

https://dspace.org/

The effect of dopamine on the interaction between the FOXP2 FHD and DNA

Symon, Joni

2020
https://hdl.handle.net/10539/31095

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

School of Molecular and Cell Biology

University of the Witwatersrand Johannesburg

The Effect of Dopamine on the Interaction Between the FOXP2 FHD and DNA

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

Supervisor: Dr Sylvia Fanucchi

20 April 2020

Joni Symon
752364

DECLARATION

I, **Joni Symon** am a student registered for the degree of MSc dissertation in the academic year 2017.

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:  Date: 16 September 2020

ABSTRACT

Transcription factors play a crucial role in the regulation of both basal and differential gene expression, interacting with specific target sequences, thereby activating or repressing transcription. The FOX family of transcription factors, characterised by the presence of a modified helix-turn-helix DNA binding domain, the forkhead domain (FHD), are key in the regulation of pathways associated with embryogenesis and development. One such member, FOXP2 has been directly associated with language acquisition. Mutations, specifically within the FHD, have been directly linked to verbal dyspraxia and cancer, as well as hypothesized links between abnormal FOXP2 function and autism and schizophrenia. It is possible that the function of FOXP2 could be regulated by the presence of small molecules, specifically dopamine. Various studies have suggested links between dopaminergic pathways and FOXP2 activity. For example, it has been proposed that FOXP2 down regulates cellular dopamine levels, as well as affects the morphologies of target neurons of dopaminergic pathways. The potential for the neurotransmitter to directly interact with FOXP2, specifically the FHD, thereby modulating the activity of the protein, specifically with regard to DNA binding, had not yet been investigated. FOXP2 FHD was overexpressed, purified and structurally and functionally characterised, before investigating whether the presence of dopamine at physiological concentrations (35 μM), as well as in excess (100 μM), affected the secondary or tertiary structure. The effect of the presence of the neurotransmitter on DNA binding was then investigated, and the K_D was determined. Molecular docking was used to investigate the potential for an interaction between the FOXP2 FHD and dopamine, as well as the FHD and various reactive products of dopamine autoxidation (namely 5,6-dihydroxyindole, 5,6-indolequinone and dopaminechrome) *in silico*. The investigation revealed a potential non-canonical binding pocket located towards the C-terminus of the FOXP2 FHD, however, it appeared that a product of dopamine autoxidation, such as 5,6-dihydroxyindole, was more likely to bind the pocket than dopamine itself. Spectroscopic investigation revealed that the presence of dopamine did not alter the fold or conformation of the protein, while fluorescence anisotropy suggested that the neurotransmitter did not significantly affect the DNA binding affinity, however a slight reduction in binding affinity was observed with increasing concentrations of dopamine (a K_D of 0,784μM was observed in the absence of dopamine, while the K_D obtained in the presence of 35 μM dopamine was found to be 1,074 μM, and in the presence of 100 μM dopamine, the K_D obtained was 1,40

μM). Supported by *in silico* observations, it was hypothesized that it is unlikely that dopamine itself directly regulates the activity of FOXP2 via interactions with the FHD, however it is possible that a product of spontaneous dopamine autoxidation may play a role in regulating FOXP2 activity physiologically.

ACKNOWLEDGEMENTS

To my supervisor, Dr Sylvia Fanucchi, thank you, for not only your endless guidance, support, expertise and brilliance throughout the course of this research project, but also for continuously reigniting my excitement and passion for biochemistry, and inspiring me throughout this process. Even after the monthly dopamine disasters.

I would also like to express my appreciation to the mentors of the PSFRU. Thank you for sharing your knowledge and expertise, and for the constant encouragement and challenges.

To my fellow members of the PSFRU, I have learnt so much from working alongside all of you. A special thank you to the FOXy group, as well as Roland Worth and Martin Bernert for your enthusiasm in sharing your knowledge, and all the time you generously dedicated to helping me, to Ashleigh Blane for continuing to be my mentor, even when it was no longer your job, and to Ilan Kirkel, Jessica Olfsen and Naadira Pahad, for all the support, sciencey ideas, laughs and coffees.

I am deeply grateful to the University of Witwatersrand Post Graduate Merit Award for the generous funding.

And lastly, thank you to my parents, brother, grandmother and friends, for dealing with my dopamine dependent moods, and always encouraging me.

TABLE OF CONTENTS

DECLARATION	**2**
ABSTRACT	**3**
ACKNOWLEDGEMENTS	**5**
LIST OF ABBREVIATIONS	**8**
LIST OF FIGURES	**11**
LIST OF TABLES	**13**
1. INTRODUCTION	**14**
1.1. Regulation of Transcription and The Role of Transcription Factors	**14**
1.1.1. Regulation of Transcription	**14**
1.1.2. Classification of TFs	**16**
1.1.3. The Zinc Finger	**17**
1.1.4. The Leucine Zipper	**18**
1.1.5. The Helix-Loop-Helix Motif	**19**
1.1.6. The Helix-Turn-Helix Motif	**20**
1.2. FOX Transcription Factors and FOXP2	**23**
1.2.1. FOX Family of Transcription Factors	**23**
1.2.2. FOX Transcription Factors and Neurodevelopment	**25**
1.2.3. FOXP2 and Prenatal Neurodevelopment (Specifically Language Acquisition	**26**
1.2.4. FOXP2 Function	**27**
1.2.5. FOXP2 Structure	**27**
1.2.6 Interaction Between the FOXP2 FHD and DNA	**29**
1.3. Dopamine	**35**
1.3.1. Signaling Molecules	**35**
1.3.2. Neurotransmitters	**35**
1.3.3. Dopamine Structure and Function	**37**
1.3.4. Autoxidation and Neuromelanin	**39**
1.4. The Potential for Dopamine to Affect the Function of FOXP2	**43**
1.4.1. The Potential for a Direct Interaction between the FOXP2 FHD and Dopamine	**44**
1.4.2. The Potential for the Presence of dopamine to disrupt DNA binding	**45**
1.5. Significance of Research	**47**
1.6. Aims and Objectives	**49**
1.6.1. Aim	**49**
1.6.2. Objectives	**49**
2. METHODS AND MATERIALS	**50**
2.1. Overexpression and Purification	**50**
2.1.1. Transformation	**50**
2.1.2. Protein Expression	**53**

2.1.3. Immobilised Metal Ion Affinity Chromatography (IMAC)	**54**
2.1.4. Size Exclusion Chromatography (SEC)	**54**
2.2. Purity Assessment, Concentration Determination and Structural Characterisation	**55**
2.2.1. Tricine Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)	**55**
2.2.2. UV-vis Absorption Spectroscopy	**57**
2.2.3. FOXP2 FHD Ligand Preparation	**59**
2.2.4. Far-UV Circular Dichroism Spectropolarimetry	**60**
2.2.5. Intrinsic Tryptophan Fluorescence	**62**
2.3. Investigating the FOXP2 FHD-Dopamine-DNA interaction	**64**
2.3.1. Electrophoretic Mobility Shift Assay (EMSA)	**64**
2.3.2. Fluorescence Anisotropy	**65**
2.4. Molecular Docking	**67**
3. RESULTS	**71**
3.1. Overexpression and Purification	**71**
3.2. Spectroscopic Structural Investigation and Functional Characterisation	**75**
3.3. Investigating the FOXP2 FHD-Dopamine Interaction	**80**
3.4. Investigating the FOXP2 FHD-Dopamine-DNA interaction	**84**
***3.5. In silico* studies**	**90**
4. DISCUSSION	**101**
5. REFERENCES	**107**
6. SUPPLEMENTARY INFORMATION	**122**

LIST OF ABBREVIATIONS

A280/A260- Absorbance at 280 nm divided by absorbance at 260 nm

Ade- Adenine

Amp^R- Ampicillin resistance

Arg- Arginine

Asn- Asparagine

BSA- Bovine serum albumin

CD- Circular dichroism

CNTP1- Contactin-associated protein 1

CtBP1- C-terminal binding protein 1

Cys- Cystine

DISC1- Disrupted in Schizophrenia 1

E. coli – Escherichia coli

EDTA- Ethylenediaminetetraacetic acid

EMSA- Electrophoretic mobility shift assay

Ets- E26 transformation-specific

FHD-Fforkhead domain

FOX- Forkhead box

FOXP2 FHD- FOXP2 forkhead domain

FPLC – Fast protein liquid chromatography

Glu- Glutamine

GSH – Glutathione

GST- Glutathione s-transferase

Gua- Guanine

GUI- Graphical user interface

HEPES- 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid

His- Histidine

HLH- Helix-loop-helix

HSF-Heat shock factor

HTH- Helix-turn-helix

HTV- High-tension voltage

IMAC- Immobilized metal ion affinity chromatography

IPTG- Isoproylthiogalactopyranoside

IRF-Interferon-regulatory factor

K_D – Dissociation constant

LB- Lysogeny broth

Lys- Lysine

MCS- Multiple cloning site

Mw- Molecular weight

NADH- Nicotinamide adenine dinucleotide + hydrogen

NKX2-1- Thyroid transcription factor 1

NTA- Nitrilotriacetic acid

NURR1-Nuclear receptor related 1

OD_{600} – Optical density at 600 nm

PAGE- Polyacrylamide gel electrophoresis

PD- Parkinson's disease

Phe- Phenylalanine

RFX-Regulatory factor binding to the X-box

RNA pol II – RNA polymerase II

ROX- 6-carboxyl-x-rhodamine

S_0 – Ground state

SDS- Sodium dodecyl sulfate

SEC- Size exclusion chromatography

SOC media- Super optimal broth with catabolite repression

TBE- Tris, Boric acid, EDTA

Tbx24- T-Box24

TF- Transcription factor

Thr- Threonine

Thy- Thymidine

Trp- Tryptophan

TSS- Transcription start site

Tyr- Tyrosine

UFF- Universal force field

UV- Ultraviolet

V_0- Void volume

V_e – Eluted volume

Vis- Visible

wt- Wild type

Θ – Elipticity

θ_{MRE} – Mean residue elipticity

LIST OF FIGURES

Figure 1.1: The composition of a prototypical TF	**15**
Figure 1.2: A single C2H2 type zinc finger (PDB: 1ZNF)	**18**
Figure 1.3: Basic leucine zipper transcription factor (PAP1) bound to DNA (PDB:1GD2)	**19**
Figure 1.4: Helix-loop-helix heterodimer (Transcription Factor E2-alpha (E47)/Neurogenic Differentiation factor 1(NeuroD1)) bound to DNA (PDB: 2QL2)	**20**
Figure 1.5: Schematic Illustrating the Helix-Turn-Helix Motif	**22**
Figure 1.6: The interaction between the winged helix-turn-helix motif and DNA, as illustrated by FOXA3 (PDB: 1VTN)	**22**
Figure 1.7: The characteristic FHD associated with FOX transcription factors (FOXA3) (PDB: 1VTN)	**24**
Figure 1.8: Schematic of the motifs and domains associated with FOXP2	**28**
Figure 1.9: Image illustrating the interaction between the FOXP2 FHD and the Wang Sequence (PDB: 2A07)	**30**
Figure 1.10: Image comparing the structural architecture of monomeric and dimeric FOXP2 FHD (PDB: 2A07)	**33**
Figure 1.11: Image illustrating the production of dopamine from tyrosine	**37**
Figure 1.12: The chemical structure of fully protonated dopamine	**38**
Figure 1.13: Image illustrating the autoxidation of dopamine to form neuromelanin	**40**
Figure 1.14: FDI-6 (inhibitor of FOXM1 FHD) **(NCGC00099374)**	**45**
Figure 2.1: Schematic illustrating the plasmid construct of pET11a	**51**
Figure 2.2: Schematic illustrating the regions upstream the FOXP2 FHD codon-optimised sequence cloned into the MCS of the pET11a	**52**
Figure 2.3: A simplified Jablonski diagram representing the singlet excited state	**72**
Figure 3.1: Purification of the FOXP2 FHD	**76**
Figure 3.2: Structural and functional characterisation of the FOXP2 FHD	**79**
Figure 3.3: Molecular surface of FOXP2 FHD	**82**

highlighting location and solvent exposure of tryptophan residues (PDB: 2A07)	
Figure 3.4: Structural characterisation of the FOXP2 FHD in the absence and presence of 35 µM and 100 µM dopamine	**86**
Figure 3.5: Evaluation of the effect of the presence of dopamine on the DNA binding function of FOXP2 FHD	**92**
Figure 3.6: Binding predictions by PyRx for monomeric FOXP2 FHD and dopamine, as well as its autoxidation products	**94**
Figure 3.7: Binding predictions by PyRx for dimeric FOXP2 FHD and dopamine, as well as its autoxidation products	**97**
Figure 3.8: Targeted binding predictions by Schrödinger suite for monomeric FOXP2 FHD and dopamine, as well as its autoxidation products	**98**
Figure 3.9: Targeted binding for predictions by Schrödinger suite for dimeric FOXP2 FHD and dopamine, as well as its autoxidation products	**103**
Figure 4.1: 5,6-dihydroxyquinone bound to NURR1 as 5,6-indolequinone (PDB: 6DDA)	**122**
Figure 6.1: Graph Illustrating the relationship between wavelength and high-tension voltage corresponding to the obtained CD spectrum for characterisation of the FOXP2 FHD (figure 3.2 b)	**123**
Figure 6.2: Graph illustrating the relationship between wavelength and high-tension voltage corresponding to the obtained CD spectrum for characterisation of the FOXP2 FHD in the absence and presence of dopamine (figure 3.4 (A))	**124**
Figure 6.3.1: Ligand diagrams generated by Schrödinger illustrating the interaction between monomeric FOXP2 FHD and ligands	**124**
Figure 6.3.2: Ligand diagrams generated by Schrödinger illustrating the interaction between dimeric FOXP2 FHD and ligands	**124**

LIST OF TABLES

Table 1: Results of student t-test	**88**
Table 2: Free energies (ΔG) associated with best poses for the blind docking of dopamine, 5,6-dihydroxyindole, 5,6-indolequinone and dopaminechrome with monomeric and dimeric FOXP2 FHD generated by PyRx	**91**
Table 3: Glidescores (G-scores) and E-model glide scores (E-model) associated with best poses for the blind docking of dopamine, 5,6-dihydroxyindole, 5,6-indolequinone and dopaminechrome with monomeric and dimeric FOXP2 FHD generated by Schrodinger	**95**

1. INTRODUCTION

1.1. Regulation of Transcription and The Role of Transcription Factors

1.1.1. Regulation of Transcription

Transcription acts as the pivotal first step in gene expression. Broadly speaking, this process results in the production of an initial mRNA transcript, the pre-mRNA, from DNA (Latchman, 1993). Following the production of this transcript, various posttranscriptional processes including mRNA splicing, translation and posttranslational modifications facilitate the formation of functional proteins from structural genes (Latchman, 1993). Specificity and tight regulation of gene transcription is crucial, as it plays a key role in orchestrating cellular function, allowing for cells to respond to both innate and external stimuli (Yilmaz and Grotewold, 2010), and allows for a defined and unique protein expression profile to be established within different cell types, thereby facilitating specialisation of different tissues (Latchman, 1993).

In the case of both basal and differential gene expression, regulation on a transcriptional level is largely facilitated by interactions between a group of regulatory proteins, known as transcription factors (TFs) and specific regulatory DNA regions (Lelli et al., 2013, Ong and Cores, 2011, Spitz and Furlong, 2012). These DNA regions, known as transcription factor binding sites contain specific motifs, short DNA sequences that TFs preferentially bind. Motifs recognised by TFs are usually small (6-12 base pairs) and typically the regulatory regions of a gene may contain multiple potential binding sites for various TFs (Wunderlich and Mirny, 2009). Many TFs also form dimers, trimers, higher order structures or require interactions with other TFs in order to be activated. Cooperativity and synergy between TFs is therefore essential for effective regulation of gene expression, and global gene regulatory networks are highly conserved (Wunderlich and Mirny, 2009, Reiter *et al.*, 2017).

In order for the transcription of a specific gene to proceed, it is essential that the pretranscription initiation complex forms on the corresponding transcription start site (TSS) of the promoter region of that gene (Luse, 2013). This large multicomponent complex is composed of RNA polymerase II (RNA pol II), as well as general and specific transcription factors and cofactors (Luse, 2013). Ultimately, TFs function by promoting or inhibiting the formation of the pretranscription initiation complex, and thus influencing the initiation of

transcription (Latchman, 1993). Interactions between DNA and a TF alone are often not sufficient to influence the formation of the pretranscription initiation complex. TFs are, therefore, often modular proteins, composed of multiple domains with varying functions, thereby allowing these proteins to interact with DNA, as well as with other proteins, such as RNA pol II (Lambert *et al.,* 2018, Ptashne, 1998, Mitchell and Tjian, 1989). This is depicted below (figure 1.1). Since TFs act predominantly as activators of transcription, following a DNA binding event, protein-protein interactions may recruit other factors, or facilitate conformational changes, to either the DNA or other proteins, thereby promoting the formation of the pretranscription initiation complex. The binding event itself between the DNA and the activator may also alter the conformation of a TF that is already bound (Roberts, 2000).

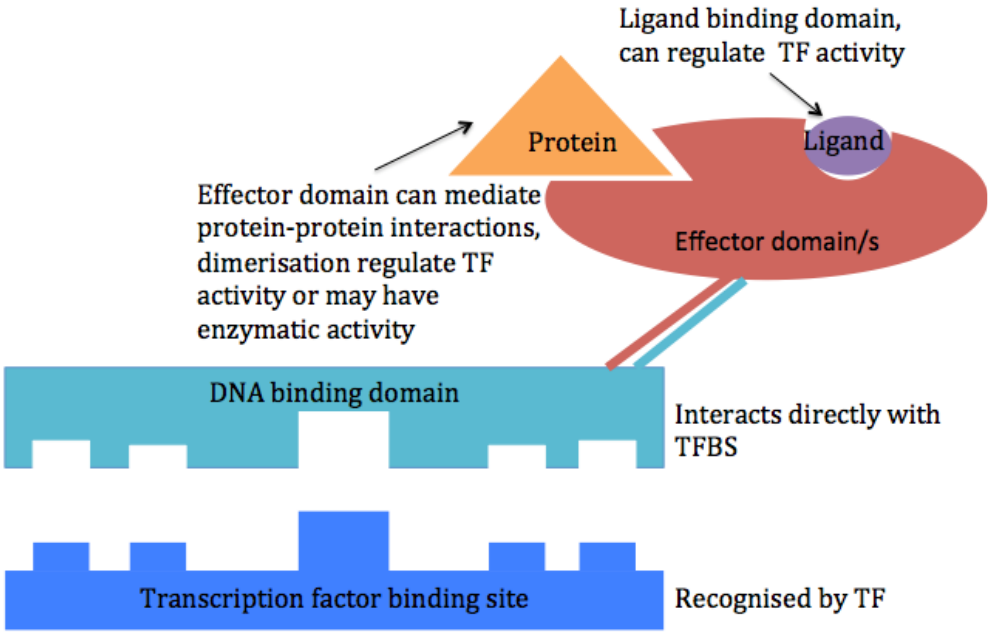


Figure 1.1: The composition of a prototypical TF TFs are typically composed of more than one functional domain. The DNA binding domain (light blue) recognises, and interacts with the target transcription factor binding site (dark blue), while the effector domain or domains (red) may interact with other proteins or a specific ligand, thereby allowing the TF to regulate transcription or facilitating regulation of the activity of the transcription factor itself.

Although less common, TFs can also act as repressors of transcription, and in some cases, TFs can act as both activators and repressors of transcriptional activity. This is spatially and temporally defined (Sharrocks, 2001; Lambert *et al.*, 2018). Differential expression, regulated through molecular mechanisms including post-translational modification, DNA

accessibility and DNA modification, facilitates the production of a unique expression profile within different tissue types, during different stages of development, or at different points within cell cycle (Alon, 2007; Meng and Alon, 2003; Simon *et al.*, 2001). The co-localisation of distinctive proteins within these different cell types, or at different stages may result in the conversion of an activator into a repressor. For example, Tbx24 (T-box24) is a member of the T-box family of transcription factors. T-box proteins play a crucial role in somite segmentation in vertebrates. Tbx24 acts predominantly as an activator of transcriptional activity, however, when associated with Ripply1, the TF has been shown to repress transcriptional activity through the recruitment of the global corepressor Groucho/Transducin-Like Enhancer of Split. The conversion of the TF from an activator to repressor is therefore dependent on coordinated colocalisation.

Generally, repressors interrupt the formation or activation of the pretranscription initiation complex. These TFs may act by blocking an essential interaction between two or more activating factors, by interacting with activator effector domains, or may competitively block the binding of an activating factor to the pretranscription initiation complex (Levine and Manly, 1989, Goodburn, 1990). Repressors have also been reported to interact with activators in solution, thereby preventing the interaction between the DNA binding domain of the activator and the corresponding transcription factor binding site (Goodburn *et al.*, 1986). A phenomenon known as quenching has also been reported, whereby the repressor binds DNA at a site adjacent to a bound activator and masks the activity of its activation domain through structural changes to the DNA (Keleher *et al.*, 1998).

The significance of TFs in the regulation of highly coordinated, essential cellular processes is highlighted in that a third of developmental disorders arise from, or are related to, dysfunctional TF activity (Boyadjiev and Jabs, 2000), and there is evidence to suggest that TFs are overrepresented among oncogenes (Furney, 2006).

1.1.2. Classification of TFs

There are several different methods by which transcription factors have been classed. Traditionally, transcription factors have been organised into different groups or families based on the sequence and structure of the associated DNA binding domain (Weirauch and Hughes, 2011). In 2013, Wingender *et al.* described a system of classification, known as TFClass for the 1558 identified human TFs, wherein the proteins were divided

into superclasses based on general topology, classes based on the structure of the DNA binding domain, families by functional similarities, and subfamilies based on similarities in transcription factor binding sites.

There are 4 predominant motifs associated with DNA-binding domains amongst the annotated transcription factors, namely, the zinc finger, the leucine zipper type, the helix-loop-helix (HLH) and the helix-turn-helix (HTH).

1.1.3. The Zinc Finger

First identified in transcription factor TFIIIA found in *Xenopus laevis*, the zinc finger motif is characterised by the presence of 1 or more zinc ions, coordinated by 4 surrounding cysteine and/or histidine residues. As much as 2% of the human genome has been reported to encode zinc finger domains (Lander *et al.*, 2001). Out of the 40 zinc finger types identified (Uniprot, 2019), the C2H2 type is the most common, and was the first to be characterised (figure 1.2). The zinc finger domain recognises and interacts with the DNA major groove, predominantly through hydrogen bonding (Klug, 2010). Traditionally, proteins contain more than one zinc finger domain, linked together in a linear, tandem manner. Since zinc fingers are relatively small, a single zinc finger will generally recognise and interact with 3-4 bps of DNA (Klug, 2010). Proteins containing multiple zinc finger domains can therefore recognise and interact with varying lengths of DNA (Klug, 2010).

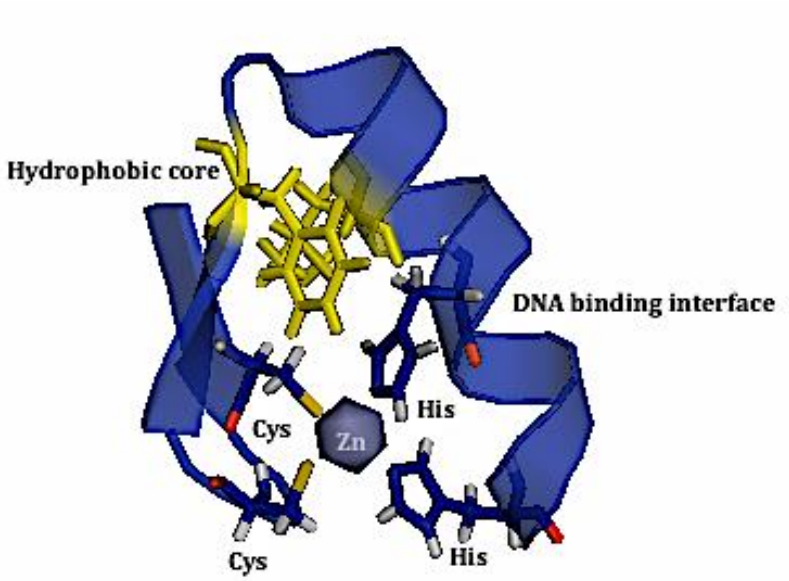


Figure 1.2: A single C2H2 type zinc finger (PDB: 1ZNF) This DNA binding domain is composed of roughly 25-30 amino acids, and is characterised by a highly conserved motif containing an inner hydrophobic core (yellow) (generally composed of tyrosine, phenylalanine and leucine), as well as 2 histidine residues and 2 cysteine residues that coordinate a single zinc ion in a tetrahedral array (Miller *et al.*, 1985). Image was generated using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC.

1.1.4. The Leucine Zipper

The leucine zipper structural motif is composed of a single extended alpha helix, roughly 30 amino acids in length, with leucine residues repeating at roughly every 7th position at the C-terminus. With regard to eukaryotic transcription factors, dimerisation is essential for function and the leucine zipper is flanked at the N-terminus by a basic region, roughly 25 amino acids in length (O'Shea *et al.*, 1991). This structure is referred to as the basic leucine zipper. The C-terminus, containing the leucine zipper, facilitates dimerisation, which is essential for the formation of the DNA binding domain at the basic N-termina (figure 1.3) Alber, 1992). Once dimerisation has occurred, the basic region interacts with short (around 8 base pairs) palindromic sequences.

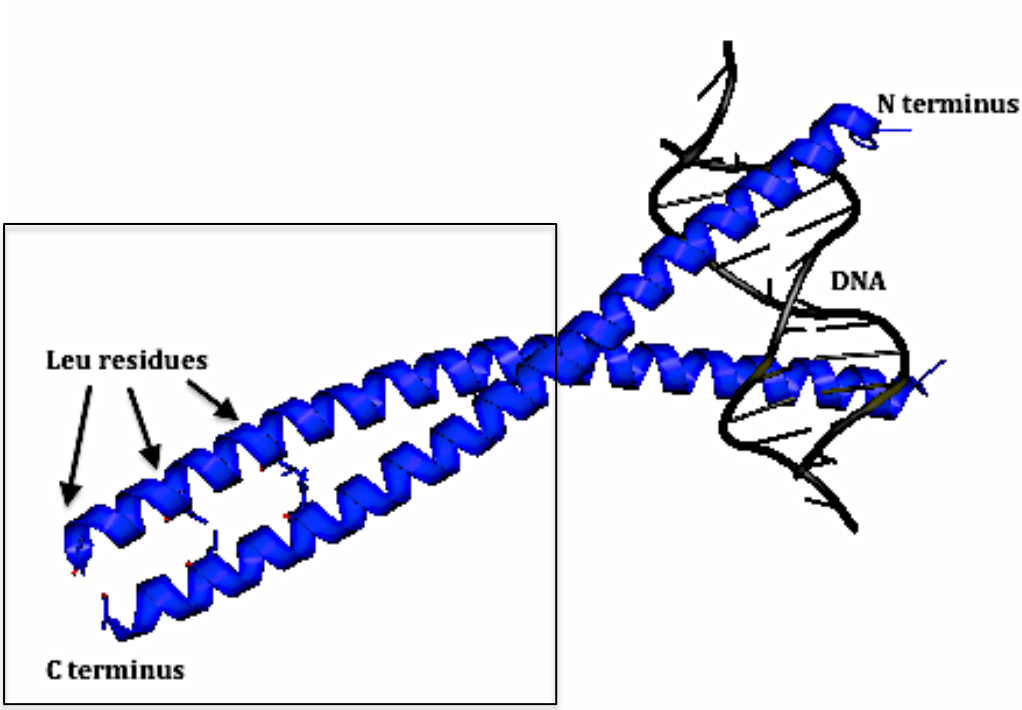


Figure 1.3: Basic leucine zipper transcription factor (PAP1) bound to DNA (PDB:1GD2) The leucine zipper region is represented within the box, while the basic region is illustrated interacting with opposite sides of the major groove of DNA in a scissor-like fashion. Two leucine zipper domains wrap around each other to form a parallel dimeric coiled coil with a slight left hand twist (Alber, 1992). The side chains of leucine residues within adjacent polypeptides interact via non-polar interactions to form a hydrophobic core that stabilises the formation of the dimer. Image was generated using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC.

1.1.5. The Helix-Loop-Helix Motif

This DNA binding motif has been identified in 125 human transcription factors (Ledent *et al.*, 2002). The helix-loop-helix motif, roughly 60-100 amino acids in length (Jindrich and Degan, 2016), is composed of two highly conserved amphipathic alpha helices, separated by a random loop region of variable length and sequence (Massari and Murre, 2000). Like the leucine zipper, the function of the helix-loop-helix motif is largely dependent on homo and heterodimerisation. In fact, higher selectivity is displayed in dimerisation than in sequence specificity (Massari and Murre, 2000). Dimerisation occurs through the faces of the alpha helices, and results in the formation of a left-handed, 4-helix bundle (figure 1.4).

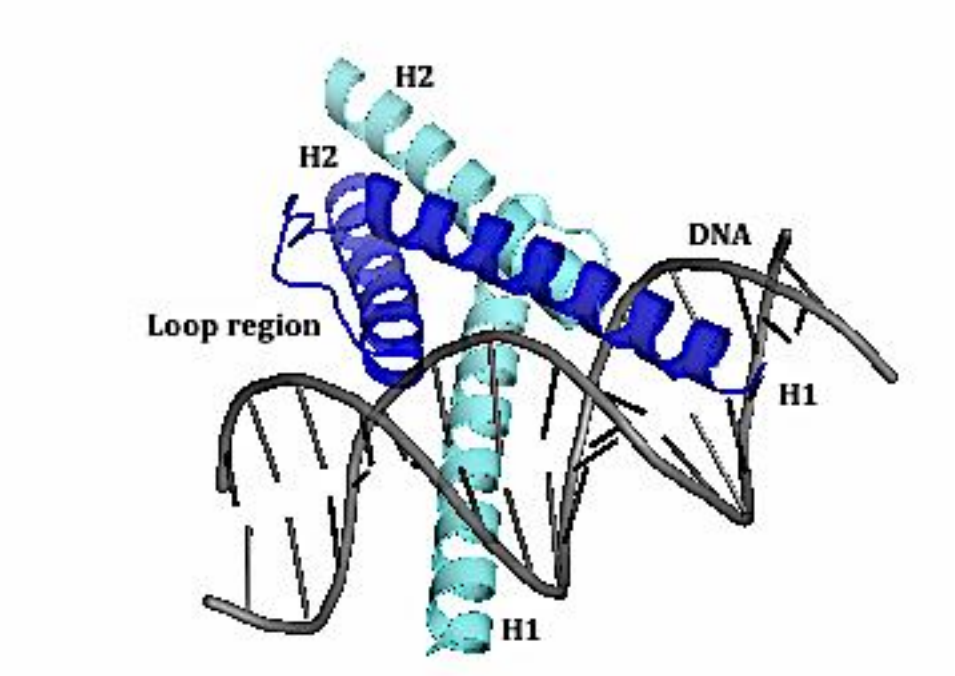


Figure 1.4: Helix-loop-helix heterodimer (Transcription Factor E2-alpha (E47)/Neurogenic Differentiation factor 1(NeuroD1)) bound to DNA (PDB: 2QL2) The helix-loop-helix motifs associated with E47 and NeuroD1 are coloured cyan and dark blue respectively. The interaction between the HLH motif and DNA is facilitated by the insertion of basic amino acids within H1 of each monomer into the major grove of DNA, while the loop region allows for increased stabilisation by interacting with the DNA backbone (Massari and Murre, 2000). Image was generated using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC.

1.1.6. The Helix-Turn-Helix Motif

The helix-turn-helix motif was first identified in in 1981 in transcription regulatory proteins from both bacteriophage λ and *E. coli* (Anderson et al., 1981; Mckay and Steitz, 1981; Pabo and Lewis, 1982). The helix-turn-helix motif, in its simplest form, is characteristically composed of 3 alpha helices, linked by random coils in a partially open configuration (figure 1.5 (A)). The three alpha helices (H1, H2, H3) allow for the formation of a highly conserved stable hydrophobic core, and the principle protein-DNA interaction is facilitated by H3 (the recognition helix), which binds the major groove of DNA (Aravind *et al.*, 2005).

The structural architecture of this motif has favoured the addition of many different extensions and modifications to the simple tri-helical bundle illustrated in figure 1.5 (A), resulting in a large amount of structural and functional diversity (Aravind *et al.*, 2005). For example, the shallow cleft between

H1 and H3 has favoured the addition of various elements, which pack in via hydrophobic interactions (Aravind *et al.*, 2005). The loop between H1 and H2 has also been observed to show a large degree of variability, and multiple diverse N- and C-terminus extensions have been reported. Extensions to the core bundle usually result in the formation of a more closed conformation (Aravind *et al.*, 2005).

Some examples of modified versions of the helix-turn-helix include the tetrahelical bundle type (characterised by the addition of an N-terminus helix to the trihelical bundle), the ribbon helix-turn-helix family (obligate dimers, defined by the presence of a single strand at the N-terminus through which dimerisation occurs, in addition to the trihelical bundle) and the winged helix-turn-helix (Aravind *et al.*, 2005).

The winged helix-turn-helix is of particular interest in this study, and is characterised by the presence of a 'wing' extension to the tri-helical bundle. The wing is comprised of a 2-stranded β-sheet (figure 1.5 (B)) (Brennan, 1993), and provides an additional interface for DNA interactions, specifically with the minor groove, thereby allowing for increased stability and specificity regarding DNA binding (Brennan, 1993). This modified motif was named based on the folding of the loop structures at the C-terminus flanking the beta sheet, appearing almost wing-like (figure 1.6).

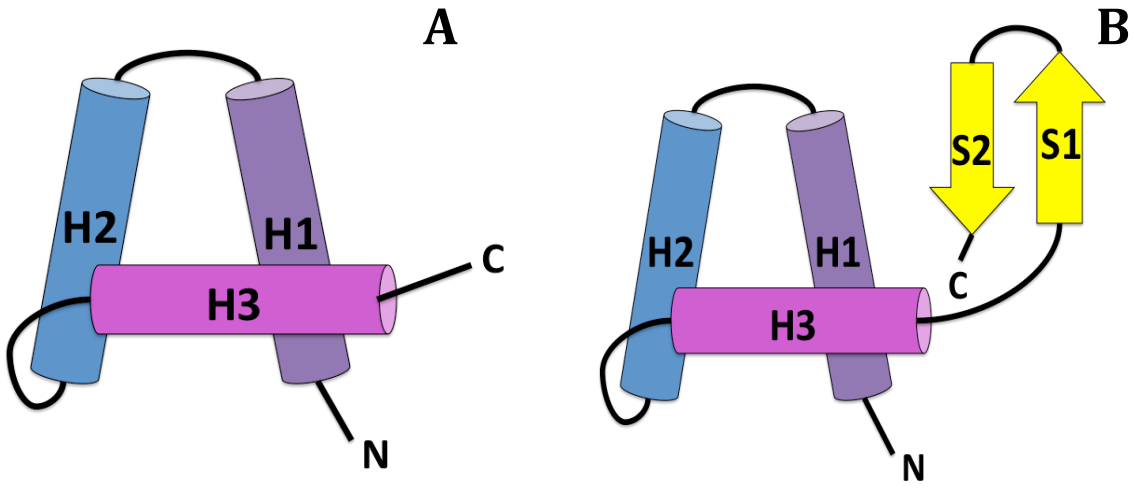


Figure 1.5: Schematic Illustrating the Helix-Turn-Helix Motif
(A) Schematic illustrating the simplest helix-turn-helix domain. The helix-turn-helix motif is composed of a trihelical bundle, composed of 3 alpha helices. When H3 (pink) is positioned in front, the domain forms a triangular shape, with a characteristic sharp turn between H3 and H2.
(B) Schematic illustrating the winged helix-turn-helix domain. The winged helix-turn-helix domain includes the core helix-turn-helix elements (illustrated in figure 1.5 (A), as well as a 2-stranded beta sheet extension at the C-terminus (yellow).

