Identification and characterisation of the interaction between FOXP2 and the ligand-binding domain of oestrogen receptor α

by

Aasiya Lakhi

(1422723)

Thesis

Submitted in fulfilment of the requirements for the degree

Philosophiae Doctor

in

Molecular and Cell Biology

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

Supervisor: Dr S. Fanucchi

October 2023

DECLARATION

I, Aasiya Lakhi (student number: 1422723), am a student registered for the degree of Doctor of Philosophy in Biochemistry and Cell Biology (MSc) at the University of the Witwatersrand.

I hereby declare the following:

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

11/10/2023

RESEARCH OUTPUTS

Conference and poster presentations

12th Cross Faculty Postgraduate Symposium:

- Poster Presentation
- The FOXP2 – ER1 association and its effect on the FOXP2's DNA's binding activity
- Aasiya Lakhi and Sylvia Fanucchi
- 26 – 30 July 2021

Molecular Biosciences Research Thrust Research Day 2021:

- GradFlash Presentation (2nd place, PhD/Post Doc division)
- The Oestrogen receptor α and FOXP2: Investigating a possible interaction and its regulation
- Aasiya Lakhi and Sylvia Fanucchi
- 9 December 2021

The South African Society for Biochemistry and Molecular Biology Virtual Congress:

- Poster Presentation
- Oestrogen receptor α and FOXP2: Investigating a possible interaction and its regulation
- Aasiya Lakhi and Sylvia Fanucchi
- 23-26 January 2022

Molecular Biosciences Research Thrust Research Day 2022:

- Poster Presentation
- The FOXP2 – ER1 association and its effect on the FOXP2's DNA-binding activity
- Aasiya Lakhi and Sylvia Fanucchi
- 2 December 2022

ABSTRACT

Forkhead box P2 (FOXP2) regulates the expression of various genes and is associated with language and speech, neural development and outgrowth, and cancer. As transcription factors rarely function in isolation, this study aims to investigate whether FOXP2 directly associates with oestrogen receptor α (ER1), a nuclear receptor responsible for sexual differentiation and cancer progression and outcome. The association between ER1 and FOXP2 was first identified in MCF-7 cells using co-immunoprecipitation. Thereafter, the interaction was characterised biophysically by overexpressing the FOXP2's DNA-binding forkhead domain (FHD) and N-terminal region (NT), and ER1's ligand-binding domain (LBD) in *E. coli* cells. Isothermal titration calorimetry and fluorescence anisotropy were used to investigate the thermodynamic parameters and regulation of interaction between FOXP2 FHD and ER1 LBD, respectively. Electrophoretic mobility shift assay was used to investigate the effect of the interaction on FOXP2's DNA binding ability. Following the successful overexpression and purification of all three proteins, ER1 LBD was found to interact with FOXP2 FHD but not with FOXP2 NT. The affinity of the ER1 LBD for FOXP2 FHD increases with an increase in salt concentration. ITC shows a similar trend and reveals that the interaction is enthalpically favoured at lower salt concentrations but enthalpically opposed at higher salt concentrations. Additionally, the FOXP2-ER1 LBD interaction remains unaffected by the inclusion of oestrogen, but addition of FOXP2 cognate DNA results in inhibition of the formation of the complex. This research identifies a novel interaction between ER1 LBD and FOXP2 FHD and shows that the interaction is regulated by salt. Moreover, FOXP2 FHD cannot bind to both ER1 LBD and DNA simultaneously suggesting a probable role of this interaction in regulating the transcriptional pathway of FOXP2. This study serves as a foundation for further investigation into the interaction between FOXP2 and ER1 in different cell lines and its relevance in FOXP2-mediated outcomes in cancer and neurodevelopmental disorders.

(Words: 303)

ACKNOWLEDGEMENTS

I am eternally grateful to Dr Sylvia Fanucchi for providing me the opportunity to pursue my post-graduate qualifications and develop a passion for research and proteins. Thank you for helping me find myself and for all the amazing and new experiences I had courtesy of you. The guidance, encouragement as well as your never-ending excitement for science went a long way in helping me complete my research. It was an absolute honour to learn from you.

To my friends and colleagues in the PSFRU, it was a pleasure working alongside you. A special thank you to Heather Donald, Mpho Setshedi, Olalaken Onisuru, and Alicia Joshua for your help, advice, and general support with my research. To Mpho Setshedi, Nozinhle Dlamini and Taryn Adams, you made the difficult days easier and for that I am grateful.

To the FOXy group, it has been a privilege to be a part of such an inspiring and dedicated group. Thank you for including me in the insightful discussions about life and research.

I would like to express my appreciation to Prof Mavri-Damelin, Dr Juan Carlos Penedo as well as Jemma Falkov, Thokozile Makhanya Mistral Sebastian, and Fillmon Kubrom for your willingness to teach me what I knew not. Thank you for your kindness, help and support throughout the collaboration.

My Saleha, anything I say will be too little, but I appreciate you for all that you do, are and your unwavering belief in me. To my mom and dad, Maryam, Soli Papa, and Amina Foi, I thank you for your support, encouragement and prayers.

Finally, I would like to express my gratitude to the University of the Witwatersrand and the National Research Foundation for the financial support.

TABLE OF CONTENTS

DECLARATION ii

ABSTRACT iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF ABBREVIATIONS x

LIST OF FIGURES xii

LIST OF TABLES xiii

1. INTRODUCTION 14

1.1) Transcriptional Regulation 14

1.1.1) Transcription factors 14

1.1.2) Protein – protein interactions 15

1.2) FOXP2 Transcription Factors 18

1.2.1) Superfamily of forkhead box transcription factors 18

1.2.2) FOXP transcription factors 19

1.2.3) Interaction of FOXP2 with DNA 20

1.3) Functions of FOXP2 21

1.3.1) Role in tumorigenesis 22

1.3.2) Speech, language, and clinical relevance 22

1.4) Oestrogen Receptor α 24

1.4.1) Nuclear receptors 24

1.4.2) Genomic and non-genomic activity of oestrogen receptor α: 27

1.4.3) Structural composition 30

1.5) Probable Association Between FOXP2 and ER1 33

1.5.1) A role for FOXP2 and ER1: sexual differentiation of language and speech 34

1.5.2) ER1 and FOXP2 in breast cancer: inference from known ER1-FOXP TF interactions 35

