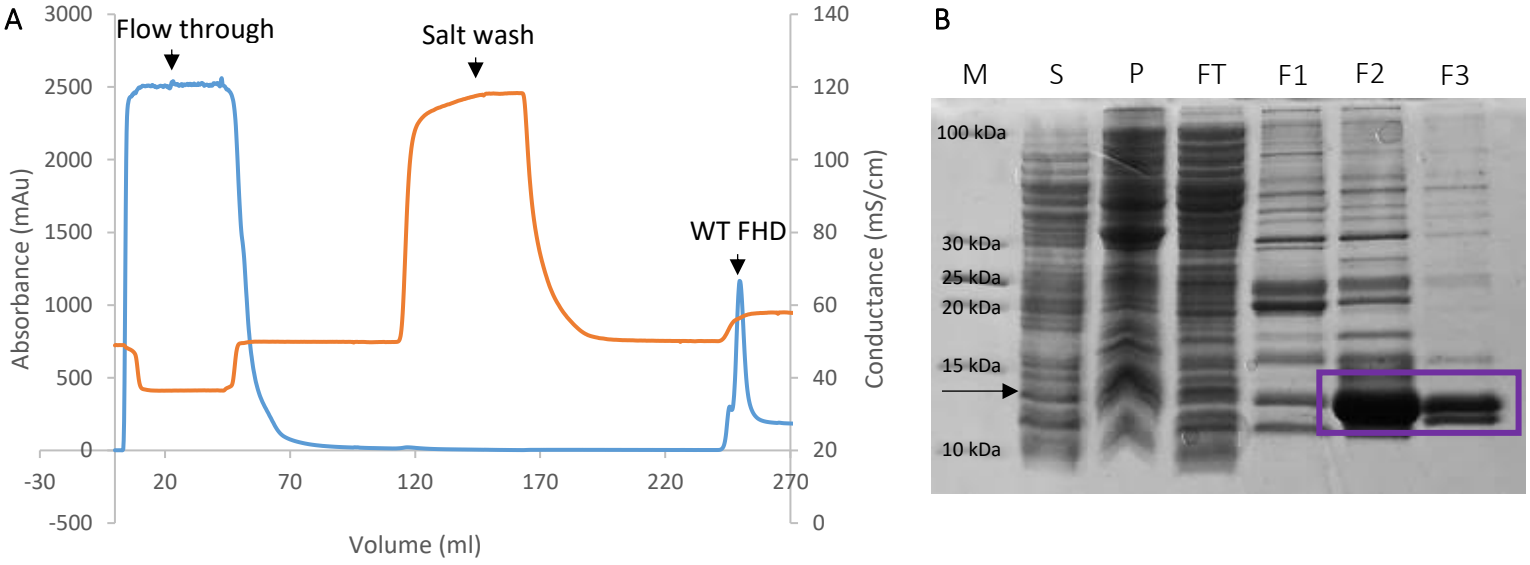

Fig 3.5. Immobilised metal affinity chromatography (IMAC) of the WT FOXP2 FHD. A: The IMAC elution profile of the WT FHD. The blue line is the absorbance (mAu) and the red line is the conductance (mS/cm). The protein was eluted using imidazole as indicated by the peak at 250 ml. **B**: 16% tricine SDS page gel of samples collected from the purification. S- supernatant. P- Pellet. FT-flow through from the IMAC column. F1-3- fractions collected of the eluted WT FHD peak. The Thermo scientific low range unstained protein ladder (100-3.4 kDa) was used.

3.3.2. His-tag cleavage

Thrombin was added to the dialysed protein and a sample was taken every half an hour and resolved on an 16% tricine SDS-PAGE gel. In order to remove the cleaved off His-tag and thrombin the protein solution was passed through a benzamidine column (to bind the thrombin) and an IMAC column (to bind the His-tag). In Fig 3.6.B it can been seen that after 3 hours there is still some uncleaved protein. F1 and F2 are the samples of the fractions of the cleaved protein eluted from the IMAC and benzamidine columns. The protein was free from protein contaminants and could be used for further study.

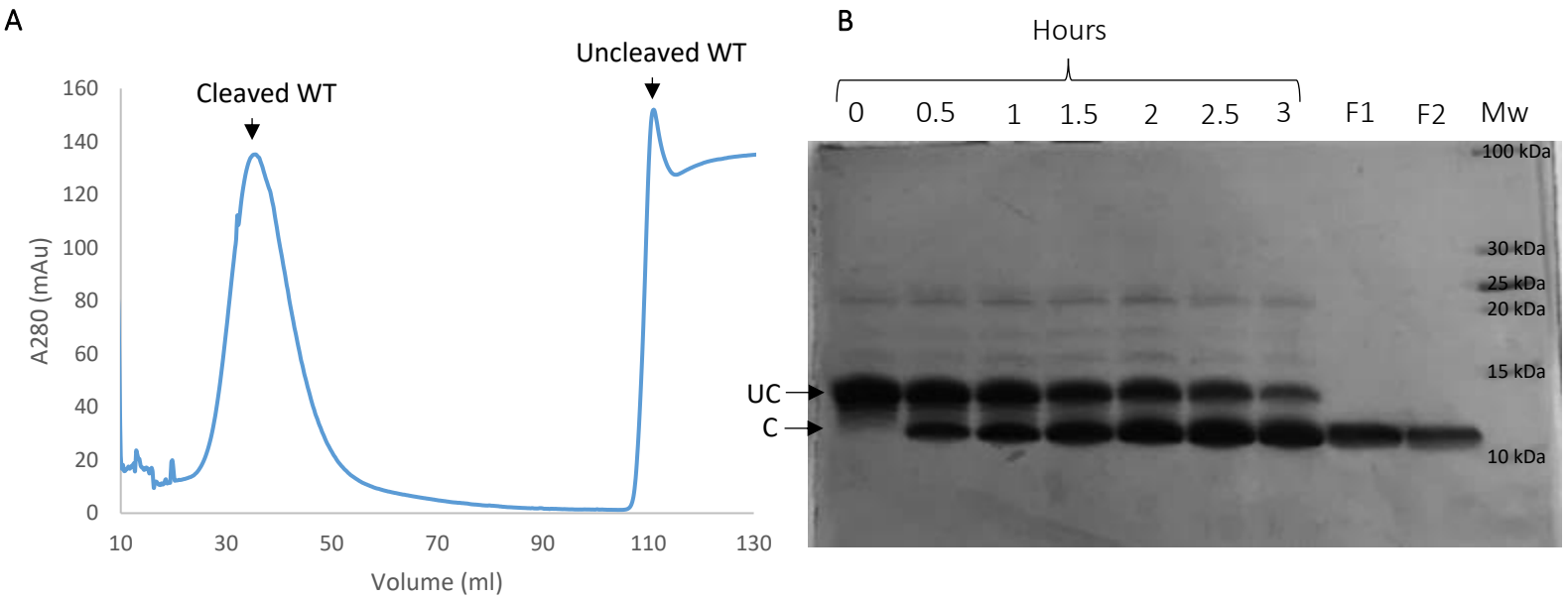

Fig 3.6. WT FOXP2 His-tag cleavage. A: The elution profile of cleaved protein from the IMAC and benzamidine column. The first peak contains the cleaved FHD which passes through both the IMAC and benzamidine columns. The second peak contains uncleaved FHD which bound the IMAC column and was eluted using imidazole. **B**: 16% tricine SDS-PAGE gel showing samples taken at time intervals during cleavage and samples of the cleaved protein eluted from the IMAC and benzamidine columns (F1 and F2). The uncleaved (UC) and cleaved (C) proteins are indicated on the gel. The Thermo scientific low range unstained protein ladder (100-3.4 kDa) was used.

3.3.3. Confirmation of purity of the WT, S557A and S557E FOXP2 FHDs

Samples of the purified WT, S557A and S557E FOXP2 FHDs were resolved on a 16% tricine SDS-PAGE gel and the size was approximated using a molecular weight marker (Fig 3.7). In each gel (WT, S557A and S557E) there is only 1 band present indicating the protein is free from protein contamination. The size of the WT, S557A and S557E proteins were calculated to be 9.9, 10 and 10.1 kDa respectively which are all close to the predicted size, determined by Expasy ProtParam (Gasteiger *et al*, 2005), of the FHD which is 10.1 kDa.