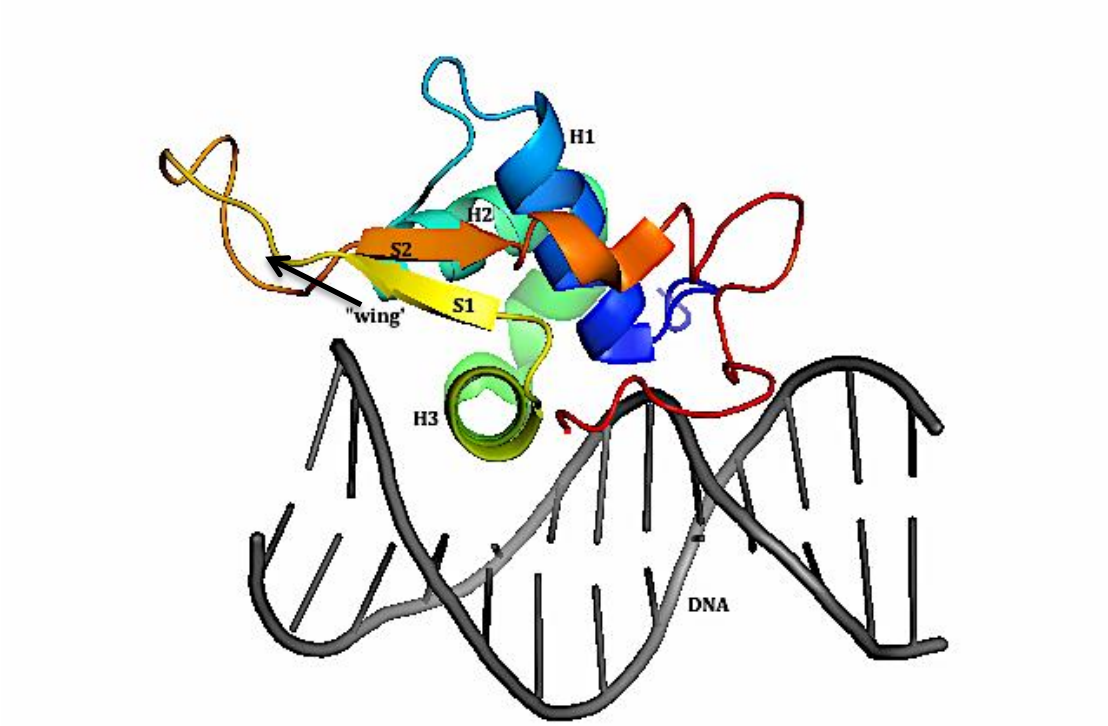


Figure 1.6: The interaction between the winged helix-turn-helix motif and DNA, as illustrated by FOXA3 (PDB: 1VTN) H3 (green) interacts directly with the major groove of DNA, while the 'wing', (yellow and orange) aids in stabilisation, and specificity (Clark *et al.*, 1993). Image was generated using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC

TFs that utilise the winged helix-turn-helix for DNA binding include the forkhead box (FOX), E26 transformation-specific (Ets)(Donaldson *et al.*, 1996), interferon-regulatory factor (IRF)(Furui *et al.*, 1998), regulatory factor binding to the X-box (RFX)(Gajiwala *et al.*, 2000) and heat shock factor (HSF)(Littlefield *et al.*, 1999) super families. The winged helix-turn-helix motif associated with each of these super families differs in topology and specific arrangement of helices, loops and beta strands (Aravind *et al.*, 2005).

This study will focus specifically on the winged helix-turn-helix motif associated with FOX transcription factors, known as the forkhead domain (FHD) (Hughes, 2011). This domain is defined by the presence of 3 N-terminus alpha helices, a 3-strand antiparallel beta hairpin element and 2 loop structures located towards the C-terminus (Weigel *et al.*, 1998). An additional helix, H4, is sometimes located between H2 and H3.

1.2. FOX Transcription Factors and FOXP2

1.2.1. FOX Family of Transcription Factors

The FOX family is a group of transcription factors that display sequence homology to a canonical DNA binding domain, the FHD (Lehmann *et al.*, 2003). First identified in *Drosophila melanogaster* (Weigel *et al.*, 1989), this domain is approximately 80 to 100 amino acids in length. As previously discussed, the FHD is a modified version of the characteristic DNA binding HTH motif, namely the winged helix motif, and takes on a butterfly-like appearance due to the presence of the 'wings' (figure 1.7).

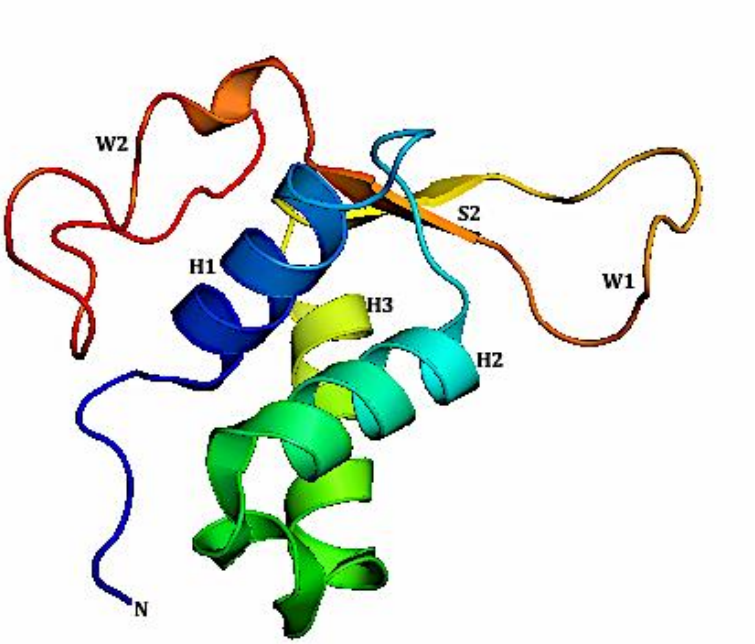


Figure 1.7: The characteristic FHD associated with FOX transcription factors (FOXA3)(PDB: 1VTN). The beta sheet (S1, S2 depicted in yellow and orange) connects the bundle of alpha helices to two loop structures, the 'wings' (depicted in orange). As clearly illustrated by the FOXA3 forkhead domain, this gives rise to the butterfly-like appearance (Clark *et al.*, 1993). H3 (yellow), the DNA recognition helix, interacts directly with the major groove. Image was generated using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC.

Overall, 19 FOX clades have been identified and are classified as FOX members A to S, based on sequence similarities within the FHD (Larroux *et al.*, 2008, Kaestner *et al.*, 2000). Due to the high degree of sequence divergence outside the FHD, phylogenic inferences are made based solely on this domain (Benayoun *et al.*, 2011). FOX proteins differ significantly in function, binding partners and spatial and temporal distribution and expression (Hannenhalli and Kaester, 2009). 50 FOX genes, as well as 2 pseudogenes have been identified within the human genome, 44 of which have a closely related murine ortholog (Jackson *et al.*, 2010).

The FOX protein family are evolutionarily conserved, and play vital roles in many processes associated with embryogenesis and development (Carlsson and Mahlapuu, 2002; Kaufmann and Knochel, 1996). They also function throughout adult life. Some processes regulated by these proteins include gastrulation and metabolism (Shih *et al.*, 1999; Kaestner *et al.*, 1999; Shen *et al.*, 2001), as well as cell cycle control, tissue specialisation and differentiation (Carlsson and Mahlapuu, 2002). Many of the FOX genes are expressed in both mature and developing brain tissue. Although little is known about the potential functions of these proteins in neurodevelopment, spatial and temporal

expression patterns, as well as gain and loss of function models in mice have illustrated significant associations between specific FOX transcription factors and neural patterning (Kaestner *et al.*, 1999).

1.2.2. FOX Transcription Factors and Neurodevelopment

It has been reported that FOX transcription factors are implicated in the development of multiple diverse neurological pathways and brain regions during embryogenesis (Kaestner *et al.*, 1999). These proteins function both independently and synergistically to regulate gene expression and signal transduction related to nodal signaling, neural patterning, cell fate determination and differentiation within the developing brain.

FOXJ1, for example, has been reported to be associated with the development of left-right asymmetry within the brain (Chen *et al.*, 1998), while FOXA1 is known to induce neural differentiation, and formation of the neural tube (Ruiz *et al.*, 1993; Sasaki and Hogan, 1993). The function of FOXA1 is regulated by FOXH1 (as well as retinoic acid) (Hoodless *et al.*, 2001; Jacob *et al.*, 1997). Expression of FOXB1 has been identified within the midbrain, spinal chord interneurons and hypothalamus, and murine models suggest that this protein is essential in the development of spatial working memory (Ang *et al.*, 1993; Johns *et al.*, 2005, Kaestner and Schutz, 1996, Labosky *et al.*, 1997). Amongst other functions, FOXD1 is associated with the development of the visual system, and is essential in optic chiasm formation (Marcus *et al.*, 1999). FOXG1 is also associated with the development of this structure, as well as the telencephalon (Pratt *et al.*, 2004). Null FOXG1 mutants are reported to die at birth, due to incomplete development of the two cerebral hemispheres (Pratt *et al.*, 2004). Members of the FOXO family have been reported to respond to insulin and neurotrophin signaling within the developing brain, regulating neural survival and cell death, as well as functioning in neuroprotection (Zheng *et al.*, 2002, Linseman *et al.*, 2002).

The FOXP subfamily is of particular interest in this investigation. This subfamily, comprised of 4 members, namely, FOXP1, FOXP2, FOXP3 and FOXP4, is largely implicated in embryonic development (Chiu *et al.*, 2014), immune disorders (Jackson *et al.*, 2010), carcinogenesis (Uva *et al.*, 2015) and angiogenesis (Myatt and Lam, 2007, Carvalho *et al.*, 2016). FOXP1 is associated with the development of B cells (Patzelt *et al.*, 2018), while FOXP4 mediates T cell development (Teufel *et al.*, 2003). FOXP3 has been linked to critical involvement in the

function and differentiation of T cells, especially relating to cancer progression (Wu *et al.*, 2006, Lu *et al.*, 2017).

The focus of this investigation is FOXP2. Amongst other functions, this protein is associated with speech acquisition, and is abundantly expressed within distinct regions of the developing brain (Ferland *et al.*, 2003, Lai *et al.*, 2003, Takahashi *et al.*, 2003, Takahashi *et al.*, 2008, Teramitsu *et al.*, 2004).

1.2.3. FOXP2 and Prenatal Neurodevelopment (Specifically Language Acquisition)

The significance of FOXP2 in neurodevelopment, specifically concerning language acquisition, was first identified when tracing the genetic lineage of the KE family. Fifteen members of this family displayed difficulty in speech articulation. Genetic studies revealed a significant heterozygous arginine-to-histidine missense mutation within the FOXP2 forkhead domain of these individuals (Fisher *et al.*, 1998; Lai *et al.*, 2001). The R553H mutation lead to disruptions in nuclear localisation signals, as well as inhibition of DNA-protein interactions. This lead to haploinsufficiency in affected individuals (Lai *et a*l., 2001). To date, no living human has been identified with a homozygous FOXP2 R533H mutation; however, murine studies suggest that homozygous mutations associated with the FHD results in immature cerebral development and the formation of FOXP2 aggregates, leading to motor impairment, and death within 3 to 4 weeks after birth (Shu *et al.,* 2005).

A FOXP2 point mutation (R328X) was later identified within an individual outside the KE lineage, referred to as CS, who was similarly affected by orofacial dyspraxia (MacDemort *et al.*, 2005). These findings allowed FOXP2 to be identified as the first genetic factor affecting specific speech and language related disorders.

Language acquisition is a highly complex phenomenon and is dependent on innate neural mechanisms (that are established genetically), as well as external stimuli. Linguists define language acquisition to encompass speaking, singing and language comprehension (Friederici, 2011). The process of speech acquisition begins prenatally, and continues throughout childhood. A study performed in 2006 highlighted that mouth movements associated with early oral, lingual and motor abilities began at around the 15th week of development (Miller *et al.,* 2006). Another study highlighted that by the 30th week of gestation, it is possible that foetuses respond to particular auditory stimulus with specific lip movements, suggesting a

high level of brain development in relation to language acquisition (Reissland *et al.,* 2016). FOXP2, as well as other members of the FOXP subfamily, is expressed differentially in a variety of central nervous system structures, including the striatum, thalamus, cerebellum, spinal cord and cerebral cortex (Takahashi *et al*., 2009). There is a spike in the expression of FOXP2 specifically at week 15 and 22 of development (Teramitsu *et al*., 2004). The expression of FOXP2 within the developing brain is generally widespread, however many areas associated with high expression levels during development appear to be linked to language acquisition. The expression of FOXP2 is developmentally regulated, and it has been reported that the relative levels of this protein in adult brain tissue are low (Lai *et al*., 2001).

1.2.4. FOXP2 Function

The functions of FOXP2 extend beyond language acquisition. This protein is associated with the specialisation and differentiation of various tissues during development, as well as morphogenesis. Mutations in FOXP2 have been linked to a variety of diseases including autism (Gong *et al*., 2004), schizophrenia (Tolosa *et al.*, 2010; Spaniel *et al*., 2011), infant onset epilepsy, dyslexia and mental retardation (Groszer *et al*, 2008).

FOXP2 acts predominantly as a repressor, but also as an activator of transcriptional activity, by recognising, and binding directly to specific DNA sequences or promoter regions, such as the SV40 promoter region (Wang *et al*., 2003). FOXP2 has been reported to interact with a variety of gene targets including contactin-asociated protein 1 (CNTP-1) (Li *et al.,* 2004; Sollis *et al*., 2016), disrupted in schizophrenia 1 (DISC1) (Walker *et al.,* 2004) and thyroid transcription factor 1 (NKX2-1) (Zhou *et al.,* 2008), and is known to associate with other proteins including other members of the FOXP subfamily (Li *et al*., 2004), as well as C-terminal binding protein 1 (CtBP1) and TBR1 (Deriziotis *et al*., 2014). The interaction between FOXP2 and these binding partners, as well as specific posttranslational modifications, influences whether transcription is repressed or activated. FOXP2 is also associated with multiple signal transduction pathways.

1.2.5. FOXP2 Structure

In addition to differences in posttranslational modification, as well as the defined spatial and temporal expression patterns of each subfamily, it is likely that sequence divergence outside the FHD contributes to facilitating functions that are specific to

each subfamily (Golson and Kaestner, 2016). The FOXP subfamily is notably divergent from other FOX families with reference to the overall domain composition, as well as the FHD itself (Stroud *et al.*, 2006).

Unlike other FOX transcription factors, members of the FOXP subfamily are shown to have both a zinc finger and leucine zipper motif in addition to the FHD (Li *et al.*, 2004). A glutamine rich region has also been identified within FOXP1/2/4 (Li *et al.*, 2004) (figure 1.8).

Each of these motifs contribute to the overall function of FOXP2. The leucine zipper is reported to enhance DNA binding affinity and specificity, as well as facilitate both homotypic and heterotypic dimerisation (Li *et al.*, 2004; Haubermann *et al.*, 2019), while the C2H2-type zinc finger is reported to be associated with specific protein-protein interactions, but only when the leucine zipper is present (Haubermann *et al.*, 2019). Polyglutamine regions are usually associated with protein-protein interactions via the formation of coiled coil domains (Chamberlain *et al.*, 1994; Flumara *et al.*, 2010), however the exact function of this region in FOXP2 has not yet been investigated. The polyglutamine region is composed of 40 uninterrupted repeating glutamine residues (Butland *et al.*, 2007), and it has been hypothesised that the presence of this region, as well as the zinc finger, facilitate the repressive function of FOXP2 (Li *et al.*, 2004).

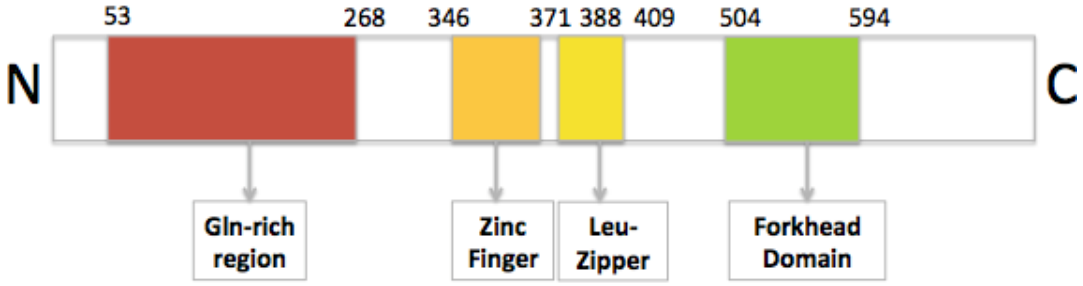


Figure 1.8: Schematic of the motifs and domains associated with FOXP2 As illustrated, FOXP2 is a modular protein, 715 amino acids in length, and composed of a glutamine rich region (red), a C2H2 type zinc finger (orange), leucine zipper (yellow) and forkhead domain separated by multiple unstructured regions.

The major function of the FHD, as previously discussed, is direct DNA interaction. The FHD is also associated with homotypic protein-protein interactions, specifically the

formation of domain swap dimers (Stroud *et al.*, 2006; Bandukwala *et al.*, 2011).

The FOXP2 FHD will be the focus of this investigation. On a structural level, there are notable differences between the wings of the FOXP2 FHD and those of other conventional FOX transcription factors (such as FOXA1). In conventional FOX proteins, a 5-7 amino acid insertion joins S3 and S2, thereby forming wing 1 (Stroud *et al.*, 2006). In FOXP2 however, this region is truncated, resulting in the formation of a type 1 turn (Stroud *et al.*, 2006). In members of the FOXP subfamily, wing 2, conventionally a loop structure, is replaced by an alpha helix, H5. When bound to DNA, this helix runs atop H1, terminating at the DNA phosphate backbone, thereby making fewer contacts with DNA compared to the conventional loop.

1.2.6 Interaction Between the FOXP2 FHD and DNA

Isolated FOXP2 FHD has been found to interact with a broad range of sequences (Wang *et al.,* 2003, Vernes *et al.*, 2007, Spiteri *et al.*, 2007, Nelson *et al.*, 2013, Webber *et al.,* 2017) in a specific manner, with varying affinities and off-rates (Webber *et al.*, 2017). The presence of sites of varying affinities throughout the genome may promote concentration of the TF within specific regions of the nucleus, and may facilitate regulation through interactions with the available location-specific binding partners (Webb *et al.*, 2017). The binding of FOXP2 to low or moderate affinity binding sites was hypothesised to facilitate scanning to locate functional binding sites, while high affinity binding sites were thought to be associated with activation of transcription (Webb *et al.*, 2017)

The exact mechanisms by which FOXP2 FHD interacts with different DNA sequences was found to vary, depending on binding affinity (Webb *et al.*, 2017). Generally speaking, TFs are largely flexible, and stability is increased with an increased number of specific DNA contacts. In the case of the FOXP2 FHD, high affinity binding interactions were found to lead to an increase in the degree of secondary structure, increased rigidity via specific DNA contacts, as well as a larger degree of tertiary structural rearrangement as compared to low affinity interactions (Webb *et al.*, 2017).

Stroud *et al.* solved a crystal structure of the FOXP2 FHD bound to DNA in 2006. The structure revealed the interaction between the FHD and a relatively low binding affinity sequence, referred to as the Wang sequence (Wang *et al.*, 2003, Webber *et al.*, 2017). Within this structure, the FHD was found to specifically

interact with the cognate sequence 5'-CAAATT-3' (Stroud *et al.*, 2006) as both monomer and domain swap dimer. It was also observed that the FOXP2 FHD was capable of looping DNA by binding two individual strands (Stroud *et al.*, 2006). The exact function of this remains unknown.

As previously discussed, the DNA interaction is facilitated by insertion of H3, the recognition helix, into the major groove of DNA. This can be clearly observed within the crystal structure (Stroud *et al.*, 2006)(figure 1.9). The interaction is stabilised by residues associated with H1, H2 and S3, which interact with both the sugar phosphate backbone of DNA, as well as H3 itself, thereby wedging the recognition helix deeper into the major groove (Stroud *et al.*, 2006)

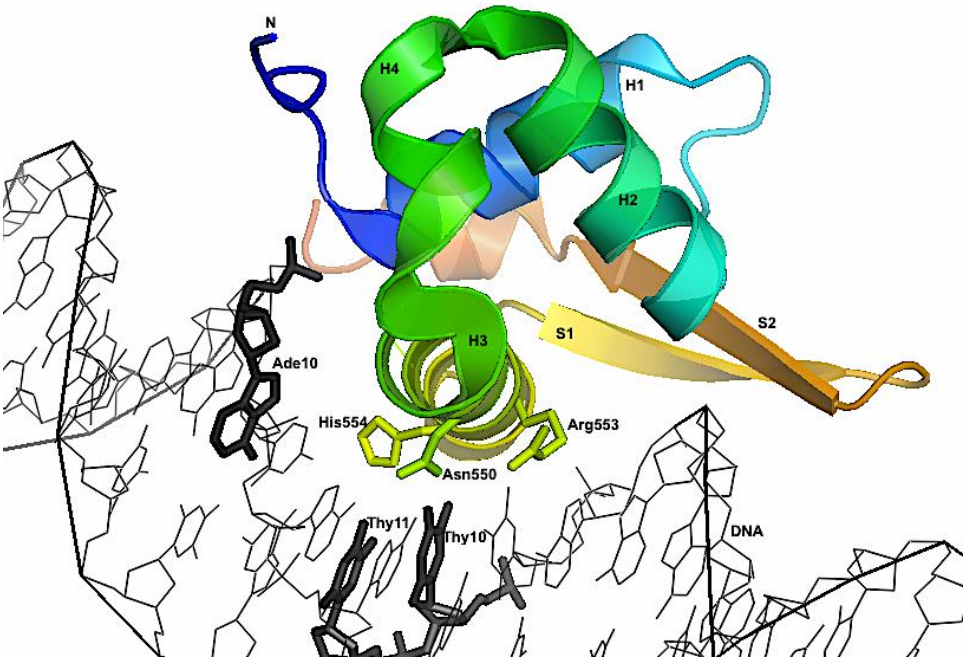


Figure 1.9: Image Illustrating the Interaction Between the FOXP2 FHD and the Wang Sequence (PDB: 2A07) H3 (light green) is wedged into the major groove. The primary residues associated with DNA recognition are Asn550, His554 and Arg553, which interact specifically with Ade10, Thy10' and Thy11' respectively with regard to the Wang sequence. Image was generated using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC.

Regarding the interaction between the Wang sequence and the FOXP2 FHD, Asn550 is reported to form a bidenate hydrogen bond with Ade10, while His554 and Arg553 interact with Thy10' and Thy11', via the formation of either water mediated or direct hydrogen bonds (Stroud *et al.*, 2006). Main chain atoms, as well

as side chains associated with Arg553, His554, Ser557, and Leu558 interact with the DNA extensively through van der Waals contacts. It is evident that the interaction is rich in van der Waals forces, rather than hydrogen bonds, or specific base pair contacts. The predominance in non-specific interactions supports the observation that this protein interacts with various target sequences (Stroud *et al.*, 2006).

In comparison to the Wang sequence, the FOXP2 FHD was computationally predicted to make more specific contacts with the high affinity binding Nelson sequence (5'-TGTTTAC-3') (Nelson *et al.*, 2013, Webb *et al.*, 2017). Asn550 was predicted to interact with Gua13, while His554 was found to be likely to interact with Thy11, Ade15 and Ade16 (Webber *et al.*, 2017). Arg504, Tyr531, Lys549, His559 and Arg583 were predicted to facilitate backbone interactions with DNA (Webb *et al.*, 2017).

Since all FOX transcription factors are associated with a similar, highly conserved DNA binding domain, they tend to interact with similar cognate DNA sequences (with different affinities)(Golson and Kaestner, 2016). Unique structural features within each FOX subfamily are likely to facilitate specificity and affinity in DNA binding (Morris *et al.*, 2018).

A region within the FOXP2 FHD located between H2 and H3, referred to as the hinge loop, has been found to be crucial in regulating the specificity and function of members of the FOXP subfamily (Morris *et al.*, 2018). In addition to the highly variable wing element, the hinge loop region has been previously reported to play a significant role in the specificity of FOX transcription factors, altering the depth of major groove insertion, as well as facilitating novel backbone interactions, essential in fine tuning the protein-DNA interaction (Overdier *et al.*, 1994, Pierrou *et al.*, 1999, Marsden *et al.*, 1998).

The hinge loop region associated with members of the FOXP subfamily is unique in comparison to other FOX TFs. Within the other FOX TFs, this region contains a highly conserved sequence, FPYF. An alanine residue, however, replaces the conserved proline at position 539 within members of the FOXP subfamily (Stroud *et al.*, 2006). This substitution reduces the overall local steric hindrance and increases the flexibility of the region, thereby allowing for an increased number of contacts with the DNA backbone (Morris *et al.*, 2018), as well as facilitating domain swapping, a function unique to members of the FOXP subfamily (Stroud *et al.*, 2006). Experimentally, an A539P substitution in FOXP2 FHD has been shown to abolish the ability of the protein to form domain swap dimers (Stroud *et al.*, 2006, Chu *et al.*, 2011).

Domain swapping describes the phenomenon whereby two or more monomers exchange identical elements, thereby facilitating the formation of dimers or multimeric proteins (Bennette *et al.*, 1994, Lui and Eisenburg, 2002). The exact mechanism by which this occurs differs between proteins, however, the presence of a flexible loop adjacent to the swapped regions is critical. In the case of the FOXP2 FHD, dimerisation occurs via the exchange of H3, H4, as well as S2 and S3 via the extension of H2 into the flexible hinge loop (Stroud *et al.*, 2006)(figure 1.10). While the majority of the domain remains unchanged, H4 and H2 merge (Stroud *et al.*, 2006). A highly conserved semi circle arch is formed, thereby resulting in burial of solvent accessible surface area, specifically various hydrophobic residues present on the surface of the monomer and thus the formation of a hydrophobic core (Stroud *et al.*, 2006). The formation of the hydrophobic core stabilises the dimer, and overall reduces the number of contacts with the DNA (Stroud *et al.*, 2006).

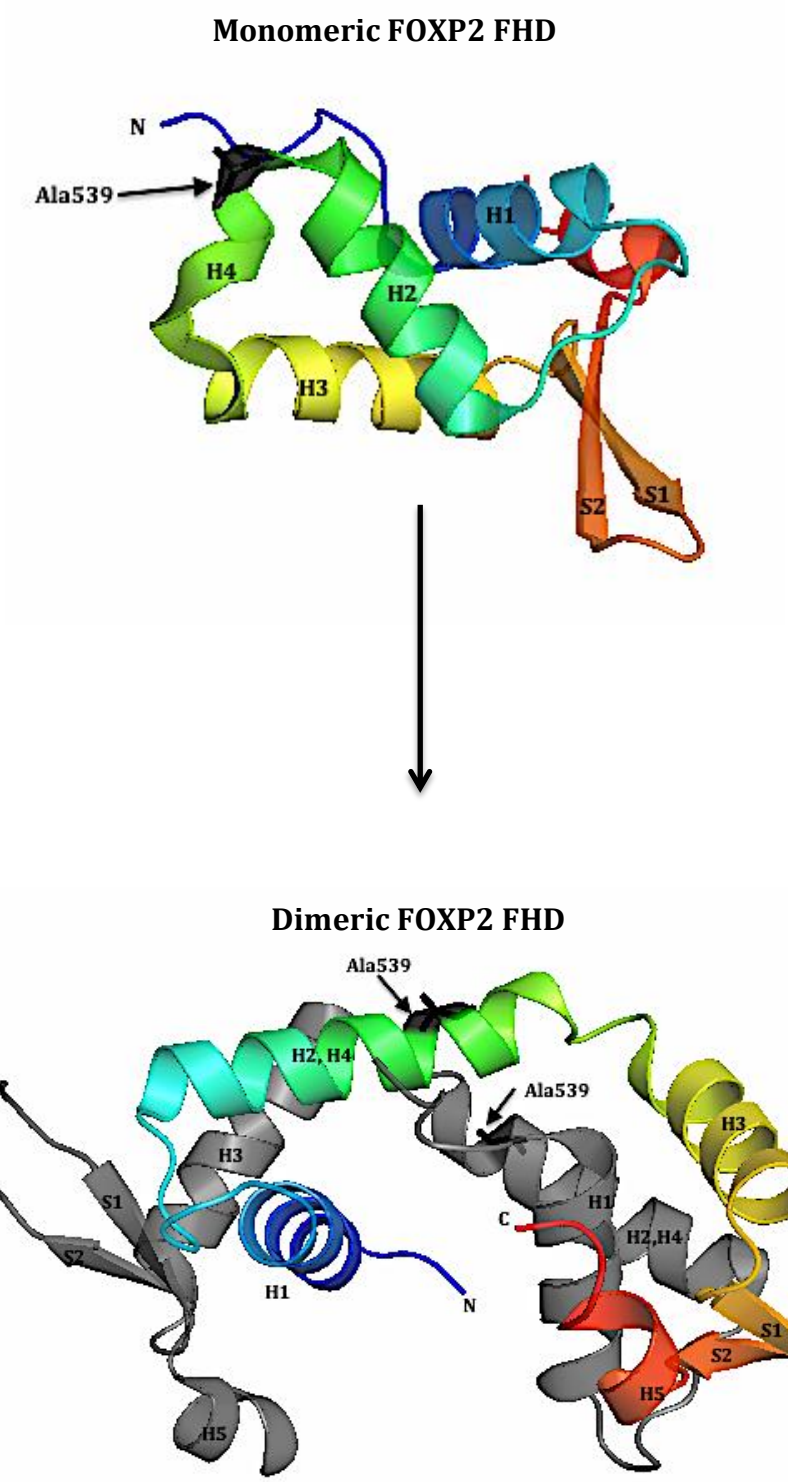


Figure 1.10: Image comparing the structural architecture of monomeric and dimeric FOXP2 FHD (PDB: 2A07) As illustrated, domain swapping creates the formation of a semi-circle arch, with pseudo 2-fold symmetry. The convex surface of the arch is composed of H2/H4 (green) and H3 (yellow), while the concave surface is composed of H5 (red) and H1 (blue). Ala539 is illustrated in black. Image was generated using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC.

It has been hypothesised that the formation of the domain swap dimer plays a significant role in the regulation of transcriptional processes controlled by FOXP2 (Morris *et al.*, 2018). It has been shown *in vivo,* through a luciferase assay, that abolishing the ability of the FOXP2 FHD to form domain swap dimers does not completely inhibit the function of the protein, but rather reduces the transcriptional activity, thereby suggesting that domain swap dimerisation is not essential for function, but rather enhances it, perhaps through selection of binding partners (Morris *et al.*, 2018). The exact physiological function of the formation of the domain swap dimer, however, remains unknown.

In addition to the importance of A539 in domain swap dimerisation, this residue, and thus the hinge loop region, has been shown to be critical in fine-tuning the interaction specifically between the FOXP2 FHD and DNA (Morris *et al.*, 2018). Morris *et al.* performed a study in 2018 comparing the dynamics and mechanism of DNA binding associated with the wild-type FOXP2 FHD, with that of the A539P mutant. The study revealed that the increased flexibility of the hinge loop in wild type FOXP2 facilitated the formation of a more open configuration, thereby allowing for an increased number of electrostatic and ionic interactions with the DNA, in comparison to the mutant, and thus higher affinity binding (Morris *et al.*, 2018). Overall, it was highlighted that the FOXP2 FHD interacts with the DNA major groove shallowly, and predominantly through non-specific interactions (Morris *et al.*, 2018).

Overall, investigating the DNA binding mechanism associated with FOXP2 FHD, in conjunction with studying the global structure of this domain, highlights that the function of the transcription factor is fundamentally sensitive to perhaps small structural or electronic changes. These changes can be induced by innate mutations or, *in vivo,* could be the product of interactions between the TF and the surrounding environment. It is therefore evident that in order to fully understand the global mechanism and physiological function, it is important to assess how co-localised compounds or molecules may affect the overall function of the TF. In other words, the spatial and temporal distribution of co-localised molecules may contribute to regulation of function. Since FOXP2 is largely expressed within neural tissue (Grozser *et al.*, 2008, Sakaguchi *et al.*, 2010), it is possible that the presence of neurotransmitters may play a role in the regulation of the function of this molecule.

1.3. Dopamine

1.3.1. Signaling Molecules

Signalling molecules are defined as a largely varied group of molecules that function as ligands, interacting with specific receptors expressed by target cells, thereby regulating cellular processes by transmitting information between locations (Miller and Davidson, 2013). Signals may act locally or over long distances, facilitating communication between neighbouring cells (Miller and Davidson, 2013).

Generally speaking, signalling molecules can be grouped based on their chemical nature, as well as mechanism of action (Miller and Davidson, 2013). The major groups include the steroid hormones, the neurotransmitters, the peptide hormones and growth factors, the plant hormones, and carbon monoxide/nitric oxide (Miller and Davidson, 2013).

This investigation will focus on the effect of the presence of a neurotransmitter, dopamine, on the binding of the FOXP2 FHD to DNA, as well as the overall structural effects the presence of dopamine may have on the FOXP2 FHD. Since FOXP2 is recognised as the 'language gene' (Lai *et al.*, 2001), this examination may provide some insight regarding the potential contribution that the presence of this molecule may have on speech acquisition.

1.3.2. Neurotransmitters

Broadly speaking, neurotransmitters are molecules that are associated with signal transduction. Neurotransmitters are released at the presynaptic axonal membrane of a neuron into the synaptic cleft, and bind to receptors located within postsynaptic neuron, thereby eliciting a biochemical response, often depolarisation of the membrane potential (Rizo *et al.*, 2018).

In order for a small molecule to be classified as a neurotransmitter, it must display five key characteristics. Firstly, it must be synthesised within, and released from, a presynaptic neuron containing both the neurotransmitter and the appropriate enzymes required for synthesis. The released molecule must be chemically or pharmacologically identifiable, and the effects of the action of the molecule must be blocked in a dose dependent manner by the presence of competitive antagonists (Rizo *et al.*, 2018). There must also be active mechanisms in place by which the biological effect of the neurotransmitter are terminated, for example, reuptake,

transport or chemical inactivation. Finally, the biological effects produced by the interaction between the signalling molecule and the postsynaptic neuron must be reproducible (Rizo *et al.*, 2018).