2. RATIONALE 37

3. AIMS 39

4. METHODS 40

4.1) Materials 40

4.2) Protein-protein interaction in cancer cells 40

4.2.1) Co-immunoprecipitation of protein-protein complex 40

4.3) Transformation 45

4.4) Overexpression of proteins 46

4.4.1) FOXP2 forkhead domain 46
4.4.2) FOXP2 N – Terminal region 47
4.4.3) Oestrogen receptor α ligand-binding domain 49
4.5) Purification of proteins 51
4.5.1) Immobilised-metal ion affinity chromatography 51
4.5.2) Size-exclusion chromatography 52
4.5.3) Thrombin Cleavage of FOXP2 53
4.6) Purity and molecular weight determination 53
4.6.1) Sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE) 53
4.7) Protein concentration determination 54
4.7.1) Bradford Assay 54
4.7.2) Measurement of UV absorbance 55
4.8) Structural Characterisation 56
4.8.1) Far-UV circular dichroism 56
4.8.2) Size exclusion chromatography 57
4.9) Fluorescence anisotropy 57
4.9.1) DNA interaction 58
4.9.2) Identification of Protein-protein interaction 58
4.9.3) Regulation of protein-protein interaction 60
4.10) Isothermal Titration Calorimetry 62
4.9.1) Functional characterisation of ER1 LBD's ligand binding activity 63
4.10.2) Analysis of protein-protein interaction 63
4.11) Electrophoretic mobility shift assay 64
4.11.1) Functional characterisation 64
4.11.2) Effect of protein-protein interaction on FOXP2 FHD's functional activity 65
4.11) Molecular Docking 66
4.11.1) AlphaFold 66
4.11.2) ClusPro 66
4.11.3) HADDOCK 67
4.10.4) Complex Structure Analysis 68
5) RESULTS 69
5.1) Protein-protein interaction in MCF-7 cells 69
5.2) Overexpression and purification of Proteins 72
5.2.1) FOXP2 FHD 72
5.2.2) FOXP2 NT 74
5.1.3) ER1 LBD 77

5.3) Protein Quantification 83
5.3.1) FOXP2 FHD and ER1 LBD 83
5.3.2) FOXP2 NT 84
5.4) Quaternary structure of the proteins 85
5.5) Functional characterisation of the proteins 88
5.5.1) FOXP2 FHD 89
5.5.2) FOXP2 NT 89
5.5.3) ER1 LBD 90
5.6) Structural Characterisation 91
5.6.1) FOXP2 FHD 91
5.6.2) FOXP2 NT 92
5.6.3) ER1 LBD 94
5.7) Identification of Protein – Protein Interaction 95
5.8) Effect of Salt on the Protein-Protein Interaction 98
5.8.1) Fluorescence Anisotropy 99
5.8.2) Isothermal Titration Calorimetry 100
5.9) Effect of binding partners on the protein-protein interaction 103
5.10) The effect of ER1 LBD-FHD interaction on FOXP2 FHD's DNA binding activity 105
5.10.1) Electrophoretic mobility shift assay 106
5.10.2) Fluorescence anisotropy 108
5.11) Molecular Docking 110
5.11.1) Docked: Predicted ER1 LBD-FOXP2 FHD structures 110
5.11.2) ER1 LBD-FOXP2 FHD complex in the presence of oestrogen 116
5.11.3) ER1 LBD-FOXP2 FHD complex in the presence of DNA 117
6) DISCUSSION 120
6.1) ER1-FOXP2 association in MCF-7 cells 120
6.2) Structure and functionality of expressed proteins 121
6.3) FOXP2 NT does not interact with ER1 LBD 122
6.4) FOXP2 FHD associates with ER1 LBD *in vitro* 124
6.5) ER1 LBD-FOXP2 FHD interaction is dependent on salt concentration 125
6.6) ER1 LBD-FOXP2 FHD interaction is only enthalpically driven at low salt concentrations 128
6.7) Not one or the other but both: effect of E2 130
6.8) One or the other but not both: effect of DNA 132
6.9) Towards a role in oncogenesis 134
6.10) Limitations of the study 135

7) CONCLUSION 136
8. REFERENCES 138
9) APPENDIX 157

LIST OF ABBREVIATIONS

AF	Activation function
AIR	Ambiguous interaction restraint
APS	Ammonium persulfate
BRCA1	Breast cancer associated protein 1
BSA	Bovine serum albumin
CBD	CREB-binding protein
CNTNAP2	Contactin-associated protein-like-2
CREB	cAMP responsive element binding protein
DBD	DNA-binding domain
DISC1	Disrupted in schizophrenia 1
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EMSA	Electrophoretic mobility shift assay
ER	Oestrogen receptor
FFT	Fast Fourier Transform
FHD	Forkhead domain
GFP	Green fluorescent protein
HADDOCK	High ambiguity driven docking
HR	Hinge region
IMAC	Immobilised-metal affinity chromatography
IP	Immunoprecipitation
ITC	Isothermal titration calorimetry
LBD	Ligand-binding domain
NDSB	Non- detergent sulfobetaine
NT	N-terminal domain
PAE	Predicted aligned error
PBS	Phosphate-buffered saline
PEG	Polyethylene glycol
Plddt	Predicted local distance difference test
PPI	Protein-protein interaction
RIP140	Receptor-interacting protein 140

ROX	RhodamineX
RSK4	X-linked ribosomal S6 kinase 4
RT	Room temperature
SNP	Single nucleotide polymorphisms
SPR	Sushi-repeat protein
TB	Terrific broth
TBST	Tris-buffered saline with tween
TCEP	Tris (2-carboxyethyl) phosphine
TCA	Trichloroacetic acid
TF	Transcription factor
TGF	Transforming growth factor
uPAR	Plasminogen activator receptor of the urokinase type
W	Wing