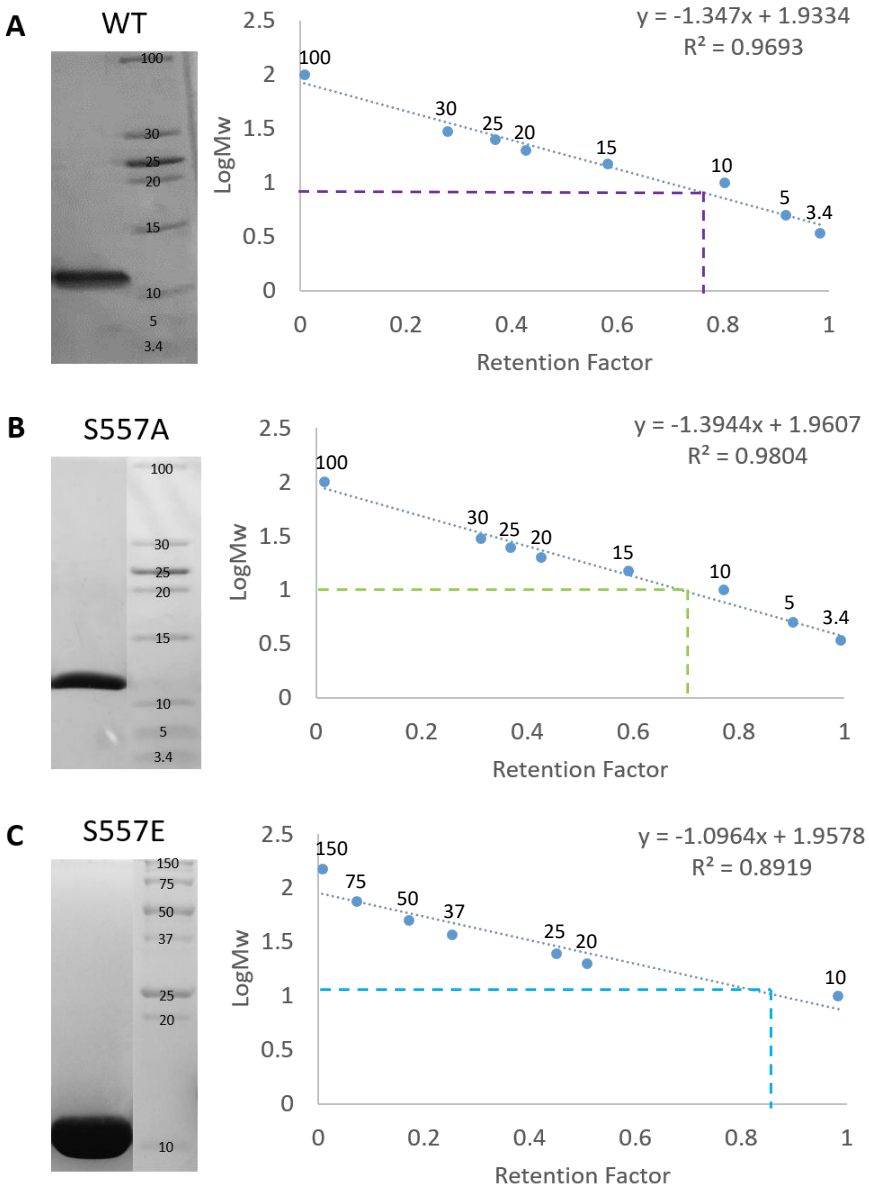

Fig 3.7. 16% tricine SDS-PAGE of the purified and cleaved WT, S557A and S557E FHDs and their calculated sizes. A sample of the purified WT(**A**), S557A(**B**) and S557E(**C**) alongside the molecular weight marker. The standard curve was used to calculate the size of the protein. The calculated sizes of the WT, S557A and S557E were 9.9, 10 and 10.1 kDa respectively. The molecular weight of each standard in the marker is indicated on the gel and on each point in the graph.

3.4. Protein purity and concentration

The concentration of the WT, S557E and S557A FOXP2 FHD was determined using a dilution series and the Beer-Lambert law (E.q. 1). An absorbance spectrum of each protein was obtained using a JASCO V-630 spectrophotometer (Fig 3.8). The peaks at 280 nm and 295 nm indicate the absorbance of tyrosine and tryptophan residues. No peak is observed at 340 nm which indicates that there is not a significant amount of aggregation. The A280/A260 ratio is 1.7 for all the proteins which indicates that the samples are free of DNA contamination.

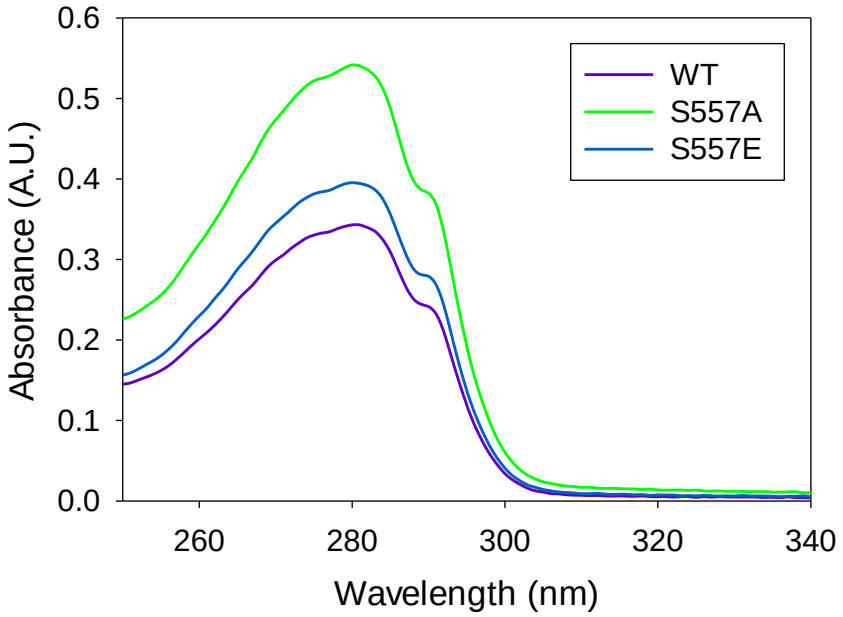

Fig 3.8. Absorbance spectra of the WT (purple), S557A (green) and S557E (blue) FOXP2 FHD. Spectra were taken on a JASCO V-630 spectrophotometer from 250 to 240 nm. All spectra peak at 280 nm, indicative of a protein sample and the A280:A260 ratio is 1.7 implying negligible DNA contamination.

3.5. Structural characterisations

Structural characterisations were performed on the WT, S557A and S557E FHD in order to see if there are any structural changes in the mutants in comparison to the WT. The structural characterisations performed included size exclusion chromatography (quaternary structure), Far-UV circular dichroism (secondary structure) and intrinsic tryptophan fluorescence (tertiary structure).

3.5.1. Size exclusion chromatography

Size exclusion chromatography was used to observe the quaternary structure of the protein in solution. The column used was a HiLoadTM 16/60 SuperdexTM 75 prep grade column (GE healthcare, Sweden). The elution and sample buffer consisted of 10 mM Tris pH 7.5, 500 mM NaCl and 30 mM imidazole.

Blue dextran (Fig. 3.9. A) was loaded first, separately from the rest of the standards as it may interact and cause incorrect sizing (Marshall, 1970). Blue dextran is larger than the fractionation range of the column and thus elutes at the column void volume. The blue dextran peak was sharp and symmetrical with exception of a small amount of tailing which indicated that the column was in good condition (Hong *et al*, 2012). The GE healthcare standards (Fig 3.9. B) resolved well and all peaks were sharp and symmetrical. The log Mw of each standard was plotted against the ratio of their respective elution volume (Ve) and the void volume (Vo). A linear regression was fitted and the equation was used to size the protein on subsequent runs. The R^2 was 0.9976 and the equation was y = -1.4479x + 3.5689.

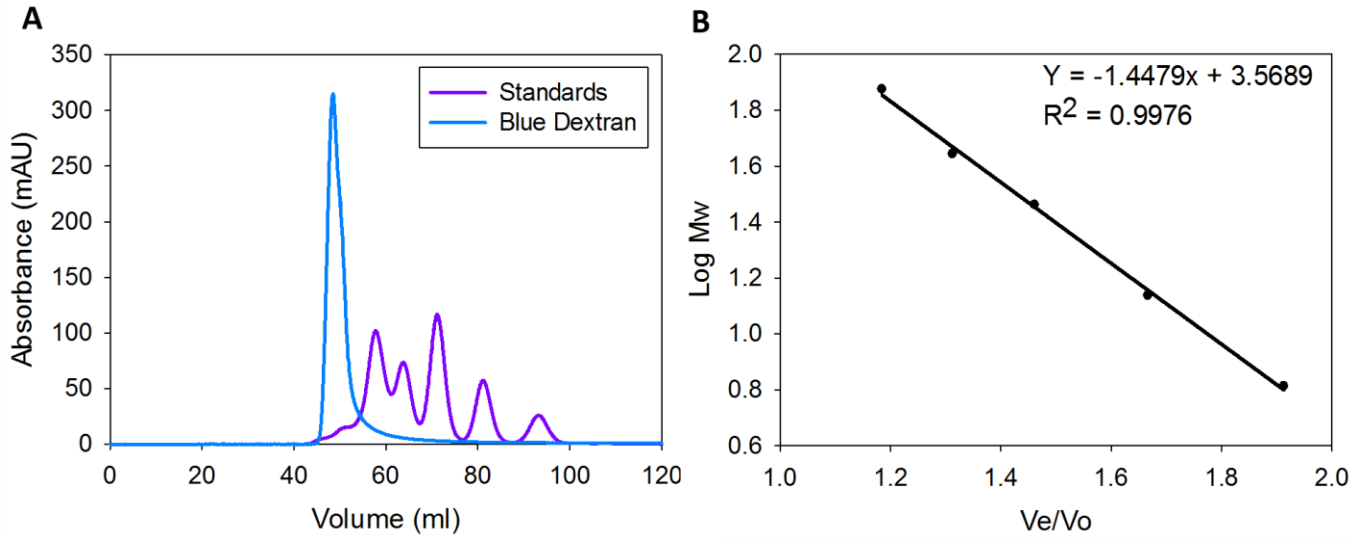

Fig. 3.9. A: Size exclusion elution profile of GE Healthcare gel filtration standards. The blue dextran (Blue) eluted at 48.3 ml which indicates the void volume. The GE healthcare standards consisted of conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa). **B: Standard curve of the gel filtration standards.** A standard curve was constructed by plotting Ve/Vo vs the Log Mw. The R^2 was 0.9976 and the equation was y = -1.4479x + 3.5689.

The WT, S557A and S557E mutants were loaded on the HiLoadTM 16/60 SuperdexTM 75 prep grade column under the same conditions as the standards. Concentrations of 50 and 100 μM were chosen to see if the mutants resulted in a change in the monomer-dimer ratio (Fig. 3.10). The monomeric FHD is 10.2 kDa (Gasteiger *et al*, 2005) and elutes at ~84 ml. Using the standard curve (Fig. 3.9), the calculated size of the monomeric FHD is 11.2 kDa. No dimer was observed for the S557A or S557E mutants but a small portion of dimer was observed for the WT at 100 μM. The dimer peak eluted at 76 ml. The calculated size of the dimer using the standard curve (Fig. 3.9) was 19.5 kDa.