In 2013, Squire *et al.* describe three major classifications of neurotransmitters, namely classic neurotransmitters, which are largely amino acid derived (including the monoamines), non-classical neurotransmitters that are neuropeptides (such as secretin, oxytocin and dynorphin), and unconventional neurotransmitters (for example nitric oxide and endocannabinoids).

Dopamine is classified as a classic neurotransmitter, specifically a biogenic amine, and is tyrosine derived (catecholamine). The process by which dopamine is synthesised in the brain, specifically within dopaminergic neurons, is illustrated below (figure 1.11).

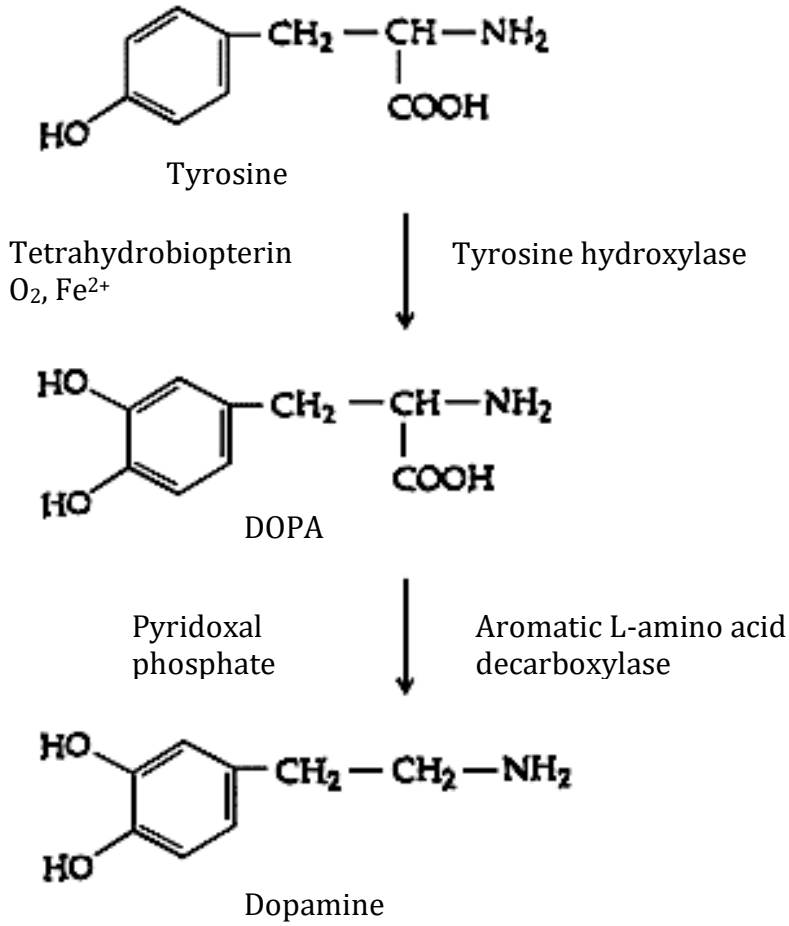


Figure 1.11: Image illustrating the production of dopamine from tyrosine (Image adapted from Cooper, 2001). As illustrated, dopamine is synthesised from tyrosine via the production of an initial DOPA intermediate. The synthesis of DOPA by tyrosine hydroxylase has been reported to be the rate-limiting step of this reaction (Cooper, 2001).

1.3.3. Dopamine Structure and Function

Dopamine is a planar molecule composed of a single benzene ring, two hydroxyl side chains and an amine group (attached via an ethyl chain) (figure 1.12). The molecule has a molecular weight of 153.181 g/mol (Pubchem Compound Database) and when protonated is freely soluble in water, with a reported solubility of 600g/L at 25 °C (Pubchem Compound Database). Dopamine therefore acts as an organic base. This molecule is also highly sensitive to oxidation, and is most stable under acidic conditions (O'Neil, 2006).

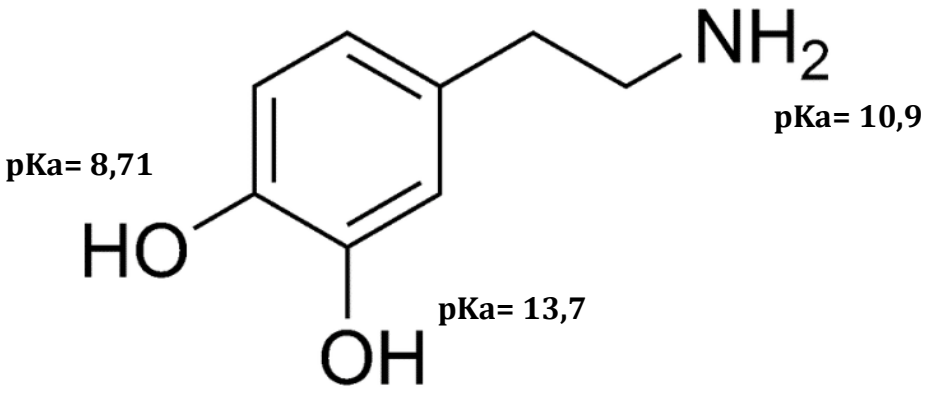


Figure 1.12: The Chemical Structure of Fully Protonated Dopamine Dopamine contains 2 hydroxyl groups and an amine group, joined to a benzene ring via an ethyl chain. Dopamine has 3 hydrogen bond donor sites, and 3 acceptor sites (Pubchem Compound Database). The pH at which ionisation occurs with regard to each functional group is also indicated (Pubchem Compound Database).

Dopamine specifically regulates neuroplasticity (Bazzari and Parri, 2019), and functions by interacting with the G-coupled dopamine receptors, namely D1, D2, D3, D4, or D5 (Missale *et al.*, 1998). Synthesis occurs within dopaminergic neurons, mainly within the cerebral cortex and subcortical layers of the brain. There is evidence to suggest that some dopaminergic regions are associated with high levels of FOXP2 expression (Haesler *et al.*, 2004).

For example, one study investigating the expression patterns of FOXP2 within the brain of Zebra Finches suggested prominent FOXP2 expression within the dopaminergic midbrain region that projects into the basal ganglia, substantia nigra and dorsal thalamus (Haesler *et al.*, 2004). It has also been reported that FOXP2 is expressed in striatal projection neurons. These neurons display evidence of dopaminergic innervation in that D1 receptors (Snyder *et al.*, 1998) as well as enzymes associated with the synthesis of biogenic amines are expressed (Haesler *et al.*, 2004).

Dopamine is functionally associated with cognition, movement and regulation of hormone release, specifically acting through four major defined circuits within the brain, the nigrostriatal pathway, the mesolimbic pathway, the mesocorticol pathway and the tuberoinfundibar pathway (Chritine and Aminoff, 2004, Bjorklund and Dunnett, 2007, Paulus and Schomburg, 2006, Ben-Jonathan and Hnasko, 2001). The nigrostriatal pathway functions in purposeful movement, motor control and

motivated behaviour (Haber, 2014, Ledonne and Mercuri, 2017), while the mesolimbic pathway is associated with the pleasure and reward centre (Adinoff, 2004). The mesocorticol pathway functions in regulating cognition and working memory (Ledonne and Mercuri, 2017), while the tuberoinfundibar pathway is predominantly associated with regulation of prolactin release (Ledonne and Mercuri, 2017)

Dopamine is likely associated with language acquisition, especially with reference to the nigrostriatal pathway and mesocorticol pathway. It is also interesting to note that dopaminergic modulations associated with cortico-basal ganglia circuits have been reported to be vital in song learning in birds, and this process is thought to be similar to certain aspects of language acquisition in humans (Hara *et al.*, 2007). It was also noted that lower levels of dopamine are linked to an increased efficiency in processing language and speech (Enard *et al.*, 2009).

1.3.4. Autoxidation and Neuromelanin

Dopamine is sensitive to oxidation under alkaline conditions, and it has been reported that this molecule undergoes spontaneous, non-enzymatic autoxidation within the brain, specifically within the cytoplasm of neurons associated with the substantia nigra, but also within the locus coeruleus and cerebral meninges (Zhang and Dryhurst, 1993). The autoxidation of dopamine results in the formation of various reactive, and in some cases, cytotoxic intermediates, and ultimately neuromelanin (Zhang and Dryhurst, 1993). This process is illustrated below (figure 1.13).

In vivo, dopamine, stored within the synaptic vesicle, is protected from autoxidation by the low pH of the surrounding environment (Sulzer *et al.*, 2000). When released into the cytoplasm, some degree of autoxidation is inevitable. Dopamine is, however, somewhat protected by the presence of antioxidants, specifically ascorbic acid (Sulzer *et al.*, 2000). Various cellular detoxification mechanisms are also exploited in preventing the accumulation of highly reactive compounds, including enzymatically driven intramolecular rearrangements to form more stable compounds (Rosengren *et al.*, 1996, Solano *et al.*, 1999). Relatively high concentrations of various detoxification enzymes (such as GSTs) also aid in reducing the accumulation of reactive intermediates by acting as nucleophiles (Solano *et al.*, 1999).

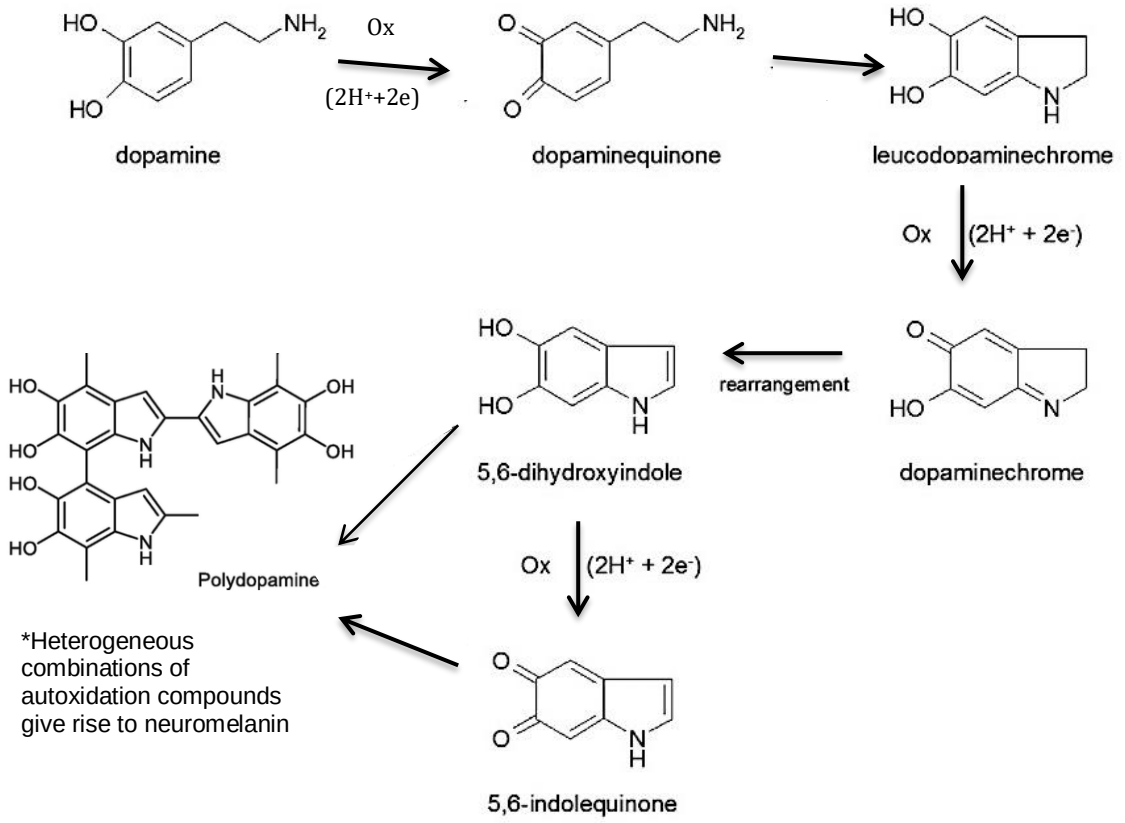


Figure 1.13:Image illustrating the autoxidation of dopamine to form neuromelanin (Image adapted from Zangmeister *et al.*, 2013). Dopamine first reacts with oxygen to form dopaminequinone, as well as superoxide anion. Dopaminequinone then cyclises to form leucodopaminechrome, which then undergoes oxidation to form dopaminechrome. Dopaminechrome is rearranged to form 5,6-dihydroxyindole, which further oxidises to form 5,6-indolequinone. These two compounds together can polymerise to give rise to polydopamine, which facilitates the formation of neuromelanin (Munoz *et al.*, 2012). This represents one of many hypothesised paths to the formation of neuromelanin.

The exact composition of neuromelanin is unknown, and the structure is not well characterised due to the numerous different combinations of various precursors, as well as multiple oxidation pathways (Solano *et al.*, 1999). This black precipitant appears to form randomly and non-enzymatically, but it has been hypothesised that it is composed of a cysteine and pheomelanin core, surrounded by eumelanin. Neuromelanin is composed predominantly of polydopamine, however it has also been reported to incorporate up to 25% cysteine in the form of benzothiazine, up to 21% cys-dopamine units and up to 6% DOPA (Wakamatsu *et al.*, 2003, Haining, 2017). Neuromelanin also sequests and chelates metal ions (Korytowski *et a*l., 1995), low molecular weight drugs, organic amines and lipids (Bush *et al.*, 2006, Haining, 2017).

The exact physiological function for neuromelanin has not yet been identified (Haining, 2017), however the presence of this

polymer results in the pigmentation and darkening of various regions of the brain. Loss of pigmentation has been directly associated with the progression of Parkinson's disease (Hirsch et al., 1988, Jenner et al., 1992, Kastner et al., 1992, Karlsson and Lindquist, 2013). One hypothesis suggests that neuromelanin facilitates storage and release of excess dopamine within the cytoplasm; another possible physiological function of neuromelanin is neuroprotection (Karlsson and Lindquist, 2013), the scavenging of free radicals. Some scientists argue that neuromelanin has no physiological function, and that it is just a product of aging as it is absent at birth, and accumulates throughout adult life (Haining and Achat-Mendes, 2017). One attractive hypothesis suggests that the function of neuromelanin is simply to prevent the accumulation of the highly reactive and toxic quinones, specifically o-quinones that form as a result of dopamine oxidation within neural tissue (Riley, 1992, Wilczok *et al.*, 1998, Hirsch *et al.*, 1991, Bustamante *et al.*, 1993, Zucca *et al.*, 2017).

O-quinones are known to be particularly harmful and have been reported to contribute to the degeneration of dopaminergic neurons (Solano *et al.*, 1999), as well as being linked to apoptosis (Offen *et al.*, 1997). Quinones and semiquinones are also known to react spontaneously with biomolecules, specifically through the nucleophillic attack of amino and thiol groups associated with amino acids (Berlett and Stadman, 1997). Binding of these compounds to proteins may result in changes to metabolic pathways via inactivation of enzymes, or may promote denaturation or structural changes in non-enzymatic proteins (Graham, 1978, Mattamal *et al.*, 1995, Prota, 1998). Three monomeric quinones are generated within the dopamine autoxidation pathway; namely dopaminequinone, dopaminechrome and 5,6-indolequinone. These compounds have been identified to be particularly cytotoxic (Haining and Achat-Mendes, 2017).

It is challenging to study the isolated role, or neurological effect, of the accumulation of a single autoxidation compound based on a globally observed neurodegenration, *in vitro* and *in vivo*. This is because the relative concentration of each compound at a particular time point is the product of multiple interconnected reactions (Bisaglia *et al.*, 2007), and due to the rapid and spontaneous intramolecular rearrangements that are constantly occurring. There is, therefore, conflicting information presented in the literature regarding which of the autoxidation compounds are the most reactive or harmful, and in some cases, proposed interactions between a specific autoxidation compound and protein of interest, or specific

pathway, have been challenged (Bisaglia *et al.*, 2007). Another limitation in studying the function of isolated autoxidation products lies in that a compound may react with a protein in one form, but may be converted to another through interaction with the protein itself (Bisaglia *et al.*, 2007, Bruning *et al.*, 2019).

Some studies suggest that dopaminequinone is the most reactive of the dopamine autoxidation products (Paris *et al.*, 2010). This quinone has been reported to bind and inactivate various proteins including mitochondrial glutathione peroxidase (Hauser *et al.*, 2013), heat shock protein 60 (Zhang *et al.*, 2018), dopamine transporter (Whitehead *et al.*, 2001) and parkin E3 lygase (LaVoie *et al.*, 2005), thereby resulting in neurotoxicity and protein degradation impairment. Dopaminequinone has also been hypothesised to promote formation of neurotoxic alpha-synuclein oligomers, thereby suggesting a potential role in the progression of Parkinson's disease (LaVoie *et al.*, 2005). Owing to its relative reactivity, this compound is highly unstable at physiological pH (dopaminequinone is stable at a pH of 2) (Zhang *et al.*, 2018), and thus following its generation, the compound almost instantly cyclises to form leucodopaminechrome (Tse *et al.*, 1976). This is followed by rapid oxidation to give rise to dopaminechrome (Paris *et al.*, 2010).

Dopaminechrome, also referred to as antichrome (PubChem), is hypothesised to be the most stable of the o-quinones produced in the dopamine autoxidation pathway (Paris *et al.*, 2010, Zhang *et al.*, 2018). This quinone has indeed been found to bind proteins *in vivo* and *vitro*, and to induce protosomal (Zafar *et al.*, 2006) and mitochondrial (van Laar *et al.*, 2009) dysfunction via the formation of adducts with associated proteins, as well as contribute to oxidative stress (Puspita *et al.*, 2017). Dopaminechrome has been found to disrupt proteins associated with the cytoskeleton (Paris *et al.*, 2010), and, much like dopaminequinone has hypothesised to promote the formation of neurotoxic alpha-synuclein oligomers (Conway *et al.*, 2001, Norris *et al.*, 2005). Interestingly, a rat study performed by Diaz-Veilz *et al.* in 2002 found that the unilateral injection of dopaminechrome into the substantia nigra of the brain resulted in impairment of motor and cognitive behaviors.

Some studies argue that 5,6-indolequinone is in fact the most reactive and cytotoxic of the autoxidation products (Bisaglia *et al.*, 2007, Bisaglia *et al.*, 2010). This has been partially attributed to the fact that this compound does not accumulate within solution, and upon formation undergoes a rapid auto-condensation reaction, reacting with 5,6-dihydroxyindole to

form a black precipitant (Bisaglia *et al.*, 2010). 5,6-indolequnione was shown to interact with alpha-synuclein in more significant concentrations than that observed with respect to the other quinones (Bisaglia *et al.*, 2007), and it has been hypothesised that perhaps some claims regarding observed interactions between dopaminechrome and proteins may actually be misinterpreted, potentially suggesting an interaction involving 5,6-dihydroxyindole (through the formation of an adduct) or 5,6-indolequinone due to spontaneous rearrangements following the addition of dopaminechrome to solution (Bisaglia *et al.*, 2010). This claim was based on a study whereby dopaminechrome and 5,6-indolequinone were reacted with NADH and GSH, and it was observed that the reaction between these molecules and 5,6-indolequinone was significantly favoured over that with dopaminechrome (Bisaglia *et al.*, 2010).

Despite a lack of clarity regarding the relative reactivity and cytotoxicity of each of the isolated autoxidation products, it is evident that the overall process of dopamine autoxidation is globally harmful to neural tissue, and that the presence of these compounds has the capacity to affect protein function.

1.4. The Potential for Dopamine to Affect the Function of FOXP2

The distribution of FOXP2 expression within the developing brain allows for a potential direct interaction between the transcription factor and dopamine. Dopamine is synthesised in multiple brain regions displaying FOXP2 expression (Co *et al.*, 2020, van Rhijin *et al.*, 2018, Enard *et al.*, 2009, Raghanti *et al.*, 2016). For example, dopamine is present in the basal ganglia during prenatal development (Haber, 2014). This region has been reported to have particularly high FOXP2 expression (Takahashi *et al.*, 2003).

Multiple links have already been identified between FOXP2 and dopamine regulated circuits. For example, a study performed by Enard *et al.* in 2009, highlighted that FOXP2 down regulates cellular dopamine levels, as well as effects the morphologies associated with medium spiny neurons- a very common target of dopaminergic neurons. Another study performed by Murugan *et al.* in 2013 reported that dopamine could influence how fast basal ganglia signals were propagated in birds expressing normal, but not mutated FOXP2. Although these, and other studies have highlighted mutant FOXP2 disrupting dopaminergic pathways, the direct interaction between the

transcription factor and signalling molecule has not yet been investigated.

Since transcription factors lack specific binding pockets, it is difficult to predict whether dopamine has the potential to bind to the FOXP2-FHD itself, thereby significantly altering its structure, and if that were the case, where exactly the small molecule would bind. It is also possible that the presence of this molecule at specific physiologically relevant concentrations may disrupt the interaction between the FOXP2 FHD and DNA, by interacting with the DNA, or with the protein-DNA complex. This study will therefore focus on firstly investigating whether dopamine interacts with the FOXP2 FHD, thereby giving rise to significant conformational changes that may affect the function of the protein, and secondly investigate whether the presence of dopamine affects the interaction between the FOXP2 FHD and its cognate binding sequence.

1.4.1. The Potential for a Direct Interaction between the FOXP2 FHD and Dopamine

It is not uncommon for small molecules to interact with transcription factors. The most well known example of the relevance of various ligands modulating the activity of TFs can be seen in the nuclear hormone receptor (NHR) family of TFs (Groove and Walhout, 2008). The NHR family, including, amongst others, the androgen receptor, oestrogen receptor, peroxisome proliferator activated receptors and retinoic acid receptor become potent activators of transcription upon natural ligand binding (Smith and O'Malley, 2004). Many of these TFs have the capacity to interact with more than one ligand of both endogenous and exogenous origins (Smith and O'Malley, 2004). Interestingly, despite being relatively well studied, many members of the NHR family are referred to as "orphan" receptors as their corresponding ligands have not yet been identified (Groove and Walhout, 2008). This highlights the gap in the current knowledge base regarding the full extent of the role of ligands in regulation of transcription.

There have been reports of small molecules, specifically inhibitors, binding other FOX transcription factors, including FOXM1 (Gormally *et al.*, 2014) and FOXO1 (Nagashima *et al.*, 2010). Although the exact binding mechanism was not eluded, the FOXM1 inhibitor, FDI-6 (figure 1.14), was found to interact directly and specifically with the FHD, thereby displacing DNA binding (Gormally *et al.*, 2014). This suggests a potential role for small molecules in the regulation of the activity of FOX transcription factors through direct interaction.

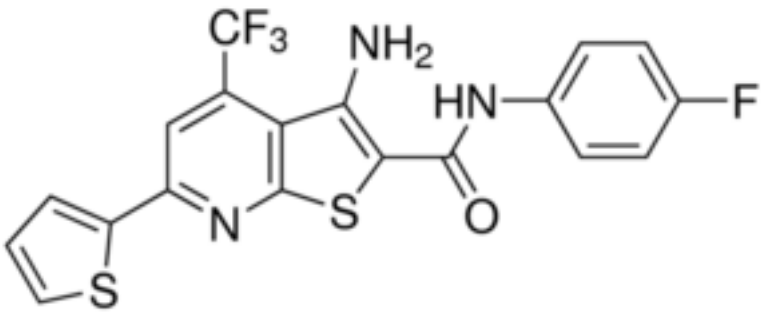


Figure 1.14: FDI-6 (inhibitor of FOXM1 FHD) (NCGC00099374)
3-amino-N-(4-fluorophenyl)-6-(thiophen-2-yl)-4-(trifluoromethyl)thieno[2,3-b]pyridine-2-carboxamide TFA is illustrated.

1.4.2. The Potential for the Presence of dopamine to disrupt DNA binding

Generally speaking, interactions between the DNA double helix and transcription factors are dependent on a variety of factors including the physical shape of the DNA itself, as well as chemical complementarity between the sequence of nitrogenous bases and amino acids (Rohs et al., 2009). The position of the alpha-carbon atoms in the polypeptide chain, as well as the nature of the side chains associated with the transcription factor also play a role in DNA recognition and binding (Pabo and Nekludova, 2000). The interaction is mediated by either van der Waals forces or hydrogen bonding. Van der Waals forces are usually associated with increasing the stability of the protein-DNA complex, while hydrogen bonds are associated with more specific interactions (Mirny, 2002). Water molecules may also be associated with mediating hydrogen bonds. Adenine usually forms hydrogen bonds with lysine and arginine, while glutamine and asparagine interact with guanine (Mandel-Gutfreund et al., 2001). Hydrogen bonds may also form between cytosine and glutamate or aspartate, however this interaction is less common (Luscombe et al., 2001).

Numerous studies have suggested that various small molecules, including dopamine (Liu *et al.,* 2002), have the capacity to bind directly (and non-covalently) to the DNA double helix. Small molecules may interact with either the major or minor groove (more commonly the minor groove), may intercalate between nitrogenous bases, may bind a single strand of the DNA, or display a form of multi-modal binding. Dopamine has been reported to bind DNA when either the neurotransmitter or DNA is present in excess (Liu *et al.,* 2002). Due to the positive charge of dopamine, and overall negative charge of DNA, the

interaction was hypothesised to be largely a result of electrostatic forces, however it was also thought that the planar nature of the small molecule may facilitate a degree of hydrogen bonding and hydrophobic interaction with the DNA core (Liu *et al.,* 2002). The specificity of this interaction has, however, not been investigated. It is possible that an electrostatic interaction between dopamine and target sequences of FOXP2 could alter the affinity or binding capacity between the FHD and DNA.

1.5. Significance of Research

The discovery of FOXP2 as a 'language gene' is still relatively new, and there is a lack of both structural and functional information surrounding this protein, especially concerning prenatal neural development.

Verbal dyspraxia, the disease most strongly and significantly associated with FOXP2, cannot presently be cured, and considerably affects the lives of those affected by it. The current 'treatment' available, usually performed by a speech therapist, is intensive, time consuming and must be personalised. Even with regular direct therapy, observable progress in patients is slow, and in some cases, impairment is so severe that functional communication may not be possible at all. The overall goal of current treatment is to maximise the ability to communicate at each stage of the disease, rather than reversal of the decline (Duffy, 2013). Investigating and fully understanding the role of FOXP2 in language acquisition may allow for potential molecular-level treatments to be developed. Similar holds true as for other diseases associated with verbal ataxia, including autism, which affects 1 in 160 children worldwide (WHO, 2017), and schizophrenia, affecting 21 million people worldwide (WHO, 2016). Understanding the molecular mechanisms behind these diseases may facilitate rational drug design in the future.

The effect of the neurotransmitter may have an organisational and fluctuating effect on language acquisition during prenatal development (Becker *et al*., 2005). There is, however, currently no evidence to clearly suggest that this is the case. One major limitation is the relatively low sensitivity of techniques that are presently available in ethically investigating prenatal neural patterning. Another challenge in studying language acquisition in its entirety is the enormous number of associated and overlapping pathways, in addition to large gaps in information. Investigating an interaction *in vivo* may also result in variation between subjects (due to innate genetic differences) and thus inconsistent results. This investigation will focus on the specific neurotransmitter-FOXP2 FHD-DNA interaction; thereby allowing these challenges to be somewhat overcome. Since experimentation will be done in isolation, interference from overlapping pathways will be eliminated, thereby potentially allowing a single gap in knowledge to be filled. This will allow clear picture of how dopamine may affect the interaction to be determined. Although this approach only provides information regarding a single interaction, studying multiple different interactions in isolation and then integrating the information will allow for an overall understanding of

language acquisition to be determined over time. It is, however, important to note that interactions observed *in vitro* are likely representative only of preliminary studies, and further investigation *in vivo* would be required to confirm the significance of the observations.

1.6. Aims and Objectives

1.6.1. Aim

The aim of this experiment is to determine the structural and functional effect of the presence of a neurotransmitter, dopamine, on the interaction between the FOXP2 FHD and a recognised cognate binding sequence, the Nelson sequence (5'-TGTTTAC-3')

1.6.2. Objectives

- To successfully overexpress and purify the WT FOXP2 FHD via heterologous recombinant protein expression (within an *E. coli* host).
- To spectroscopically characterise the overexpressed product, thereby ensuring that the protein is natively folded.
- To functionally characterise the overexpressed product
- To investigate whether the presence of dopamine at physiological concentrations, as well as in excess has a significant effect on the secondary and tertiary structure of the FOXP2 FHD.
- To investigate whether the presence of dopamine at physiological concentrations, as well as in excess has a significant effect on the ability of FOXP2 FHD to bind the cognate high affinity binding sequence, the Nelson sequence.
- To investigate the potential for dopamine, as well as products of dopamine autoxidation to bind the FOXP2 FHD *in silico* through molecular docking.

2. METHODS AND MATERIALS

2.1. Overexpression and Purification

2.1.1. Transformation

When studying the structure and function of a specific protein in vitro, there is often an experimental demand for large amounts of pure protein. It is therefore important to design a system of protein expression whereby high yields can be obtained. Heterologous protein expression facilitates this, in that the gene of interest is cloned into a host organism and can be overexpressed. In this case, an *Escherichia coli* (*E. coli*) host was exploited, however alternative heterologous expression systems include yeast or plant cells (Demain and Vaishnav, 2009, Adrio and Demain, 2010). Cloning is achieved by transforming the host with a vector that has been designed to contain the gene of interest, as well as various regulatory regions, thereby allowing for overexpression to be controlled (Makrides, 1996).

The pET-11a vector (GensScript, USA) was used in this study to overexpress the FOXP2 FHD. The plasmid construct is illustrated below (figure 2.1).

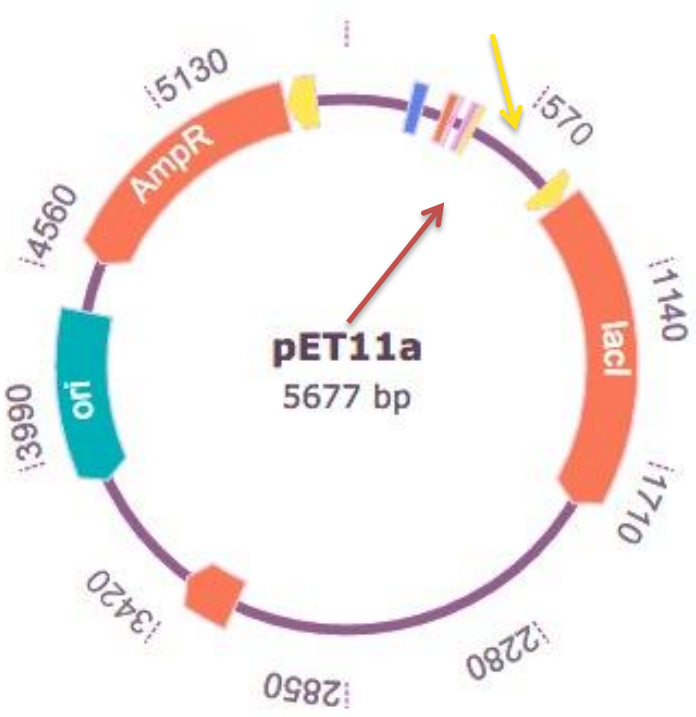


Figure 2.1: Schematic illustrating the plasmid construct of pET11a. Promoter regions are illustrated in yellow, while the regions shown in orange represent sequences encoding specific protein products. Due to the location of the multiple cloning site (MCS)(red arrow) downstream of the T7 promoter region (yellow arrow), overexpression of the insert is controlled by the activity of bacteriophage T7 RNA polymerase. The lac operon, including the presence of lac1 and promoter, as well as the Lac operator region (purple) located downstream of the T7 promoter allows for overexpression to be tightly controlled. (Image generated with GenSmart Design)

The lac1 region encodes the Lac repressor. This protein binds the Lac operator, and inhibits transcription by blocking the activity of T7 polymerase. Transcription can be activated by the addition of isoproylthiogalactopyranoside (IPTG), a galactose analog that interacts with the lac repressor, thereby altering its conformation and inhibiting interaction with the lac operator and allowing RNA polymerase to bind and transcription to proceed (Marbach and Bettenbrock, 2012).

Selection of only transformed *E. coli* cells can be achieved by conferring antibiotic resistance. The Amp^R region encodes β-lactamase. The expression of this enzyme facilitates specific antibiotic resistance in bacteria, as it catalyzes the hydrolysis of the β-lactam ring associated with penicillin, thereby inactivating the action of the antibiotic (Abraham and Chain, 1940).

For the overexpression of FOXP2 FHD, the following insert was cloned into the MCS of the pET11a vector (figure 2.2). When designing the insert, the nucleic acid sequence for the expression of this protein was codon optimised (Genscript). Codon optimisation is essential in heterologous protein expression as different organisms tend to display bias towards certain codons over others encoding the same amino acid. In this study, a bacterial host was utilised in overexpressing a human protein. Selecting codons that are preferentially utilised by bacteria can therefore increase the chance of expression success (Burgess-Brown *et al.*, 2008).



Figure 2.2: Schematic illustrating the regions upstream the FOXP2 FHD codon-optimised sequence cloned into the MCS of the pET11a vector The insert includes an N-terminus histidine tag (purple) to aid in affinity purification, as well as a thrombin recognition sequence (red) to allow for removal of the Histag. The specific cleavage site is depicted by a dotted line. The dark blue region represents the T7 phage gene 10 leader. This sequence acts as a ribosomal binding site, and functions to enhance expression of foreign genes within an E coli host (Olins *et al.*, 1988).

Competent T7 *E. coli* cells (50 µL) (New England Biolabs, USA) was thawed on ice, before the addition of 2 µL of the vector containing the FOXP2 FHD insert (solubilised in milliQ water, around 50 ng). The mixture was incubated on ice for 30 minutes, before heat shocking the cells at 42 °C for 10 seconds, thereby destabilising the membrane and facilitating the entry of DNA (Chen and Dubnau, 2004). The cells were then placed on ice for 5 minutes, before the addition of 950 µL of SOC media (Super Optimal broth with Catabolite repression, composed of 2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 20 mM glucose). The cells were grown at 37 °C for 60 minutes with vigorous shaking (250 rpm), before plating on agar containing 0.1 mg/mL ampicillin (Melford, UK). The plates were grown overnight at 37 °C and single, isolated colonies were selected. Each colony was grown for 16 hours in Lysogeny broth (LB) (composed of 1% w/v tryptone, 0.5% w/v yeast extract and 1% NaCl) with 0.1 mg/mL ampicillin. Glycerol stocks were prepared by mixing 80% glycerol (80% v/v glycerol, 20% v/v distilled water) with the culture in a 1:1 ratio and flash frozen, before storing at -80 °C.

The vector was extracted from the leftover culture using the GeneJet plasmid miniprep kit (Thermofisher Scientific, EU) as per manual, and sent for sequencing with Inquaba Biotech

(RSA), as to ensure that the transformation was successful, and there were no mutations introduced in the insert.