LIST OF FIGURES

FIGURE 1.1: GENERALISED REPRESENTATION OF THE PROTEIN-PROTEIN INTERACTIONS GOVERNING NUCLEAR RECEPTOR SIGNALLING. 17
FIGURE 1.2: SCHEMATIC REPRESENTATION OF THE FUNCTIONAL DOMAINS OF FOXA, FOXO AND FOXP PROTEINS. 19
FIGURE 1.3: SCHEMATIC REPRESENTATION OF THE FHD STRUCTURE OF A) FOXP2 AND B) FOXA3. 20
FIGURE 1.4: CARTOON REPRESENTATION OF FOXP2 FHD BOUND TO DNA. 21
FIGURE 1.5: GENERAL STRUCTURE OF NUCLEAR RECEPTORS. 26
FIGURE 1.6: OESTROGEN SIGNALLING PATHWAYS. 29
FIGURE 1.7: STRUCTURE OF ER1 LBD. 31
FIGURE 1.8: CARTOON REPRESENTATION OF THE ER1 LIGAND-BINDING DOMAIN IN A) UNLIGANDED AND B) LIGANDED FORM. 32
FIGURE 1.9: INTERACTION OF THE ER1 LIGAND-BINDING DOMAIN WITH A COACTIVATOR. 33
FIGURE 1.10: NUCLEAR RECEPTOR MOTIF SEQUENCE OF THE FOXP1 AND FOXP2 TRANSCRIPTION FACTORS. 36
FIGURE 4.1: PROTEIN-PROTEIN DETECTION WORKFLOW. 43
FIGURE 4.2: SCHEMATIC REPRESENTATION OF THE PET-11A PLASMID. 45
FIGURE 4.3: FOXP2 STRUCTURAL ORGANISATION AND CONSTRUCTS USED FOR RECOMBINANT PROTEIN EXPRESSION. 46
FIGURE 4.4: STRUCTURAL INFORMATION OF THE ER1 LBD 49
FIGURE 4.5: SET UP AND DENSITOMETRIC ANALYSIS OF EMSA GELS. 65
FIGURE 5.1: FOXP2-ER1 INTERACTION IN CELLS. 71
FIGURE 5.2: PURIFICATION PROFILE OF FOXP2 FORKHEAD DOMAIN. 73
FIGURE 5.3: EVALUATION OF FOXP2 NT EXPRESSION TRIALS ON SDS-PAGE GELS. 75
FIGURE 5.4: PURIFICATION OF FOXP2 NT. 77
FIGURE 5.5: EXPRESSION TRIALS OF ER1 LBD. 79
FIGURE 5.6: EXPRESSION TRIAL OF ER1 LBD IN BL21(DE3) CELLS. 81
FIGURE 5.7: IMAC PROFILES OF ER1 LBD LYSED IN DISRUPTION BUFFER A) WITHOUT ANY ADDITIVES AND B) WITH NDSB. 82
FIGURE 5.8: SIZE EXCLUSION PROFILE OF ER1 LBD. 83
FIGURE 5.9: ABSORBANCE SPECTRUM OF A) FOXP2 FHD AND B) ER1 LBD. 84
FIGURE 5.10: QUANTIFICATION OF FOXP2 NT CONCENTRATION. 85
FIGURE 5.11: QUATERNARY STRUCTURE OF A) FOXP2 FHD, B) FOXP2 NT AND C) ER1 LBD. 87
FIGURE 5.12: ROLE OF DISULFIDE BONDS IN DETERMINING ER1 LBD QUATERNARY STRUCTURE. 88
FIGURE 5.13: EMSA EVALUATING THE DNA BINDING ACTIVITY OF A) FOXP2 FHD AND B) FOXP2 NT. 89
FIGURE 5.14: INTERACTION BETWEEN ER1 LBD AND E2. 91
FIGURE 5.15: SECONDARY STRUCTURE CHARACTERISATION OF FOXP2 FHD. 92
FIGURE 5.16: FAR-UV CIRCULAR DICHROISM SPECTRA OF FOXP2 NT. 93
FIGURE 5.17: FAR-UV CD STRUCTURE OF ER1 LBD. 95
FIGURE 5.18: INTERACTION BETWEEN ER1 LBD AND FOXP2. 96
FIGURE 5.19: INTERACTION BETWEEN ER1 LBD AND FOXP2. 98
FIGURE 5.20: EFFECT OF SALT ON THE FOXP2 FHD-ER1 LBD INTERACTION. 100
FIGURE 5.21: THERMODYNAMICS OF THE FOXP2 FHD-ER1 LBD INTERACTION. 102
FIGURE 5.22: COMPARATIVE ANALYSIS OF A) DISSOCIATION CONSTANTS AND B) THERMODYNAMIC PARAMETERS AT DIFFERENT SALT CONCENTRATION. 103
FIGURE 5.23: EFFECT OF THE INCLUSION OF BINDING PARTNERS (E2 OR DNA) ON THE ER1 LBD-FOXP2 FHD INTERACTION. 105
FIGURE 5.24: FOXP2 FHD'S ABILITY TO BIND TO DNA FOLLOWING ITS INCUBATION WITH ER1 LBD. 107
FIGURE 5.25: THE EFFECT OF FOXP2'S INTERACTION WITH NELSON DNA ON THE FOXP2 FHD -ER1 LBD INTERACTION. 109
FIGURE 5.26: ALPHAFOLD STRUCTURE PREDICTION OF THE FOXP2 FHD: ER1 LBD COMPLEX. 111
FIGURE 5.27: STRUCTURE-BASED DOCKING OF ER1 LBD AND FOXP2 FHD. 113

FIGURE 5.28: ALIGNMENT OF ER1 LBD – FOXP2 FHD STRUCTURE PREDICTED USING DIFFERENT DOCKING PROGRAMMES. 114
FIGURE 5.29: ER1 LBD CRYSTAL STRUCTURE IN ALIGNMENT WITH PREDICTED COMPLEX STRUCTURE FROM A) ALPHAFOLD, B) CLUSPRO AND C) HADDOCK. 117
FIGURE 5.30: INTERACTION BETWEEN FOXP2 FHD AND NELSON DNA.. 118
FIGURE 5.31 ANALYSIS OF THE FOXP2 FHD – ER1 LBD INTERACTION IN THE PRESENCE OF DNA USING THE PREDICTED STRUCTURES FROM A) ALPHAFOLD B) CLUSPRO AND C) HADDOCK. 119
FIGURE 6.1: THE DISORDERED REGION OF FOXP2. 123
FIGURE 6.2: ER1 LBD STRUCTURE. A) A FOCUS ON CYSTEINE RESIDUES. 125
FIGURE 6.3: WATER-MATCHING AFFINITIES IN REGULATION OF PROTEIN-PROTEIN INTERACTION IN SALT SOLUTIONS. 127
FIGURE 6.4: CARTOON STRUCTURE OF ER1 LBD. 131
FIGURE S1: INDUCTION TRIALS OF FOXP2 NT AT DIFFERENT OPTICAL DENSITY.. 157
FIGURE S2: ER1-FOXP2 INTERACTION IN THE ABSENCE AND PRESENCE OF DNA. 158

LIST OF TABLES

Table 4.1: Conditions investigated for optimal FOXP2 NT protein overexpression…….…….. 48

Table 4.2: Conditions investigated for optimal ER1 LBD overexpression…….……..…… 50

Table 4.3: Buffers used in the study…….…….……...…… 52

Table 4.4: Fluorescence anisotropy experiments to investigate protein-DNA and protein-protein interaction……...…….…… 62

Table 5.1: Residues involved in ER1 LBD-FOXP2 FHD interaction obtained using different docking programs……...…….…… 115

1. INTRODUCTION

1.1) Transcriptional Regulation

Multicellular organisms are comprised of a large array of cell types with diverse functions even though all cells contain the same genetic information. This diversity in cell types and functions is attributed to differential gene expression, the process by which genetic information encoded in DNA is first transcribed to RNA and then translated to synthesise proteins that serve numerous functions. Tissue-specific gene expression and differential gene activity in response to stimuli is primarily mediated by regulation of gene transcription, a process that is activated or repressed by the activity of transcription factors (TFs), chromatin remodelling proteins and histone-modifying enzymes (Featherstone *et al.*, 2012).

1.1.1) Transcription factors

Transcription factors (TFs) are DNA-binding proteins that regulate transcription by influencing the recruitment of RNA polymerase II at promoter sites. They can be grouped into general and specific TFs. General TFs recognise specific DNA sequences within the core promoter site and are constituents of the pre-initiation complex responsible for basal transcription. General TFs in eukaryotes include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Thomas and Chiang, 2006). Most TFs fall in the category of specific TFs, however, which are required for condition-specific regulation of gene expression. These TFs bind specifically to their co-ordinate DNA sequence located at the promotor or enhancer region of particular genes. They contribute significantly towards determining the identity of a cell as their expression is more restricted as opposed to general TFs that are universally expressed in all cell types.

Specific TFs usually contain a DNA-binding domain (DBD) that is able to recognise short (6-12 bps), specific DNA sequences that are repeated numerously within the cis-regulatory, non-coding enhancer and promoter regions. TFs interact with DNA through non-covalent interactions, primarily hydrogen bonds and van der Waals forces. The interaction between the TF and DNA can be both specific and non-specific. The latter occurs between particular TF residues and the nucleotide bases located within the DNA's core binding sites (Blainey *et al.,* 2009). Non-specific interactions, in contrast, occur between TF residues and the DNA phosphate backbone and allows the TF to slide along the DNA, facilitating the search for the target site (Record *et al.,* 1976; Afek *et al.,* 2014). Both specific and non-specific interactions are necessary for finetuning gene

regulation. Notably, the multilayer transcriptional regulatory system does not only involve direct binding between the TF and the target gene sequence, but also the interaction between TFs and TF-binding proteins. This interaction includes co-factors, chromatin remodelling proteins, inhibitors, and TF modulating proteins (Auboeuf *et al.,* 2002; Rhee *et al.,* 2014; Carnesecchi *et al.,* 2017).