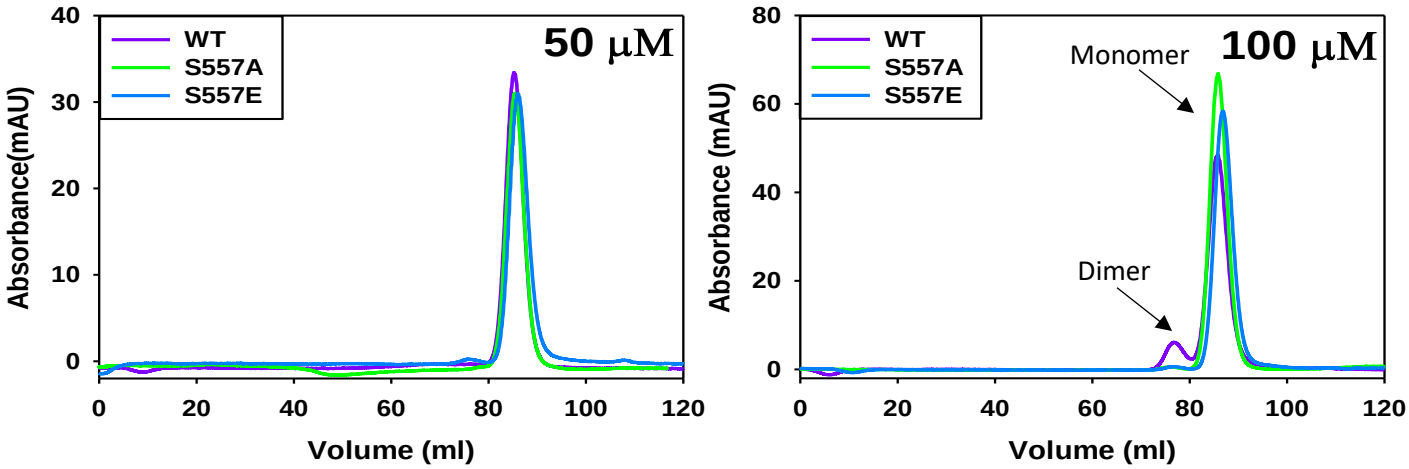

Fig. 3.10. Size exclusion chromatography profiles of the WT (purple), S557A (green) and S557E (blue) FHDs at 50 and 100 μM. The buffer used consisted of 20 mM Tris pH 7.5, 500 mM NaCl and 30 mM Imidazole. The column was a a HiLoadTM 16/60 SuperdexTM 75 prep grade column.

Because some dimer was observed for the WT at 100 μM, a range of WT protein concentrations was passed through the size exclusion column to calculate the K_d *of* dimerisation (Fig 3.11). This was found to be 896 μM using equation 3.

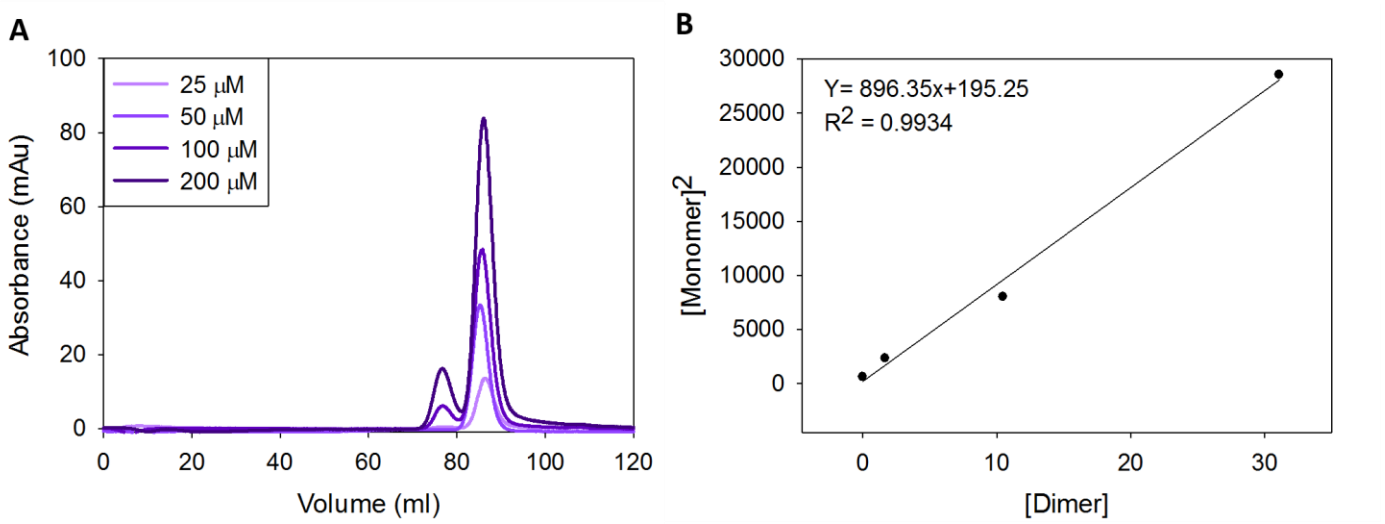

Fig. 3.11. A: Size exclusion chromatography of the WT FHD at varying concentrations (25 to 200 μM). B: WT K_d of dimerisation curve. The K_d of dimerisation of the WT FHD was calculated using the equation K_d = [Monomer]^2/[Dimer]. The K_d was calculated as 896 μM using the regression curve (Y = 896.35x + 195.25). The R^2 value of the curve was 0.9934.

3.5.2. Far-UV circular dichroism

Far-UV circular dichroism was done to observe the effect of the mutation on the overall secondary structure of the FHD. According to the crystal structure (2A07), the WT FOXP2 FHD bound to DNA (Stroud *et al*, 2006) is predominantly helical. A predominately alpha helical protein will display troughs at 222 and 208 nm and a peak at 190 nm on a far-UV circular dichroism spectrum (Kelly *et al*, 2005). In figure 3.12A the WT, S557A and S557E FHDs in the presence and absence of DNA all display the characteristic signal of an alpha helical protein. Furthermore, it can be seen in Fig 3.12C that the spectra are very similar and overlay quite well. The difference in the signal obtained for S557A is not significantly different. DNA does appear to affect the secondary structure of the WT and both mutants. In Fig 3.12A and B it is apparent that DNA acts by decreasing the helical content. Figure 3.12D and E are plots of the difference in mean residue ellipticity, at 222 and 208 nm, between the protein alone and the protein in the presence of DNA. There is no change in the difference in mean residue ellipticity at 222 nm or 208 nm between the WT and S557A. The S557E FHD however displays a greater difference between the θMRE_{222} of the protein and the protein in the presence of DNA. This indicates that the DNA reduces the alpha helical content of the

S557E protein more than it does the WT and S557A protein. The difference however is small and may not be significant.

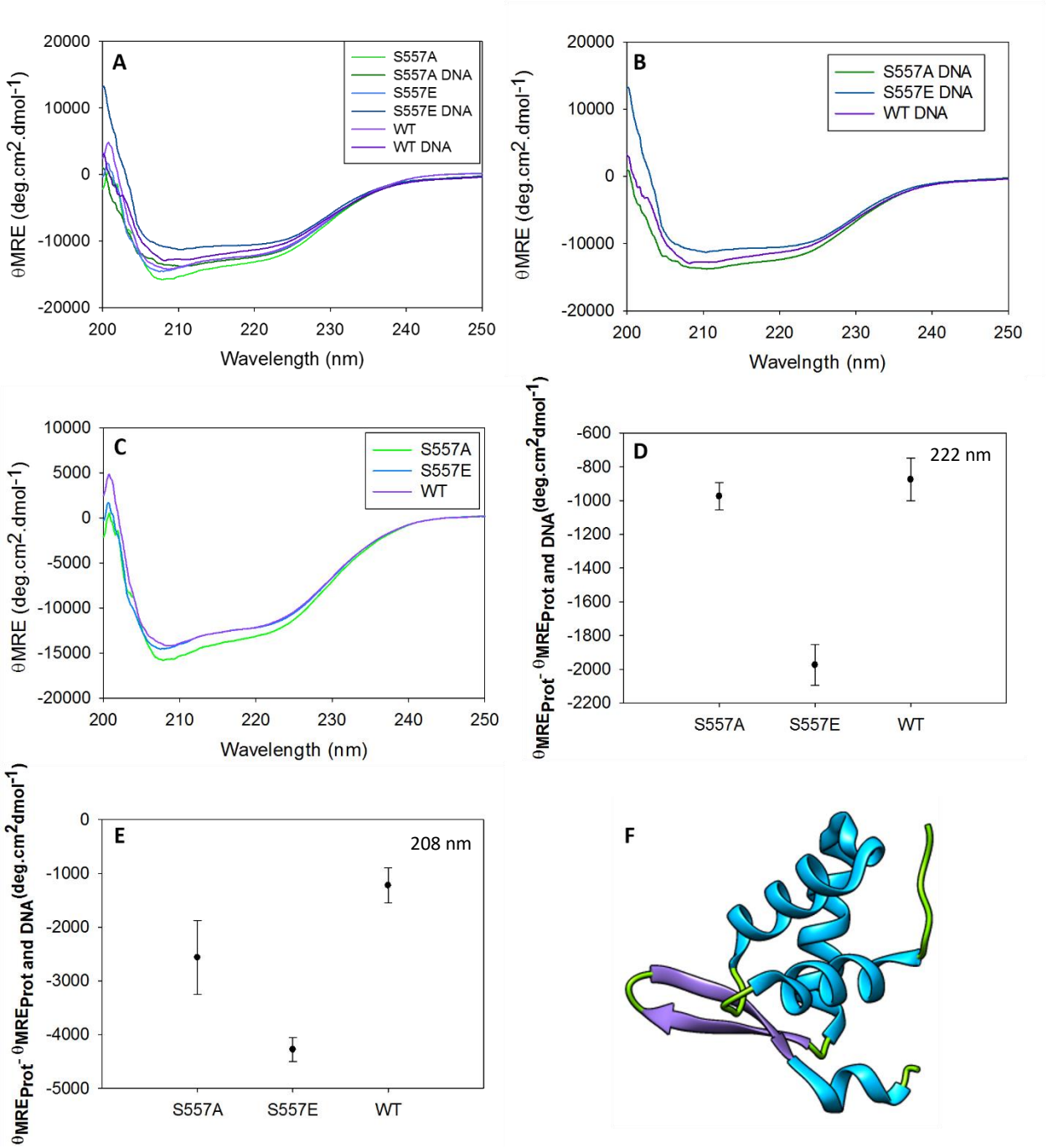

Fig. 3.12. Far-UV circular dichroism of the WT, S557A and S557E FHDs. A: Averaged Far-UV CD spectra of the S557A (green), S557E (blue) and WT (purple) FHD in the presence and absence of DNA. **B**: Averaged Far-UV CD spectra of the S557A (green), S557E (blue) and WT (purple) FHD in the presence of DNA. **C**: Averaged Far-UV CD spectra of the S557A, S557E and WT FHD in the absence of DNA. **D**: The difference in θMRE_{222} between the protein alone and in the presence of DNA. **E**: The difference in θMRE_{208} between the protein alone and in the presence of DNA.**F**: The WT FOXP2 FHD (2A07) coloured by secondary structure. Helices (blue) make up most of the structure. There is a beta-sheet shown in purple and some random coils in green.