2.1.2. Protein Expression

The protocol for overexpression and purification of WT FOXP2 FHD had been optimised within our lab (Blane and Fanucchi, 2015). Expression and purification trials were therefore not required.

The T7 glycerol stocks were used to prepare an initial overnight culture by performing a 1000-fold dilution with freshly autoclaved 2XYT media (1.6% w/v tryptone, 1% w/v yeast extract, 0.5% w/v NaCl) containing 0.1 mg/mL ampicillin. The antibiotic was added to facilitate the growth of only successfully transformed cells. The culture was incubated for 16 hours at 37 °C, with vigorous shaking (230 rpm) to allow for aeration and bacterial growth. This initial overnight culture was used to inoculate a larger volume (roughly 6 L) of freshly autoclaved 2XYT media (1 in 50 dilution), containing 0.1 mg/mL ampicillin. This secondary culture was grown at 37 °C with vigorous shaking until an OD_{600} of 0.6-0.8 was observed (typically around 2 hours). The cells were cold shocked at 4 °C for 10 minutes, before inducing overexpression with the addition of 0.6 mM IPTG. The temperature was reduced to 20 °C, and the culture was grown for a further 22 hours. The cells were then harvested via centrifugation for 20 minutes at 5000 x g at 4 °C. The supernatant was discarded, and pellet was resuspended in IMAC equilibration buffer (10 mM Tris-HCl, 30 mM Imidazole, 500 mM NaCl, pH 7.5). Resuspended cell lysates were stored at -20 °C until required for purification.

The resuspended cell lysates were thawed completely by rotation at room temperature, and 1 mg/mL chicken egg lysozyme (Sigma-aldrich, USA) was added. The formation of ice crystals when freezing the lysate aids in disrupting the cell membranes (Johnson and Hect, 1994), while the addition of lysozyme facilitates chemical cell lysis. The cells were further lysed via multiple rounds of sonication (Sonica Q55) at 60% power, alternating sonication for 30 seconds, with 30 seconds rest on ice to prevent protein degradation due to increasing temperature. The sonication process also disrupts interactions between the overexpressed product and contaminating DNA. DNase I (Merck, Germany) (0.01 mg/mL) was then added to the lysate to further reduce DNA contamination. After incubation at room temperature for roughly 20 minutes, the lysate was centrifuged at 23 000 x g for 25 minutes at 4 °C, and the supernatant was collected.

2.1.3. Immobilised Metal Ion Affinity Chromatography (IMAC)

IMAC facilitates the isolation and purification of a tagged recombinant protein of interest from a heterogeneous mixture of bacterial proteins (following overexpression). The separation is based on affinity, specifically the interaction between the electron donating polyhistidine tag (fused to the protein) and the stationary phase, chelated metal ions (in this case, specifically Ni^{2+}) immobilised on the matrix of the column (Porath *et al.*, 1975). The metal ions are immobilised through the use of a support molecule. In this case, nitrilotriacetic acid (NTA) was used. When the recombinant protein travels through the column, it is retained, and can be eluted by increasing concentrations of imidazole (as this molecule competes for the histag binding site on the column)(Bornhorst and Falke, 2000).

IMAC was performed using the *GE-healthcare* (Sweden*)* 5 mL Histrap® column (charged with nickel sulfate as per manufacturers instructions), and fast protein liquid chromatography (FPLC) (with the AKTA Pure Chromatography System). All IMAC steps were performed at a flow rate of 5 mL/min. The nickel-NTA column was first equilibrated with 10 column volumes of equilibration buffer, before loading the collected supernatant containing the His-tagged overexpressed protein. To remove DNA bound to the immobilised FOXP2 FHD, as well as to reduce non-specific interactions, the column was washed with 5 column volumes of salt wash buffer (10 mM Tris-HCl, 30 mM imidazole, 1.5 M NaCl, pH 7.5). After equilibrating the column, FOXP2 FHD was eluted via a single step elution with an increased imidazole concentration (10 mM Tris-HCl, 300 mM imidazole, 500 mM NaCl, pH 7.5).

2.1.4. Size Exclusion Chromatography (SEC)

SEC is used as a polishing step in purification, ensuring that the protein of interest is at maximum purity. SEC also facilitates separation of different oligomeric states of the protein column (Lathe and Ruthven, 1955). In this chromatographic technique, separation is based on differences in the hydrodynamic radius of molecules travelling through the column (Lathe and Ruthven, 1955). The column is packed with a porus matrix, which acts as the stationary phase. Molecules that are larger than the pores are unable to be retained, and travel through the column with the mobile phase. Smaller molecules, depending on their relative size, will diffuse into the pores at varying degrees. The smallest molecules will thus elute last (Moldoveneanu and David, 2013).

A sephadex S75 16/60 SEC column (GE Healthcare, Sweden) was equilibrated with 1 column volume (120 mL) of IMAC equilibration buffer, at a flow rate of 0.2 ml/min. Fractions collected from IMAC containing FOXP2 FHD (identified by SDS-PAGE) were concentrated to ±2 mL. All aggregation was removed from the sample by centrifugation at 13 000 x g for 10 minutes at 4 °C, before loading onto the equilibrated column. The flow rate was set at 1 mL/min for 120 minutes. This protocol was repeated with the GE healthcare standards, thereby allowing for the molecular weight of eluent to be determined.

2.2. Purity Assessment, Concentration Determination and Structural Characterisation

2.2.1. Tricine Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE is an analytical technique that exploits the migration of charged biomolecules through a porus gel within an electric field. Under native conditions, the rate at which a molecule migrates is dependent on its electrophoretic mobility, a function of the size, shape and net charge (Laemmli, 1970).

With regard to PAGE, the porus gel matrix is composed of a three-dimensional cross-linked network of acrylamide and N,N'-methylene-bis-acrylamide. The pore size can be manipulated by altering the ratio of these two components. Polyacrylamide gels are favoured over agarose gels due to the smaller pore size, and thus increased sensitivity for the separation of small molecules (Svasti and Panijpan, 1997).

After treatment with SDS, the rate at which biomolecules migrate through the gel is exclusively dependent on size. SDS is a denaturant and anionic detergent that eliminates the factors of net charge and molecular shape (Reynolds and Tsnford, 1970). The treatment of biomolecules with SDS results in linearisation, and overall, a constant charge to mass ratio is established (Reynolds and Tsnford, 1970). SDS is composed of a 12-carbon hydrophobic tail that interacts with the protein, disrupting non-covalent bonds, while the hydrophilic sulfate-based head interacts with the solvent, and imparts the negative net charge (Svasti and Panijpan, 1997).

Smaller molecules will migrate rapidly through the gel, while larger molecules are retarded by the pores and thus migrate more slowly. By running samples in conjunction with a molecular weight marker (composed of protein extracts of known weights), the size of a sample can be estimated based on

the distance of migration within a set period of time (Svasti and Panijpan, 1997. Proteins are visualised on the gel by staining with coomassie brilliant blue dye. Under acidic conditions, the negatively charged sulfonic acid groups associated with the dye interact with positively charged amino acids, predominantly arginine (Bradford, 1976; Chial *et al.*, 1993).

In this investigation, the tricine gel electrophoresis protocol published by Schägger (2006) was followed. In comparison to glycine gel electrophoresis (Laemmli, 1970), this protocol allows for a higher resolution separation and facilitates the visualisation of proteins smaller than 20 kDa (Schägger, 2006). The protocol developed by Laemlli, as well as that designed by Schägger, exploit a discontinuous buffering system to facilitate separation. This means that proteins migrate through two separate gel layers, a stacking layer (pH 6.8) and a separating layer (pH 8.8). The function of the stacking layer is to allow proteins to be stacked into a tight band before separation. This facilitates improved resolution, and complete separation. The pores of the stacking gel are smaller than those associated with the separating gel, and stacking is largely facilitated by the presence of two ions differing in electrophoretic mobility. With reference to the protocol developed by Laemmli, the ions are glycine and chloride. At pH 6.8, chloride is negatively charged, relatively small and migrates quickly through the gel, more rapidly than the protein sample, acting as a leading ion. Glycine exists as a zwitterion at this pH, and thus migrates more slowly than the negatively charged protein sample, thereby acting as a trailing ion. A narrow but steep voltage gradient is therefore created. The protein sample is concentrated within this zone and stacks. When the sample reaches the separating gel, the pH increases resulting in the formation of glycinate anions. The negatively charged glycine molecules migrate rapidly through the separating gel, thereby resulting in the deposition of the sample as sharp band at the resolving layer. This concept applies to tricine gel electrophoresis. The increased resolution in separation of small molecules by tricine gel electrophoresis lies in differences in the pKa, and thus electrophoretic mobility of tricine and glycine. The pKa of tricine is 8,15, while that of glycine is 9,6. This means that the use of tricine as a trailing ion results in a steeper voltage gradient and thus the deposition of a sharper protein band at the resolving layer.

One cannot, however, rely solely on this technique to determine whether the protein of interest has been successfully overexpressed and purified. This technique is indicative only of the size of the protein, and not the primary structure or fold. It is therefore important to employ further characterisation techniques to confirm the identity of the protein.

SDS samples were prepared by mixing 50 µL of each fraction/sample (collected from various purification steps) with 16 µL of reducing sample buffer (150 mM Tris-HCl, 12% w/v SDS, 6% w/v β-mercaptoethanol, 30% w/v glycerol, 0.005% coomassie blue G250, pH 7.5) to chemically denature the proteins. Coomassie G250 was included in the reducing sample buffer to act as a tracking dye. Samples were then boiled at 95 °C for 5 minutes to aid in protein denaturation and reduce degradation (Grabski and Burgess, 2001), before briefly vortexing. This was followed by centrifugation for a couple of seconds at maximum speed to remove insoluble matter within the sample (Grabski and Burgess, 2001), and sonication at 50% for 3 seconds. This was to ensure that samples were completely denatured and combined within each tube. The samples were resolved on a 10% discontinuous gel at room temperature.

The gel was cast using the BioRad miniprotean electrophoresis system, and was prepared using an acrylamide/bis-acrylamide stock of 49.5% T/ 3% C, as well as gel buffer (3M Tris, 1M HCl, 0.3% SDS, pH 8.45). The prepared samples were run alongside the SDS-PAGE ruler low range unstained protein ladder 100-3.4 kDa (Thermo Fisher Scientific, EU), initially at 45 V until the bands had migrated through the stacking gel, and then at 160 V for roughly 90 minutes. The anode buffer was composed of 1 M Tris at pH 8.9 (adjusted with HCl), while the cathode buffer was comprised of 1 M Tris, 1 M tricine, 1% SDS at approximately pH 8.25 (not adjusted with HCl). Following electrophoresis, the gel was stained overnight with coomassie G250, followed by destaining with 10% acetic acid, 50% methanol and 40% distilled water. The gel was analysed using the BioRad Gel Doc viewing system.

2.2.2. UV-vis Absorption Spectroscopy

Absorption spectroscopy is based on the principle that absorbance of light by a particular compound is quantal and is dependent on the specific electronic arrangement of the molecule in its ground state (Burns, 1993). The energy provided, in the form of a specific wavelength of light irradiated onto the substance, should be equal to the amount of energy required for an electron transition to an orbital of higher energy, an antibonding orbital (Burns, 1993). Therefore, specific compounds will absorb predictably at specific wavelengths under defined conditions. The aromatic amino acids, peptide bond, and to a small degree, disulfide bonds, found in proteins absorb light within the UV-vis spectrum (200-700 nm)(Prasad *et al.*, 2017). Tyrosine, and predominantly tryptophan absorb strongly, with an absorption

maximum observed at 280 nm (Wetlaufer, 1962). A second absorbance maximum is generally observed at 205 nm, corresponding to the presence of the peptide bond (Goldfarb et al, 1951)

Absorption spectroscopy is a popular technique exploited in molecular biology as it allows for the concentration of a particular chromophore (absorbing species) in solution to be determined. The concentration of a protein sample in solution can be determined using the Beer-Lambert law (McNaught and Wilkinson, 2006). This principle relates the absorption of light to the properties of the material through which it passes, as well as, in the case of determining protein concentration, primary structure (through the molar absorptivity constant, the extinction coefficient). The Beer-Lambert law describes that the observed absorbance of monochromatic radiation within a homogenous isotropic medium is proportional to the path length, as well as concentration. Or, simply it is expressed as follows:

$$A = C.l.\varepsilon$$

Where **A** denotes the observed absorbance at 280 nm (A/U), **C** represents protein concentration (M); **ε** denotes the extinction coefficient (that can be calculated as represented below)(M^{-1} cm^{-1}) and **l** symbolises the path length (cm).

$$\varepsilon \text{ protein} = (\text{number of tryptophan residues} \times \varepsilon \text{ tryptophan}) + (\text{number of tyrosine residues} \times \varepsilon \text{ tyrosine}) + (\text{number of cysteine residues} \times \varepsilon \text{ cysteine})$$

Where **ε protein** represents the theoretical extinction coefficient of the protein at 280 nm in water (M^{-1} cm^{-1}) when the calculation is performed using the extinction coefficients corresponding to each amino acid experimentally determined under those same conditions (Perkins, 1986).

Since this constant is calculated based on the amino acid composition of a protein, it is representative of an intrinsic property of that protein, and is indicative of how strongly the compound absorbs light at a specific wavelength. In this experiment, the theoretical extinction coefficient for the FOXP2 FHD at 280 nm in water (Gill and von Hippel, 1989) was calculated using EXPASY ProtParam (Gasteiger *et al.*, 2005) and was found to be 22460 M^{-1} cm^{-1}.

Absorbance spectroscopy also allows for the purity of a sample to be assessed. The heterocyclic rings associated with DNA

nucleotides act as a chromophore, exhibiting an absorption maximum at 260 nm (Greenstein, 1944), thus DNA contamination within the sample can be detected, specifically by measuring the ratio of A280/A260. By reading the absorbance of a protein sample at 340 nm, it can be ascertained whether there is a large amount of aggregation present. Measuring wavelengths above 340 nm can be indicative of light scattering (Birdsall *et al.*, 1983).

Fractions containing pure FOXP2 FHD were pooled together and dialysed against HEPES storage buffer (10 mM HEPES, 100 mM NaCl, pH 7.5). The protein was then centrifuged at 13 000 x g for 10 minutes to remove any aggregation before obtaining an absorption spectrum between 250 and 350 nm. A JASCO V630 UV/VIS absorbance spectrophotometer was used for all measurements. Absorbance readings were taken in triplicate at 280 nm, 260 nm and 340 nm using five serial dilutions of protein. After subtracting the readings at 340 nm from those at 280 nm for each sample, linear regression in conjunction with the Beer-Lambert law was used to determine the concentration of the undiluted sample. Subtracting the reading at 340 nm acted as a means by which the spectrum could be 'zeroed' as the sample should be free of aggregates, and thus a signal should not be observed at this wavelength. The A280/A260 ratio was also determined.

2.2.3. FOXP2 FHD Ligand Preparation

Since the focus of this investigation was exploring a potential interaction between the FOXP2 FHD and dopamine, and the potential effect this may have on DNA binding affinity, the ligands used in this study were dopamine and DNA. Ligand stocks were prepared as described for all subsequent experiments as required.

Dopamine stocks were prepared by solubilising dopamine hydrochloride (Sigma-Aldrich, South Africa) in MilliQ water. Owing to the rapid rate of dopamine oxidation in the absence of ascorbic acid (Zhang and Dryhurst, 1993) (Sulzer *et al.*, 2000), stocks were prepared before each experiment to reduce the accumulation of the reactive oxidation intermediates in solution.

The potential interaction between the FOXP2 FHD and dopamine was investigated at two different concentrations throughout this study, one of which was selected to mimic the physiological dopamine concentration within the brain, while the other was representative of excess. The concentration of dopamine *in vivo* is constantly fluctuating, and differs

significantly between different regions of the brain. It has, however, been reported that within the basal ganglia, a region of the brain displaying particularly high dopamine levels, that the dopamine concentration oscillates around 35 µM (Mefford *et al.*, 1980). FOXP2 is also expressed within this region of the brain (Haesler *et al.*, 2007). The effect of dopamine on the structure and function of the FOXP2 FHD at this concentration was therefore investigated for physiological relevance. 35 µM is a relatively high physiological concentration of dopamine, so excess was represented by 100 µM. Although in ligand studies, excess is generally considered 25-100 fold (Motiejunas and Wade, 2007) increasing the concentration of dopamine above around 300 µM was found, in this investigation, to result in rapid formation of neuromelanin within the sample.

It was found that the FOXP2 FHD recognises and binds the sequence 5'-AGGT**GTTTA**CTTTCATAG-3' with high affinity (Nelson, 2013) (Webb *et al.*, 2017). This sequence was used in all investigations surrounding the interaction between the FOXP2 FHD, dopamine and DNA, as well as studies involving only FOXP2 FHD and DNA.

Unlabelled, as well as 5' ROX (6-carboxyl-x-rhodamine) labeled lyophilised DNA was obtained from Integrated DNA Technologies- Whitesci Scientific, South Africa, and was prepared by dilution with milliQ water.

2.2.4. Far-UV Circular Dichroism Spectropolarimetry

Many biomolecules are asymmetric and are thus optically active. Optically active molecules display differences in the extent of absorbance of left and right-handed circularly polarised light. This phenomenon is exploited in CD (Miscsonai *et al.*, 2015).

Owing to the chirality of polypeptides, as well the aromatic amino acids, CD can be used to investigate both the secondary and tertiary structure of a protein. Exposing a protein to monochromatic circularly polarised light between 180 nm and 260 nm (Far-UV) allows for the secondary structure of a protein to be studied, as the major contribution to observed energy (associated with electronic transitions) within this wavelength range is attributed to the peptide backbone. In contrast, the wavelength range between 260 nm and 350 nm (near-UV) can be used to probe the tertiary structure of a protein using the aromatic amino acids as chromophores (Greenfield, 2006).

In this study, CD was used only to investigate the secondary structure of the overexpressed FOXP2 FHD in the presence and

absence of dopamine. Generally speaking, natively folded polypeptide chains contain a regular and repeating sequence of amide groups (Holzwarth and Doty, 1965). These amide groups form various arrangements, dictated by the protein backbone conformation. The geometric relationship between amide groups in various secondary structural elements determines the CD spectrum associated with a specific protein. Regarding α-helices and β-sheets specifically, successive amide groups are orientated identically along the chain. This allows these structural elements to display characteristic CD patterns. Generally speaking, characteristically, in proteins rich in α-helices, a positive band is observed at 192 nm, while negative bands are usually detected at 208 nm, as well as 222 nm groups (Holzwarth and Doty, 1965). A positive band is observed at around 215 nm, and a negative band is characteristic at 198 nm in proteins composed predominantly of β-sheets (Greenfield and Fasman, 1969).

The purified FOXP2 FHD was dialysed against CD buffer (10 mM HEPES, 50 mM Na_2SO_4 pH 7.5). Chloride ions absorb UV light maximally at 180 nm, so buffer exchange was imperative in preventing noise (Miles and Wallace, 2016). Samples were diluted to a concentration of 10 uM with MilliQ water after centrifugation at 13000 x g for 10 minutes. The CD spectrum for the protein was obtained between 190 nm and 260 nm using a JASCO J-1500 with a bandwidth of 1 nm and a scanning speed of 100 nm/s at 20 °C. Samples were prepared in triplicate, and 10 accumulations were measured for each sample. The high-tension voltage (HTV) was observed, and measurements displaying high HTV were discarded. This protocol was then repeated in the presence of 35 µM dopamine, as well as 100 µM dopamine.

The data was normalised by converting the ellipticity measurements, θ (m.deg), representing the difference between left and right-handed circularly polarized light absorption, to mean residue ellipticity (θ_{MRE}) (deg.cm^2.dmol^{-1}) using the following equation:

$$\boldsymbol{\theta_{MRE} = \frac{\theta\ (\mathrm{m.deg}) x\ 100}{c\ (M) x\ n\ x\ l\ (mm)}}$$

Where c refers to the concentration of the protein, n refers to the number of peptide bonds and *l* refers to the path length. In this case, a path length of 2 mm was used.

2.2.5. Intrinsic Tryptophan Fluorescence

Luminescence is broadly defined as the emission of light from a substance as an electron returns to the ground state from an electronically excited state (Lakowicz, 2006). Generally, the excited state follows the absorption of a specific amount of energy. There are two types of luminescence, fluorescence and phosphorescence. The distinction is determined by the nature of the excited state of the substance (Lakowicz, 2006).

Fluorescence occurs when an electron is excited within a substance in the singlet state. In other words, when the electron is excited, it is still paired with the electron in the ground state, in that the two electrons have opposite spin. The return to the ground state, with the release of energy, is therefore rapid (Lakowicz, 2006). Phosphorescence, on the other hand, occurs when an electron is excited from the triplet state (Lakowicz, 2006). In this case, the excited electron has the same spin as the electron in the ground state. Transitions back to the ground state are therefore forbidden, and emissions are thus slow. Phosphorescence generally does not occur in fluid solutions at room temperature, however fluorescence is exploited in protein biochemistry in performing both structural and functional investigations (Lakowicz, 2006). The Jablonski diagram illustrated below provides a more detailed explanation for the phenomenon of fluorescence (figure 2.3)

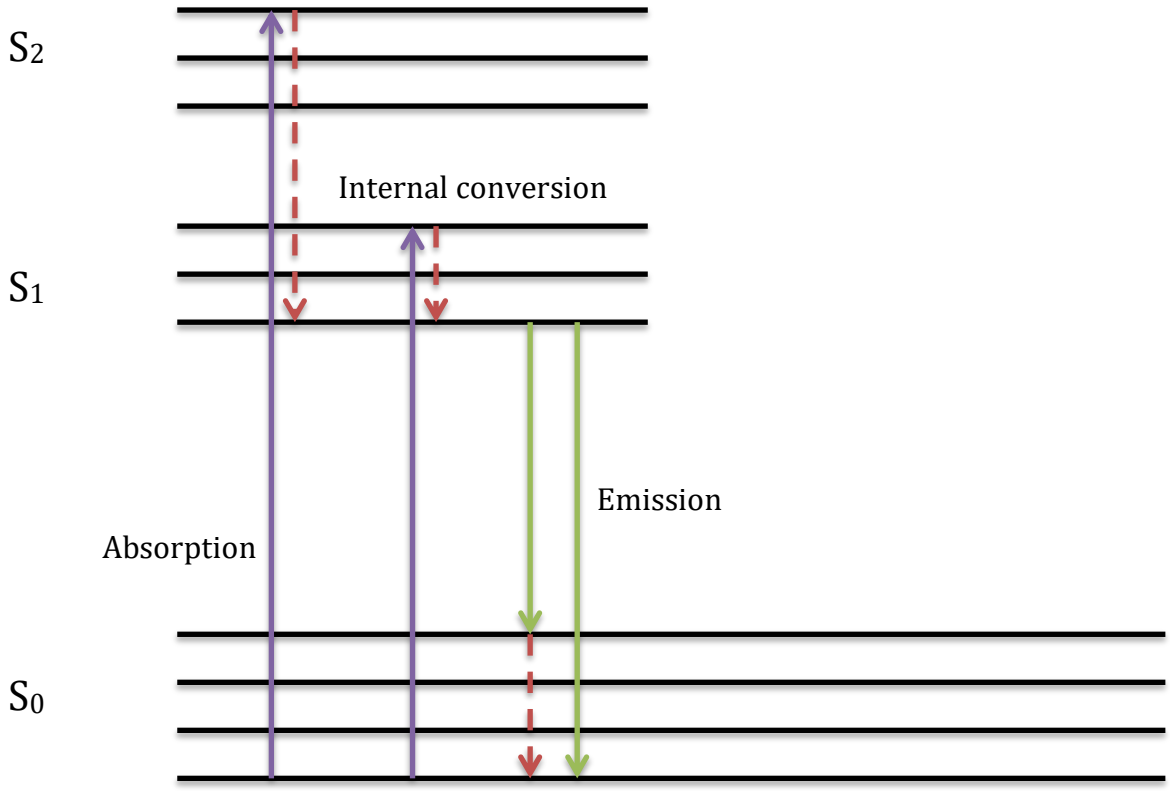


Figure 2.3: A simplified Jablonski diagram representing the singlet excited state. Broadly speaking, a specific amount of energy is absorbed by a fluorophore (represented by the purple arrow). This facilitates the excitation of an electron from the ground state (S_0) to an excited electronic state (S_1 or S_2). As the electron falls back to the ground state, energy is emitted in the form of fluorescence (represented by the green arrow). It is evident that the energy of emission is less than the energy initially absorbed by the fluorophore. This can be somewhat explained by internal conversion (represented by the red dotted arrow), as well as loss of energy to the solvent. Following excitation, the electron relaxes to the lowest vibrational level of the excited electronic state, resulting in the loss of energy. This means that emission occurs at thermal equilibrium. Energy can also be lost if the electron returns to a higher vibrational level within the ground state following emission, and thermal equilibrium is once again reached due to the loss of heat energy.

Typically aromatic compounds display fluorescence due to the presence of an even number of electrons in the π orbital, thereby allowing for electron pairing (Lakowicz, 2006). With regard to proteins, the aromatic residues, predominantly tyrosine and tryptophan, display fluorescent properties. Since both tyrosine and tryptophan are excited at 280 nm, this wavelength is often used in protein studies. The emission spectrum is however dominated by tryptophan. This is partially due to energy transfer from the tyrosine to tryptophan residues within the protein. Tryptophan, however, can be exclusively excited at 295 nm (Lakowicz, 2006). The emission spectrum for tryptophan can be observed between around 310 and 350 nm, depending on the polarity of the surrounding microenvironment (Lakowicz, 2006).

The sensitivity of tryptophan, specifically the indole group, to the polarity of the surrounding microenvironment can be exploited in probing the tertiary structure of a protein (Lakowicz, 2006). When a tryptophan residue is exposed to a polar environment, the residue is likely solvent exposed, and the emission maximum is observed at longer wavelengths. In other words, a bathochromic, or red shift occurs. When the residue is buried within the protein, and the microenvironment is usually less polar, a blue or hypsochromic shift is observed (Lakowicz, 2006). This phenomenon can be partially explained by stabilisation of the excited state by a polar environment. The excited state of a fluorophore has a larger dipole moment than the ground state. Polar solvent molecules reorientate and relax around the excited fluorophore, resulting in stabilisation and an overall lowering of the energy of the excited state (Lakowicz, 2006). This results in a lower energy fluorescence emission.

Samples were prepared in triplicate containing 2 µM FOXP2 FHD in HEPES storage buffer. A JASCO FP-6300 spectrofluorometer was used to excite samples at both 280 nm, as well as 295 nm with a bandwidth of 5 nm and 2.5 nm respectively, at 20 °C. The emission spectra were read between 280 nm and 450 nm with a scanning speed of 500 nm/min. This protocol was then repeated in the presence of 35 µM dopamine, as well as 100 µM dopamine.

2.3. Investigating the FOXP2 FHD-Dopamine-DNA interaction

2.3.1. Electrophoretic Mobility Shift Assay (EMSA)

The EMSA is a technique that exploits the use of native gel electrophoresis to investigate either a protein, or sequence dependent, specific binding interaction. The rate at which non-denatured biomolecules migrate through the gel matrix is dependent on the electrophoretic mobility, and thus is partially dependent on the molecular weight and shape of the molecule. Since unbound DNA has a higher electrophoretic mobility than the protein-DNA complex, fragments of free DNA are readily separated from those that are associated with protein (Garner and Revzin, 1981).

If binding occurs, the overall distance travelled by the DNA-protein complex is reduced in comparison to the free DNA. This is referred to as an electrophoretic shift. An increase in the number or size of molecules bound to free DNA will result in a greater shift relative to free DNA. Since the FOXP2 FHD is capable of independent binding to the cognate DNA binding

sequence, the Nelson sequence, this assay can be used to confirm that the overexpressed protein is natively folded and functional. In the presence of varying concentrations of dopamine, this assay can provide an indication as to whether the presence of this signalling molecule significantly affects the ability of the FOXP2 FHD to interact with its high affinity binding DNA target.

This technique however does not provide high-resolution data, due to rapid dissociation of the protein-DNA complex within the gel (as the complex is not at chemical equilibrium within the gel), as well as band smearing (Fried and Lui, 1994, Fried and Bromberg, 1997). Although it is possible to determine kinetic parameters such as the dissociation constant (K_D) using the EMSA, other techniques, such as ITC and fluorescence anisotropy, are more accurate and are thus favoured. This technique is predominantly used to indicate whether or not complex formation occurs.

Samples were prepared after centrifugation (at 13000 x g for 10 minutes at 4 °C) of the purified FOXP2 FHD (in HEPES storage buffer) to remove aggregates. The protein was then combined at increasing concentrations (between 0 µM and 25 µM) with 0.5 µM Nelson DNA, as well as binding buffer (10 mM HEPES, 100 mM KCl, 1 mM EDTA, 1 mM $MgCl_2$, 0.1 mg/mL BSA (bovine serum albumin), 10% v/v glycerol, pH 7.5 in MiliQ water). Samples were placed at 4 °C on ice for 30 minutes, thereby allowing binding to occur and equilibrium to be reached. TBE running buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3), and a 10% polyacrylamide gel (containing 20% triethylene glycol and prepared with TBE buffer) were also cooled to 4 °C. Electrophoresis was performed for 90 minutes at 120 V at 4 °C. The gel was then stained with a 1:10000 dilution of SYBR® Gold (Invitrogen, USA) in TBE buffer for 5 minutes, before visualisation using the BioRad Gel Doc viewing system.

After determining the ratio of DNA:protein at which complex formation occurs, the assay was repeated using a constant ratio of 1:5 DNA:Protein in each lane. Increasing concentrations of dopamine (between 0 µM and 200 µM), were included in each lane, alongside the free DNA.

2.3.2. Fluorescence Anisotropy

Within an isotropic solution, fluorophores in the ground state are randomly orientated, and preferentially absorb polarised light within a specific plane (Lakowicz, 2006). The orientation

of the excited population is therefore not random, as maximum absorption occurs when the axis of the dipole moment of the fluorophore lies parallel to the electric vector of the incoming excitation radiation (Lakowicz, 2006).

Emissions following excitation are also polarised; however the excitation and emission dipoles are not parallel. There is, therefore, partial depolarisation of emission relative to the excitation plane. This is referred to as intrinsic depolarization (Lakowicz, 2006). Since molecules are sensitive to rotational diffusion, further depolarisation can occur due to rotation of the molecule within the lifetime of the excited state. This is referred to as extrinsic depolarisation. Smaller molecules tend to have increased mobility, and thus generally reorientate more rapidly than larger molecules within the lifetime of the excited state (Lakowicz, 2006).

The extent of depolarisation, or the relative angle between the absorbed and emitted polarised radiation, is defined in terms of anisotropy, and can be described using the following equation:

$$A=(I_{II}-I_{\perp})/(I_{II}+2GI_{\perp})$$

Where A represents anisotropy, I_{II} and $I_{\perp}$ represent the observed emission intensity parallel and perpendicular to the excitation radiation, respectively. The expression $(I_{II}+2GI_{\perp})$ describes the total intensity, and thus anisotropy is normalised, and independent of the total fluorescence intensity of the sample (Lakowicz, 2006). G describes the G factor of the instrument. It is therefore evident that A is high when the degree of depolarisation is low.

Anisotropy can be used as a basis for studying the formation of macromolecular complexes in solution, specifically, in this case, the association between a transcription factor and DNA. Although this technique does not provide insights regarding the intrinsic structural changes that occur upon association, the binding parameters such as the K_D can be obtained.

When performing an anisotropy experiment, the oligonucleotide is usually labeled with a fluorescent tag. When a protein-DNA complex forms, there is an increase in the overall molecular weight and specific volume of the fluorescently labeled molecule. Since molecular size is directly related to rotational correlation time, a change in anisotropy can be observed. An increase in anisotropy is associated with an increase in molecular size, so the maximum anisotropy can be

observed when all molecules in solution are bound (Gijsvers, Nishigaki and Sanchez-Puig, 2016).

Samples were prepared in triplicate by incubating increasing concentrations of FOXP2 FHD (final concentrations ranging from 0 uM to 5 uM protein) in HEPES storage buffer with 2 nM 6-carboxyl-x-rhodamine (ROX) 5'- labeled Nelson DNA (Integrated DNA technologies-Whitesci Scientific, South Africa). Before use, the purified FOXP2 FHD was centrifuged at 13000 x g for 10 minutes.

The anisotropy of each sample was measured in triplicate using the PerkinElmer LS-50B spectrofluorometer fitted with an anisotropy filter, with the excitation and emission wavelengths set to 580 nm and 605 nm respectively. The bandwidth for both excitation and emission was set to 5 nm, and the integration time was 3 seconds. Ten accumulations of each sample were taken. The relative anisotropy for each sample was then plotted against protein concentration, and fitted using SigmaPlot 12.0 to a single site binding model. This was repeated with samples that included either 35 uM or 100 uM dopamine. The significance of differences in K_D were analysed and compared using the student t-test, performed using SigmaPlot 12.0.

2.4. Molecular Docking

Molecular docking is a computational method by which the predominant binding model between a ligand and protein with a known 3D structure can be predicted (McConkey *et al.*, 2002). Generally, this technique is exploited in atomic level investigations to predict the behavior of a small molecule within the binding site of a protein. Molecular docking can be used to perform virtual screenings of large libraries of ligands or to propose structural hypothesises regarding how ligands bind or inhibit a specific target (Moitessier *et al.*, 2008).

Molecular docking is achieved through two interrelated steps. The first is prediction of ligand position, orientation and conformation relative to the protein of interest, while the second is prediction of the binding affinity, thereby allowing predicted positions to be ranked against each other. These steps are achieved through various sampling and scoring algorithms (Meng *et al.*, 2011).

Since there are 6 degrees of rotational and translational freedom, as well as the conformational degrees of freedom associated with both ligand and protein (Meng *et al.*, 2011), there are a large number of possible binding modes. It would

be too computationally expensive to generate every possible conformation associated with a protein-ligand complex, thus various sampling algorithms are widely used in molecular docking such as matching algorithms and incremental construction based algorithms (Meng *et al.*, 2011) .

The scoring function of molecular docking programs facilitate the calculation of binding affinity based on a number of estimates, simplifications and assumptions. Scoring functions are based on the force field, a defined set of parameters and functions that allow for the prediction of energies associated with specific coordinates of the system (Kitchen *et al.*, 2005). Different programs utilise different force fields, namely classical, empirical or knowledge based force fields. Classical force fields estimate the binding energy by calculating the sum of electrostatic forces and van der Waals interactions (Kollman, 1993). The van der Waals forces are defined by the Lennard-Jones potential function, while the contribution of charge-charge based interactions is calculated by the distance dependent dielectric function. Empirical force fields are calculated based on contributions from hydrogen bonding, ionic and hydrophobic interactions, and entropy (Bohm, 1998). Each of these components is multiplied by a specific coefficient to create a weighting. These components are calculated by using regression analysis from ligand-protein complexes with known binding affinity. Knowledge based force fields exploit statistical analysis of known protein-ligand complexes (Feher et a.l., 2003). Atomic contact frequencies are obtained, and more frequent interactions are taken to be more favorable. The number of favorable and repulsive interactions dictates the final score. Some molecular docking programs use consensus scoring. Consensus scoring combines a variety of scores obtained from different force field algorithms, thereby improving the overall prediction (Gohlke *et al.*, 2000).