1.1.2) Protein – protein interactions

Proteins, in general, do not act in isolation when performing their function *in vivo* as over 80% of proteins operate in complexes (Berggård *et al.,* 2007). In particular, enzymes and TFs are the kinds of proteins that are involved in many of the interactions among proteins (Dunker *et al.,* 2005; Sarmady *et al.,* 2011). The functional activities of these proteins which are integral for all cellular processes in the body, are either dependent on or influenced by interaction with other proteins. Protein-protein interactions (PPIs) are thus involved in numerous cellular processes including replication, transcription, signal transduction, membrane transport and cell-cell communications (Keskin *et al.,* 2016). Proteins are repeatedly found to interact with each other when they are involved in the same cellular processes (Von Mering *et al.,* 2002). Hence, the study of PPIs can be useful for inferring the functionality of unidentified proteins by using the evidence of its interaction with a protein that has a known function. Additionally, they may be useful in gleaning insights into the molecular mechanisms of cellular processes by expediating the modelling of functional pathways (Rao *et al.,* 2014).

PPIs can be characterised based on the type of oligomeric state formed, the stability of the complex, and the duration of complex formation. PPIs can be classified as forming either homo-oligomeric or hetero-oligomeric complexes. Most homo-oligomers are symmetrical and stable. In contrast, the stability of hetero oligomers varies (Keskin *et al.,* 2016). Additionally, PPIs can be differentiated into either obligate or non-obligate interactions based on the affinity of the interaction. If the individual protein components of a complex exist independently, the interaction is considered to be non-obligate. In contrast, an obligate interaction involves protein components that do not exist as individual stably folded structures. PPIs may also be differentiated into permanent and transient interactions based on the lifetime of the interaction. The interacting proteins of permanent complexes have a greater tendency to be co-expressed and colocalised compared to the interacting proteins in transient complexes (Jansen et al., 2002). Chaperone-

assisted protein folding, signal transduction as well as associations between hormone receptors are classified as transient non-obligate interactions as they allow the cell to respond to extracellular stimuli only when required. Conversely, interactions between antibody and antigens are classified as a permanent non-obligate interaction (Keskin *et al.*, 2016).

Notably, interactions between proteins display temporal and spatial heterogeneity, complicating the mapping of an organism's interactome. The interactions may be transient, with proteins temporarily associating with and dissociating from each other (Nooren and Thornton, 2003). Proteins may also require chemical modifications, such as phosphorylation, for the interaction to occur and are also prone to modifications in their structure in response to their environment. This could directly interfere with the protein's binding preferences. Moreover, proteins that are likely to interact on the basis of their physiochemistry may be constrained by their localisation. This is because proteins found in distinct locations and are not transported to other locations are prevented from interacting (Keskin *et al.*, 2016). Additionally, the expression levels of proteins are tissue-specific and hence their interaction may vary between different cell types.

1.1.2.1) Properties of protein-protein interactions: Interfaces and hotspots

Proteins interact through their surfaces and hence the geometric, physiochemical, and dynamic aspects of protein-protein interfaces have been the focus of numerous studies. Interactions between proteins results in burial of parts of their accessible surface area making the area, known as the interface, inaccessible to solvents (Chen *et al.*, 2013). Interfaces have a standard size ranging between 1200–2000 $A^{\circ 2}$ and this size corresponds to the energy of the protein–protein binding interaction (Keskin *et al.*, 2005; Moreira *et al.*, 2007). The interfaces are frequently comprised of hydrophobic residues including phenylalanine, tryptophan, tyrosine and histidine (Conte *et al.*, 1999). Hydrophobic interactions between nonpolar regions of amino acids ensure tight packing of the residues and cause the water molecules in the interface to be expelled, increasing the entropy, and favouring the formation of protein-protein complexes (Dill, 1990). The energy associated with desolvation (entropy), and the energy produced by each of the van der Waals interactions (enthalpy) results in a substantial increase in free energy, which in turn enhances the stability of the protein-protein complex (Fernández and Scheraga, 2003). PPIs are also mediated by electrostatic forces and hydrogen bonds.

PPIs are essential for fine-tuning gene expression and can do so by regulating TF activity. Differential gene expression is partly mediated through signalling cascades that control the activity of the downstream TFs. Specific TFs can be activated or inhibited through association or dissociation with inhibitory proteins, dimerisation, changes in localisation, ligand-binding, post-transcriptional modifications, or a combination of these approaches (Plevin *et al.*, 2005, Smutny *et al.*, 2013). The importance of the interaction of TFs with other proteins is depicted in Figure 1.1 where a nuclear receptor, dimerised in the presence of ligands, is able to interact with different proteins, thereby mediating the activation and repression of transcription. In the presence of an antagonist, the nuclear receptor binds to corepressor proteins, like the receptor-interacting protein 140 (RIP140), and thus indirectly interacts with histone deacetylases, resulting in the repression of transcription (Plevin *et al.*, 2005). Conversely, the nuclear receptor, when bound to agonists, recruits coactivator proteins, such as cAMP responsive element binding protein (CREB)-binding protein (CBP), to a specific DNA sequence to mediate activation of gene transcription through formation of a transcription initiation complex with general TFs and DNA polymerases.

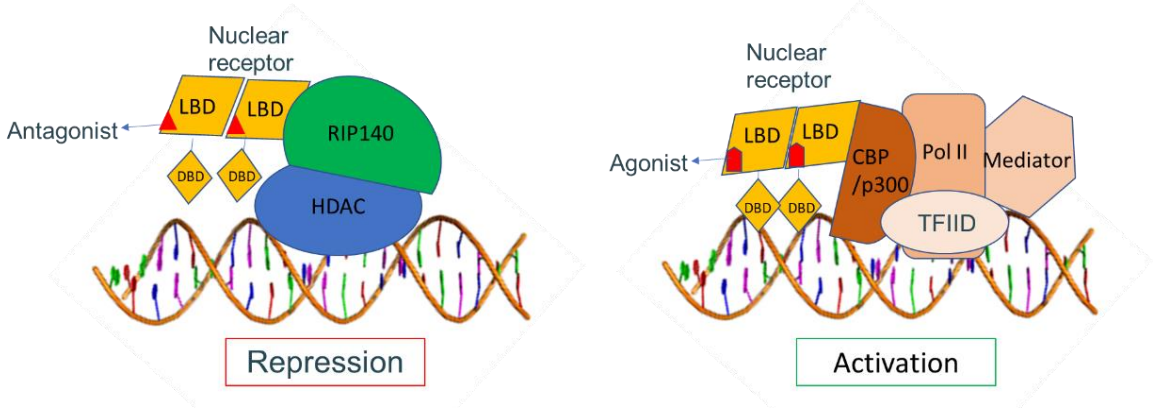


Figure 1.1: Generalised representation of the protein-protein interactions governing nuclear receptor signalling. Ligands induce homo-dimerisation of the nuclear receptor, allowing the nuclear receptor to interact with other proteins. Ligands are associated with distinct conformational changes in the structure of the nuclear receptor as interactions with the antagonist represses transcription through associations with corepressors (RIP140) and histone-modifying enzymes (HDAC). In contrast, agonists promote the recruitment of coactivators (CBP/p300) and ultimately general transcription factors (TF11D), DNA polymerases (Pol II) and mediators to activate gene transcription. CBP: cAMP responsive element binding protein (CREB)-binding protein; DBD; DNA-binding domain; HDAC: histone deacetylases; LBD: ligand-binding domain; RIP140; receptor-interacting protein 140 (Adapted from: Plevin *et al.*, 2005)

Considering the importance of PPIs in transcriptional regulation, the focus of this study is the probable PPI between the FOXP2 TF and the oestrogen receptor α since this interaction could have a role in governing the transcriptional regulation of either of the TFs.