3.5.3. Intrinsic fluorescence

Intrinsic tryptophan fluorescence was done to observe if there is a change in overall tertiary structure between the WT and Ser557 mutants. The Fluorescence spectra were recorded with excitation wavelengths at both 280 and 295 nm in the presence and absence of DNA (Fig 3.13). In Figure 3.13A and B there are two peaks visible for all the FHDs, in the absence of DNA, at the same maximum emissions (330 and 336 nm). There are no wavelength shifts in the spectra of any of the proteins in the absence of DNA at both 280 and 295 nm. The lower fluorescence intensity observed for the WT is within error and is negligible. In the presence of DNA there was no wavelength shifts of the peaks in the WT, S557A or S557E FHDs (Fig 3.13C and D). The protein fluorescence is quenched by the DNA. The most quenching is observed for the WT followed by the S557A and S557E FHDs which can be seen in Fig3.13E and F.

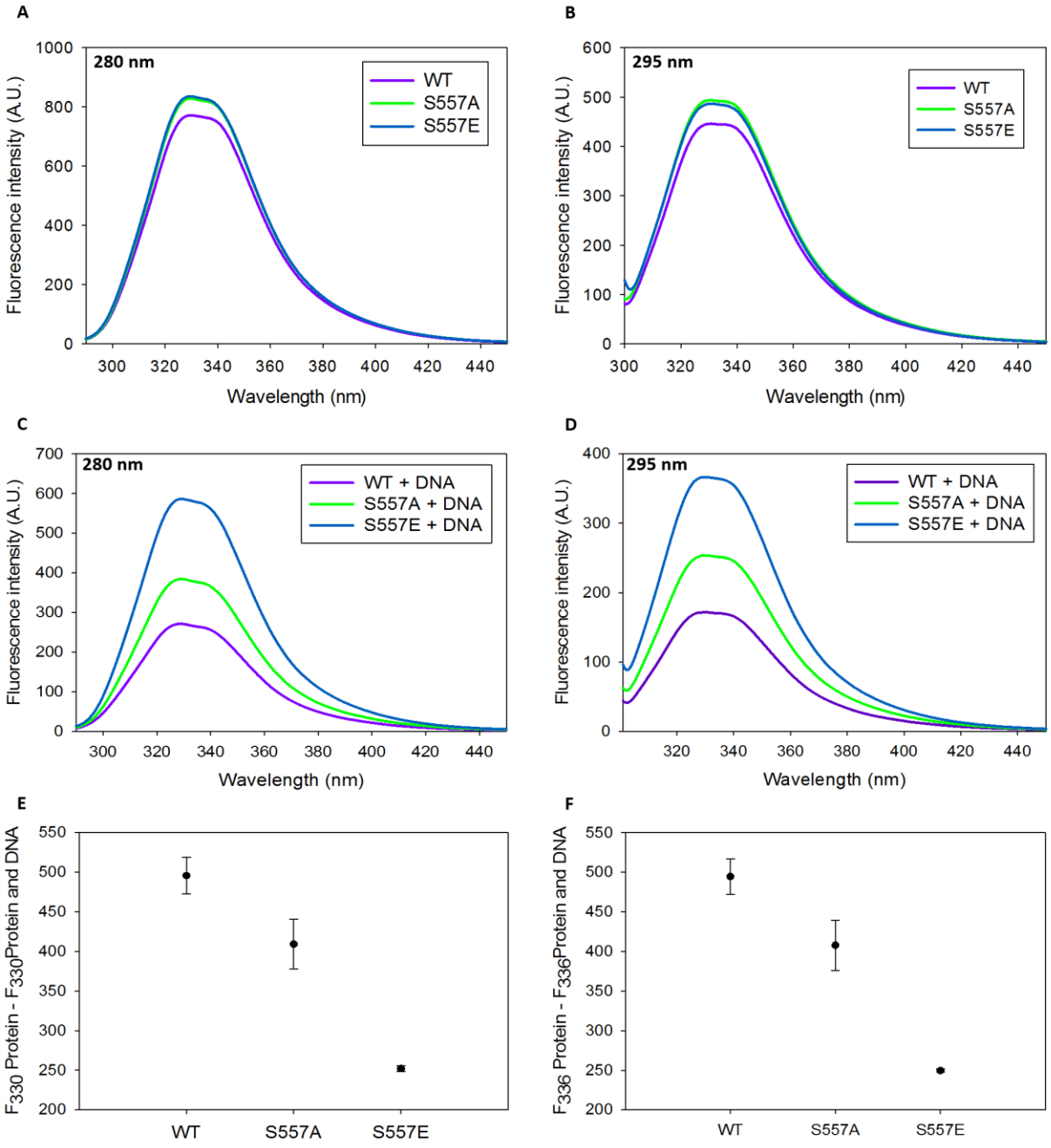

Fig 3.13. Intrinsic tryptophan fluorescence of the WT, S557E and S557A FHDs. A: Intrinsic fluorescence spectra of the WT (purple), S557A (green) and S557E (blue) with excitation wavelengths of 280 nm. **B**: Intrinsic fluorescence spectra of the WT (purple), S557A (green) and S557E (blue) with excitation wavelengths of 295 nm. **C**: Intrinsic fluorescence spectra of the WT (purple), S557A (green) and S557E (blue) in the presence of DNA with excitation wavelengths of 280 nm. **D**: Intrinsic fluorescence spectra of the WT (purple), S557A (green) and S557E (blue) in the presence of DNA with excitation wavelengths of 295 nm. **E**: The difference in fluorescence intensity (excitation 280 nm and 330 nm peak) between the WT, S557E and S557A in the absence and presence of DNA. **F**: The difference in fluorescence intensity (excitation 280 nm and 336 nm peak) between the WT, S557E and S557A in the absence and presence of DNA. Fluorescence spectra were recorded with a JASCO FP-6300 spectrofluorometer.

3.6. DNA binding studies

3.6.1. Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSA) are a good qualitative tool to observe if there is formation of a complex. EMSAs were done to see if the WT, S557A and S557E proteins are capable of binding DNA. In Fig.3.14 EMSAs of the WT, S557E and S557A FHDs are displayed. Binding for the WT was evident at 3:1 protein to DNA ratio (purple block). The S557E and S557A mutants displayed slight smears at 1:40 and 1:20 protein to DNA respectively, indicating very weak binding. Since there are bands present on the S557A and S557E FHDs EMSA, it shows that they can bind DNA and are correctly folded.

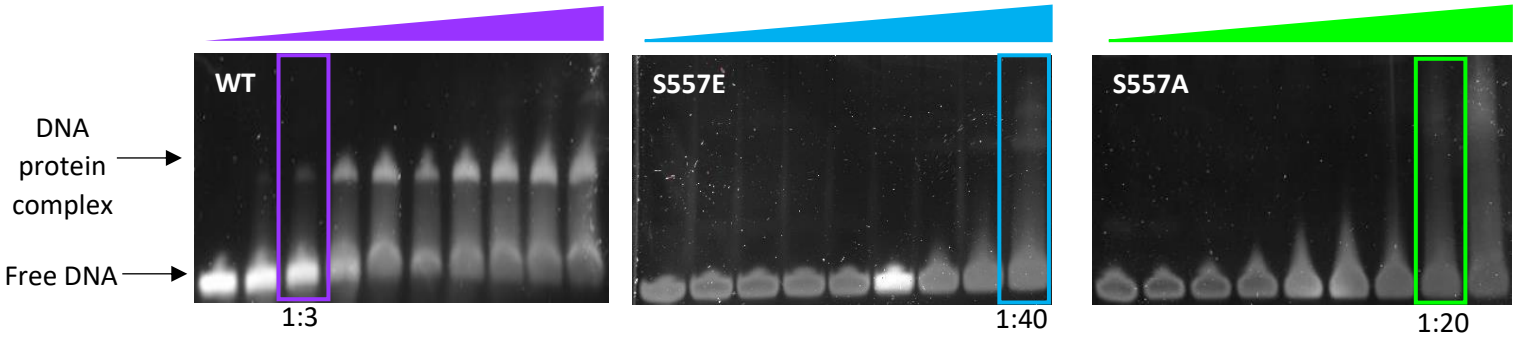

Fig. 3.14. Electrophoretic mobility shift assay of the WT (purple), S557E (blue) and S557A (green). The band present at the bottom of each gel and in each well, is the free DNA band. The band present higher up the gels (smeared) is the protein-DNA complex.