In this study, two different molecular docking programs were used to investigate the potential for an interaction between either monomeric or dimeric FOXP2 and dopamine, 5,6-dihydroxyindole, 5,6-indolequinone or dopaminechrome, namely PyRx Virtual Screening Software (Dallakyan and Olson, 2015) and Maestro 11,2, Schrodinger (Schrodinger Suite, LCC, NY, 2017).

With reference to both the blind docking, and site targeted docking investigations, the 3D structures for dopamine (PubChem identifier: CT10005), 5,6-dihydroxyindole (PubChem identifier: 114683), 5,6-indolequinone (PubChem identifier: 440728) and dopaminechrome (PubChem identifier: 170262) were retrieved in SDF format. The crystal structure for the

FOXP2 FHD was retrieved off PDB (2A07)(Stroud *et al.*, 2006). The structure was investigated using Swiss-PDB viewer, and the monomeric and dimeric structures were separated and saved separately in the absence of DNA.

Pyrx (Dallakyan and Olson, 2015) was used to perform blind docking studies. This virtual screening tool runs multiple open source programs to facilitate docking studies. Autodock4 (Morris *et al.*, 2009) and AutoDock Vina (Trott and Olson, 2009) were used as the docking software, while AutoDockTools (Morris *et al.*, 2009) was used to generate the input files. Python is the programming language, and wxPython facilitates cross-platform graphical user interface (GUI). Opal Toolkit allows docking studies to be performed remotely through web services, while matplotlib (Hunter, 2007) is used for 2D plotting. Open Babel is used to import SDF files, remove salts and for energy minimization. The monomer structure was uploaded in PDB format to AutoDock after the addition of hydrogen molecules, energy minimisation and preparation using AutoDockTools. The four ligand SDF files were imported to AutoDock using OpenBabel. After selecting the universal force field (UFF) (Rappe *et al.*, 1992) and conjugate gradient method, energy minimisation of the ligands was performed. The entire protein surface was selected as the search space, and the docking wizard was used to facilitate the scoring process with reference to each ligand. The poses displaying the most negative scores with respect to each ligand were selected and visualised using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC. This process was repeated with dimeric FOXP2 FHD.

With regard to the site targeted docking studies, Maestro 11.2, Schrödinger (*Schrödinger Suite, LCC, NY, 2017)* was used. The crystal structure of monomeric FOXP2 FHD was imported in PDB format and the Protein-Prep Wizard (Schrodinger Release 2017-3, Sastry *et al.*, 2013) was used for energy minimisation. The OPLS3 force field (Harder *et al.*, 2016) was selected before importing the dopamine structure in SDF format. The LigPrep (Schrodinger Release 2017-3, Farid *et al.*, 2006) tool was used for energy minimisation of the ligand. SiteMap (Schrodinger Release 2017-3,Halgren *et al.*, 2009) was used to identify potential binding pockets within the protein. The identified potential binding pocket was selected as the search space and the InducedFit (Schrodinger Release 2017-3, Farid *et al.*, 2006) docking tool was used to begin the scoring process, thereby generating 20 different poses. The best pose generated by the program was selected and visualised using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC. This process was then repeated with each of the dopamine

autoxidation ligands, and then with the structure of dimeric FOXP2 FHD.

3. RESULTS

Following successful overexpression and purification of the FOXP2 FHD, the recombinant protein was structurally characterised by spectroscopic methods, specifically UV-Vis absorbance, CD and fluorescence spectroscopy. The FOXP2 FHD was then functionally characterised through the EMSA, before investigating whether the presence of dopamine affected the structure or DNA binding function of the protein. CD and intrinsic tryptophan fluorescence were exploited in observing whether the presence of 35 μM or 100 μM dopamine significantly affected the secondary or tertiary structure of the protein. The EMSA, as well as fluorescence anisotropy were utilised in investigating whether the presence of dopamine significantly affected the binding affinity between the FOXP2 FHD and a high affinity binding target sequence. Molecular docking was then used to investigate the potential for an interaction between the FOXP2 FHD and dopamine, as well as various products of dopamine autoxidation.

3.1. Overexpression and Purification

Recombinant FOXP2 FHD was overexpressed within T7 *E. coli* cells, transformed with the pET-11a vector. The soluble fraction was collected through centrifugation. The incorporation of an N-terminus Histag within the FOXP2 FHD construct allowed for IMAC to be used in the first step of purification. The elution profile is illustrated in figure 3.1(A).

Fractions containing purified overexpressed WT FOXP2 FHD were concentrated before performing SEC. SEC was used as a polishing purification step, allowing for any contamination (due to non-specific electrostatic interactions between bacterial proteins and the IMAC column) to be separated from the recombinant protein. The elution profile for the collected fractions, as well as standards is illustrated in figure 3.1(D) and (B) respectively. The retention time of the standards was used to construct a standard curve, thereby allowing the molecular weight of the FOXP2 FHD eluent to be calculated, as illustrated in figure 3.1(C).

Fractions collected from both IMAC and SEC were run on a 10% tricine SDS-PAGE gel (figure 3.1(E)), such that the fractions containing recombinant protein could be identified and purity of the samples could be assessed. The molecular weight of the overexpressed protein was calculated by plotting a standard curve (figure 3.1(F)).

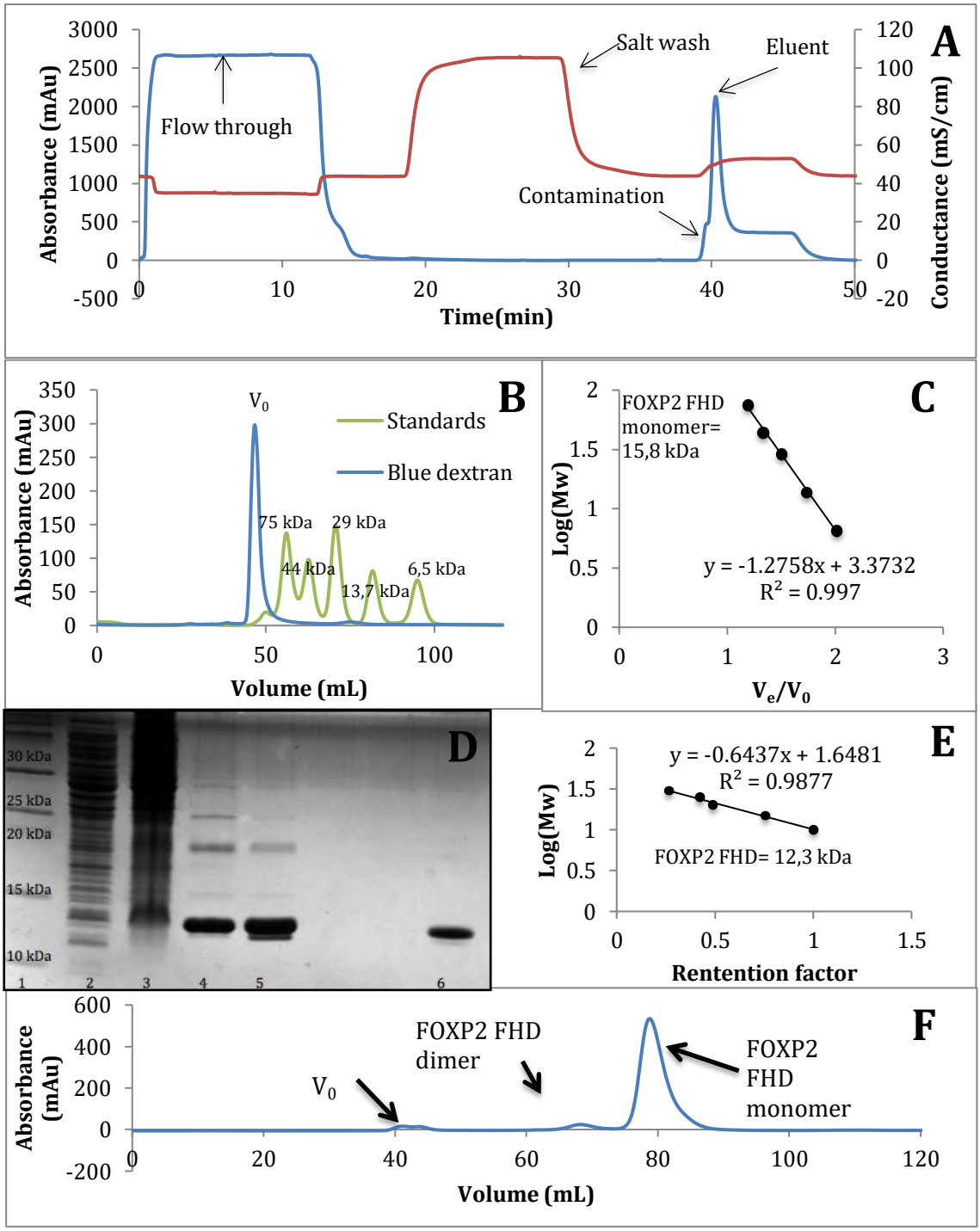


Figure 3.1: Purification of the FOXP2 FHD **(A)** IMAC purification of FOXP2 FHD **(B)** SEC elution of blue dextran and standards (Mw of standards indicated in kDa) **(C)** Standard curve representing the linear relationship between the Mw of standards and the V_e/V_0 **(D)** Preparative SEC of FOXP2 FHD **(E)** Tricine SDS-PAGE of FOXP2 FHD purification fractions. Lane 1 contains the molecular weight marker, lane 2 contains the pellet and lane 3 contains the IMAC flow through. Lane 4 represents the IMAC contamination shoulder, lane 5 contains the FOXP2 fraction collected from IMAC and lane 6 contains the purified fraction collected from SEC. **(F)** Standard curve illustrating the relationship between retention factor on the SDS-PAGE gel and log (Mw).

With regard to the IMAC purification profile (figure 3.1(A)), there is a broad peak in absorbance (blue) between 0 and 15 minutes. This is representative of the elution of flow-through, including the majority of bacterial cellular proteins and contaminants (figure 3.1(D), lane 3). The second, sharper, peak in absorbance is representative of elution of the tagged FOXP2 FHD with 300 mM imidazole (figure 3.1(D), lane 5). The small, broad shoulder following the elution of the FOXP2 FHD is descriptive of the high imidazole buffer interacting with the detector. The eluent peak has a small shoulder that eluted at around 40 minutes. This shoulder contains contaminants, as well as the purified FOXP2 FHD (figure 3.1(D), lane 4) The broad peak in conductance (red) between 20 and 30 minutes is representative of a salt wash with 1 M NaCl. The salt wash aids in disrupting electrostatic interactions between the tagged recombinant TF and DNA, as well as aids in removing contaminants interacting non-specifically with the IMAC column. Inspection of figure 3.1 (D) suggests that although the recombinant protein was found to be relatively pure following IMAC, a few faint contaminating bands can be observed within lane 5. The most significant contaminating band appears to be around 18 kDa.

SEC was performed to remove the faint contaminating bands observed in figure 3.1 (D), lane 5 (specifically that observed at roughly 20 kDa). It is evident that this polishing step was successful. Only a single band is present in lane 6, at a Mw corresponding to that of the FOXP2 FHD, thereby suggesting the sample is free from contaminating proteins.

The molecular weights for peaks eluted from the SEC column (figure 3.1 (F)) were calculated using standards (figure 3.1 (B)), and constructing a standard curve (figure 3.1 (C)). Blue dextran (figure 3.1 (B)) eluted at 46,6 mL. This is indicative of the void volume of the column. The standards included conalbumin (75 kDa, eluted at 55,6 mL), ovalbumin (44 kDa, eluted at 62 mL), carbonic anhydrase (29 kDa, eluted at 70,2 mL), ribonuclease A (13.7 kDa, eluted at 81 mL) and aprotinin (6.5 kDa, eluted at 94 mL). As illustrated in figure 3.1 (E), the molecular weights of the three peaks were calculated using the equation $y=1,2758x+3,3732$ ($R^2=0,97$). The first peak, observed between 42 and 45 mL. Based on observations of the elution of blue dextran, it was hypothesised that this peak corresponded to elution of the void volume. The sephadex S75 can resolve proteins between the size of 3 and 70 kDa (GE Healthcare Life Sciences). It is therefore likely that this peak corresponded to the elution of contaminants within the IMAC sample larger than 70 kDa. Contaminants that are larger than 70 kDa, with regard to this column, will elute within the void volume, as these

molecules will not be retained within the resin, but rather flow straight through the column. The second peak at 66,5 mL was calculated to correspond to the Mw 31,6 kDa, while the peak eluted at 78,6 mL was calculated to represent the elution of a molecule of 15,8 kDa. In combination with the results observed from SDS-PAGE (figure 3.1 E), the peak eluting at 78,6 mL was found to contain the FOXP2 FHD. Since the peak eluting at 66,5 mL was found to be double the molecular weight of that eluted at 78,6 mL, it is likely that this is representative of elution of dimeric FOXP2 FHD. It has been reported that overexpressed FOXP2 FHD remains monomeric at concentrations below 300 µM (Morris and Fanucchi, 2016). The elution of a dimer peak on the SEC profile is likely due to concentrating the protein within a small volume of buffer (2 mL). In performing successful preparative SEC, preparing a small concentrated sample is key for good resolution. It can therefore be assumed that when performing subsequent experiments at significantly lower concentrations, only monomeric protein will be present. Evidence for the presence of dimeric FOXP2 FHD would not be present on the SDS-PAGE gel as this technique is performed under denaturing conditions, and thus would result in loss of quaternary structure.

Figure 3.1(D) suggests that the recombinant overexpression and purification of the FOXP2 FHD was successful. This is evident in that thick dark bands, free from contamination, were observed on the tricine gel at a Mw corresponding to that of the FOXP2 FHD. The Mw of the overexpressed band was calculated, based on the standard curve (figure 3.1 (E)) to be 12,3 kDa. This was calculated using the equation y=-0,6437+1,6481 (R^2 =0,99). The calculated Mw is very close to the expected size of the FOXP2 FHD, 12, 97 kDa (as predicted by Expasy ProtParam) (Gasteiger *et al.*, 2005).

Although the calculated Mw of 15,8 is larger than the predicted Mw of the FOXP2 FHD (12,97 kDa)(Gasteiger *et al.*, 2005) it is evident that the blue dextran, and thus void volume, eluted 2-3 minutes later in figure 3.1 (B) compared to that in figure 3.1 (F). This would mean that the calculated size of the FOXP2 FHD might be overestimated.

Overall, with reference to figure 3.1, it is evident that the overexpressed protein corresponded to the predicted size of the FOXP2 FHD. Although the size, and potentially identity of the overexpressed product was confirmed through SDS-PAGE, the observations could not confirm that the protein was natively folded. Aggregated, misfolded or denatured protein would have the same molecular weight as the folded product visualised using SDS-PAGE. A spectroscopic investigation, as

well as confirming the functionality of the protein was therefore required before assessing the effect of dopamine on the structure or function of the FOXP2 FHD.

3.2. Spectroscopic Structural Investigation and Functional Characterisation

Following centrifugation of the purified sample to remove aggregation, the protein sample was assessed in terms of purity and concentration using absorbance spectroscopy, such that subsequent experiments could be reliably performed. The absorbance spectrum was measured between 250 and 350 nm (figure 3.2 (A)). The concentration of the overexpressed protein was determined using the Beer-Lambert law, and the absorbance at 260 nm was observed in order to ensure that the protein stock was free from DNA contamination, while the observation of the absorbance at 340 nm provided insights regarding the presence of aggregates in the stock.

When investigating the native fold of the protein, the primary, secondary, tertiary and quaternary structure should be considered (Rehman and Botelho, 2019). Investigating the secondary and tertiary structure of the recombinant product will both reinforce the idea that the correct protein has been overexpressed and indicate that the protein is soluble and exists within its native conformation. The integrity of the secondary structure was investigated by measuring the CD spectrum of the purified FOXP2 FHD between 200 nm and 250 nm (figure 3.2 (B)), while the tertiary structure of the overexpressed product was investigated using intrinsic tryptophan fluorescence. Excitation was performed at 280 nm, and emission was read between 295 and 450 nm. For comparison, the overexpressed product was thermally unfolded, and the fluorescence spectrum was read between 290 and 450 nm with excitation at 280 nm (figure 3.2 (C)).

The activity of the overexpressed product was then investigated. Since structure is directly related to function, the observation that a protein is fully functional can be indicative of native fold. Functional characterisation also further reinforces the identification of the overexpressed product, since proteins have specific physiological functions. The specific physiological function of transcription factors is dependent on the ability to bind DNA, specifically cognate target sequences. Since the FOXP2 FHD has been reported to bind the Nelson sequence with high affinity (Nelson *et al.*, 2013, Webb *et al.*, 2017), an EMSA was used to functionally characterise the overexpressed product (Figure 3.2 (D)). The

ratio of DNA:protein at which binding occurs was also determined.

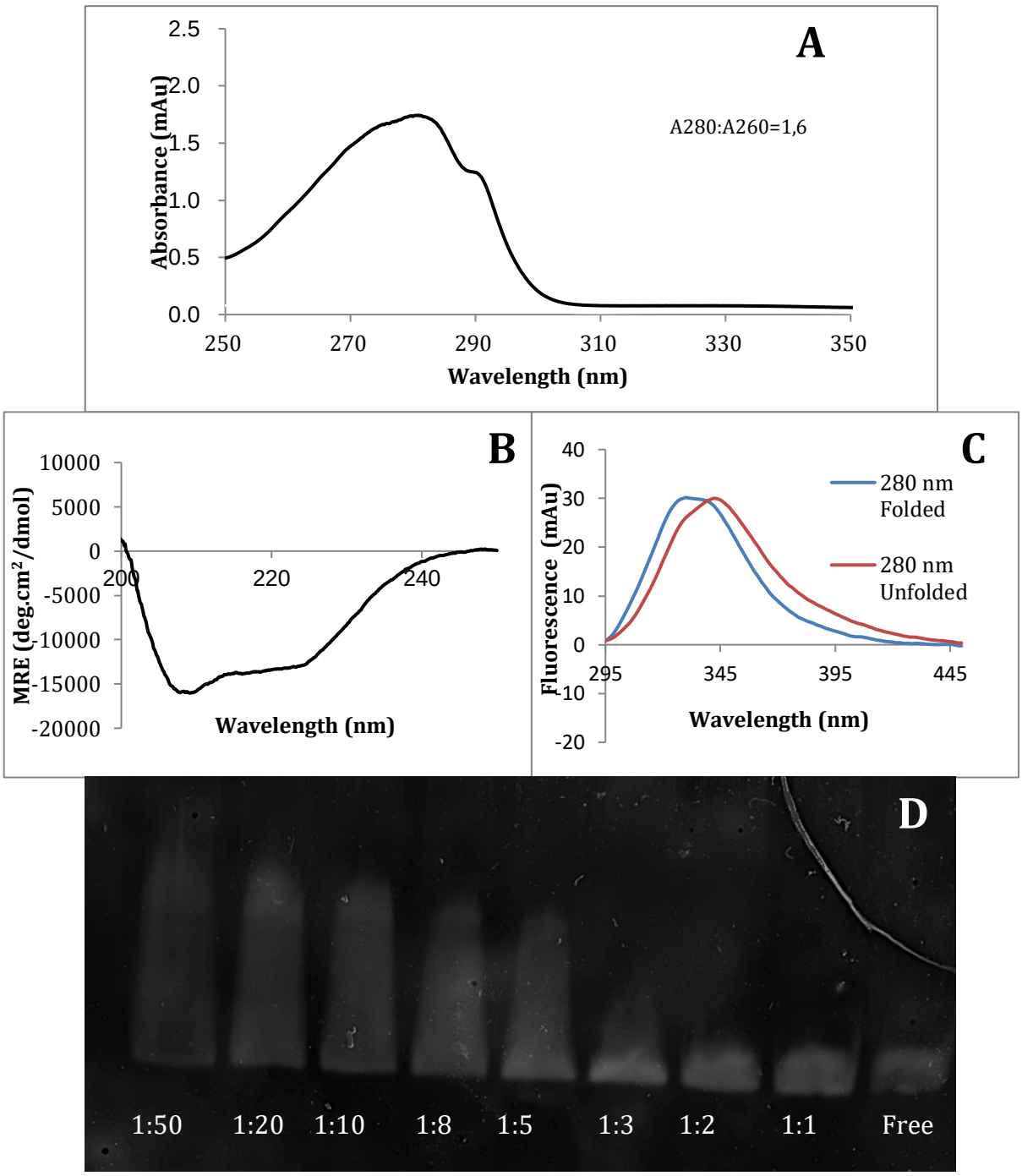


Figure 3.2: Structural and Functional Characterisation of the FOXP2 FHD (A) UV-Vis Absorbance spectrum of the FOXP2 FHD **(B)** CD spectrum of 10 µM FOXP2 FHD at 20 °C. The HTV was found to be 451 V at 200 nm (supplementary figure 6.1). Observations below this wavelength were disregarded **(C)** Intrinsic fluorescence spectrum of 2 µM folded and thermally unfolded FOXP2 FHD with excitation at 280 nm **(D)** EMSA illustrating complex formation at increasing ratios of protein:DNA with a DNA concentration of 0,5 µM. The reduction in electrophoretic mobility at DNA:protein ratios from 1:5 indicates DNA binding.

Figure 3.2 (A) illustrates a broad peak at 280 nm, as well as a shoulder at 295 nm. The peak at 280 nm is representative of the absorption by the aromatic residues, predominantly tyrosine and tryptophan, while the shoulder is exclusively a result of absorbance by tryptophan. The average A280: 260 was found to be 1,65. This suggests negligible DNA contamination. Due to the scattering of light, absorbance at 340 nm is generally associated with aggregation. In this case, the lack of a visible absorbance signal at 340 nm suggests that the stock is free of aggregates.

The secondary structure of a protein arises due to the interactions between atoms of the polypeptide backbone, thereby forming characteristic folds and coils. These folds and coils, beta sheets and alpha helices respectively, contribute to the overall characteristic shape of the protein (Reham and Botelho, 2019). Characterisation of the secondary structure of an overexpressed protein is therefore important in determining whether the molecule exists in its native conformation. Although CD observations below 200 nm are beneficial in investigating the secondary structure (for example, with regard to alpha helical proteins, a peak is generally observed at 192 nm, Correa *et al.*, 2009), certain cuvettes, as well as buffer components, absorb circularly polarised light at wavelengths below 200 nm, thus causing noisy and unreliable observations. Based on the crystal structure solved by Stroud *et al.* in 2006, the FOXP2 FHD is predominantly alpha helical in structure. It is therefore expected that a trough be observed at both 208 and 222 nm (Correa *et al.*, 2009). Figure 3.2 (B) illustrates the presence of a clear trough at 208 nm, as well as at 222 nm. This is indicative of natively folded FOXP2 FHD. The observed spectrum is in agreement with both the crystal structure, as well as other published work illustrating the secondary structure of this protein (Blane and Fanucchi, 2015, Morris and Fanucchi, 2016).

The tertiary structure is dependent on interactions between side chains of amino acids. The chemical and physical properties of amino acid side chains essentially dictate the tertiary fold of the protein, such that non-polar residues cluster at the core of the protein via hydrophobic interactions, buried from the hydrophilic environment (Cooper, 2000). Hydrogen bonds, ionic interactions and van der Waals forces between polar side chains on the surface of the protein contribute to facilitating the formation of a unique shape, and thereby, contribute to the function of a specific protein (Cooper, 2000). Excitation at a wavelength of 280 nm excites both tyrosine and tryptophan residues, while excitation at 295 nm excites exclusively tryptophan (Lakowicz, 2006). As evident

in figure 3.2 (C), fluorescence of natively folded FOXP2 FHD with excitation at 280 nm resulted in a broad emission peak between roughly 320 and 345 nm. The emission spectrum corresponding to the thermally unfolded protein displayed a red shift, and an overall change in the shape of the peak.

It is evidently essential to investigate the spatial arrangement, and distribution of tryptophan residues within the protein of interest when exploiting intrinsic fluorescence in probing protein tertiary structure (figure 3.3).

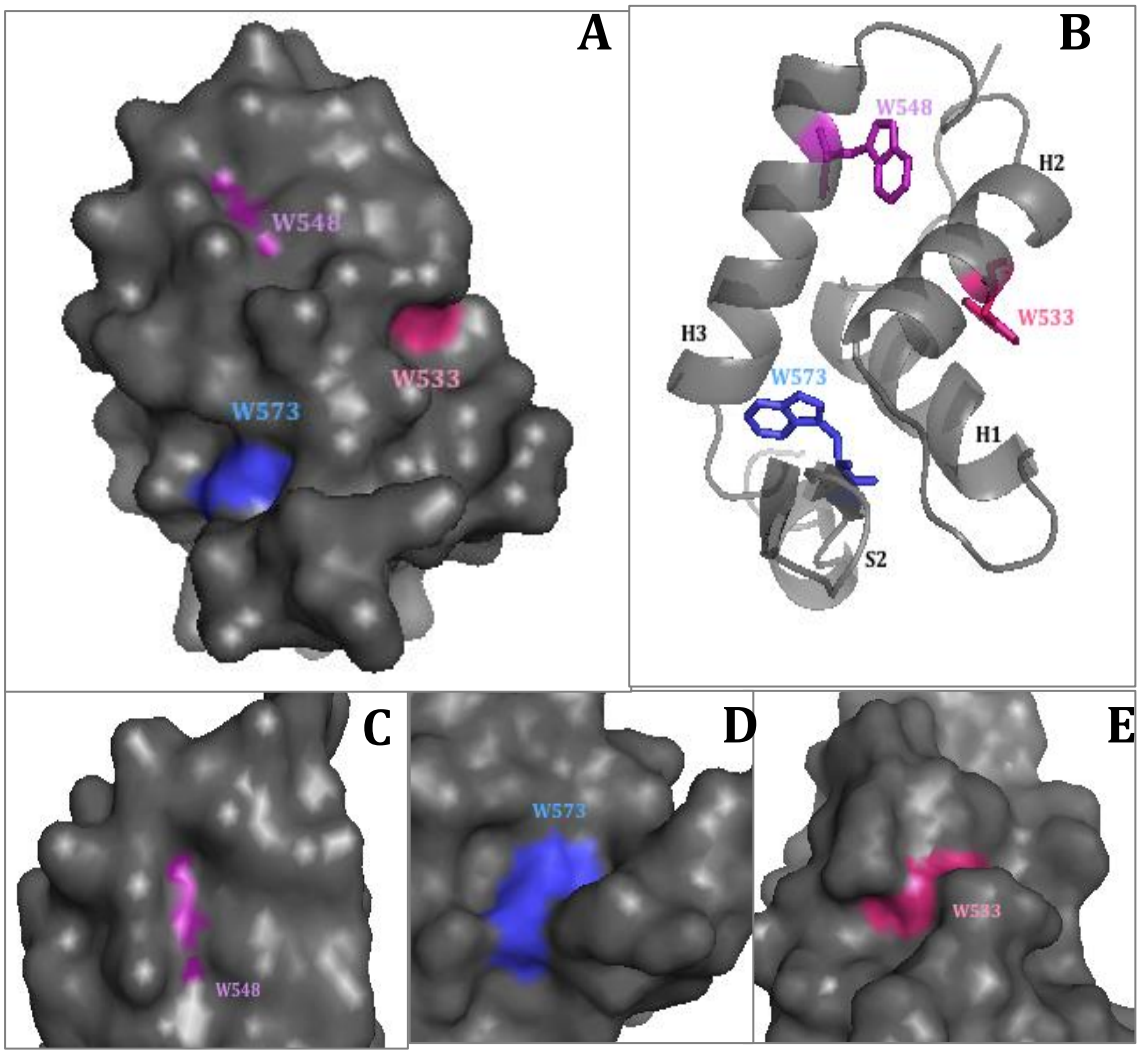


Figure 3.3 Molecular Surface of FOXP2 FHD Highlighting Location and Solvent Exposure of Tryptophan Residues (PDB: 2A07) (A) Molecular surface of FOXP2 FHD highlighting the location of tryptophan residues relative to each other **(B)** Ribbon structure of FOXP2 FHD highlighting location of tryptophan residues relative to protein structure **(C)** Molecular surface of FOXP2 FHD highlighting the solvent exposure of Trp548 **(D)** Molecular surface of FOXP2 FHD highlighting the solvent exposure of Trp573 **(E)** Molecular surface area of FOXP2 FHD highlighting the solvent exposure of Trp533

As illustrated in figure 3.3 (A), FOXP2 FHD has 3 tryptophan residues (Trp553, Trp573 and Trp548), located within H2, H3 and S3 (Stroud *et al.*, 2006). Inspection of figure 3.3 overall suggests that W548 is less solvent exposed than W574 and W533. When analysing the fluorescence emission spectra in figure 3.2 (C) and (D), with this observation in mind, it is evident that the overexpressed FOXP2 FHD is both natively folded and has been correctly identified. The emission maximum corresponding to the folded FOXP2 FHD excited at both 280 nm, as well as 295 nm, illustrates both a peak and shoulder. This is characteristic of the FOXP2 FHD and is in

agreement with other published studies (Blane and Fanucchi, 2015, Morris and Fanucchi, 2016). Since tryptophan fluorescence emissions are sensitive to the surrounding microenvironment, and red shifts are generally associated with increased solvent exposure, a difference in emission maximum corresponding to tryptophan residues with varying solvent exposure is expected (Lakowicz, 2006). This is further supported by the spectra corresponding to thermally unfolded FOXP2 FHD. Unfolding of the protein results in increased solvent exposure due to destabilisation of the hydrophobic core, and thus a red shift in emission is observed. The overall shape of the characteristic FOXP2 FHD emission maximum is also abolished in unfolding, likely due to changes in relative tryptophan solvent exposure.

In combination with spectroscopic observations, the EMSA, observed in figure 3.2 (D), suggests that not only was the protein natively folded and free of aggregation, but also capable of binding the high affinity target sequence, the Nelson sequence. Since DNA is labeled with Sybr gold, an electrophoretic shift, indicated by the presence of a faint second band, or smear, within each lane from a minimum DNA: protein ratio of 1:5 is indicative of complex formation. As protein and DNA interact, the overall molecular weight of the complex increases, and thus the complex is retarded by the gel matrix and migrates slower than the free DNA. The observation that complex formation occurs at various protein concentrations again confirms that the FOXP2 FHD is correctly folded, since correct fold is required for protein function.

After confirming the structural and function integrity of the protein, the direct interaction between the FOXP2 FHD and dopamine could be investigated. The structural and functional characterisation ensured that any significant observations made in subsequent experiments were the result of the presence of dopamine, rather than misfolded or impure protein stocks. Investigating any potential structural changes to the FOXP2 FHD that occur in the presence of dopamine could be indicative of an interaction, due to the protein conformational changes that may be induced upon binding. Since structure is directly related to function, structural changes that occur upon exposure to dopamine could suggest the potential for the function of FOXP2, specifically DNA binding, to be altered.

3.3. Investigating the FOXP2 FHD-Dopamine interaction

The direct effect of dopamine on the FOXP2 FHD was investigated by examining whether the secondary or tertiary structure of the protein was significantly altered in the

presence of the neurotransmitter. This was performed at both a physiologically relevant concentration (35 μM), as well as in excess (100 μM).

CD was performed to investigate the effect of the presence of dopamine on the secondary structure of the FOXP2 FHD. The spectra were measured between 200 and 250 nm (figure 3.4 (A)).

The effect of the neurotransmitter on the tertiary structure of the protein was then investigated using intrinsic tryptophan fluorescence. The excitation wavelength was set at 295 nm, such that exclusively tryptophan residues were excited. Shifts in emission are therefore indicative of changes in the microenvironment surrounding only the tryptophan probes, rather than all aromatic residues. The emission spectra for FOXP2 in the absence of dopamine, as well as in the presence of dopamine were recorded between 290 and 540 nm (figure 3.4 (B)).

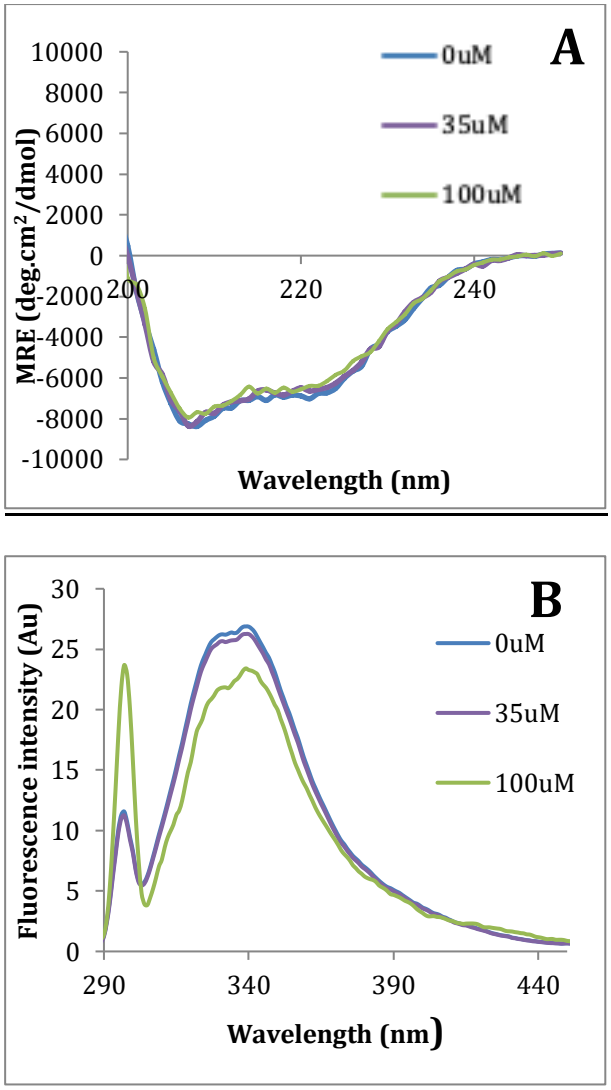


Figure 3.4: Structural Characterisation of the FOXP2 FHD in the absence and presence of 35 µM and 100 µM dopamine (A) CD spectra of 10 µM FOXP2 FHD measured in the presence of dopamine at 20 °C. The spectra overlay and the presence of dopamine does not appear to affect the secondary structure of the protein **(B)** Intrinsic tryptophan fluorescence spectra of 2 µM FOXP2 FHD measured in the presence and absence of dopamine at 20 °C.