1.2) FOXP2 Transcription Factors

The forkhead box P2 (FOXP2) transcription factor functions predominantly as a transcriptional repressor and is associated with language and speech, and neural development and outgrowth (Shu *et al.,* 2001; Vernes *et al.,* 2008). The *FOXP2* gene was identified in 2001 as the first gene to be implicated in a hereditary speech and language disorder. The TF is expressed in several neuronal populations as well as the lung epithelium, cardiac tissue, and intestine (Shu *et al.,* 2001; Lai *et al.,* 2003).

1.2.1) Superfamily of forkhead box transcription factors

FOXP2 belongs to the evolutionarily conserved family of forkhead box (FOX) proteins. The genomes of mammals are comprised of numerous *FOX* genes as evidenced by the 50 genes identified in the human genome and 44 in the mouse genome (Jackson *et al.,* 2010). FOX proteins have diverse functions in biological processes including organogenesis, metabolism, apoptosis, and cell differentiation, proliferation and longevity (Myatt and Lam, 2007; Hannenhalli and Kaestner, 2009). Mutation and changes in expression profiles of these TFs are thus associated with numerous diseases, including human congenital and neurodevelopmental disorders and cancers (Zhang *et al.,* 2014).

On the basis of structural similarity, FOX proteins are categorised into 19 subclasses (FOXA-FOXS) (Jackson *et al.,* 2010) all of which contain a highly conserved DBD, called the forkhead domain (FHD). The FOXP family, with its four members (FOXP1-FOXP4), is distinct from other FOX proteins. Although the FHD is common to all FOX TFs, it is located at the C-terminal end of the FOXP proteins while it is found in the centre to the N-terminus of the other FOX proteins (Chu *et al.,* 2011). FOXP proteins are the only members of the superfamily to contain a glutamine-rich region and a zinc-finger and leucine-zipper motif (Golson and Kaestner, 2016). The glutamine rich region likely functions as a transcriptional activation/ repression domain and the zinc-finger and leucine-zipper motif is associated with formation of homo- and heterodimers of the FOXP proteins (Shu *et al,* 2001; Wang *et al.,* 2003). The ability of FOXP proteins to form homo and heterodimers sets them apart from most other members of the superfamily. The similarities and

differences between the organisation of the domains in the sequence of FOXA, FOXO and FOXP TFs are shown in Figure 1.2.

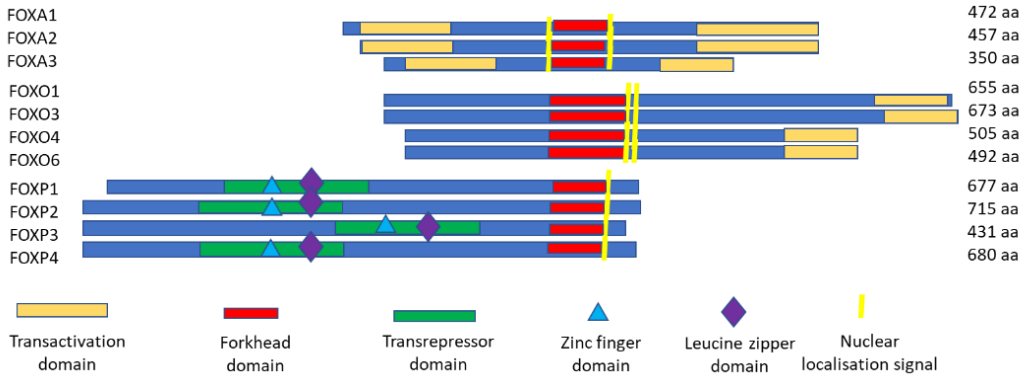


Figure 1.2: Schematic representation of the functional domains of FOXA, FOXO and FOXP proteins. All members of the superfamily contain a forkhead domain (FHD). The FOX proteins are categorised into subfamilies based on their structural similarity in regions that extend beyond the FHD. (Adapted from: Golson and Kaestner, 2016)

1.2.2) FOXP transcription factors

The conserved FOX FHD is largely comprised of three α-helices, three β-sheets and two loops, or wings, that form a helix-turn helix-like motif (Clark *et al.*, 1993). As shown in Figure 1.3, FOXP FHD differs from other FOX proteins at the winged regions as one of the wings (W1) of FOXP2 FHD is truncated while the other (W2) forms a short helix. This is different to the more canonical structure of FOXA3, which represents the general structure of the FHD of FOX proteins (Clark *et al.*, 1993; Stroud *et al.*, 2006). This has implications on the way in which the FOXP2 FHD interacts with DNA.

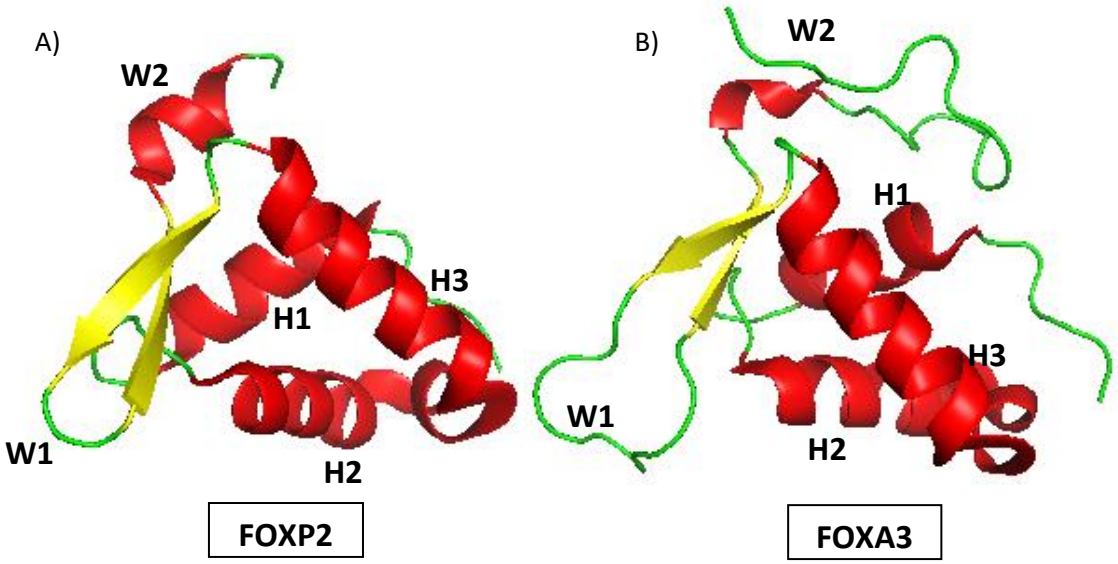


Figure 1.3: Schematic representation of the FHD structure of A) FOXP2 and B) FOXA3. Both the FHDs have a similar topology comprised of three α- helices (red) and a β- sheet (yellow). However, the winged region of FOXP2 FHD (PDB ID:2A07) is either truncated (W1) or forms a short helix (W2) when compared with the FOXA3 FHD (PDB ID: 1VTN). Images rendered using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC. (Clark *et al.,* 1993; Stroud *et al.,* 2006)