3.6.2. Isothermal titration calorimetry

Isothermal titration calorimetry was done to obtain quantitative binding parameters of the FHD binding to DNA. S557A and S557E mutants did not display any binding under the same conditions used for the WT (Fig. 3.15). The parameters obtained from fitting the WT ITC (Table 3.1) are similar to values obtained from previous work (Morris & Fanucchi, 2016). The WT FHD binds to Nelson DNA with a large enthalpic term (-27.14 kJ/mol). The entropic term is also favourable (TΔS is 10.96 kJ/mol/K). The ΔG is -38.1 kJ/mol and the K_d was found to be 0.1628 μM. The stoichiometry of 1.423 however was unexpected. This could be a result of a monomer: dimer mixture where the dimer could possibly bind two strands of DNA or there may be an additional binding site on the DNA.

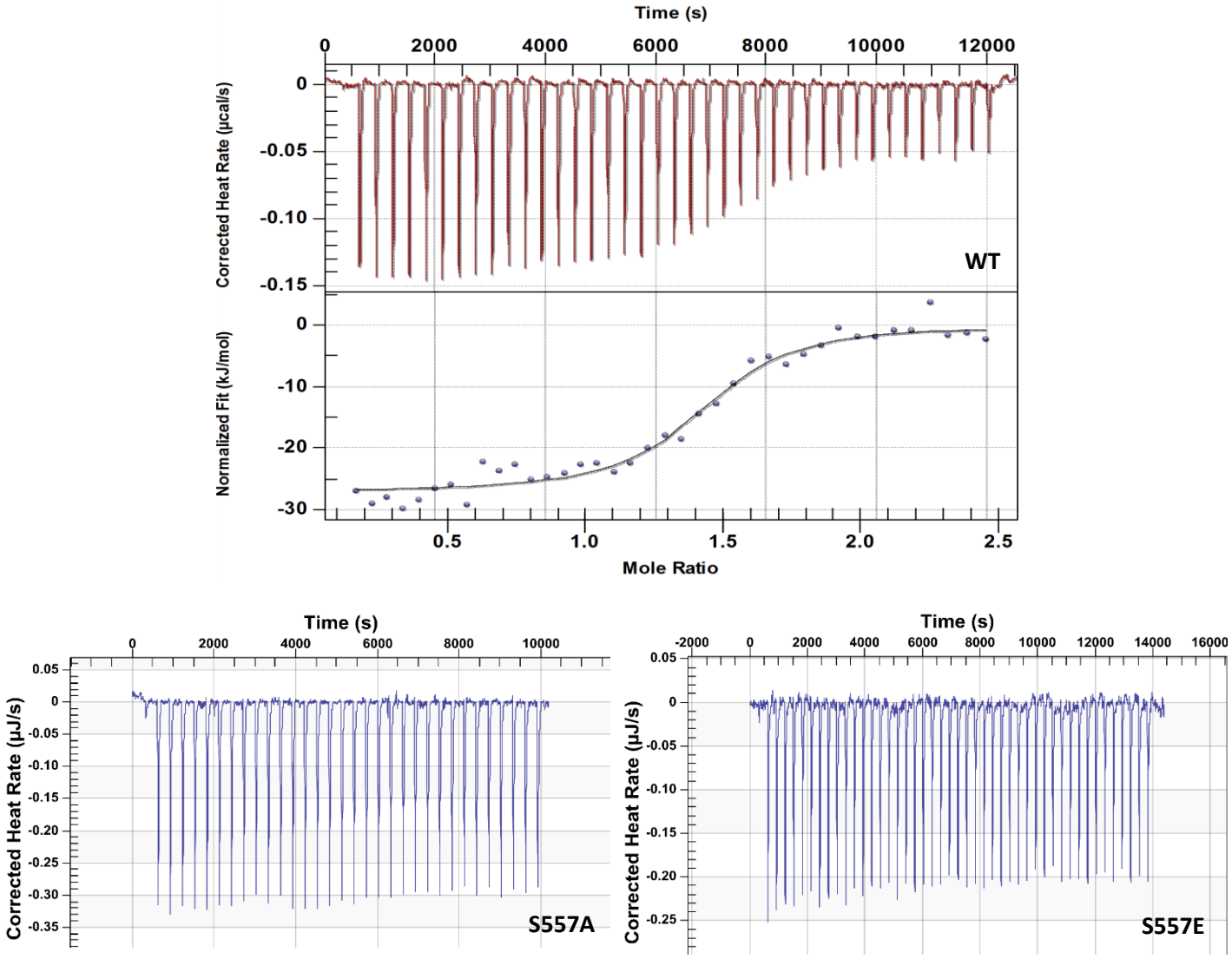

Fig. 3.15. ITC titrations of the WT, S557A and S557E FHD into Nelson DNA. The isotherms were generated by titrating multiple small volumes of protein (90-120 µM) into DNA (8- 10 µM). Only the WT shows binding with a K_d of 162.8 nM.

Table 3.1 Thermodynamic parameters obtained from titrating WT into Nelson DNA.

Variable	Value	Error
K_d (μM)	0.1628	±0.062
n	1.423	±0.04
ΔH (kJ/mol)	-27.14	±1.282
ΔG (kJ/mol)	-38.1	
TΔS (kJ/mol/K)	10.96	

3.6.3. Molecular docking

Modelled Nelson DNA was docked to the structure of the WT, phosphorylated WT (Phos-WT) and the Ser557 mutants using HADDOCK. The S557A and S557E protein models were made by mutating Ser557 to alanine and glutamate respectively using Chimera 1.7 (Pettersen *et al*, 2004). The Phos-WT model was made by phosphorylating the WT crystal structure using PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC with the PyTMs plugin. Docking to DNA was done with HADDOCK. HADDOCK scores the models using a score comprised of the electrostatic energy, van der Waals energy, desolvation energy and restraints violation energies (Kastritis & Bonvin, 2010). The HADDOCK scores however are not comparable between different complexes. In general docking scores are not correlated to binding affinity and their purpose is the comparison of different docked models of the same structure (Kastritis & Bonvin, 2010). For this reason Chimera 1.7 (Pettersen *et al*, 2004) was used to analyse binding of each model by looking at the number of hydrogen bonds formed, the number of contacts and the surface area buried upon binding (Fig 3.16).

The WT has 6 potential hydrogen bonds whereas the Phos-WT, S557A and S557E models have 2, 4 and 4 respectively (Table 3.2, Fig. 3.17). The WT in addition to having the most hydrogen bonds also has the greatest buried surface area upon DNA binding and forms the most contacts. The S557A mutant model has fewer hydrogen bonds than the WT but exhibits similar buried surface area upon DNA binding and number of contacts. The S557E mutant and phosphorylated WT models display the fewest contacts and the smallest buried surface area upon binding. This suggests that the WT and S557A FHDs may have the greatest number of

interactions with the DNA and likely greater binding affinities than the S557E mutant and Phos-WT.

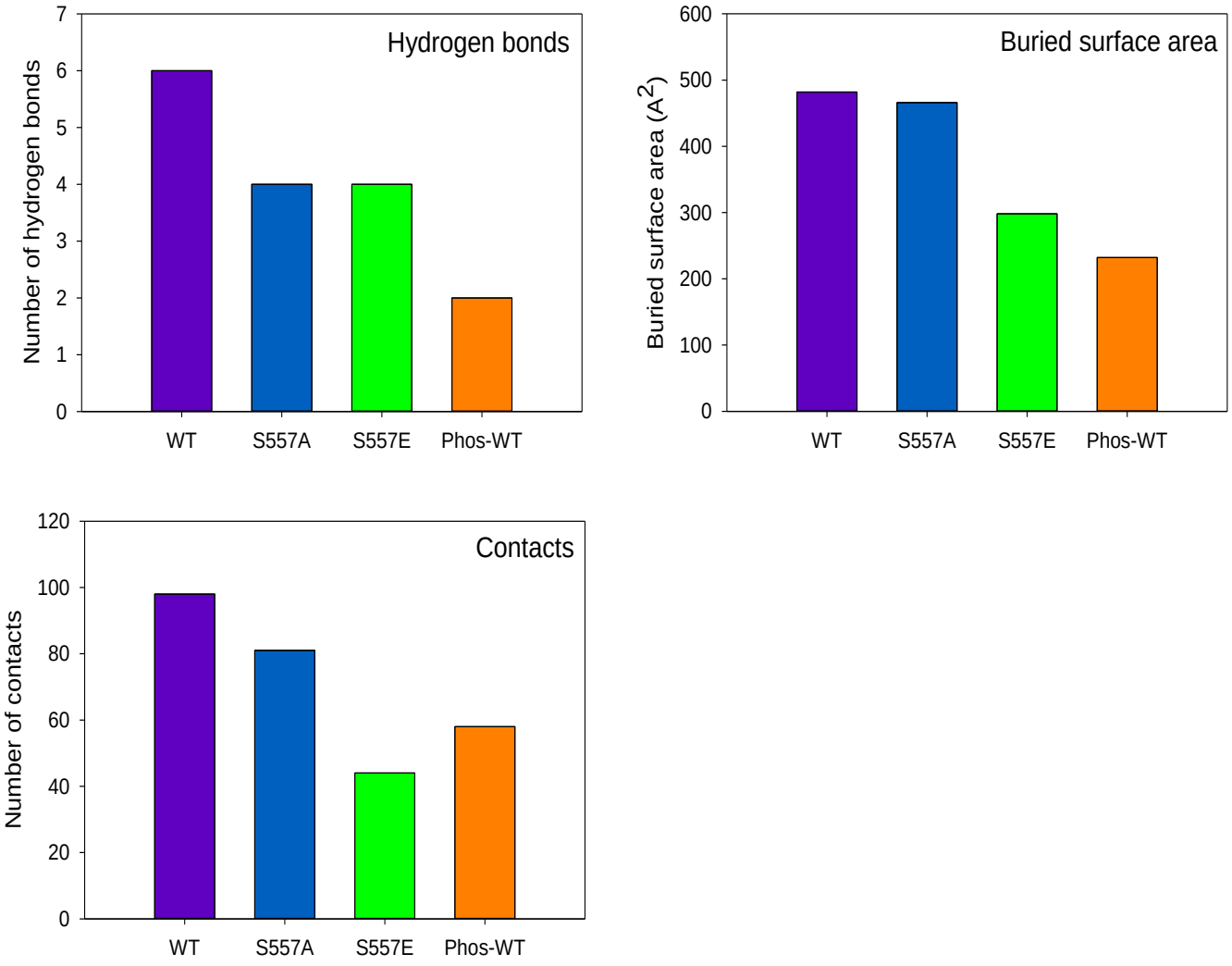

Fig. 3.16. Number of hydrogen bonds (A), buried surface area upon DNA binding (B) and number of contacts (C) of the docked WT, S557A, S557E and Phos-WT models. The models were docked using HADDOCK to a model of Nelson DNA. The number of hydrogen bonds, buried surface area and number of contacts was determined using Chimera 1.7 (Pettersen *et al*, 2004).