With reference to figure 3.4 (A), the CD spectra suggest that the presence of dopamine at both the physiological concentration, as well as in excess, has negligible effect on the overall secondary structure of the FOXP2 FHD. The protein appears to remain predominantly alpha helical (as indicated by the clear troughs present at both 208 and 222 nm), and natively folded in the presence of the neurotransmitter. This observation not only suggests the lack of a significant disruption in the secondary structural elements of the protein due to a binding event, but also implies that the presence of dopamine at 35 µM and 100 µM is not toxic or denaturing to the recombinant protein in solution, thereby suggesting that these concentrations were appropriate to use subsequently in

investigating the potential interaction between the FOXP2 FHD and DNA in the presence of the neurotransmitter. It is evident that the relative HT(V) reading at a specific wavelength increases as the concentration of dopamine increases. This suggests that this molecule absorbs circularly polarised light. Since the HT(V) reading at 200 nm for 100 µM dopamine was approaching 700 V (supplementary figure 6.2), the data points at wavelengths lower than 200 nm were discarded. The absorbance of light by dopamine is visually evident in that the data becomes visibly noisier as the concentration of dopamine increases.

With regard to the tertiary structural investigation, despite excitation at a wavelength of 295 nm, the presence of 3 probes within different regions of the protein (figure 3.4) provides a relative global view of the effect of a potential interaction on the tertiary structure of the protein. It was not possible to excite the FOXP2 FHD at 280 nm in the presence of dopamine, as the neurotransmitter, structurally similar to tyrosine, is excited at 279 nm (Wang *et al.*, 2002), and dominated the emission spectrum.

The fluorescence emission spectrum (figure 3.4 (B)) corresponding to FOXP2 FHD in the absence of dopamine (blue), overlays almost perfectly with that of the protein in the presence of 35 µM dopamine (purple) (after adjusting for the inner filter effect), despite small differences in fluorescence intensity. This could be attributed to small differences in protein concentration.

The emission spectrum of FOXP2 FHD in the presence of 100 µM dopamine (green), however, appears to differ quite significantly from the other two spectra. The overall fluorescence intensity appears to be quenched, and there is large peak at 295 nm, representing a larger proportion of Raleigh Scattering. It is inevitable that even with a fresh stock of dopamine, autoxidation will begin to occur. The observations could, therefore, be attributed to a higher proportion of reactive oxidised dopamine species in solution. Intermediates such as dopaminechrome, for example, have been reported to increase aggregation propensity in proteins associated with PD. It is therefore possible that there is an increase in aggregated protein in the solution, thereby resulting in an increase in the scattering of light, and thus larger peak observed at 295 nm. This would also account for the reduced fluorescence intensity observed. The increase in protein aggregation would reduce the concentration of soluble protein, and thus fluorescent signal.

Lakowics describes a phenomenon whereby fluorophores can form non-fluorescent complexes with other molecules in solution (quenchers). This can occur in either the ground or excited state, thereby resulting in a decrease in absorption or emission. Interestingly, it has been observed that when tyrosine is excited, the pKa drops to roughly 4, usually resulting in ionisation (Lakowicz, 2006). The ionised molecule is reported to be far more reactive, and more likely to quench other fluorescent probes in solution. Although dopamine is not excited at a wavelength of 295 nm, the pKa of dopamine is has been hypothesised to be concentration dependent (Armstrong and Barlow, 1976). Increased dopamine concentrations have been reported to lower the pKa (Armstrong and Barlow, 1976). Since dopamine is tyrosine derived, it can be hypothesised that perhaps the ionisation of dopamine too could result in increased reactivity and thus when dopamine is at a concentration of 100 µM, perhaps there is an increase in the propensity to form non-fluorescent complexes and thus significant quenching is observed.

There does not, however, appear to be a significant blue or red shift when the FOXP2 FHD is exposed to dopamine. This suggests that it is unlikely that there are changes in the microenvironment surrounding the fluorescent probes. A binding event would result in conformational changes (due to induced fit ligand binding), or direct interactions between dopamine and the region surrounding the probe.

It is therefore clear, based on the observations illustrated in figure 3.4, there is not sufficient evidence to suggest a direct interaction between the FOXP2 FHD and dopamine.

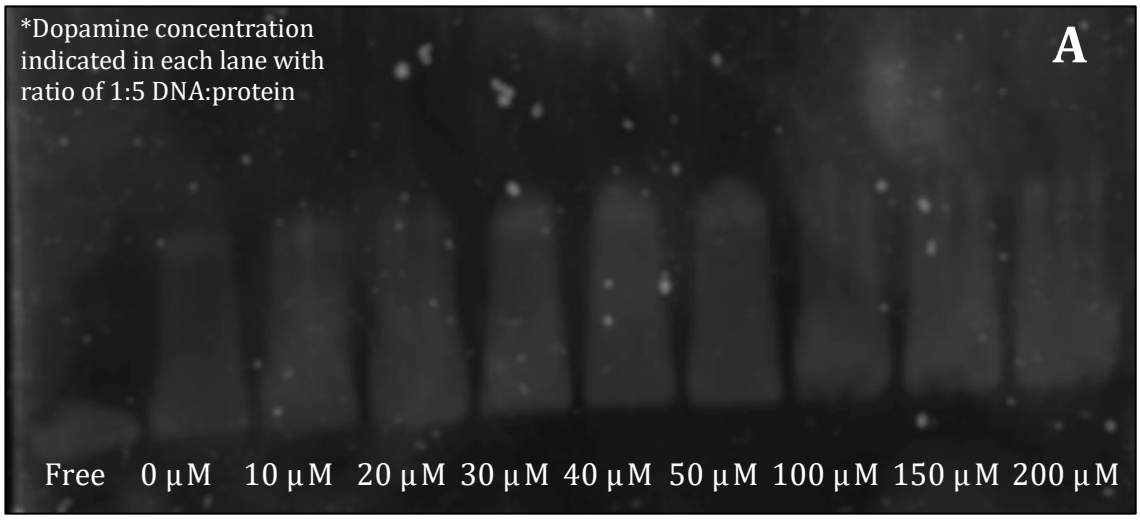
This does not, however, mean that the presence of dopamine will not affect DNA binding *in vivo.* It is possible that the presence of dopamine could interact with the protein-DNA complex, or interact with DNA, such that the activity of FOXP2 could be enhanced or inhibited.

3.4. Investigating the FOXP2 FHD-Dopamine-DNA interaction

Ultimately, the function of FOXP2 is to interact with DNA, thereby facilitating regulation of gene expression, specifically via the FHD. Investigating whether the presence of dopamine affects the function of the FOXP2 FHD would therefore involve investigating whether the ability of the FOXP2 FHD to bind DNA is affected by the presence of dopamine. The Nelson sequence (5'-TGTTTAC-3*)* was used as the binding target due to the reported high affinity binding.

Due to its low resolution, the EMSA was used primarily to investigate whether the presence of dopamine would significantly inhibit or enhance complex formation between the FOXP2 FHD and DNA. The ratio of DNA: protein at which binding occurred was confirmed in the absence of dopamine (figure 3.2 (D)) and was found to be 1:5. This ratio was used in the investigation as to whether the presence of dopamine would affect the binding affinity of the transcription factor for DNA (figure 3.5 (A)). The minimum concentration of protein required for evident DNA binding in an EMSA was selected based on figure 3.2 (D). Increasing the protein concentration in the presence of dopamine and DNA resulted in the blackening of samples over time, indicating an increase in the rate at which neuromelanin formed within samples, thereby indicating an increase in the rate of autoxidation of the neurotransmitter.

Since EMSAs are a relatively low-resolution method by which binding affinity can be investigated, the effect that the presence of dopamine has on the K_D of binding between the FOXP2 FHD and DNA was investigated using fluorescence anisotropy. Nelson DNA, labeled with a ROX tag, was used as a fluorescent probe in this investigation. Samples were therefore excited at 580 nm (the excitation wavelength of ROX) while emission was read at 605 nm. The relative anisotropy was measured in the presence of 35 µM dopamine, as well as 100 µM dopamine (figure 3.5 (B)).



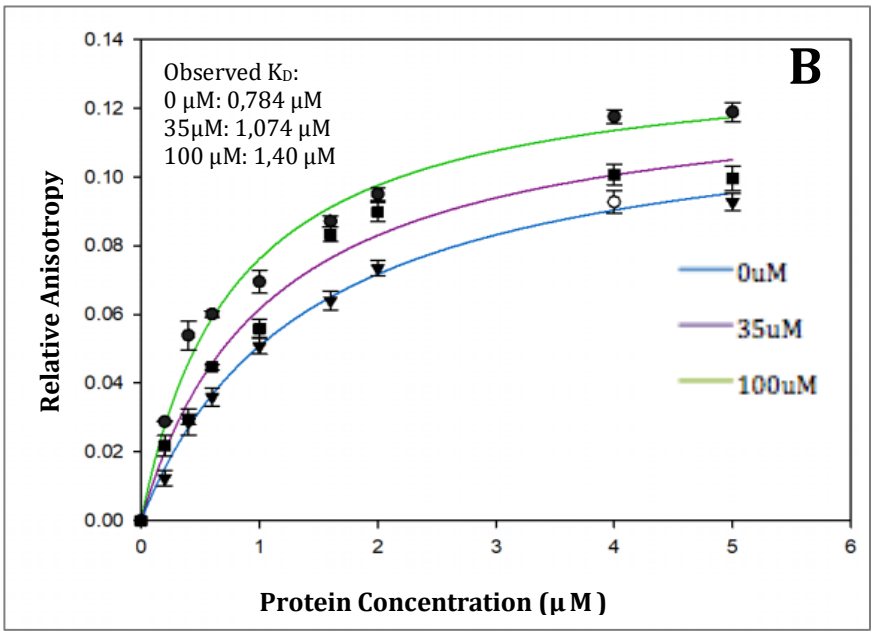


Figure 3.5: Evaluation of the Effect of the Presence of Dopamine on the DNA Binding Function of FOXP2 FHD (A) EMSA illustrating complex formation between 2,5 µM FOXP2 FHD and 0,5 µM Nelson DNA (ratio of 1:5) in the absence and presence of varying concentrations of dopamine **(B)** Relative anisotropy observed with regard to the interaction between the FOXP2 FHD and ROX-labeled DNA in the presence and absence of dopamine.

With reference to figure 3.5 (A), an electrophoretic shift appears to occur at all concentrations of dopamine evidenced by the presence of a smear, and faint second band in addition to that representing free DNA in each lane. The smearing made it difficult to discern whether there is a significant difference in binding affinity as the concentration of dopamine increases. This occurs due to rapid association and dissociation of the complex as the TF and DNA exist within chemical equilibrium within the gel (Hellman and Fried, 2007). Since an electrophoretic shift appears to occur at all concentrations of dopamine between 10 and 200 µM, it is evident that the

presence of dopamine at these concentrations does not completely inhibit the interaction between the FOXP2 FHD and DNA.

Through fluorescence anisotropy performed in triplicate (figure 3.5 (B)), the K_D for the interaction in the absence of dopamine was found to be 0,7835 µM (blue), with a standard error of 0,0562, while the K_D for the interaction in the presence of 35 µM dopamine was found to be 1,0736 µM (purple). The standard error was found to be 0,0982. In the presence of 100 µM dopamine, the K_D was found to be 1,4040 µM (purple), with a standard error of 0,0821. The observed K_D in the absence of dopamine was found to be very similar to that previously reported in the literature for FOXP2 FHD DNA binding, specifically 0,892 µM (Morris and Fanucchi, 2016).

The student t-test was used to investigate the significance of observed differences in K_D.

The student t-test is a parametric hypothesis test, used to establish whether observed differences between two samples occurred by chance, and thus whether the differences are statistically significant and reproducible. Since the t-test is a parametric test, the requirement for data to follow a normal distribution must be satisfied (Kim, 2015). In this experiment, this requirement was satisfied. All 3 experiments passed the Shapiro-Wilk test for normality at 5% significance.

The results for the t-test are described in terms of a t-score. The t-score represents the ratio of the difference between 2 groups to the difference within the groups, and can be described by the equation below:

$$t = \frac{(x_1 - x_2)}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}}$$

Where t represents the t-score, X represents the mean, S symbolizes the variance and n indicates the sample size. The subscripts 1 and 2 are indicative of the two different data sets (Kim, 2015).

The smaller the t-score, the more similarity there is between the groups, while a larger t-score is indicative of greater differences between groups. Every t-score has an associated p-value. The p-value is indicative of the probability that the results from the sample occurred by chance. Lower p-values, therefore, represent more reproducible data. The p-value

represents a percentage of 100% and is expressed as a decimal (Kim, 2015).

The statistical summary for the 2-tail t-test comparing the observed K_D values obtained is summarised below (table 1). The null hypothesis when performing a t-test is that there is no observable difference between groups.

Table 1: Results of student t-test

Experiments compared		K_D (µM)	t-score	p-value
0 µM	35 µM dopamine	0,7835 and 1,0736	0,653	0,523
35 µM dopamine	100 µM dopamine	0,7835 and 1,4040	1,160	0,263
0 µM dopamine	100 µM dopamine	1,0736 and 1,4040	0,498	0,625

It is evident that with regard to all 3 comparisons, and at a 95% confidence interval, there is not sufficient variation between the groups to suggest that differences in the observed binding affinity is not due to chance and random sampling, and the null hypothesis cannot be rejected. Differences in K_D are therefore not found to be significant.

This is evident through examination of the p-values. With regard to all 3 comparisons, the p-values are greater than 0,05, and represent a high probability of variation occurring by chance (52,3 % with regard to the comparison between 0 µM and 35 µM dopamine, 26,3 % corresponding to the comparison between 0 µM and 100 µM dopamine, and 62,5% with reference to the comparison between 35 µM and 100 µM dopamine).

It is, however, evident that there appears to be a trend of a small, statistically non-significant reduction in binding affinity between the FOXP2 FHD and the Nelson sequence as the concentration of dopamine increases.

To investigate the likelihood of type II error (failing to reject the null hypothesis when there is in fact significant difference between groups), the power of performed 2-tail test was utilised (with alpha=0,05). This test combines factors such as sample size, effect size and significance to statistically represent the possibility a false negative. The higher the statistical power statistic, the lower the likelihood of a type II error. Power statistics are reported as decimals to represent a

percentage of 100%. Generally, the desired power statistic is 0,8 (or 80%) (Uttley, 2019).

The power statistics for this study at alpha=0,05 were found to be 0,094 when comparing the K_D at 0 µM and 35 µM dopamine, as well as when comparing the K_D at 0 µM and 100 µM dopamine. When comparing the K_D at 35 µM and 100 µM dopamine the power statistic was found to be 0,076. These scores are significantly lower than the desired score of 0,8. This suggests that although the t-test did not find the differences in K_D to be significant, there is a high chance that this could be due to type II error. The low power score is likely the result of the small sample size (Uttley, 2019).

Although, theoretically statistically insignificant, the observation that the binding affinity of the FOXP2 FHD for DNA decreased in the presence of increasing concentrations of dopamine suggests that the presence of dopamine could potentially still have an effect on the overall cellular function of the FOXP2 FHD without causing detectable disruption to the secondary or tertiary structure, or significantly disrupting the DNA binding function. One such explanation for this could be that dopamine itself does not affect the structure or function of the FOXP2 FHD, however the presence of a dopamine autoxidation intermediate may. The observed small statistically non-significant reduction in binding affinity supports this hypothesis in that it is inevitable that these compounds would begin to accumulate in the sample over time while performing the experiment, thereby potentially reducing the observed binding affinity. Since precautions were taken to minimise the accumulation of these reactive compounds (including preparing dopamine stocks fresh, keeping them on ice and ensuring samples were used within 5 minutes after mixing), the relative concentration of autoxidation products would theoretically be low compared to the concentration of dopamine. Compounds that formed as a result of autoxidation would also exist in random proportions and concentrations within the sample.

It is challenging to test this hypothesis *in vitro,* as intermediates including dopaminechrome, dopaminequinone and 5,6-indolequinone are highly unstable and reactive. It would therefore be difficult to trap a single intermediate and analyse the effect of its presence on the interaction between the FOXP2 FHD and DNA, or on the structure of the protein itself. Studies were therefore performed in *silico* to investigate the potential for specific autoxidation intermediates to interact with the FOXP2 FHD.

3.5. In silico studies

Molecular docking was used to investigate the potential for an interaction between the FOXP2 FHD and dopamine, as well as between the protein and 3 products of dopamine autoxidation, specifically 5,6-dihydroxyindole, 5,6-indolequinone and dopaminechrome. These compounds were selected based on previous studies highlighting their potential to affect the function of various proteins *in vitro* and *vivo* (Segura-Aguilar, 2019, Graham, 1978, Mattamal *et al.*, 1995, Prota, 1998, Hauser *et al.*, 2013, Zhang *et al.*, 2018, Whitehead *et al.*, 2001, LaVoie *et al.*, 2005m Zafar *et al.*, 2006, van Laar *et al.*, 2009, Puspita *et al.*, 2017, Conway *et al.*, 2001, Norris *et al.*, 2005). The potential for these compounds, however, to interact specifically with FOXP2 had not yet been investigated. Despite numerous reports of dopaminequinone affecting the function of various proteins, kinetic studies have suggested that the rate at which the compound undergoes intramolecular cyclisation is too rapid to allow for the majority of interactions with surrounding biomolecules to occur (Tse *et al.*, 1976, Bisaglia *et al.*, 2006). It is therefore unlikely that this compound would interact with the FOXP2 FHD, and thus this reactive compound was omitted from this study.

The goal of docking is to attempt to predict the position, as well as orientation of a ligand when bound to a protein of interest. Blind, as well as focused, site targeted docking were used to investigate the potential interactions between the FOXP2 FHD and ligands.

Blind docking refers to the process by which a ligand is docked to the entire surface of a protein. This is usually performed when *in silico* studies are implemented with no prior knowledge of a specific binding pocket or site within a protein of interest. Since no canonical binding site has been identified with regard to the FOXP2 FHD, nor have specific ligand interactions yet been identified, blind docking allowed potential binding sites to be identified.

PyRx was used to perform the blinding docking studies. The FOXP2 FHD monomer, as well as dimer, were docked with dopamine, 5,6-dihydroxyindole, 5,6-indolequinone and dopaminechrome after preparation with Autodock Tools. The best pose (that which scores highest with regard to thermodynamic and steric favourability) (out of 9) for each ligand was selected based on the PyRx scoring tool and free energies, and the results are summarised below (table 2).

Table 2: Free energies (ΔG) associated with best poses for the blind docking of dopamine, 5,6-dihydroxyindole, 5,6-indolequinone and dopaminechrome with monomeric and dimeric FOXP2 FHD generated by PyRx

ΔG (kcal/mol)				
	Dopamine	**5,6-dihydroxy indole**	**5,6-indolequinone**	**Dopaminechrome**
Monomeric FOXP2 FHD	-4,7	-5,2	-5,2	-5,1
Dimeric FOXP2 FHD	-5,2	-5,3	-5,4	-5,3

The ΔG value is representative of a combination of estimates of the van der Waals forces, hydrogen bonds, and electrostatic and hydrophobic interactions, and thus is illustrative of the total estimated energy of intermolecular forces driving a binding event. It is important to note that the calculation of ΔG with regard to PyRx relies on several approximations and thus the reported binding free energies should not be viewed as an accurate report of the energy of binding *in vivo*, but rather act as a means by which results can be compared.

Since a lower ΔG score is indicative of a higher binding affinity, it is evident that 5,6-dihydroxyindole and 5,6-indolequinone are predicted to bind monomeric FOXP2 FHD with higher affinity than dopaminechrome and dopamine. 5,6-indolequinone is predicted to bind the dimer with the highest affinity, while dopamine is predicted to bind with the lowest. The ΔG scores overall suggest that the autoxidation products are more likely to bind the protein than dopamine. The predicted binding location and positions were visualised using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC (figure 3.6).

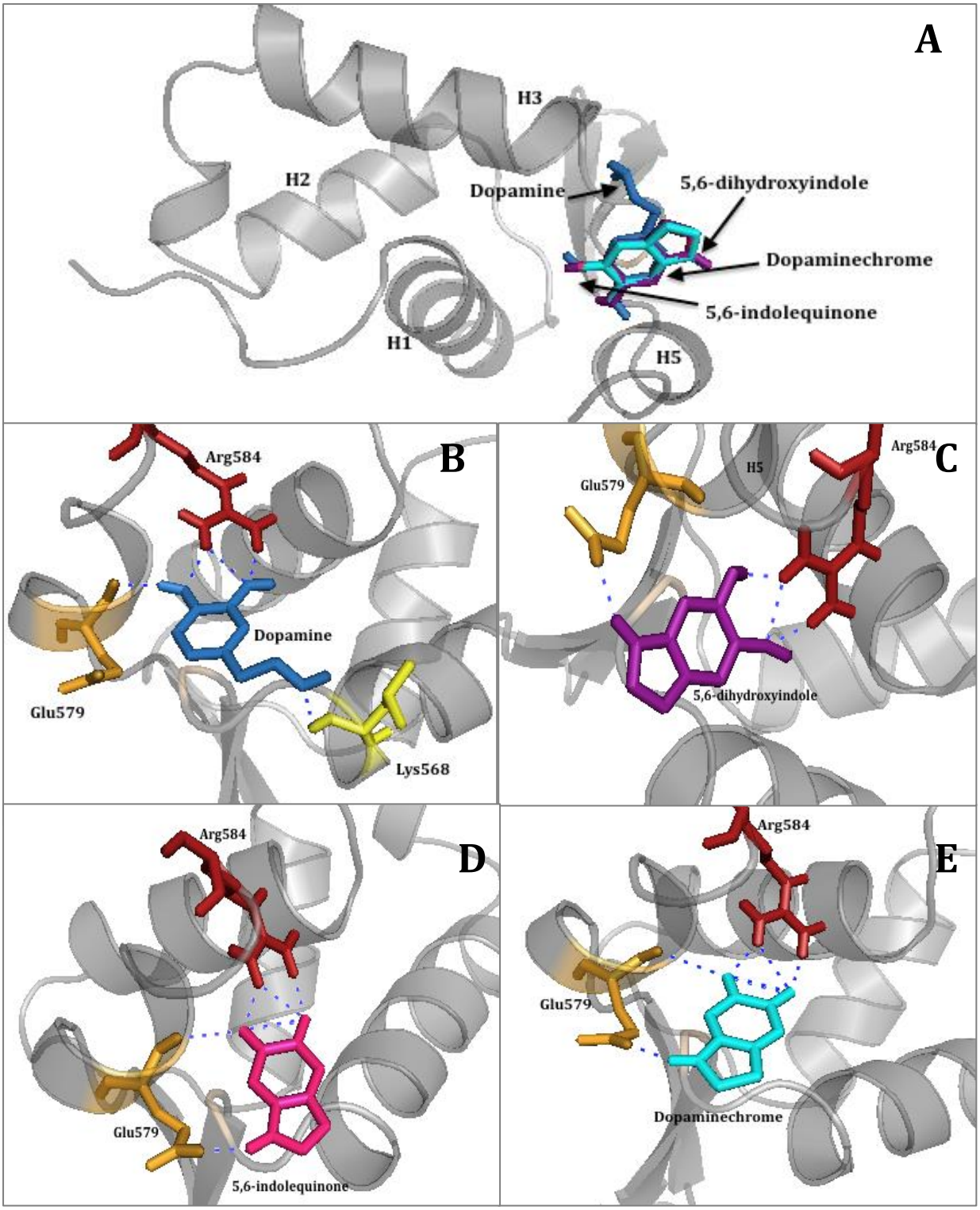


Figure 3.6: Binding Predictions by PyRx for Monomeric FOXP2 FHD and Dopamine, as well as its Autoxidation Products. Predicted H bonds are illustrated by dotted blue lines **(A)** Global view of monomeric FOXP2 FHD with all ligands bound in predicted positions **(B)** Predicted interaction between monomeric FOXP2 FHD and dopamine (blue) **(C)** Predicted interaction between monomeric FOXP2 FHD and 5,6-dihydroxyindole (purple) **(D)** Predicted interaction between monomeric FOXP2 FHD and 5,6-indolequinone (pink) **(E)** Predicted interaction between monomeric FOXP2 FHD and dopaminechrome (cyan)

It is evident that all 4 ligands share a similar predicted binding site. This binding site is located towards the C-terminus of the protein, and contacts are made predominantly with amino acids associated with H5. The key residues associated with the potential interactions appear to be Arg584 and Glu579. These residues are predicted to interact with all 4 ligands by facilitating the formation of a number of hydrogen bonds thus aiding in stabilisation.

Visualisation of the potential interactions between residues associated with dimeric FOXP2 FHD and dopamine, as well as the autoxidation intermediates are illustrated below (figure 3.7). The physiological function of the domain swap dimer still remains unknown. It is therefore of value to investigate whether these small molecules have the potential to interact with the FOXP2 FHD in this form, as it may reveal valuable information regarding function *in vivo*.

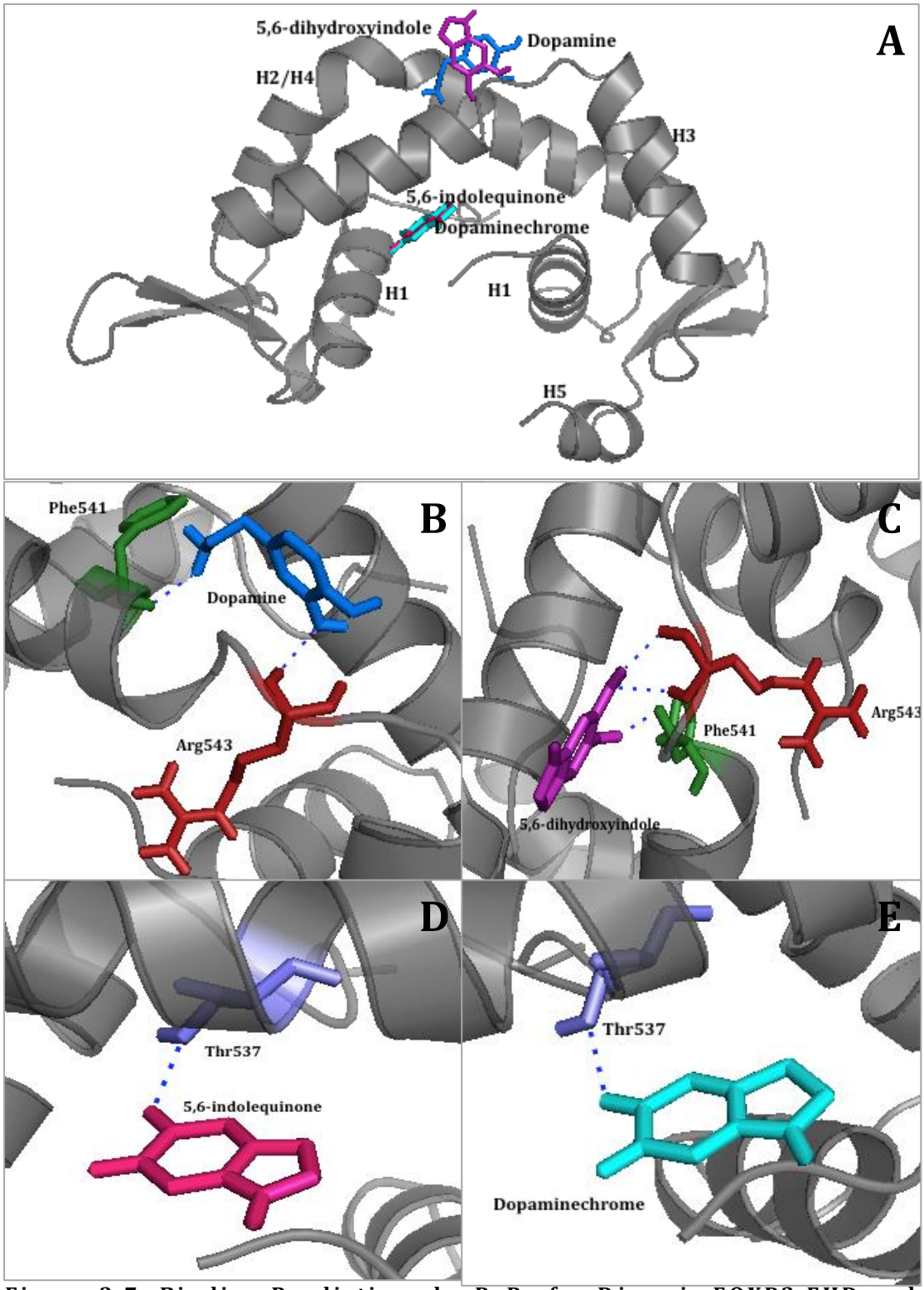


Figure 3.7: Binding Predictions by PyRx for Dimeric FOXP2 FHD and Dopamine, as well as its Autoxidation Products. Predicted H bonds are illustrated by dotted blue lines **(A)** Global view of dimeric FOXP2 FHD with all ligands bound in predicted positions. Two different binding sites were identified **(B)** Predicted interaction between dimeric FOXP2 FHD and dopamine (blue) **(C)** Predicted interaction between monomeric FOXP2 FHD and 5,6-dihydroxyindole (purple) **(D)** Predicted interaction between dimeric FOXP2 FHD and 5,6-indolequinone (pink) **(E)** Predicted interaction between dimeric FOXP2 FHD and dopaminechrome (cyan).

It is evident that dopamine and 5,6-dihydroxyindole share a predicted binding site, while 5,6-indolequinone and dopaminechrome are bound within a similar region. Since dopamine is structurally similar to its autoxidation products, it is not unexpected for a similar potential binding site to be identified within both monomer and dimer regarding all four ligands. With regard to the potential binding site for dopamine and 5,6-dihydroxyindole, the key residues in stabilizing the interaction are predicted to be Phe541 and Arg543, while Thr537 appears to be largely associated with stabilisation of 5,6-indolequinone and dopaminechrome.

After performing blind docking, a more directed approach was used to investigate the potential for interactions between the FOXP2 FHD and dopamine (and its autoxidation products). Schrodinger's Maestro 11.2 was used to perform site targeted docking. After energy minimisation, Sitemap algorithm was used to identify potential binding sites within both monomeric and dimeric FOXP2 FHD. This algorithm is specifically designed to locate regions within the protein that are rich in hydrogen bond donors and acceptors, or suitable to accommodate ring structures and hydrophobic groups, or both. After energy minimisation and preparation through LigPrep, dopamine, 5,6-dihydroxyindole, 5,6-indolequinone and dopaminechrome were docked to the identified potential binding regions using the Schrodinger Induced Fit protocol. The best poses (out of 12) were selected for each ligand predominantly based on the observed E-model scores. The results are illustrated below (table 3).

Table 3: Glidescores (G-scores) and E-model glide scores (E-model) associated with best poses for the blind docking of dopamine, 5,6-dihydroxyindole, 5,6-indolequinone and dopaminechrome with monomeric and dimeric FOXP2 FHD generated by Schrodinger

	Dopamine	**5,6-dihydroxy indole**	**5,6-indole quinone**	**Dopamine chrome**	**Score type (kcal/mol)**
Monomeric FOXP2 FHD	-6,14	-6,49	-5,69	-5,72	**G-score**
	-38,99	-45,95	-37,29	-36,41	**E-model**
Dimeric FOXP2 FHD	-5,262	-5,70	-4,73	-4,51	**G-score**
	-31,99	-32,40	-29,87	-27,42	**E-model**

The G-score represents the ligand free binding energy. Much like the ΔG score (used by PyRx to describe free energy of binding), the G-score should not be taken as an indication of free binding energy *in vivo*, but rather act as a score by which different ligands or poses can be ranked against each other. The

G-score is calculated based on a combination of factors, including the approximated electrostatic and van der Waals contributions, as well as other factors known to influence ligand binding. The more negative the observed G-score, the more tightly a ligand is predicted to bind. It is therefore evident that based on the G-score, 5,6-dihydroxyindole is predicted to bind the FOXP2 FHD with the highest affinity, while 5,6-indolequinone was predicted to bind the least tightly. 5,6-dihydroxyindole was also predicted to display the highest binding affinity for the FOXP2 dimer, while dopaminechrome appeared to be the least likely to bind.

The E-model score is calculated similarly to the G-score; however, the weighting towards force field components (namely the electrostatic and van der Waals energies) is significantly larger. For this reason, it provides a means by which different poses corresponding to the same ligand can be ranked and the best selected.

The potential interactions between monomeric (figure 3.8) and dimeric FOXP2 FHD (figure 3.9) and the 4 ligands were visualised using PyMol Molecular Graphics System, version 1.7.4.5, Schrödinger, LLC,, as well as the generated 2D ligand interaction diagrams (supplementary information figure 6.3)

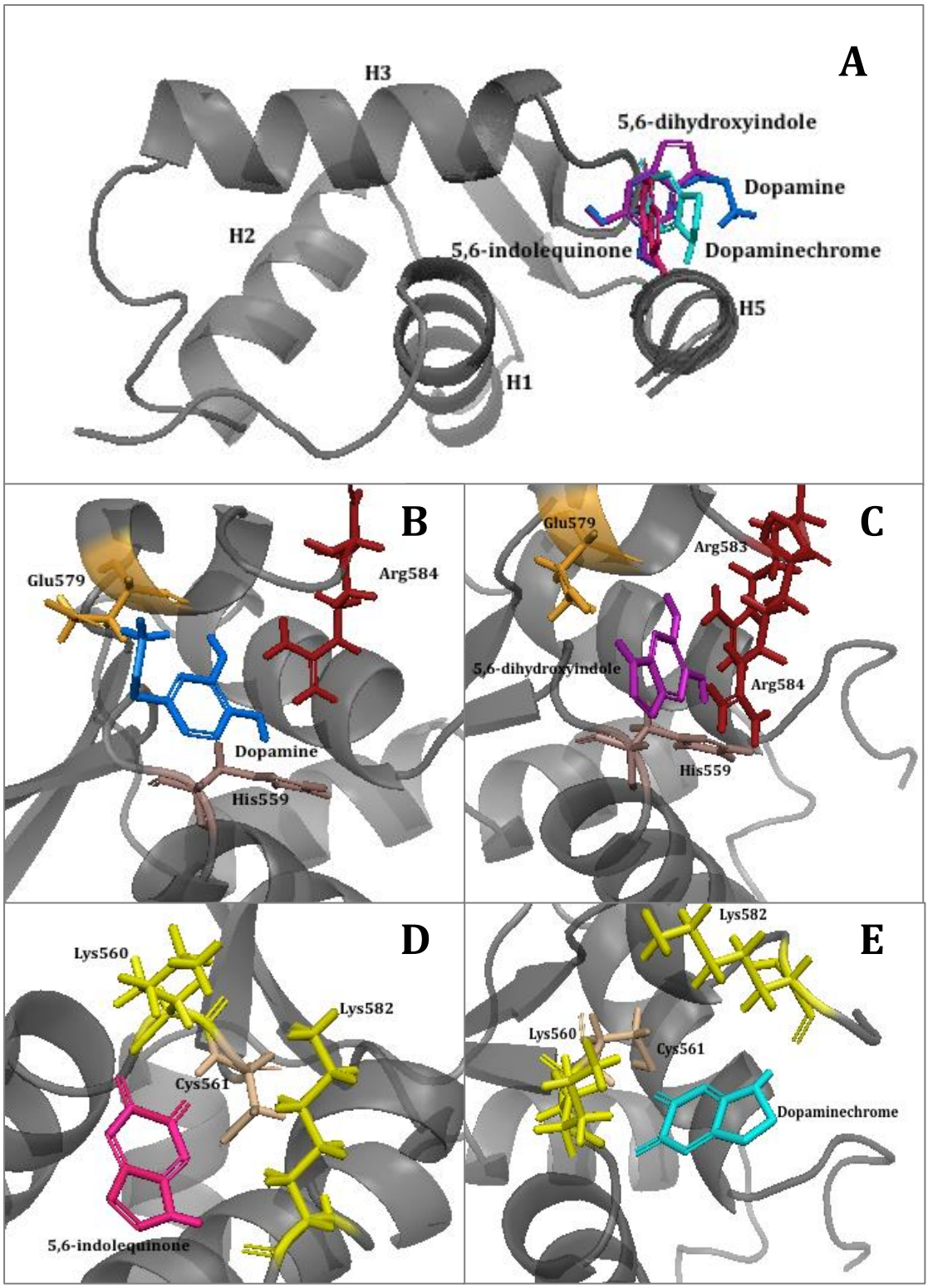


Figure 3.8: Targeted Binding Predictions by Schrödinger Suite for Monomeric FOXP2 FHD and Dopamine, as well as its Autoxidation Products. **(A)** Global view of monomeric FOXP2 FHD with all ligands bound in predicted positions **(B)** Predicted interaction between monomeric FOXP2 FHD and dopamine (blue) **(C)** Predicted interaction between monomeric FOXP2 FHD and 5,6-dihydroxyindole (purple) **(D)** Predicted interaction between monomeric FOXP2 FHD and 5,6-indolequinone (pink) **(E)** Predicted interaction between monomeric FOXP2 FHD and dopaminechrome (cyan).