FOXP proteins bind to the core consensus sequence, 5′-(G/A)(T/C)(A/C)AA(C/T)A-3′, which is found on promoter regions (Benayoun *et al.,* 2011). DNA sequence adjacent to the consensus sequence is important for differential affinity of binding between different FOXP proteins and DNA (Nakagawa *et al.,* 2013). The third helix and second wing of the FHD allow for interaction with DNA by binding to the major and minor groove of DNA, respectively. The second wing is less conserved and contributes towards the specificity of DNA binding (Weigelt *et al.,* 2001).

1.2.3) Interaction of FOXP2 with DNA

The crystal structure of FOXP2 FHD with DNA, as depicted in Figure 1.4 (PDB ID: 2A07), shows that FOXP2 interacts with DNA by inserting the third helix of its FHD into the major groove of DNA which is similar to all other FOX proteins. This interaction with DNA is mediated by hydrogen bonds and van der Waals forces (Stroud *et al.,* 2006). However, unlike other FOX proteins, the winged regions of FOXP2 FHD do not contribute majorly to DNA binding due to the aforementioned differences in the winged regions between FOXP2 FHD and the FHD of other FOX proteins. The latter shows extensive contacts between the winged regions and DNA. The

relatively limited DNA contacts that FOXP2 forms with DNA could mean that FOXP2 FHD has lower affinity for DNA than some other FOX proteins (Stroud *et al.*, 2006)

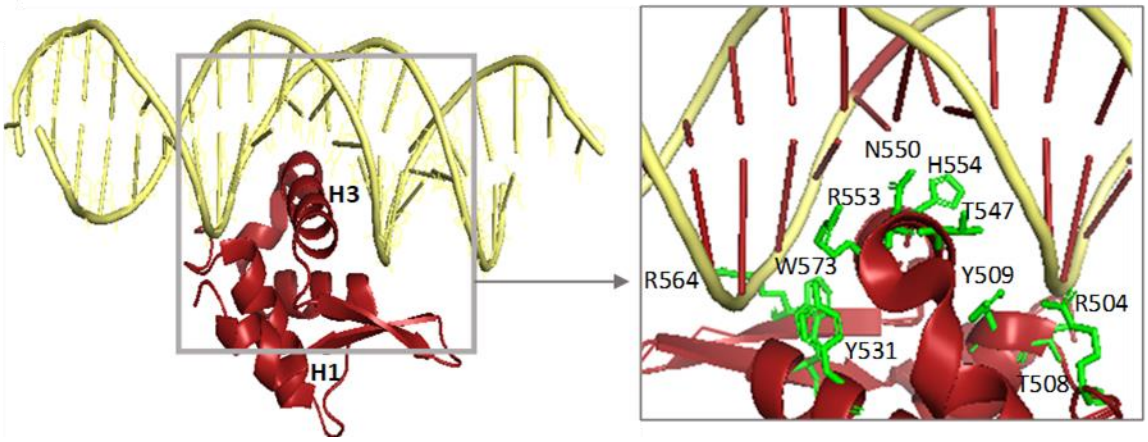


Figure 1.4: Cartoon representation of FOXP2 FHD bound to DNA. A) Structure of the monomeric FOXP2 FHD (red) bound to double-stranded DNA. The residues involved in the interaction with DNA are shown in green. The recognition helix or helix 3 inserts into the major groove of the cognate DNA sequence, whereas helix 1 forms stabilising contacts with the DNA (PDB ID: 2A07; Stroud *et al.*, 2006). Images are rendered using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC.

Additionally, FOXP2 FHD can recognise and bind to a broad range of sequences with differing rates and affinities. FOXP2 FHD was crystallised in the presence of the core consensus sequence of 5'- AAATA-3'sequence (Stroud *et al.*, 2006). However, FOXP2 also binds to several other divergent sequences but tends to favour the DNA sequence 5'-TGTTTAC-3' for which it has a high affinity (Enard *et al.*, 2009: Nelson *et al.*, 2013, Webb *et al.*, 2017). The ability of the FOXP2 FHD to bind to several DNA sequences can serve as a regulatory mechanism as different binding sites could promote site-dependent binding with other specific TFs at neighbouring response elements. This is evidenced by the difference observed between DNA binding of the monomeric and dimeric forms of the FOXP2 FHD.

1.3) Functions of FOXP2

As a TF, FOXP2 mediates a variety of interactions physiologically through interactions with DNA and other proteins (Herrero and Gitton, 2018). Hence, both abnormal expression and mutations of FOXP2 have been implicated in a variety of disorders and diseases (Vernes *et al.*, 2008; Bach *et al.*, 2018).

1.3.1) Role in tumorigenesis

FOXP2 has been found to be implicated in tumorigenesis. FOXP2 interacts with various signalling molecules and proteins. This includes transcriptional repressors such as the C-terminal binding protein 1 (CTBP1) that regulates the expression of p16, BAX, PTEN, and other tumour suppressors (Bach *et al.,* 2018). FOXP2 may also be associated with cell – signalling pathways (Bach *et al.,* 2018). This includes the Wnt/b-catenin pathway which is associated with numerous cancer types due to its role in cellular proliferation, differentiation, and apoptosis. FOXP2 was observed to be upregulated in glioma, a type of brain cancer wherein FOXP2 levels corresponded to the pathological grade (Li *et al.,* 2019). It was also found that silencing FOXP2 promoted apoptosis and impeded cell proliferation, migration, and invasion, hence proving its role in tumorigenesis (Li *et al.,* 2019). Therefore, in cancers like glioma, multiple myeloma, monoclonal gammopathy of unknown significance, lymphoma, and neuroblastoma, FOXP2 has been found to be upregulated with pro-oncogenic activity (Herrero and Gitton, 2018).

In contrast, FOXP2 has also been found to have a tumour suppressor role in specific cancer types. In a specific study, the role of FOXP2 in metastatic breast cancer was considered (Chen *et al.,* 2018). They found that FOXP2 has a role in the inhibition of the epithelial-to-mesenchymal transition of cells through the transforming growth factor (TGF) β/SMAD signalling pathway which is activated in tumour metastasis. This pathway is responsible for sending signals from the cell surface to the nuclear transcription factors in elevated amounts in tumorigenic cells. FOXP2 could influence this pathway, considering its downregulation corresponds not only to a decrease in mesenchymal biomarkers but also to elevated levels of TGFβR1, phosphorylated-SMAD3, SMAD4, Snail and other TGFβ pathway related proteins. Additionally, Cuiffo and Karnoub (2016) showed that FOXP2 inhibited cancer stem cell- associated factors like c-Myc, CD44 and Oct-4 which halted breast cancer initiation and development. FOXP2 therefore, has a definitive role in cancer, however the mechanisms discovered thus far are not exhaustive. FOXP2 mutations, aberrant upstream regulation and its effect on downstream targets have not yet been fully investigated.