Table 3.2: Hydrogen bonds formed between protein residues and nucleotides in the docked models of the WT, phosphorylated WT, S557A and S557E.

	Residue	Nucleotide
WT	Asn550	Thy29
	Arg553	Thy30 **
	Asn555	Ade12, Ade13
	Tyr509	Ade13
Phos-WT	Asn555	Ade13
	Tyr509	Ade 13
S557A	Arg553	Thy30 ***
	Tyr509	Ade12
S557E	Arg 553	Thy30 **
	Asn550	Ade13
	Tyr509	Ade13

The shaded nucleotide indicates at hydrogen bond formed between a specific DNA base rather than a backbone interaction. * represent interactions where multiple hydrogen bonds may be shared among the same residue and nucleotide.

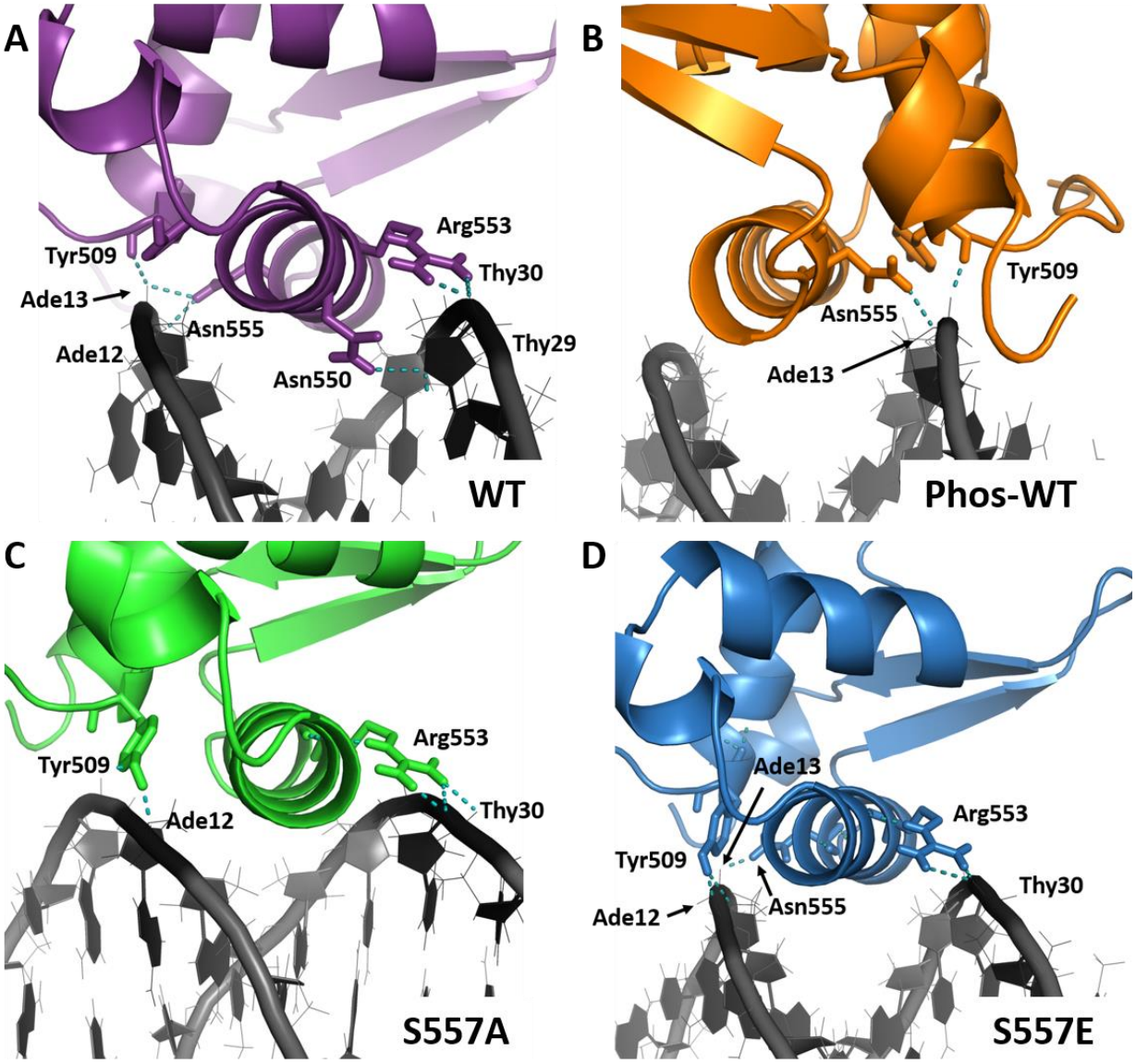

Fig. 3.17. WT (A), Phos-WT (B), S557A (C) and S557E (D) docked to Nelson DNA using HADDOCK. The protein models were modified from the 2A07 crystal structure (Stroud *et al*, 2006) and the DNA models were made in the Acceryls discovery suite 4. The hydrogen bonds between the protein and DNA in each model are displayed with blue dashes. The residues involved in hydrogen bonding are labelled.

A structural alignment of the Phos-WT, S557A and S557E docked models to the WT docked model was done using PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC (Fig 3.18). The RMSD values for the alignment of the WT to Phos-WT, S557A and S557E were 0.495, 0.339 and 0.371 respectively. It is apparent that the proteins align well with the biggest differences occurring in the random coil regions. The major difference between the models is the DNA. This

suggests that the phosphomimetic, and phosphorylation, does not significantly alter the structure of the protein but rather the interactions and position of the DNA when binding occurs. It is apparent from the alignment that the phosphorylated model is not bound as deeply in the major groove of the DNA as the WT model. This could account for the decreased binding affinity observed with the EMSA and ITC experiments.

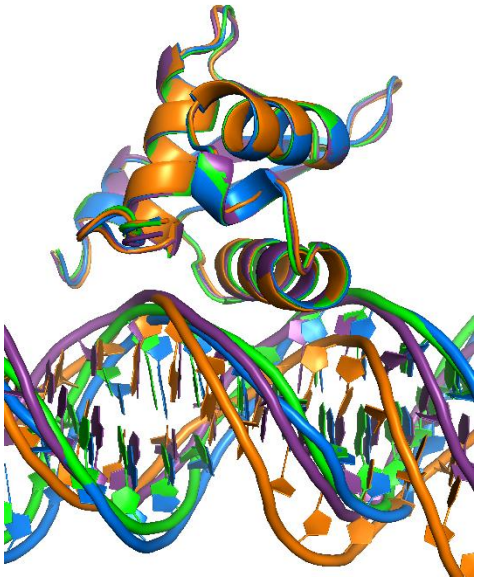

Fig. 3.18. Structural alignment of the docked models of the WT (purple), Phos-WT (orange), S557A (green) and S557E (blue). The Phos-WT, S557A and S557E complexes were aligned to the WT structure and the RMSD values were 0.495, 0.339 and 0.371 respectively. Alignment was done using the Schrodinger PyMol molecular graphics system.

CHAPTER 4 Discussion

The aim of the study was to establish the possible effects of phosphorylation of Ser557 on the structure and DNA binding of the FOXP2 FHD. A phosphate group adds considerable charge and bulk to a residue. This has been known to cause conformational, structural and electrostatic changes to proteins (Johnson & Barford, 1993). In some cases, this PTM changes the function of a protein and even acts as an "on" or "off" switch. In FOXO1, phosphorylation of the equivalent serine (Ser218) inhibits DNA binding (Brent *et al*, 2008). To observe if this is the case for Ser557 in FOXP2, a glutamate mutant was made so as to mimic phosphorylation. An alanine mutant was also made to act as a control to ensure changes observed were not due to disruption of the structure of the protein. Since Ser557 is located at the C-terminal end of helix 3 in the FOXP2 FHD, it is likely that the addition of a phosphate group (bulky and negatively charged) will electrostatically and sterically inhibit DNA binding. If phosphorylation of Ser557 does inhibit DNA binding of FOXP2, phosphorylation may be a mechanism by which FOXP2 is regulated *in vivo*.

As phosphorylation can induce structural changes in proteins (Johnson & Barford, 1993), Far-UV circular dichroism and intrinsic tryptophan fluorescence were performed to establish if the S557E mutation causes secondary and tertiary structural changes. The Far-UV CD spectra (Fig 3.14C) of the S557E and S557A mutants overlay well the WT which indicates that the secondary structure of the FHDs is the same. This implies that the secondary structure of the FHD is not altered by mutation of serine 557 or the phosphomimetic. In addition to this, the shape of the spectra indicates that the proteins are predominantly alpha helical which corresponds with the crystal structure (2A07) which gives evidence that the WT, S557A and S557E FHDs are correctly folded.