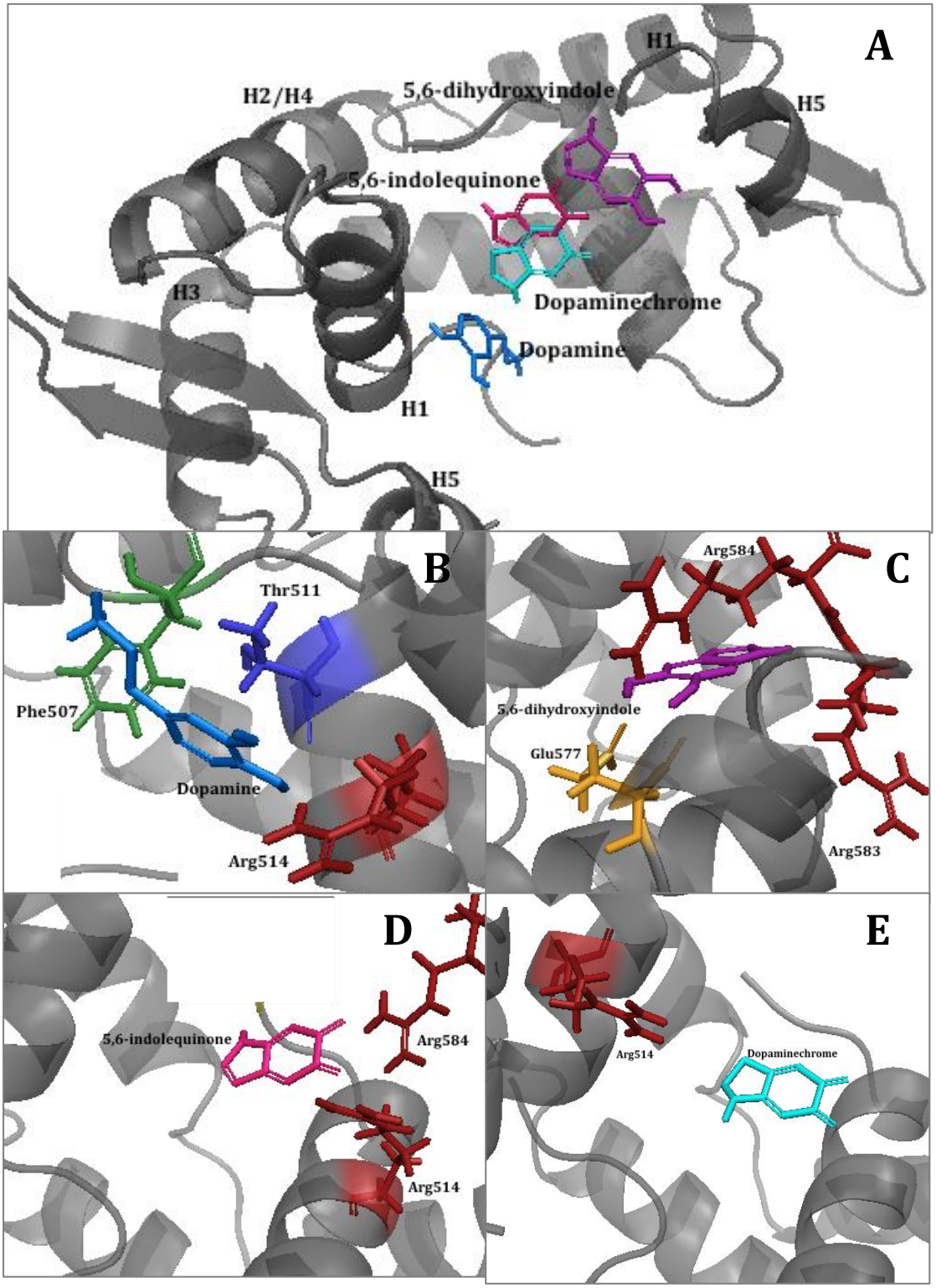


Figure 3.9: Targeted Binding for Predictions by Schrödinger Suite for Dimeric FOXP2 FHD and Dopamine, as well as its Autoxidation Products. (A) Global view of dimeric FOXP2 FHD with all ligands bound in predicted positions **(B)** Predicted interaction between dimeric FOXP2 FHD and dimeric (blue) **(C)** Predicted interaction between monomeric FOXP2 FHD and 5,6-dihydroxyindole (purple) **(D)** Predicted interaction between dimeric FOXP2 FHD and 5,6-indolequinone (pink) **(E)** Predicted interaction between dimeric FOXP2 FHD and dopaminechrome (cyan).

It is evident that, with reference to monomeric FOXP2 FHD (figure 3.8), Sitemap identified a single potential binding site located towards the C-terminus of the protein, and all 4 ligands were found to make contacts with residues associated with H5. With regard to dopamine and 5,6-dihydroxyindole, the key residues associated with the potential interaction were found to be Arg584, Glu577 and His559, while Lys582, Lys560 and Cys561 were largely associated with the binding of 5,6-indolequinone and dopaminechrome.

The potential binding site identified within the FOXP2 FHD dimer (figure 3.9) appeared to be less defined and larger than the monomer binding site, as the ligands appeared to make numerous contacts located between the two H1 regions, as well as with residues associated with H5. The G-scores associated with the dimer appear to be overall lower than those associated with ligand-monomer interactions. This could be attributed to the overall reduction in the number of hydrogen bonds (supplementary information figure 6.3).

Interestingly, integrating the results obtained from the blind docking studies, as well the site targeted docking investigation suggests that, with regard to monomeric FOXP2, a consistent binding 'pocket' is identified. With reference to the blind docking study, all four ligands appeared to bind in a similar conformation, predominantly interacting with Arg584 and Glu579 (figure 3.6). These residues are similarly found to interact with dopamine and 5,6-dihydroxyindole in the site targeted docking (figure 3.8). Although 5,6-indolequinone and dopaminechrome were not found to interact specifically with Arg584 and Glu579, but rather Lys582, Lys560 and Cys561, this interaction occurred at a similar location within the monomeric protein. It is evident that this region provides sufficient hydrogen bond donor and acceptor groups to facilitate ligand binding, as well as satisfies the spatial and electrostatic requirements to accommodate aromatic rings. Because this 'binding pocket' was identified numerous times through molecular docking, it can be hypothesised that if dopamine or one of its autoxidation products does in fact interact with the FOXP2 FHD, this would likely be the location.

Regarding the integrated results for docking studies relating to dimeric FOXP2 FHD, there appears to be a degree of randomness in the location of ligand binding. There is no single binding pocket, site or specific combination of residues that consistently appears between studies. Since all ligands under investigation are structurally similar, this may suggest a lower degree of specificity in binding with regard to the dimer, despite the observation of a more negative ΔG value in blind

docking studies when compared to those associated with monomer-ligand interactions. Quinones are known to be particularly reactive compounds, and are known to bind promiscuously (Solano *et al.*, 1999). It is therefore important to be skeptical of observations that appear to be very non-specific when attempting to investigate the potential for a functional interaction between an autoxidation product of dopamine and the FOXP2 FHD using molecular docking methods. The dimerisation of FOXP2 FHD results in burial of a number of aromatic residues, thereby giving rise to a hydrophobic dimer core (Stroud *et al.*, 2006). This could possibly increase the potential to accommodate ring structures, and the overall reduction in accessible surface area in dimerisation may facilitate the formation of new potential binding pockets, allowing for an increase in the number of non-specific binding events.

It can, therefore, be hypothesised that a specific interaction between one of the ligands and the FOXP2 FHD is more likely to occur when the protein exists in monomeric rather than in dimeric conformation.

With regard to comparing the approximated free energies of binding related specifically to monomeric FOXP2 FHD in the blind and site directed experiments, the consistent observation is that 5,6-dihydroxyindole appears to bind with the highest affinity. The binding free energies are however not largely different from each other.

4. DISCUSSION

Overall, the *in vitro* experimentation suggested that it is unlikely that dopamine interacts directly with the FOXP2 FHD, nor is it likely that dopamine significantly affects the DNA binding affinity. The only observed indication of the potential for an interaction was the small statistically non-significant reduction in binding affinity observed through fluorescence anisotropy (figure 3.4 (B)).

Although experiments *in vitro* aim to mimic physiological conditions, it is important to consider the fact that the experimental procedures described above have been performed under conditions that are, in a sense, thermodynamically ideal. In other words, dilute solutions containing low concentrations of isolated macromolecules and moderate salt (Kuznetsova *et al.*, 2014). In contrast to this, the intracellular environment is an incredibly complex, crowded, heterogeneous milieu, with limited available water and unoccupied space. The cellular concentration of biomolecules is 80-400 mg/mL (Zimmerman, 1991), and the volume occupancy can range between 5 and 40% (Ellis and Minton, 2003). The behaviour of some molecules in a crowded environment may differ from that observed in solution. This is known as macromolecular crowding and is partially due to the phenomenon of excluded volume, whereby a reaction might differ *in vivo* due to spatial restraints or steric hindrance created by the presence of 'crowder' molecules (Minton, 1997, Minton, 2000, Ralston, 2005). In other words, the available volume is reduced and interactions with surrounding molecules are unavoidable (Kuznetsova, 2014).

A crowded solution results in limited space, which means that there is a general decrease in the randomness of distribution of molecules. This results in a decrease in entropy of solution, which gives rise to an increase in the free energy, and thus an increase in thermodynamic activity due to the drive of a system to increase the entropy (Ralston, 2005). Macromolecular crowding, therefore, has significant effects on the stability, reactivity and structural properties of various biomolecules, specifically proteins, and it is important to consider this when analysing observations made within the laboratory (Minton, 2000, Ralston, 2005).

This investigation focused on investigating the effect of a small molecule on the interaction between a protein and nucleic acid. The literature suggests that depending on the chemical nature of all molecules involved, macromolecular crowding can reduce the binding affinity between a ligand and target, due to weak attractive interactions with crowding agents (Aumiller *et al.*,

2014). It has also been reported that macromolecular crowding has pronounced effects on the interaction between DNA binding proteins and nucleic acid, and that the fastest kinetics of sequence recognition and binding is achieved at physiological concentrations of crowders (Dey and Bhattacherjee, 2018). This has been attributed to the idea that frequent collisions between the protein and surrounding molecules aid in orientating the protein thereby facilitating specific complex formation. The constant collisions are also thought to aid the protein in "hopping on" DNA, thereby allowing for rapid target sequence screening (Dey and Bhattacherjee, 2018). It is therefore evident that since the presence of crowder molecules has a significant effect on the association state, as well as chemical equilibrium, it can be thought that although the observations here suggests that there is no observed interaction between dopamine and FOXP2, it is possible that this does not hold true physiologically, especially with regard to observations surrounding the DNA binding affinity.

In silico observations appear to suggest that with regard to monomeric FOXP2 FHD, a small non-canonical ligand-binding pocket may exist towards the C-terminus of the protein, surrounding Lys568-Arg584 (including residues associated H5).

A recent study, and crystal structure, published by Bruning *et a*l. in 2019 illustrated that the activity of a TF, NURR1, was modulated by a covalent interaction with 5,6-dihydroxyindole within a non-canonical ligand-binding pocket. 5,6-dihydroxyindole was reported to interact with the TF via the formation of an adduct from a cysteine residue between helices, resulting in the formation of bound 5,6-indolequinone. The formation of this adduct was hypothesised to affect DNA binding or homo and heterodimerisation. The binding pocket of NURR1 containing bound 5,6-indolequinone is illustrated below (figure 4.1).

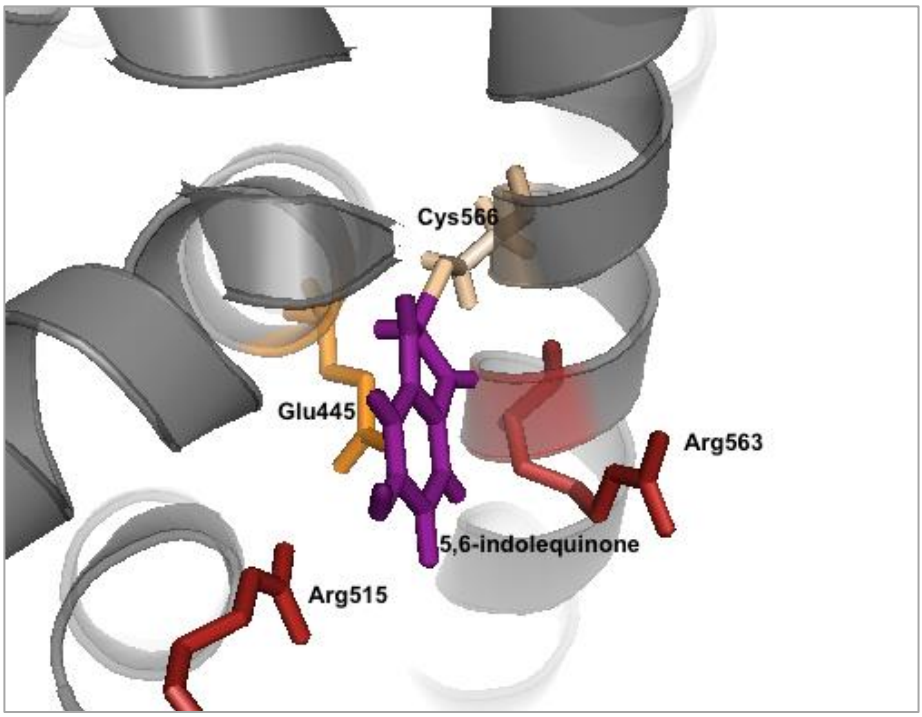


Figure 4.1: 5,6-dihydroxyquinone bound to NURR1 as 5,6-indolequinone (PDB: 6DDA) 5,6-dihydroxyquinone binds NURR1 via the formation of an adduct with Cys566. The interaction is stabilised via the formation of hydrogen bonds between the quinone and Arg515, as well as Glu445. Cation-π interactions between the indole and NH group of 5,6-indolequinone and Arg515 and/or Arg563 were hypothesised to stabilise the interaction.

The residues acting as hydrogen bond donors and acceptors within the binding pocket associated with NURR1 closely resemble those identified within the potential FOXP2 FHD ligand binding pocket identified through the *in silico* studies presented in this investigation. Although consistently, the *in silico* observations suggested that 5,6-dihydroxyindole was the most likely of the four intermediates to bind the monomeric FOXP2 FHD, the differences in the approximated free ligand binding energies were not large. The docking, however, does not account for the possibility of covalent interaction through adduct formation. The presence of a cysteine residue within the predicted monomeric FOXP2 FHD ligand-binding site indicates that an interaction similar to that observed in NURR1 with regard to 5,6-dihydroxyindole may still be possible, despite the lack of clear evidence through molecular docking. Further *in vitro and in vivo* experimentation would be required to investigate further whether this binding site is truly valid, and whether the presence of 5,6-dihydroxyindole specifically would alter the function of the FOXP2 FHD.

It is important to note that dopamine autoxidises rapidly in solution (the reported $t_{1/2}$ at pH 7.4 in solution was found to be 4.95 minutes)(Umek *et al.*, 2018), so although dopamine stocks

were freshly prepared for each experiment, it is likely that these stocks began to oxidise throughout experimental procedures, and thus likely contained low concentrations of various autoxidation products, including 5,6-dihydroxyindole. It was also observed that over time, a black precipitant would form in the presence of FOXP2, suggesting complete autoxidation of dopamine in the presence of protein if left at room temperature for extended periods of time. *In vivo* dopamine is protected from autoxidation by the presence of ascorbic acid (Neal *et al.*, 1999), however this compound could not be incorporated into experiments due to its propensity to interact non-specifically with DNA (Lui *et al.*, 2006).

It is therefore possible that an interaction between the FOXP2 FHD and a product of dopamine autoxidation, perhaps 5,6-dihydroxyindole, gave rise to the small differences between the observed K_D values at increasing concentrations of dopamine. Although according to the student t-test these differences were not statistically significant, it is evident that the binding affinity decreases slightly with increasing concentration of dopamine. If the hypothesis holds true that 5,6-dihydroxyindole or one of the other products of autoxidation has the capacity to bind the FOXP2 FHD within the site identified through molecular docking, it is likely that the binding event occurs within the region associated with W2. With regard to FOXP2, H5 is reported to influence the specificity and affinity of binding, however this wing makes limited contacts with the DNA (Stroud *et al.*, 2006). It is therefore possible that an interaction with 5,6-dihydroxyinole has a small but not insignificant effect on DNA binding. It is however, also possible, that the relative concentration of autoxidation products in solution was too low to show significant effect.

The lack of clear evidence for this potential binding event in the CD and intrinsic tryptophan fluorescence investigations can perhaps be attributed to low concentrations of products of autoxidation relative to dopamine, as well as to the location of the potential binding site. It is possible that an interaction occurring at the C-terminus of the protein does not cause global structural disruption, and thus the CD spectrum would not be visisbly affected. In fact, it has been reported that in 85% of ligand-protein interactions, global changes to the protein backbone are negligible, and generally only 3 or less side chains undergo conformational changes (Najmanovich *et al.*, 2000). It is therefore evident that due to the location of the binding site, the polarity of the microenvironment surrounding all intrinsic probes would not be altered. It can therefore be further hypothesised that perhaps physiologically, the potential

interaction is not associated with regulation of DNA, but rather regulation of localisation of FOXP2.

Two nuclear localisation signals identified for FOXP2 are located within the FHD, namely NLS1: (^{476}RRRHS^{480}) and NLS2: (^{582}KRRSQK^{587}) (Mizutani *et al.*, 2007). The presence of both signals is described to be essential in translocation of FOXP2 into the nucleus (Mizutani *et al.*, 2007). If the hypothesis holds true in that 5,6-dihydroxyindole has the potential to bind the FHD, it is possible that the concentration of dopamine within the cytoplasm may be associated with regulating the translocation of FOXP2 into the nucleus. The binding of 5,6-dihydroxyindole is likely to disrupt NLS2 due to the recurring hypothesised interaction with Arg584. This may in turn regulate the levels of FOXP2 within the nucleus, thereby modulating the activity of the TF in a concentration dependent manner, rather than affecting the ability of the TF to interact with the DNA directly. It is possible that excess FOXP2 localised in the cytoplasm is sequested within neuromelanin, as it has been reported that this polymer contains units of quinone bound protein (Haining, 2017).

Although the observations made in this study did not conclusively provide an answer to the question of whether or not the presence of dopamine has the potential to affect the DNA binding function of the FOXP2 FHD, the observations made here have brought to light the potential for many new hypothesises and ideas to be developed surrounding the regulation of FOXP2 *in vivo*, specifically with regard to dopamine and its autoxidation products. Previous studies focused on investigating the potential links between FOXP2 and dopamine have predominantly investigated the effect of mutant or reduced levels FOXP2 on dopaminergic circuits (Co *et al.*, 2020, van Rhijin *et al.*, 2018, Enard *et al.*, 2009, Raghanti *et al.*, 2016). This study begins to explore the question of whether alterations in cellular dopamine levels have the potential to affect the function of FOXP2 *in vivo*.

Conclusions

In conclusion it is evident that the *in vitro* investigation suggested that the presence of dopamine does not observably affect the structure or the function of the FOXP2 FHD, however, a small statistically non-significant reduction in binding affinity was observed through fluorescence anisotropy as the concentration of dopamine increased. *A* potential ligand-binding site within monomeric FOXP2 FHD was identified through *in silico* studies, specifically through both blind and site targeted docking. This site was located towards the C-terminus of the protein and is comprised of residues associated with W2. Dopamine, 5,6-dihydroxyindole, 5,6-indolequinone and dopaminechrome were found to bind this site with varying affinities. 5,6-dihydroxyindole was consistently found to bind with the highest affinity in both blind and targeted docking, and it was hypothesised that the accumulation of this compound in solution, through the autoxidation of dopamine, gave rise to the slight reduction in the binding affinity between the FOXP2 FHD and dopamine observed in fluorescence anisotropy. It was then further hypothesised that the function of ligand-FOXP2 FHD interaction was not associated with DNA binding, but rather with cellular localisation of the protein due to the location of the identified binding site. Further investigations both in *vitro* and in *vivo* are, however, required to confirm these claims. Comparative *in silico* studies with previously identified protein targets may provide insights regarding the specificity of this interaction, while investigating the effect of increasing dopamine concentrations on FOXP2 activity within cell culture may provide some evidence to substantiate the hypothesis. Mass spectroscopy, or protein crystallography would also be beneficial in investigating the potential for covalent modification of the protein.

5. REFERENCES

Abraham, E.P., Chain, E. (1940) An enzyme from bacteria able to destroy penicillin. *Nature* 146, 837. doi:10.1038/146837a0

Adinoff, B. (2004). Neurobiologic processes in drug reward and addiction.*Harvard review of Psychiatry 12*, 305–320. https://doi.org/10.1080/10673220490910844

Adrio, J. L., Demain, A. L. (2010). Recombinant organisms for production of industrial products. *Bioengineering Bugs* 1, 116–131. doi: 10.4161/bbug.1.2.10484

Alon, U. (2007) Network motifs: theory and experimental approaches. *Nature Reviews Genetics* 8, 450–461

Armstrong, J., Barlow, R. B. (1976). The ionization of phenolic amines, including apomorphine, dopamine and catecholamines and an assessment of zwitterion constants. *British journal of pharmacology 57*, 501–516. https://doi.org/10.1111/j.1476-5381.1976.tb10377.x

Ayano, G. (2016) Dopamine: Receptors, functions, synthesis, mental disorders: review of literatures. *Journal of Mental Disorders and Treatment 2*, 120

Bazzari, A. H., Parri, H. R. (2019) Neuromodulators and long-term synaptic placticity in learning and memory: a steered-glutamatergic perspective. *Brain Science* 9, 300-320

Ben-Jonathan, N., Hnasko, R. (2001) Dopamine as a prolactin (PRL) inhibitor. *Endocrine Reviews* 22, 724-763

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

Bennett, M.J., Choe, S., Eisenberg, D. (1994) Domain swapping: entangling alliances between proteins. *Proceedings of the National Academy of Sciences*, 91

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

Bisaglia, M., Mammi, S., Bubacco, L. (2007) Kinetic and structural analysis of the early oxidation products of dopamine: analysis of the interactions with alpha-synuclein. Journal of Biological Chemistry 282, 15597-15605

Bisaglia, M., Soriano, M.E., Arduini, I., Mammi, S. Bubacco, L. (2010) Molecular characterization of dopamine-derived quinones reactivity toward NADH and glutathione: Implications for mitochondrial dysfunction in Parkinson disease. *Molecular Basis of Disease* 1802, 699-706

Bjorklund, A., Dunnett, S.B. (2007) Dopamine neuron systems in the brain: an update. *Trends in Neurosciences 30*, 194-202

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

Bohm, H.J. (1998) Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from *de novo* design or 3D database search programs. *Journal of Computer Aided Molecular Design* 12, 309–323

Bornhorst, J. A., Falke, J. J. (2000) Purification of proteins using polyhistidine affinity tags. *Methods in enzymology*, *326*, 245–254. doi:10.1016/s0076-6879(00)26058-8

Boyadjiev, S.A., Jabs, E.W. (2000) Mendelian Inheritance in Man (OMIM) as a knowledgebase for human developmental disorders. *Clinical Genetics* 4, 253–266

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72, 248–254

Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., Gileadi, O. (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expression and Purification* 59, 94–102

Burns, D. T. (1993). Principles of spectrophotometric measurements with particular reference to the UV-visible region. In UV Spectroscopy: Technique, Instrumentation, Data Handling, (B. J. Clark, T. Frost, and M. A. Rusell, eds.) 1–15. Chapman and Hall, London

Bush, W.D., Garguilo, J., Zucca, F.A., Albertini, A., Zecca, L., Edwards, G.S., Nemanich, R.J., Simon, J.D. (2006) The surface oxidation potential of human neuromelanin reveals a spherical architecture with a pheomelanin core and a eumelanin surface. Proceedings of the National Academy of Sciences 103, 40

Bustamante, J., Bredeston, L., Malanga, G., Mordoh, J. (1993) Role of melanin as a scavenger of active oxygen species. *Pigment Cell Research* 6, 348–353

Butland, S.L., Devon, R.S., Huang, Y., Mead, C.L., Meynert, A.M., Neal,S. J., Lee, S. S., Wilkinson, A., Yang, G.S., Yuen, M.M.S., Hayden, M.R., Holt, R.A., Leavitt, B.R., Ouellette, B.F.F.(2007) Physical chemistry of polyglutamine: Intriguing tales of a monotonous sequence. *BMC Genomics* 8, 126

Carlsson, P., Mahlapuu, M. (2002) Five years on the wings of forkhead. *Mechanisms of development* 57, 3-20

Carvalho, M.I., Pires, I., Prada, J., Gregorio, H., Lobo, L., Queiroga, F.L. (2016) Intratumoral FoxP3 expression is associated with angiogenesis and prognosis in malignant canine mammary tumors. *Veterinary Immunology Immunopathology* 178, 1–9

Chamberlain, N.L., Driver, E.D., Misefeld, R.L. (1994) The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acid Research 11*, 3181-3186

Chen, J., Knowles, H.J., Hebert, J.L., Hackett, B.P. (1998) Mutation of the mouse hepatocyte nuclear factor/forkhead homologue 4 gene results in an absence of cilia and random left-right asymmetry. *Journal of Clinical Investestigation* 102, 1077–1082

Chen, Y., and M. D. Barkley. (1998). Toward understanding tryptophan fluorescence in proteins. *Biochemistry* 37, 9976–9982

Chial, H.J., Thompson, H.B., and Splittgerber, A.G. (1993) A spectral study of the charge forms of Coomassie Blue G. *Analytical Biochemistry* 209, 258–266

Chiu, Y.C., Li, M.Y., Liu, Y.H., Ding, J.Y., Yu, J.Y., Wang, T. W. (2014) Foxp2 regulates neuronal differentiation and neuronal subtype specification. *Developmental Neurobiology* 74, 723–38

Clark, K.L., Halay, E.D., Lai, E., Burley, S.K. (1993) Co-Crystal Structure of the HNF-3/Fork Head DNA-Recognition Motif Resembles Histone H5. *Nature* 364, 412-420

Co., M., Hickey, S.L., Kulkarni, A., Harper, M., Konopka, G. (2020) Cortical foxp2 supports behavior flexibility and developmental dopamine D1 receptor expression.*Cerebral Cortex* 30, 1855-1870

Conway, K. A., Rochet, J. C., Bieganski, R. M., Lansbury, P. T., Jr. (2001) Kinetic stabilization of the alpha-synuclein protofibril by a dopamine-alpha-synuclein adduct. *Science 294*, 1346–1349

Cooper, G. M. (2000) The Cell: A Molecular Approach. 2nd edition. Sunderland(MA): Sinauer Associates. Protein Folding and Processing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9843/

Corrêa, D. H. A., Ramos, C. H. I. (2009) The use of circular dichroism spectroscopy to study protein folding, form and function. African Journal of Biochemistry Research 3,164–173
Crow, T. J. (2000). Schizophrenia as the price that Homo sapiens pays for language: A resolution of the central paradox in the origin of the species. *Brain Research Reviews*, 31, 118–129

Demain, A. L., Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances* 27, 297–306. doi:10.1016/j.biotechadv.2009.01.008

Díaz-Véliz, G., Mora, S., Dossi, M. T., Gómez, P., Arriagada, C., Montiel, J., Segura-Aguilar, J. (2002). *Behavioral effects of aminochrome and dopachrome injected in the rat substantia nigra. Pharmacology Biochemistry and Behavior, 73, 843–850.*doi:10.1016/s0091-3057(02)00923-1

Donaldson, L. W., Petersen, J. M., Graves, B. J., McIntosh, L. P. (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *The EMBO journal 15*, 125–134

Enard, W., Gehre, S., Hammerschmidt, K., Holter, S. M., Blass, T., Somel, M., (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137, 961–971

Farid, R., Day, T., Friesner, R. A., Pearlstein, R. A. (2006) New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorganic and Medicinal Chemistry 14*, 3160-3173

Feher, M., Deretey, E., Roy, S. (2003) BHB: a simple knowledge-based scoring function to improve the efficiency of database screening. *Journal of Chemical Information and Modelling* 4, 1316–1327

Ferland, R.J., Cherry, T.J., Preware, P.O., Morrisey., E.E., Walsh, C.A. (2003) Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *Comparative Neurology* 460, 266-279.

Files, J.A., Ko, M.G., Pruthi, S. (2011). Bioidentical hormone therapy. *Mayo Clinic. Proceedings* 86, 673–80

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., and Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *National Genetics* 18, 168–170

Fiumara, F., Fioriti, L., Kandel, E.R., Hendrickson, W.A. (2010) Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* 143, 1121–1135

Fried, M.G., Bromberg, J.L.(1997) Factors that affect the stability of protein-DNA complexes during gel electrophoresis. Electrophoresis 18, 6–11

Fried, M.G., Liu, G. (1994) Molecular sequestration stabilizes CAP-DNA complexes during polyacrylamide gel electrophoresis. Nucleic Acids Research 22, 5054–5059

Friederici, A.D. (2011) The brain basis of language processing: from structure to function. *Physiological Reviews* 91, 1357-1392

Furney, S.J., Higgins, D.G., Ouzounis, C.A., Lopez-Bigas, N. (2006) Structural and functional properties of genes involved in human cancer. *BMC Genomics* 7, 3

Furui, J., Uegaki, K., Yamazaki, T., Shirakawa, M., Swindells, M.B., Harada, H., Taniguchi, T., Kyogoku, Y. (1998) Solution structure of the IRF-2 DNA-binding domain: a novel subgroup of the winged helix-turn-helix family. *Structure* 6, 491-500

Gajiwala, K.S., Chen,H., Cornille, F., Rogues, B.P., Reith, W., Mach, B., Burley, S.K. (2000) Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403, 916-921

Gohlke, H., Hendlich, M., Klebe, G. (2000) Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology* 2, 337–356
Goldfarb, R.A, Saidel, L.J., Mosovich, E. (1951) The ultraviolet absorption spectra of proteins. *Journal Biological Chemistry* 193, 397-404

Gong, X., M. Jia, Y. Ruan, M. Shuang, J. Liu, S. Wu, Y. Guo, J. Yang, Y. Ling, X. Yang and D. Zhang (2004). Association between the FOXP2 gene and autistic disorder in Chinese population. *American Journal of Medical Genetics B Neuropsychiatry 127B*, 113-116

Gormally, M., Dexheimer, T., Marsico, G., Sanders, D.A.,Lowe, C., Matak-Vinkovic, D.,Michael, S., Jadhav, A., Rai, G., Simeonoy, A., Balasubramanian, S. (2014) Suppression of the FOXM1 transcriptional programme via novel small molecule inhibition. Nature Communications 12, 5165

Graham, D.G. (1978) Oxidative pathways for catecholamines in the genesis of neuromelanin and cytotoxic quinones. *Molecular Pharmocology* 14, 633-643

Greenfield, N. (2009) Using circular dichroism to estimate protein secondary structure. *Nature Protocols* 1, 2876-2890
Greenfield, N., Fasman, G.D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* 8, 4108–4116

Groszer, M., Keays, D.A., Deacon, R.M., de Bono. J.P., Prasad- Mulcare, S., Gaub, S., Baum, M.G., French, C.A., Nicod, J., Coventry, J.A. (2008) Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Current Biology 18*, 354-362

Haber, S. N. (2014). The place of dopamine in the cortico-basal ganglia circuit. *Neuroscience* 282, 248–257. doi: 10.1016/j.neuroscience.2014.10.008

Haesler, S., Wada, K., Nshdejan, E., Morrisey, E., Lints, T., Jarvis, E.D., Scharff, C. (2004) FoxP2 expression in avian vocal learners and non-learners. Journal of Neuroscience 24, 3164-3175

Haining, R. L. Achat-Mendes, C. (2017) Neuromelanin, One of the most overlooked molecules in modern medicine, in not a spectator. *Neural Regeneration Research* 12, 372-375.