1.3.2) Speech, language, and clinical relevance

FOXP2 has also been implicated in language and speech disorders. Interactions of FOXP2 with various DNA sequences and proteins regulates the expression of various genes associated with

speech and language (Vernes *et al.,* 2008; Graham and Fisher, 2013; Oswald *et al.,* 2017). Mutational losses of function in *FOXP2* or haploinsufficiency results in a condition called verbal dyspraxia, which is characterised by a decrease in the ability to accurately sequence speech sounds. Other notable phenotypes in humans include dysarthria, oral dyspraxia, and impaired linguistic abilities (Crespi *et al.,* 2017; Reuter *et al.,* 2017). Moreover, FOXP2 regulates a variety of speech and language associated genes including Disrupted in schizophrenia 1 (*DISC1),* Contactin-associated protein 2 (*CNTNAP2)* and Sushi-repeat protein/ plasminogen activator receptor of the urokinase type (*SRPX2/uPAR)* (Vernes *et al.,* 2008; Walker *et al.,* 2012). FOXP2 targets like the MET receptor tyrosine kinase gene and *DISC1* are implicated in autism and schizophrenia, respectively; this could indicate an involvement of FOXP2 in neurodevelopmental diseases associated with language and speech development (Herrero and Gitton, 2018). Additionally, various single nucleotide polymorphisms (SNPs) of FOXP2 are linked with phenotypic effects in cognition and psychiatric diagnoses associated with speech and language. This includes lateralised activation of the frontal cortex, verbal fluency, dyslexia, and risk of schizophrenia (Sanjuán *et al.,* 2006; Pinel *et al.,* 2012; Mozzi *et al.,* 2017). These common SNPs of FOXP2 may affect aspects of speech, including fluency. Hence, FOXP2 is associated with language and speech impediments as well as certain neurodevelopmental disorders.

1.3.2.1) Differentiation of language and language-associated diseases between males and females

Notably, speech and language are features that differ between males and females. Women are better at spelling and grammar, have a faster rate of speech acquisition in childhood and have a greater verbal fluency as opposed to males (Wallentin, 2009). Women also require activation of bilateral hemispheres to process language, as opposed to males who make greater use of their left hemisphere (Bitan *et al.,* 2010). Additionally, speech and language-related disorders in which FOXP2 is suggested to play a role also show variation between the sexes (Herrero and Gitton, 2018). Autism, which affects 1 % of the population, favours males by a 5:1 ratio (Bowers *et al.,* 2014). Men also have a higher incidence, earlier onset and even contrasting symptoms for schizophrenia, a neuropsychiatric disorder which affects comprehension, attention, and cerebral lateralisation of language in patients that suffer from it (Bergemann *et al.,* 2007; Kulkarni *et al.,* 2013). Women face fewer of the negative effects associated with the disorder including depression, aggression, elevated psychopathology, and substance abuse which are found to more commonly

afflict men (Crider and Pillai, 2017). Hence, the variable presentation of FOXP2's functions in males and females could suggest that FOXP2 behaves differentially in the two sexes.

FOXP2 expression levels has been previously reported to vary in response to hormone levels. It was found in animal studies that hormones regulate FOXP2 expression seasonally and influence the singing of zebrafinch (Wohlgemuth *et al.*, 2014). They also found that FOXP2 levels were the highest in certain areas when testosterone was present in low amounts, causing the zebrafinch to sing variable, non-stereotyped songs (Wohlgemuth *et al.*, 2014). Additionally, a study on mice found that males and females have a different reaction to ultrasonic vocalisations, a form of communication, after FOXP2 knockdown. This led to a differential response of the other mice to the distressed pups (Lepp *et al.*, 2013). These results suggest that sex hormones do play a role in sex differences observed in FOXP2 function. Therefore, differentiation of language and speech between the two sexes could also be attributed to sex hormones. In this study, only the effect of oestrogens will be the focus. However, to understand the link between oestrogens, FOXP2 and sex differences in FOXP2 functions, it is imperative to first understand how oestrogens mediate their effect in the body.

1.4) Oestrogen Receptor α

Oestrogen receptors (ERs) are intracellular transcription factors that mediate most of the biological effects of oestrogens at the level of gene regulation. ERs belong to the family of nuclear receptors (Auwerx *et al.*, 1999).

1.4.1) Nuclear receptors

ERs are members of the ligand-regulated nuclear receptor superfamily, one of the largest group of TFs in eukaryotes. Nuclear TFs are distributed throughout the body and have several physiological functions, including cell proliferation, metabolism, and reproduction (Mazaira *et al.*, 2018). They have also been implicated in neurological syndromes, cancers, and several other pathological processes (Evans, 1988). Nuclear TFs directly regulate transcription when activated by various lipid-soluble ligands, including steroids, retinoids, and thyroid hormones so that these extracellular signals can regulate specific intracellular events. Notably, the functional activity of nuclear receptors is not only regulated by ligands but also by coactivators, corepressors, and other regulatory proteins (Oñate *et al.*, 1995; Plevin *et al.*, 2005).

Most nuclear receptors have a common modular structure that comprises of 5 domains (designated A-E) as shown in Figure 1.5. The A/B region spans the highly disordered N-terminal domain that exhibits great variability in sequence and size between the nuclear receptors. The N-terminal domain contains the activator function (AF) -1 surface, a region that can interact with coregulators to activate transcription (Weikum *et al.,* 2018). The C region consists of the DBD, the most conserved domain among the nuclear receptor domains (Weikum *et al.,* 2018). The hallmark of the DBD is the two zinc-finger motifs, in which a zinc ion is tetrahedrally coordinated by four cysteine residues as shown in Figure 1.5B (Gronemeyer and Moras, 1995). Additionally, the DBD folds into two perpendicular α-helices, with the first helix mediating DNA recognition by binding to the major groove of specific DNA sequences. The second helix makes non-specific contacts with the DNA backbone and in some proteins, the minor groove of DNA. It also contains the D-box which can mediate dimerisation of some nuclear receptors (Weikum *et al.,* 2018). The D region is the least conserved and consists of a short and flexible hinge region which links the DBD and ligand-binding domain (LBD) (Germain *et al.,* 2006). This region also contains a nuclear localisation signal. The E region is the allosteric LBD that can interact with both ligands and coregulator proteins. Coregulators are factors that are able to interact with nuclear receptors in either a ligand- dependent or independent manner and influence their transcriptional activity. As shown in Figure 1.5C, the LBD has a conserved α-helical sandwich structure comprising 12 α-helices and two β-strands (Germain *et al.,* 2006). A ligand-binding pocket is found at the base of the nuclear receptor, which is the most variable region of the LBD and accounts for interaction with a diverse array of ligands (Wurtz *et al.,* 1996). The LBD contains a ligand-activated AF-2 surface, which is conformationally dynamic upon ligand binding, and consequently enhances the stability of the LBD and facilitates interaction with various co-regulator proteins (Wurtz *et al.,* 1996).