The intrinsic tryptophan fluorescence spectra of the WT, S557A and S557E (excitation at 295 nm and 280 nm) have an emission maximum at 330 nm with a shoulder at 336 nm (Fig 3.13). The FHD contains three tryptophan residues so the presence of the shoulder implies that one of the tryptophans is in a different polar environment to the others thus has a shifted emission. The spectra of the mutants do not overlay with the WT; however, the lack of wavelength shifts indicates that the fluorophores are in the same environment as they are in the WT. Amino acids with charged side chains, electron acceptors (protonated carboxyl groups), disulfides and amines have been

shown to quench fluorescence (White, 1958; Christov *et al*, 2004; Lakowicz, 2006). However, the decrease in fluorescence intensity of the WT is unlikely to be due to quenching as the ratio between the peak and the shoulder is the same in the WT (1.01) and mutants (1.01). The decrease in fluorescence intensity, in the WT, is therefore likely to be caused by a slight difference in concentration of the protein rather than a tertiary structural change. The conserved secondary and tertiary structure in the glutamate mutants shows that if there is a change in the protein's DNA binding upon phosphorylation of Ser557 it is unlikely to be due to conformational changes but more likely to be due to electrostatic or steric interference.

Size exclusion chromatography was used to establish at what concentrations the WT, S557A and S557E FHDs were monomeric and assess any changes that phosphorylation of Ser557 may have on the quaternary structure of the FHD. Both serine 557 mutants exhibited a decreased propensity to dimerise at concentrations at which the WT FHD showed a small portion of dimer (Fig 3.10). This indicates that Ser557 does play a role in dimerisation of FOXP2, and the mutation of Ser557 inhibits dimerisation since both the serine mutants (S557A and S557E) do not dimerise at concentrations below 100 μM. When the FOXP2 FHD forms a domain swapped dimer, helix 2 is extended as the turn between helix 2 and helix 3 opens (Fig 1.6). The domain interface consists of multiple hydrophobic residues with charged and polar residues facing away from the interface (Stroud *et al*, 2006). According to the crystal structure (2A07) Ser557, located on the C-terminal end of helix 3, does not appear to play any role in the dimer interface or be involved in DNA binding in the dimeric form. The mutation and phosphorylation of Ser557 may affect the formation of the dimer by preventing the reshuffling of the domains or even stabilisation of the helix in the domain swapped form. Interestingly, although not much is known about DNA binding of the dimeric form of the FOXP2 FHD, it has been postulated that the dimeric form of the FHD may bind DNA at a higher affinity than the monomer (Morris & Fanucchi, 2016). Furthermore, the presence of DNA may induce dimerisation of the FHD and that the purpose of dimerisation may be chromosomal looping or DNA bending (Stroud *et al*, 2006).

An electrophoretic mobility shift assay was initially used to probe if the mutant protein could bind DNA and give an indication of the K_d in relation to the WT. The EMSAs showed that the S557E and S557A mutants could bind DNA however the K_d appeared to be significantly lower than that of the WT. The quality of the S557A and S557E EMSAs did not match up to that of the WT as the

protein- DNA complex could not be distinguished. The smearing of bands in EMSAs is a common occurrence when the complex has a high K_d and the complex is thus not very stable and dissociates (Hellman & Fried, 2007). The presence of the band corresponds to a protein- DNA complex on the EMSA's (Fig 3.15) of the serine mutants along with the Far-UV circular dichroism and intrinsic fluorescence data all indicate that the mutant proteins are correctly folded and are capable of binding DNA. However, no binding for either mutant was observed when using ITC under the same conditions as the WT. This discrepancy can be explained because when using ITC, the high protein to DNA ratio used in EMSAs was not reached. In addition to this, in EMSAs there is a caging effect, which may stabilise the protein: DNA complex. This is because as the complex dissociates, the binding partners cannot diffuse away due to the gel matrix and may associate again quickly (Sidorova *et al*, 2010). ITC being a technique where molecules are free in solution rather than in a gel would not experience this phenomenon.

Although S557A binding DNA was not observed using ITC the EMSA and fluorescence studies in the presence of DNA suggests that the S557A mutant binds with a greater affinity than the S557E mutant. In the EMSAs the protein-DNA complex is present for the S557A mutant at half the protein concentration required of the S557E mutant. Fluorescence studies in the presence of DNA resulted in quenching for all three proteins. The quenching effect was most noted in the WT, followed by the S557A mutant. DNA is known to quench fluorescence statically via direct base stacking or indirectly as a result of conformational changes (Lakowicz, 2006). As there are no tryptophans that interact directly with the DNA it is likely that the quenching is a result of a conformational change in the protein upon DNA binding. The amount of quenching may be related to DNA binding as more protein molecules bound to DNA would result in a greater quenching effect. The greater quenching of the S557A mutant in comparison to the S557E mutant indicates that the S557A mutant binds DNA with a greater affinity.

The significantly decreased DNA binding affinity of the mutants compared to the WT alludes to the importance of Ser557 upon binding of the FHD to DNA. In the crystal structure of the WT FHD bound to Wang DNA (2A07), Ser557 on the monomeric FHD forms a hydrogen bond with the DNA backbone. The loss of this hydrogen bond in the S557A mutant would reduce the binding affinity to some degree which is observed. The S557E, mutant with its increased negative charge,

likely causes electrostatic repulsion which could prevent helix 3 from entering the major groove of the DNA to some extent. If Ser557 were phosphorylated, the bulk and greater negative charge would further inhibit DNA binding. This would be due to electrostatic repulsion between the negatively charged oxygen ions and the negatively charged DNA backbone. The increased bulk of a phosphate group would sterically inhibit helix 3 from entering the major groove of the DNA.

To complement the DNA binding studies and try to obtain more detailed information about the possible changes in DNA binding of the FOXP2 FHD when Ser557 is phosphorylated, molecular docking studies were performed. The FHD in the crystal structure (2A07) was modified to produce S557E, S557A and phosphorylated WT models which were all subsequently docked to modelled Nelson DNA. The interactions in the interface of the models were analysed by looking at the number of hydrogen bonds, contacts and buried surface area. The docking studies complemented the EMSA and ITC data as they indicated that the S557E mutant has a decreased DNA binding affinity in comparison to the WT. The S557A mutant and the WT had the greatest buried surface areas and highest number of contacts compared to the S557E mutant and Phos-WT (Fig 3.16). This could indicate that the use of the phosphomimic, in this case, could give reliable information about phosphorylated FOXP2. Furthermore, the number of contacts, hydrogen bonds and buried surface area obtained from the docked models indicate that it is likely that the WT and S557A mutant bind at a stronger K_d than the S557E and Phos-WT. Interestingly a structural alignment of the docked models revealed that helix 3 in the phosphorylated model was not docked as deeply into the major grove of the DNA when compared to the WT. This suggests that the bulk and charge of the phosphate group prevents helix 3 from entering the major groove when serine 557 is phosphorylated thus inhibiting DNA binding.

CONCLUSION

Serine 557 is a residue in the DNA binding helix of the FOXP2 FHD. A glutamate mutant (S557E) was used, alongside a control (S557A) to determine the possible effects that phosphorylation of this serine would have on the structure and DNA binding of the FOXP2 FHD. Both the glutamate and alanine mutants had conserved secondary and tertiary structure in comparison with the WT indicating that it is likely phosphorylation will not alter the secondary or tertiary structure of the FHD. Mutation of Ser557 does however prevent dimerisation of the FHD which highlights not only importance of the serine in maintaining the dimer but also indicates that phosphorylation could prevent domain swapping. Most notably, the S557E mutation showed significantly reduced DNA binding when compared to the WT FHD. This indicates that phosphorylation of Ser557 could disrupt DNA binding likely to due to electrostatic and steric hindrance. This shows that phosphorylation of Ser557 in the FOXP2 FHD could act as a control mechanism for FOXP2 and ultimately could be involved in regulation of transcription.

CHAPTER 5 References

Akira, S. (1997) IL-6-regulated transcription factors. *International journal of Biochemistry and Cell Biology*, **29,** 1401–1418.

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

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

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

Banham, A.H., Beasley, N., Campo, E., Fernandez, P.L., Fidler, C., Gatter, K., Jones, M., Mason, D.Y., Prime, J.E., Trougouboff, P., Wood, K. & Cordell, J.L. (2001) The FOXP1 winged helix transcription factor is a novel candidate tumor suppressor gene on chromosome 3p1. *Cancer Research*, **61,** 8820–8829.

Bannister, J., Miska, E., Görlich, D. & Kouzarides, T. (2000) Acetylation of importin-alpha nuclear import factors by CBP/p300. *Current biology*, **10,** 467–470.

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

Berry, F.B., Tamimi, Y., Carle, M. V., Lehmann, O.J. & Walter, M.A. (2005) The establishment of a predictive mutational model of the forkhead domain through the analyses of FOXC2 missense mutations identified in patients with hereditary lymphedema with distichiasis. *Human Molecular Genetics*, **14,** 2619–2627.

Biggs, W.H., Meisenhelder, J., Hunter, T., Cavenee, W.K. & Arden, K.C. (1999) Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix

transcription factor FKHR1. *Proceedings of the National Academy of Sciences of the United States of America*, **96,** 7421–7426.

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

Blurnenfeld, F. & Gardner, J. (1985) Gel filtration-an innovative separation technique. *Journal of Chemical Education*, **62,** 715–717.

Brahms, S. & Brahms, J. (1980) Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *Journal of Molecular Biology*, **138,** 149–178.

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

Brunet, A., Bonni, A., Zigmond, M.J., Lin, M.Z., Juo, P., Hu, L.S., Anderson, M.J., Arden, K.C., Blenis, J. & Greenberg, M.E. (1999) Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell*, **96,** 857–868.

Burley, S.K. (1997) Eukaryotic transcription factor-DNA complexes. *Annual Review of Biophysics and Biomolecular Structure*, **26,** 289–325.

Carlsson, P. & Mahlapuu, M. (2002) Forkhead transcription factors: key players in development and metabolism. *Developmental biology*, **250,** 1–23.