Hannenhalli, S., Kaestner, K. H. (2009). The evolution of Fox genes and their role in development and disease. *Nature reviews. Genetics*, *10*, 233–240. https://doi.org/10.1038/nrg2523

Hara, E., Kubikova, L., Hessler, N.A., Jarvis, E.D. (2007) The role of the midbrain dopaminergic system in modulation of vocal brain activation by social context. *European Journal of Neuroscience* 25, 3406-3416

Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y., Wang, L., Lupyan, D., Dahlgren, M.K., Knight, J.L., Kaus, J.W., Cerutti, D., Krilov, G., Jorgensen, W.L., Abel, R., Friesner, R.A. (2016) OPLS3: a force field

providing broad coverage of drug-like small molecules and proteins. *Journal of Chemical Theory and Computation* 12, 281-296

Hauser, D. N., Dukes, A. A., Mortimer, A. D., Hastings, T. G. (2013) Dopamine quinone modifies and decreases the abundance of the mitochondrial selenoprotein glutathione peroxidase 4. *Free Radical Biology and Medicine* 65, 419–427

Hellman, L. M., Fried, M. G. (2007) Electrophoretic mobility shift assay (EMSA) for protein-nucleic acid interaction. Nature Protocols 2, 1849-1861

Hirsch, E., Graybiel, A.N., Agid, Y.A. (1988) Melanized dopaminergic neurons are differentially suceptible to degeneration in Parkinson's disease. *Nature* 334, 345-348

Hirsch, E.C., Brandel, J.P., Galle, P., Javoy-Agid, E., Agid, Y. (1991) Iron and aluminium increase in the substantia nigra of patients with Parkinson's disease: a X-ray microanalysis. *Journal of Neurochemistry* 56, 446-451

Holzwarth, G., Doty, P. (1965) The ultraviolet circular dichroism of polypeptides. *Journal of the American Chemical Society* 87, 218–228
Jackson, B.C., Carpenter, C., Nebert, D., Vasiliou, V. (2010) Update of human and mouse forkhead box (FOX) gene families. *Human Genomics 4*, 324

Jackson, B.C., Carpenter, C., Nebert, D.W., Vasiliou, V (2010). Update of human and mouse forkhead box (FOX) gene families. *Human Genomics* 4, 345–52

Jacob, A., Budhiraja, S., Reichel. R.R. (1997) Differential induction of hnf-3 transcription factors during neuronal differentiation. *Experimental Cell Research* 234, 277–284

Jaenisch, R., Bird, A. (2003) Epigenetic regulation of gene expression: how the gemone integrates intrinsic and extrinsic signals. *Nature Genetics* 33, 245-254

Jenner, P., Dexter, D.T., Sian, J., Schapira, A.H.V., Marsden, C.D. (1992) Oxidative stress as a cause of neuronal deathin Parkinson's disease and incidental Lewy body disease. *Annals of Neurology*. 32, 82-87

Johns, S.A., Soullier, S., Rashbass, P., Cunliffe. V.T. (2005) Foxn1 is required for tissue assembly and desmosomal cadherin expression in the hair shaft. *Developmental Dynamics* 232, 1062–1068

Johnson, B., Hecht, M. (1994) Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Nature Biotechnology* 12, 1357–1360. https://doi.org/10.1038/nbt1294-1357

Kaestner, K.H., Katz, J., Liu, Y., Drucker, D.J., Schutz, G. (1999) Inactivation of the winged helix transcription factor hnf3alpha affects glucose homeostasis and islet glucagon gene expression in vivo. *Genes and Development 13*, 495– 504

Kaestner, K.H., Knochel, W., Martinez, D.E. (2000) Unified nomenclature for the winged helix/forkhead transcription factors. *Genes and Development* 14, 142-146

Kaestner, K.H., Schutz, G., Monaghan, A.P. (1996) Expression of the winged helix genes fkh-4 and fkh-5 defines domains in the central nervous system. *Mechanisms of Development* 55, 221–230

Kastner, A., Hirsch, C., Lejeune, O., Javoy-Agid, F., Rascol, O. Agid, Y. (1992) Is the vulnerability of neurons in thesubstantia nigra of patients with Parkinson's disease related to their neuromelanin content? *Journal of Neurochemistry* 59, 1080-1089

Kaufman, E., Knochel, W. (1996) Forkhead transcription factors: key players in development and metabolism. *Developmental Biology* 250, 1-23

Klotz, L.O., Ramos, C.S., Prieto-Arroyo, I.P., Urbanek, P.,Steinbrenner, H., Monsalve, M. (2015) Redox regulation of FoxO transcription factors. *Redox Biology* 5, 51-72

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

Lai, E., Prezioso, V.R., Smith, E., Livin, O., Costa, R.H., Damell, J.E. (1990) HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes and Development* 4, 1427-1436

Larroux, C., Luke, G.N., Koopman, P., Rokhsar, D.S., Shimeld, S.M., Degnan, B.M. (2008) Genesis and expansion of metazoan transcription factor gene classes. *Molecular Biology and Evolution 25*, 980-996

Lathe, G. H., Ruthven, C. R. J. (1955) The separation of substances on the basis of their molecular weights, using columns of starch and water. *Biochemical Journal* 60, 4

LaVoie, M. J., Ostaszewski, B. L., Weihofen, A., Schlossmacher, M. G., Selkoe, D. J. (2005) Dopamine covalently modifies and functionally inactivates parkin. *Natmed Medical Defence* 11, 1214–1221

Lehmann, O.J., Sowden, J.C., Carlsson, P., Jordan, T., Bhattacharya, S.S. (2003) Foxs in development and disease. *Trends in Genetics 19*, 339-344

Lelli, K.M., Slattery, M., Mann, R.S (2014). Disentangling the many layers of eukaryotic transcriptional regulation. *Annual Review of Genetics 46, 43-68*

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 interaction. *Molecular Cell Biology 24*, 809-822

Lightfoot, D. (2010) Language acquisition and language change. *Wiley Interdisciplinary Reviews: Cognitive Science* 1, 677-684

Littlefield, O., Nelson, H.C. (1999) A new use for the 'wing' of the 'winged' helix-turn-helix motif in the HSF-DNA cocrystal. *Nature Structural Biology* 6, 464-470. doi:10.1038/8269

Liu, J., W, Z.H., Luo, G.A., Li, Q.W., Sun, H.W. (2002) The interaction of DNA with Dopamine by spectroscopic and electrochemical methods. *Analytical Sciences 18*, 751-755

Luscombe, N. M., Laskowski, R.A., Thornton, J.M. (2001). Amino acid-base interactions: a threedimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Research* 29, 2860- 2874.

Luscombe, N.M., Austin, S.E., Berman, H.M., Thornton, J.M. (2001) An overview of the structures of protein-DNA complexes. *Genome Biology* 1, 1

Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiology Reviews* 60, 512–538

Mandel-Gutfreund, Y., O. Schueler and H. Margalit (1995). Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. *Journal of Molecular Biology* 253, 370-382
Mangan, S., Alon, U. (2003) Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America* 100, 11980–11985

Marcus, R.C., Shimamura, K., Sretavan, D., Lai, E., Rubenstein, J.L., Mason, C.A. (1999) Domains of regulatory gene expression and the developing optic chiasm: correspondence with retinal axon paths and candidate signaling cells. *Journal of Comparative Neurology 403*, 346–58

McConkey, B.J., Sobolev, V., Edelman, M. (2002) The performance of current methods in ligand-protein docking. *Current Science* 83, 845–855
McEwen, B.S., Alves, S.E. (1999) Estrogen actions in the central nervous system. *Endocrinology Review* 20, 279–307

Meng, X. Y., Zhang, H. X., Mezei, M., Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design 7*, 146–157.
https://doi.org/10.2174/157340911795677602

Miles, A. J., Wallace, B. A. (2016) Circular dichroism spectroscopy of membrane proteins. *Chemical Society Reviews 54,* 4859-4872.
doi: 10.1039/C5CS00084J

Miller, C.J., Davidson, L.A. (2013) The interplay between cell signalling and mechanics in developmental processes. *Nature Reviews Genetics* 14, 733–744

Miller, J.L., Macedonia, C., Sonies, B.C. (2006) Sex differences in prenatal oral-motor function and development. *Developmental Medicine and Child Neurology* 48, 465-470

Mirny, L. A. (2002). Structural analysis of conserved base pairs in protein-DNA complexes. *Nucleic Acids Research* 30, 1704-1711

Missale C, Nash SR, Robinson SW, Jaber M, Caron MG. (1998) Dopamine receptors: from structure to function. *Physiological Reviews* 78, 189–225

Moitessier, N., Englebienne, P., Lee, D., Lawandi J., Corbeil, C.R. (2008) Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British Journal of Pharmocology* 153, 7–26

Moldoveanu, C. S., David, V. (2013) Essentials in Modern HPLC Separations, Elsevier, Chapter 3: Equilibrium Types in HPLC, 85-114.
https://doi.org/10.1016/B978-0-12-385013-3.00003-3

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *Journal of Computational Chemistry* **16**, 2785-91.

Motiejunas, D., Wade, R.C. (2007) Computer-assisted drug design. *Comprehensive Medicinal Chemistry II*

Murugan, M., Harward, S., Scharff, C., Mooney, R. (2013) Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. *Neuron* 6, 1464-1476

Myatt, S.S., Lam, E.W. (2007) The emerging roles of forkhead box (Fox) proteins in cancer. *Nature Reviews Cancer 7*, 847–59

Nagashima, T., Shigematsu, N., Maruki, R., Urano, Y., Tamaka, H., Shimava, A., Shimokawa, T., Shibasaki, M. (2010) discovery novel forkhead box O1 inhibitors for treating type 2 Diabetes: Improvement of Fasting Glycemia in Diabetic ***db/db*** Mice.Molecular Pharmocology 78, 961-970

National Center for Biotechnology Information. PubChem Compound Database;CID=5757, https://pubchem.ncbi.nlm.nih.gov/compound/5757

National Center for Biotechnology Information. PubChem Compound Database; CID=6013, https://pubchem.ncbi.nlm.nih.gov/compound/6013

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

Norris, E. H., Giasson, B. I., Hodara, R., Xu, S., Trojanowski, J. Q., Ischiropoulos, H., Lee, V. M. (2005) Reversible inhibition of alpha-synuclein fibrillization by dopaminochrome-mediated conformational alterations. *Journal of Biological Chemistry* 280, 21212–21219

Nuñez, J.L.,, Alt, J.J., McCarthy, M.M. (2003) A new model for prenatal brain dadma. I. GABAA receptor activation inducescell death in developing rat hippocampus. *Experimental Neurology 181, 258-269*

O'Neil, M.J. (2006) The Merck Index - An Encyclopedia of Chemicals, Drugs, and Biologicals. Whitehouse Station, NJ: Merck and Co., Inc., 635

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

Ong, C.T., Corces, *V.G. (2011)* Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Review Genetics 12, 283-293*

Pabo, C. O. Nekludova, L. (2000). Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *Journal of Molecular Biology 301*, 597-624.

Paris, I., Perez-Pastene, C., Cardenas, S., Iturria, P., Munoz, P., Couve, E., Caviedes, P., Segura-Aguilar, J. (2010) Aminochrome induces disruption of actin, alpha-, and beta-tubulin cytoskeleton networks in substantia-nigra-derived cell line. *Neurotoxic Research* 18, 82– 92

Patzelt, T., Keppler, S.J., Gorka, O., Thoene, S., Wartewig, T., Reth, M., Forster, I., Lang, R., Buchner, M., Ruland, J. (2018) Foxp1 controls mature B cell survival and the development of follicular and B-1 B cells. *Proceedings of the National Academy of Science U S A* 115, 3120–3125

Paulus, W., Schomburg, E.D. (2006) Dopamine and the spinal cord in restless leg syndrome: does the spinal cord physiology reveal a basis for augmentation. *Sleep Medicine Reviews* 10, 185- 196

Perkins, S.J. (1986) Protein volumes and hydration effects. The calculations of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *European Journal of Biochemistry* 157, 169–180

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

Prasad, S., Mandal, I., Singh, S., Paul, A., Mandal, B., Venkatramani, R., Swaminathan, R. (2017) Near UV-Visible electronic absorption originating from charged amino acids in a monomeric protein. *Chemical science*, *8*, 5416–5433. doi:10.1039/c7sc00880e

Pratt, T., Tian, N.M., Simpson, T.I, Mason, J.O., Price, D.J. (2004) The winged helix transcription factor foxg1 facilitates retinal ganglion cell axon crossing of the ventral midline in the mouse. *Development 131*, 3773– 3784

Prota, G. (1988) Progress in the chemistry of melanins and related metabolites. *Medical Research Reviews* 8, 525-556

Prota, G. (1992) Melanin and Melanogenesis (Academic Press, San Diego, California).

Prota, G., D'Ischia, M. Napolitano, A. (1998). The chemistry of melanin and related metabolites (Chapter 24). The pigmentary system. Oxford Univ. Press, pp. 307-332.

Puspita, L., Chung, S. Y., Shim, J. W. (2017) Oxidative stress and cellular pathologies in Parkinson's disease. *Molecular Brain* 10, 53

Raghanti, M.A., Edler, M.K., Stephenson, A.R., Wilson, L.J., Hopkins, W.D., Ely, J.J. Erwin, J.M., Jacobs, B., Hof, P.R., Sherwood, C.C. (2016). Human-specific increase in dopaminergic innervation in a striatal region associated with speech and language: a comparative analysis of the primate basal ganglia. Journal of Comparative Neurology 524, 2117-2129

Rappe, A. K., Casewits, C. J., Colwell, K. S., Goddard, W. A. III Skidd, W. M. (1992) UFF, a full periodic table force field for molecular mechanics

and molecular dynamics simulations. *Journal of the American Chemical Society 14*, 10024-10039

Reynolds, J. A., Tanford, C. (1970). Binding of dodecyl sulfate to proteins at high binding ratios. Possible implications for the state of proteins in biological membranes. *Proceedings of the National Academy of Sciences of the United States of America 66*, 1002–1007. https://doi.org/10.1073/pnas.66.3.1002

Riley, P.A. (1992) Materia melanica: Further Dark Thoughts. *Pigment Cell Research* 5, 101-106

Rizo, J. (2018) Mechanism of neurotransmitter release coming into focus. *Protein Science* 27, 1364-1391

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

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

Ruiz, A. Altaba, I., Prezioso, V.R., Darnell, J.E., Jessell, T.M. (1993) Sequential expression of hnf-3 beta and hnf-3 alpha by embryonic organizing centers: the dorsal lip/node, notochord and floor plate. *Mechanisms of Development* 44, 91– 108

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

Sasaki, H., Hogan, B.L. (1993) Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development 118*, 47–59

Sastry, G.M., Adzhigirey, M., Day, T., Annabhimoju, R., Sherman, W., (2013) Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* **27,** 221-234

Schulster, M., Bernie, A. M., Ramasamy, R. (2016). The role of estradiol in male reproductive function. *Asian Journal of Andrology*, *18*, 435–440. Seeman, M.V. *(1996)* The role of estrogen in schizophrenia. *Journal of Psychiatry and Neuroscience 21, 123-7*

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

Shen, W., Scearce, L.M., Brestelli, J.E., Sund, N.J., Kaestner, K.H. (2001) Foxa3 (hepatocyte nuclear factor 3gamma) is required for the regulation of hepatic glut2 expression and maintenance of glucose homeostasis during a prolonged fast. *Journal of Biological Chemistry 276,* 42812-42817

Shih, D.Q., Navas, M.A., Kuwajima, S., Duncan, S.A. Stoffel, M. (1999) Impaired glucose homeostasis and neonatal mortality in hepatocyte

nuclear factor 3apha-deficient mice. *Proceedings of the National Academy of Sciences 96*, 10152-19157

Shu, W., Cho, J.Y., Jiang, Y., Weisz, D., Elder, G.A., Schmeidler, J., De Gasperi, R., Sosa, M.A., Rabidou, D., Santucci, A.C., Peri, D., Morrisey, E., Buxbaum, J.D. (2005) Altered ultrasonic vocalization in mice with a disruption in FOXP2 gene. *Proceedings of the National Academy of Science USA* 102, 9643-9648

Simon, I., Barnett, J., Hannett, N., Harbison, C. T., Rinald, N. J., Volkert, T. L., Wyrick, J. J., Zeitlinger, J., Gifford, D. K., Jaakkola, T. S., Young, R. A. (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* 106, 697–708

Smith, C.L., O Malley, B.W (2004) Coregulator function: a key to understanding tissue specificity of selective receptor modulators. *Endocrinology Reviews* 25, 45-71

Solano, F., Hearing, V. J., García-Borrón, J. C. (1999). *Neurotoxicity due to o-Quinones: Neuromelanin formation and possible mechanisms for o-Quinone detoxification. Neurotoxicity Research 1, 153–169.* doi:10.1007/bf03033287

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

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

Spiteri, E., Konopka, G., Coppola, G. (2007) Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *American Journal of Human Genetics 81*, 1144-1157

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., Vernes, S.C., Fisher, S.E., Ren, B., Geschwind, D.H. (2007). Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *American Journal of Human Genetics* 8, 11144–1157

Spitz, F., Furlong, E.E. *(2012)* Transcription factors: from enhancer binding to developmental control. *Nature Review Genetics13, 613-26*

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

Svasti, J., Panijpan, B. (1977). *SDS-polyacrylamide gel electrophoresis. A simple explanation of why it works. Journal of Chemical Education 54, 560.* doi:10.1021/ed054p560

Takahashi, H., Takahashi, K., Liu, F.C. (2009) FOXP genes, neural development, speech and language disorders. *Madame Curie Bioscience Database Austin (TX)* https://www.ncbi.nlm.nih.gov/books/NBK7023/

Takahashi, K., Liu, F.C., Hirokawa, K., Takahashi, H. (2003) Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. *Journal of Neuroscience Research* 73, 61-72

Teichmann, M., Dumay-Odelot, H. Fribourg, S. (2012) Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes, *Transcription* 3, 2-7, doi: 10.4161/trns.3.1.18917

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

Teufel, A., Wong, E.A., Mukhopadhyay, M., Malik, N., Westphal, H. (2003) FoxP4, a novel forkhead transcription factor. *Biochimica Biophysica Acta*. 1627, 47–52

Tolosa, A., J. Sanjuan, A. M. Dagnall, M. D. Molto, N. Herrero, de Frutos, R. (2010). FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. *BMC Medical Genetics* 11, 114

Trott, O., Olson, A.J. (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31, 455-461

Tse, D.C., McCreery, R.L., Adams, R.N. (1976) Potential oxidative pathways of brain catecholamines. Journal of Medicinal Chemistry 19, 37-40

Uttley, J. (2019) Power analysis, sample size, and assessment of statistical assumptions—improving the evidential value of lighting research. LEUKOS 15:2-3, 143-162, DOI:10.1080/15502724.2018.1533851

Uva, V., Sfondrini, L., Triulzi, T., Casalini, P., Tagliabue, E., Balsari, A. (2015) FOXP3 expression in tumor cells and its role in cancer progression. *Atlas of Genetics Cytogenetics Oncology and Haematology 19*, 234–9

Van Laar, V. S., Mishizen, A. J., Cascio, M., Hastings, T. G. (2009) Proteomic identification of dopamine-conjugated proteins from isolated rat brain mitochondria and SH-SY5Y cells. *Neurobiology of Disease* 34, 487–500

Van Rhijin, J.R., Fisher, S.E., Venes, S.C., Nadif Kasri, N. (2018) Foxp2 loss of function increases striatal direct pathway inhibition via increased GABA release. Brain Structure and Function 223, 42111-4226. doi: 10.1007/s00429-018-1746-6

Vernes, S., Spiteri, E., Nicod, J. (2007) Highthroughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *American Journal of Human Genetics* 81, 1232-1250

Walker, R.M., Hill, A.E., Newman, A.C., Hamilton, G., Torrance, H.S., Anderson, S.M., Ogawa, F., Derizioti, P., Nicod, J., Vernes, S.C., Fisher, S.E., Thomson, P.A., Porteous, D.J., Evans, K.L. (2012) The DISC1 promotoer: characterization and regulation by FOXP2. *Human Molecular Genetics* 13, 2862-2872

Wang, B., Lin, D., Li, C., Tucker, P. (2003) Multiple domains define the expression and regulatory properties of FoxP1 forkhead transcriptional repressors. *Journal Biological Chemistry 278*, 24259-24268

Wang, H. Y., Sun, Y., Tang, B. (2002) Study on the fluorescent property of dopamine and determination of dopamine by fluorimetry. *Talanta* 57, 899-907

Wang, H.U., Sun, Y., Tang, B. (2002) Study on fluorescence property of dopamine and determination of dopamine by fluorimetry. *Talanta,* 57, 899-907

Wang, J., Lu, J., Gu, G., and Liu, Y. (2011) In vitro DNA-binding profile of transcription factors: methods and new insights. *Journal of Endocrinology 210*, 1527

Webb, H., Steeb, O., Blane, A., Rotherham, 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* 162, 45-54

Weigel, D., Jackle, H. (1990) The forkhead domain, a novel DNA-binding motif of eukaryotic transcription factors? *Cell 63*, 445-456

Weigel, D., Jurgens, G., Kuttner, F., Seifert, E., Jackle, H. (1989) The homeotic gene forkhead encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo*. Cell 57*, 645-658

Wetlaufer, D.B. (1962) Ultraviolet spectra of proteins and amino acids. *Advances in Protein Chemistry* 17. 303-391

Whalen, R.E., Olsen, K.L *(1981).* Role of aromatization in sexual differentiation: effects of prenatal ATD treatment and neonatal castration. *Hormones and Behavior, 107-22.*

Wilczok, T., Stepien, K., Dzierzega-Lecznar, A., Zajdel, A. Wilzok, A. (1998) Model neuromelanin as antioxidativeagents during lipid peroxidation. INABIS'98 Internet World Congress, Neuroscience Section.

Wingender, E., Schoeps, T., Donitz, J. (2013) TFClass: an expandable hierarchical classification of human transcription factors. *Nucleic Acids Research* 41, 165–70.

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A.D., Stroud, J.C., Bates, D.L., Guo, L., Han, A., Ziegler, S.F., Mathias, D., Benoist, C., Chen, L., Rao, A. (2006). FOXP3 controls regulatory T cell function through cooperation with NFAT. Cell 126, 375–87

Yilmaz, A., Grotewold, E. (2010) Components and mechanisms of regulation of gene expression. *Methods in Molecular Biology* 674, 23-32

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

Zafar, K. S., Siegel, D., Ross, D. (2006) A potential role for cyclized quinones derived from dopamine, DOPA, and 3,4- dihydroxyphenylacetic acid in proteasomal inhibition. *Molecular pharmacology* 70, 1079–1086

Zhang, M., Aman, P., Grubb, A., Panagiopoulos, I., Hindemith, A., Rosengren, E. and Rorsman, H. (1995) Cloning and sequencing of a cDNA encoding rat D-dopachrome tautomerase. *FEBS Letters* 373, 203-206

Zhang, S., Wang, R., Wang, G. (2018) *Impact of dopamine oxidation on dopaminergic neurodegeneration. ACS Chemical Neuroscience 10, 945-953.*doi:10.1021/acschemneuro.8b00454

Zheng, W.H., Kar, S., Quirion. R. (2002) Insulin-like growth factor-1-induced phosphorylation of transcription factor fkhrl1 is mediated by phosphatidylinositol 3-kinase/akt kinase and role of this pathway in insulin-like growth factor-1-induced survival of cultured hippocampal neurons. *Molecular Pharmacology 62*, 225–233

Zhou, B., Zhong, Q., Minoo, P., Li, C., Ann, D.K., Frenkel, B., Morrisey, E.E., Crandall, E.D., Borok, Z. (2008) Foxp2 inhibits Nkx2.1-mediated transcription of SP-C via interaction with the Nkx2.1 homeodomain. *American Journal of Respiratory Cell and Molecular Biology* 38, 750-758

Zucca, F. A., Segura-Aguilar, J., Ferrari, E., Munoz, P., Paris, I., Sulzer, D., Sarna, T., Casella, L., Zecca, L. (2017) Interactions of iron, dopamine and neuromelanin pathways in brain aging and Parkinson's disease. *Progress in Neurobiology* 155, 96–119

6. SUPPLEMENTARY INFORMATION

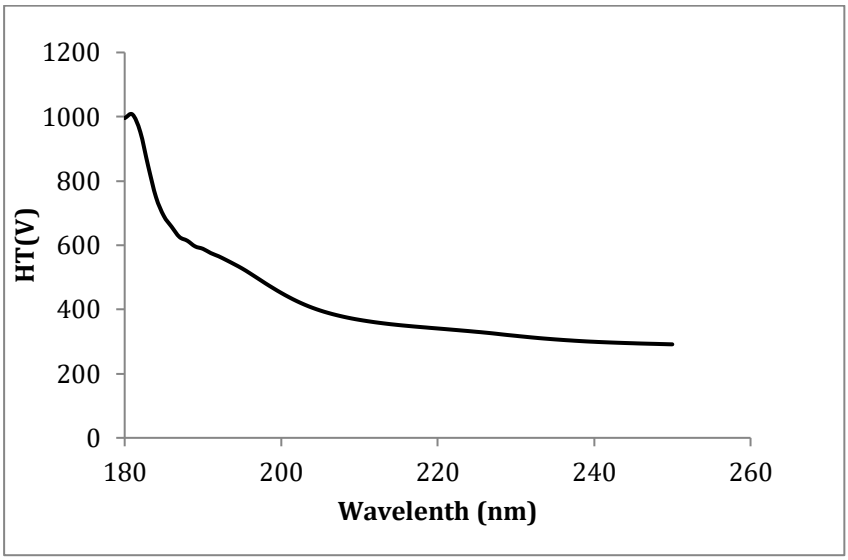


Figure 6.1: Graph Illustrating the Relationship Between Wavelength and High Tension Voltage Corresponding to the Obtained CD Spectrum for Characterisation of the FOXP2 FHD (figure 3.2 b)

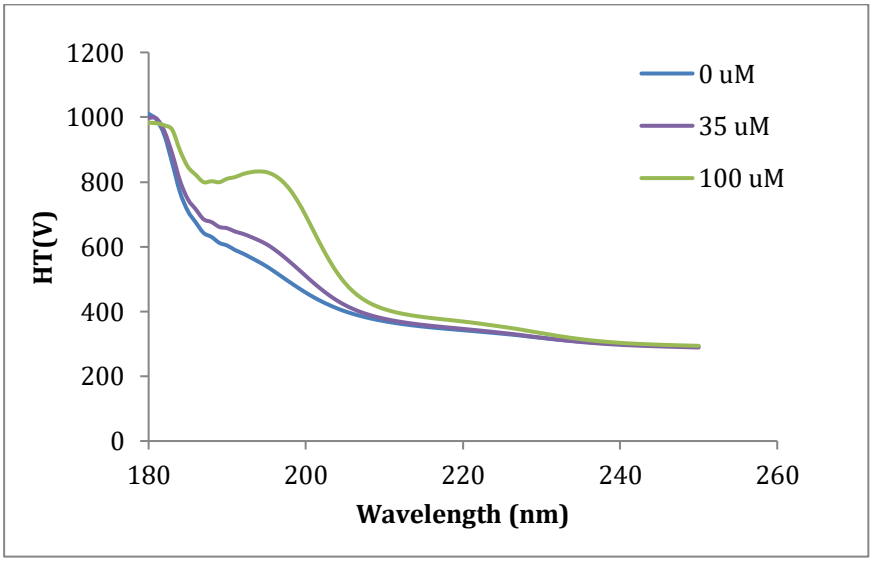


Figure 6.2: Graph Illustrating the Relationship between Wavelength and High Tension Voltage Corresponding to the Obtained CD Spectrum for Characterisation of the FOXP2 FHD in the Absence and Presence of Dopamine (figure 3.4 (A)) The spectrum corresponding to the absence of dopamine is represented in blue, while the spectra obtained in the presence of 35 µM dopamine and 100 µM dopamine are illustrated in purple and green respectively

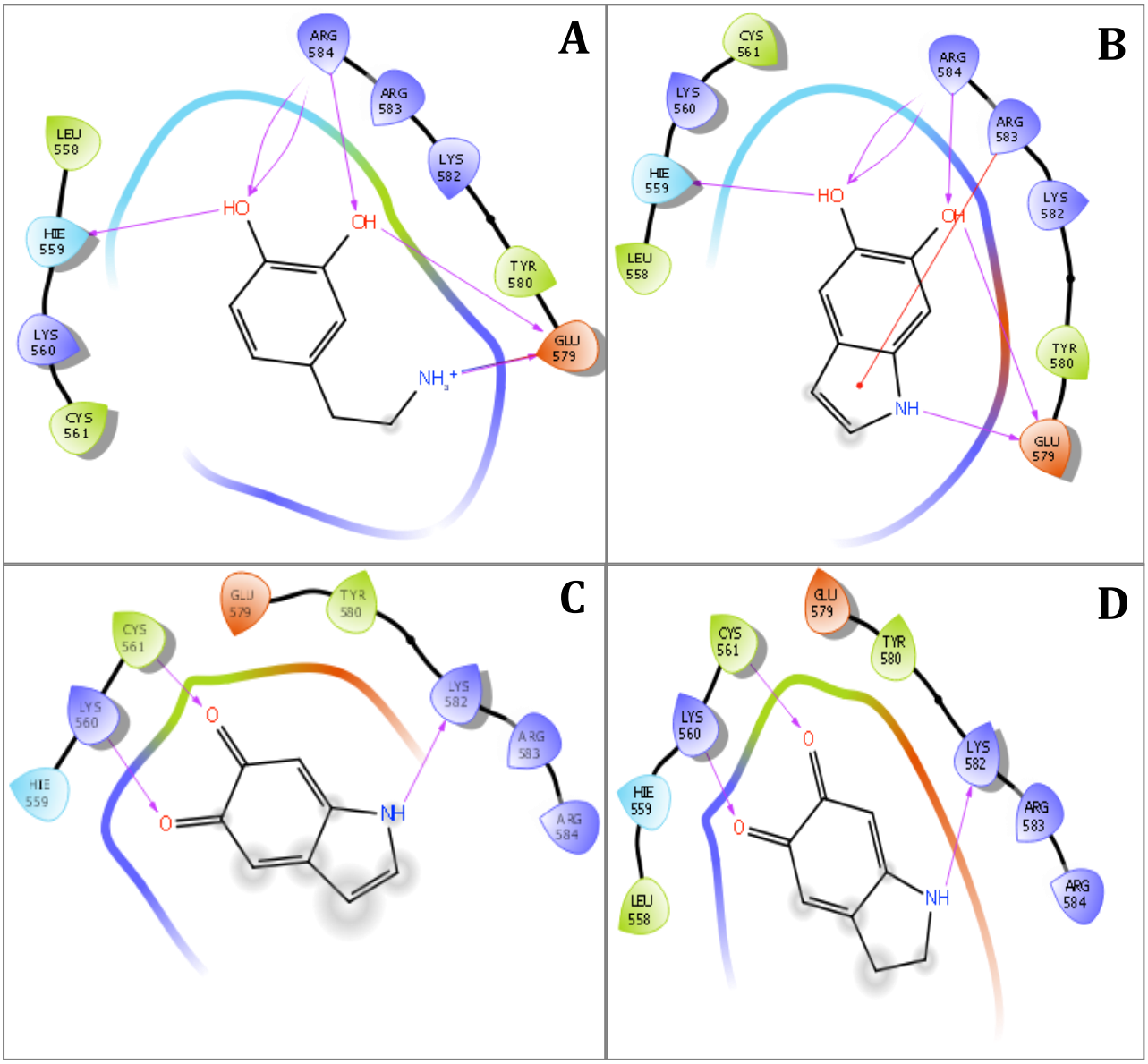


Figure 6.3.1: Ligand Diagrams generated by Schrödinger illustrating the interaction between monomeric FOXP2 FHD and ligands. Purple arrows indicate predicted H-bonds **(A)** Predicted key residues and bonds associated with the interaction between monomeric FOXP2 FHD and dopamine **(B)** Predicted key residues and bonds associated with the interaction between monomeric FOXP2 FHD and 5,6-dihydroxyindole **(C)** Predicted key residues and bonds associated with the interaction between monomeric FOXP2 FHD and 5,6-indolequinone **(D)** Predicted key residues and bonds associated with the interaction between monomeric FOXP2 FHD and dopaminechrome.

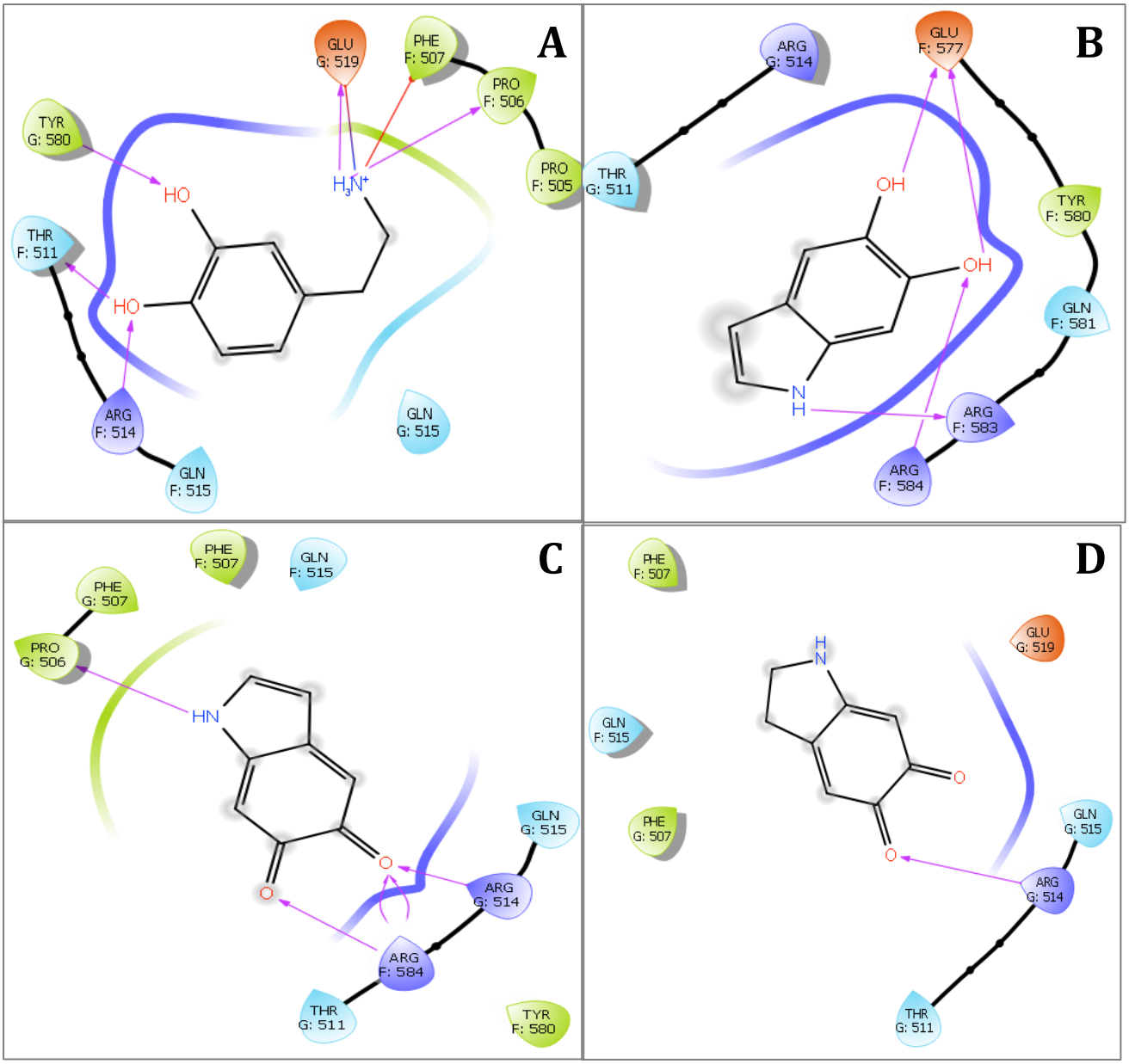


Figure 6.3.2: Ligand Diagrams generated by Schrödinger illustrating the interaction between dimeric FOXP2 FHD and ligands. Purple arrows indicate predicted H-bonds **(A)** Predicted key residues and bonds associated with the interaction between dimeric FOXP2 FHD and dopamine **(B)** Predicted key residues and bonds associated with the interaction between dimeric FOXP2 FHD and 5,6-dihydroxyindole **(C)** Predicted key residues and bonds associated with the interaction between dimeric FOXP2 FHD and 5,6-indolequinone **(D)** Predicted key residues and bonds associated with the interaction between dimeric FOXP2 FHD and dopaminechrome.