Chen, Y.J., Dominguez-Brauer, C., Wang, Z., Asara, J.M., Costa, R.H., Tyner, A.L., Lau, L.F. & Raychaudhuri, P. (2009) A conserved phosphorylation site within the forkhead domain of FoxM1B is required for its activation by cyclin-CDK1. *Journal of Biological Chemistry*, **284,** 30695–30707.

Christov, C., Ianev, D., Shosheva, A. & Atanasov, B. (2004) pH-Dependent quenching of the fluorescence of tryptophan residues in class A -lactamase from E . coli (TEM-1). *Zeitschrift für Naturforschung C*, **59,** 824–827.

Chu, Y.-P., Chang, C.-H., Shiu, J.-H., Chang, Y.-T., Chen, C.-Y. & Chuang, W.-J. (2011)

Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: insight into its domain swapping and DNA binding. *Protein Science*, **20,** 908–24.

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

Conaway, R.C. & Conaway, J.W. (1993) General initiation factors for RNA polymerase II. *Annual review of biochemistry*, **62,** 161–190.

Cowell, I.G. (1994) Repression versus activation in the control of gene transcription. *Trends in Biochemical Sciences*, **19,** 38–42.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.M., Taly, J.F. & Notredame, C. (2011) T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, **39,** 1–5.

Dominguez, C., Boelens, R. & Bonvin, A.M.J.J. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, **125,** 1731–1737.

Fisher, S.E. & Scharff, C. (2009) FOXP2 as a molecular window into speech and language. *Trends in Genetics*, **25,** 166–177.

Friedman, J.R. & Kaestner, K.H. (2006) The Foxa family of transcription factors in development and metabolism. *Cellular and Molecular Life Sciences*, **63,** 2317–2328.

Garner, M.M. & Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of protteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Research*, **9,** 3047–3060.

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

Geiss-friedlander, R. & Melchior, F. (2007) Concepts in sumoylation: a decade on. *Nature Reviews Molecular Cell Biology*, **8,** 947–956.

Gill, G. (2004) Review SUMO and ubiquitin in the nucleus : different functions, similar mechanisms? *Genes and development*, **18,** 2046–2059.

Gill, S.C. & von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Analytical biochemistry*, **182,** 319–26.

Greenfield, N.. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, **1,** 2876–2890.

Grünberg, S. & Hahn, S. (2013) Structural insights into transcription initiation by RNA polymerase II. *Trends in Biochemical Sciences*, **38,** 603–611.

Hellman, L. & Fried, M. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein- nucleic acid anteractions. *Nature protocols*, **2,** 1849–1861.

Hong, P., Koza, S. & Bouvier, E.S.P. (2012) Size-Exclusion Chromatography for the analysis of protein biotherapeutics and their aggregates. *Journal of liquid chromatography & related technologies*, **35,** 2923–2950.

Huffman, J.L. & Brennan, R.G. (2002) Prokaryotic transcription regulators: More than just the helix-turn-helix motif. *Current Opinion in Structural Biology*, **12,** 98–106.

Johnson, L. & Barford, D. (1993) The effects of phosphorylation on the structure and function of proteins. *Annual review of biochemistry*, **22,** 199–232.

Kastritis, P.L. & Bonvin, A.M.J.J. (2010) Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *Journal of Proteome Research*, **9,** 2216–2225.

Kelly, S.M., Jess, T.J. & Price, N.C. (2005) How to study proteins by circular dichroism. *Biochimica et Biophysica Acta - Proteins and Proteomics*, **1751,** 119–139.

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

Research, **35,** 43–46.

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

Ladbury, J.E. & Chowdhry, B.Z. (1996) Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chemistry & biology*, **3,** 791–801.

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

Lakowicz, J.. (2006)Principles of fluorescence spectroscopy 3rd Edition. Baltimore, USA: Springer.

Latchman, S. (1997) Transcription factors : An overview. *International journal of Biochemistry and Cell Biology*, **29,** 1305–1312.

Lehtinen, M.K., Yuan, Z., Boag, P.R., Yang, Y., Villén, J., Becker, E.B.E., Di Bacco, S., de la Iglesia, N., Gygi, S., Blackwell, T.K. & Bonni, A. (2006) A conserved MST-FOXO signaling pathway mediates oxidative-stress responses and extends life span. *Cell*, **125,** 987–1001.

Li, S., Weidenfeld, J. & Morrisey, E.E. (2004) Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Molecular and cellular biology*, **24,** 809–22.

Lowy, A.M. (1996) Mutation of phosphoserine 389 affects p53 function in vivo. *Journal of Biological Chemistry*, **271,** 29380–29385.

Marshall, J. (1970) Comments on the use of blue dextran in gel chromatography. *Journal of Chromatography*, **53,** 379–380.

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

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

Nie, H., Zheng, Y., Li, R., Guo, T.B., He, D., Fang, L., Liu, X., Xiao, L., Chen, X., Wan, B., Chin, Y.E. & Zhang, J.Z. (2013) Phosphorylation of FOXP3 controls regulatory T cell function and is inhibited by TNF-α in rheumatoid arthritis. *Nature medicine*, **19,** 322–8.

Notredame, C., Higgins, D.G. & Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, **302,** 205–217.

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

Papworth, C., Bauer, J., Braman, J. & Wright, D. (1996) Site-directed mutagenesis in one day with >80% efficiency. *Strategies***,** 3–4.

Pedersen, A.G., Baldi, P., Chauvin, Y. & Brunak, S. (1999) The biology of eukaryotic promoter prediction-a review. *Computers & chemistry*, **23,** 191–207.

Perumal, K., Dirr, H.W. & Fanucchi, S. (2015) A single amino acid in the hinge loop region of the FOXP Forkhead domain is significant for dimerisation. *Protein Journal*, **34,** 111–121.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. & Ferrin, T.E. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25,** 1605–1612.

Potter, L.R. & Hunter, T. (1998) Phosphorylation of the kinase homology domain is essential for activation of the A-type natriuretic peptide receptor. *Molecular and cellular biology*, **18,** 2164–2172.

Roberts, S.G.E. (2000) Mechanisms of action of transcription activation and repression domains. *Cellular and Molecular Life Sciences*, **57,** 1149–1160.

Roeder, R.. (1991) The complexities of eukaryotic transcription initiation: Regulation of preinitiation complex assembly. *Science*, **16,** 402–408.

Roeder, R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences*, **21,** 327–335.

Ruthven, C.R.J., Bernhard, T., Memorial, B. & Charlotte, Q. (1954) The separation of substances and estimation of their relative molecular sizes by the use of columns of starch in water. *Biochemical Journal*, **62,** 665–674.

Sainsbury, S., Bernecky, C. & Cramer, P. (2015) Structural basis of transcription initiation by RNA polymerase II. *Nature reviews. Molecular cell biology*, **16,** 129–143.

Schägger, H. (2006) Tricine–SDS-PAGE. *Nature Protocols*, **1,** 16–22.

Schägger, H. & Gebhard, V.J. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical biochemistry*, **379,** 368–379.

Sharma, S., Yang, S., Zhu, L., Reckamp, K. & Gardner, B. (2005) Tumor cyclooxygenase-2 / prostaglandin E2 − dependent promotion of FOXP3 expression and CD4 + CD25 + T regulatory cell activities in lung cancer. *Cancer Research*, **65,** 5211–5220.

Sharrocks, A.D. (2001) The ETS-domain transcription factor family. *Nature reviews. Molecular cell biology*, **2,** 827–837.

Shiromizu, T., Adachi, J., Watanabe, S., Murakami, T., Kuga, T., Muraoka, S. & Tomonaga, T. (2013) Identification of missing proteins in the neXtProt database and inregistered phosphopeptides in the PhosphoSitePlus database as part of the chromosome-centric human proteome project. *Journal of Proteome Research*, **12,** 2414–2421.

Shu, W., Lu, M.M., Zhang, Y., Tucker, P.W., Zhou, D. & 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. & Morrisey, E.E. (2001) Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. *The Journal of biological chemistry*, **276,** 27488–97.

Sidorova, N.Y., Hung, S. & Rau, D.C. (2010) Stabilizing labile DNA-protein complexes in polyacrylamide gels. *Electrophoresis*, **31,** 648–653.

Śmiechowski, M. (2010) Theoretical pKa prediction of O-phosphoserine in aqueous solution. *Chemical Physics Letters*, **501,** 123–129.

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

Tarrant, M.K. & Cole, P.A. (2009) The chemical biology of protein phosphorylation. *Annual review of biochemistry*, **78,** 797–825.

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

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

Van Dyke, M.W., Roeder, R.G. & Sawadogo, M. (1988) Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science*, **241,** 1335–1338.

Verger, A., Perdomo, J. & Crossley, M. (2003) Modification with SUMO. A role in transcriptional regulation. *EMBO reports*, **4,** 137–142.

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

Vosseller, K., Wells, L. & Hart, G.W. (2001) Nucleocytoplasmic O-glycosylation : O-GlcNAc and functional proteomics. *Biochimie*, **83,** 575–581.

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

Wasylyk, C., Bradford, a P., Gutierrez-Hartmann, a & Wasylyk, B. (1997) Conserved mechanisms of Ras regulation of evolutionary related transcription factors, Ets1 and pointed P2. *Oncogene*, **14,** 899–913.

White, B.Y.A. (1958) Effect of pH on Fluorescence of Tyrosine , Tryptophan and Related Compounds. *Biochemical Journal*, **7,** 217–220.

Young, C.L., Britton, Z.T. & Robinson, A.S. (2012) Recombinant protein expression and purification : A comprehensive review of affinity tags and microbial applications. *Biotechnology Journal*, **7,** 620–634.

Zhang, X., Gan, L., Pan, H., Guo, S., He, X., Olson, S.T., Mesecar, A., Adam, S. & Untermant, T.G. (2002) Phosphorylation of serine 256 suppresses transactivation by FKHR (FOXO1) by multiple mechanisms: Direct and indirect effects on nuclear/cytoplasmic shuttling and DNA binding. *Journal of Biological Chemistry*, **277,** 45276–45284.