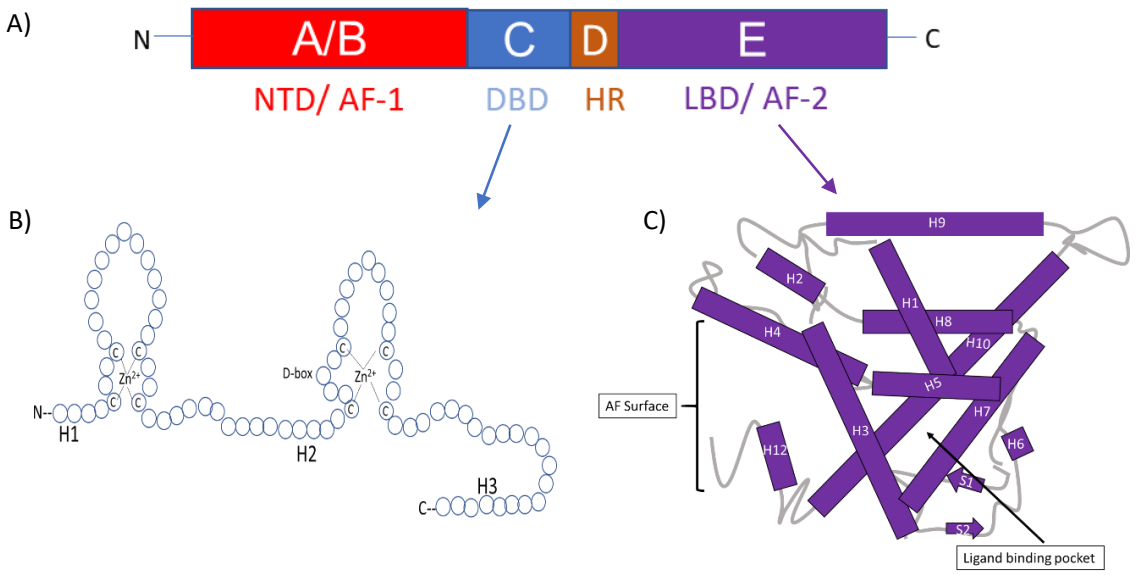


Figure 1.5: General structure of nuclear receptors. A) Structural organisation of nuclear receptors. The nuclear receptors are divided into 5 modular domains (A-E) that comprises of the N-terminal domain (NTD), DBD, a hinge region (HR) and the ligand-binding domain (LBD). Numerous nuclear receptors have at least two activation domains: activation functions (AF)-1 and -2 which are located at the N-terminal and ligand-binding domains, respectively. B) Schematic representation of the DBD. This domain contains two zinc finger subdomains. The residues of the first zinc finger subdomain interact with the major groove of DNA, whereas the residues of the other subdomain interact non-specifically with the DNA backbone and participate in DBD dimerisation. C) Schematic representation of the LBD. This domain is composed of 12 α-helices which fold to form a hydrophobic ligand binding pocket. This domain also contains the AF-2 surface, composed of helices (H) 3, 4, and 12, which interacts with coregulator proteins (Adapted from: Weikum *et al.*, 2018).

Nuclear receptors are classified into 7 subfamilies, designated NR0 – NR6, based on the evolution of the DNA- and ligand- binding domains (Auwerx *et al.*, 1999). The NR3 subgroup is comprised of steroid receptors, which include the glucocorticoid, mineralocorticoid, progesterone, androgen, and oestrogen receptors. Steroid receptors are present in an inactive form in the cytoplasm and are activated upon direct association with ligands. Notably, steroid receptors can also be activated independently of ligand-binding through interaction with various molecules, including growth factors, peptide hormones, neurotransmitter agonists, and various signalling pathways (Blaustein, 2004). Ligand-dependent activation of steroid receptors is mediated by interaction with steroid hormones, which are cholesterol-derived lipophilic molecules such as cortisol and oestrogen, synthesised in the adrenal cortex, testes, ovary, and placenta. Consequently, the SR-ligand

complex can mediate transcription of target genes through interaction with short palindromic DNA sequences, called the hormone response elements, or composite response elements though interaction with other TFs (Beato and Sánchez-Pacheco, 1996; McEwan *et al.,* 2004; Marino *et al.,* 2006). The androgen, glucocorticoid, mineralocorticoid, and progesterone receptors bind to the half-site sequence, 5'-AGAACA-3', which is separated by 3 nucleotides; the ERs recognise the half-site sequence 5'-AGGTCA-3', which is also separated by 3 nucleotides and is termed the oestrogen response element (McEwan *et al.,* 2004).

In particular, the ERs exist predominantly as two subtypes, ERα (ER1) and ERβ (ER2). ERs partly mediate the physiological effect of oestrogens, which includes oestrone, oestriol and the biologically active metabolite 17β-oestradiol (E2). Both the ER subtypes have similar affinities for oestrogen despite only having 56 % sequence similarity in their LBDs (Katzenellenbogen and Katzenellenbogen, 2000). Additionally, the ER subtypes have distinct cellular distribution and biological roles (Lee *et al.,* 2012; Yaşar *et al.,* 2017). ER1 appears to be predominantly expressed in the uterus, pituitary glands, skeletal muscle, adipose tissue, and bone. Conversely, ER2 is considered to be more predominant in the ovary, prostate, lung, central nervous system, and cardiovascular system (Desouza *et al.,* 2019). Moreover, the distribution of ERs varies even within the same tissue as is the case in the ovaries, where ER1 is more abundant in the theca cells and ER2 in the granulosa cells (Yaşar *et al.,* 2017). This differential expression of ER1 and ER2 could mediate a tight spatial regulation of oestrogenic effects (Desouza *et al.,* 2019). ER1 is the focus of this study.

ER1 mediates the majority of the activities of oestrogens, which have functions in a diverse range of biological processes including reproduction, skeletal and cardiovascular health, circulation, bone development and neuroprotection. Consequently, oestrogens are implicated in many pathological conditions, including cancers, osteoporosis, and cardiovascular diseases (Arao and Korach, 2018; Patel *et al.,* 2018). ER1 has thus been the focus of numerous studies.

1.4.2) Genomic and non-genomic activity of oestrogen receptor α:

ERs function as ligand-activated transcription factors (Rosenfeld *et al.,* 2006). The transcriptional activity of ER1 occurs as a multistep process since ERs are typically kept inactive through interactions with specific inhibitory proteins. Interaction with a ligand induces conformational changes in ER1, causing it to disassociate from its chaperones/nuclear matrix-associated binding

proteins as shown in Figure 1.6. The allosteric change in structure causes the unmasking of the domains responsible for receptor dimerisation, nuclear localisation, and DNA and protein-interactions (Pratt and Toft, 1997). Hence, once ER1 is activated, it dimerises and is translocated to the nucleus where it interacts with the inverted palindromic oestrogen response elements and oestrogen response elements - like sequences directly. Notably, a third of the genes ERs interact with do not contain the oestrogen response element sequence. This interaction between ER1 and DNA is mediated indirectly through a tethered mechanism. As shown in Figure 1.6, this mechanism of ER1 activity is mediated through crosstalk between the ER and other TFs or DNA-binding proteins. One of the predominant mediators of tethered ER transcriptional activity is stimulating protein -1 (Marino *et al.*, 2006).

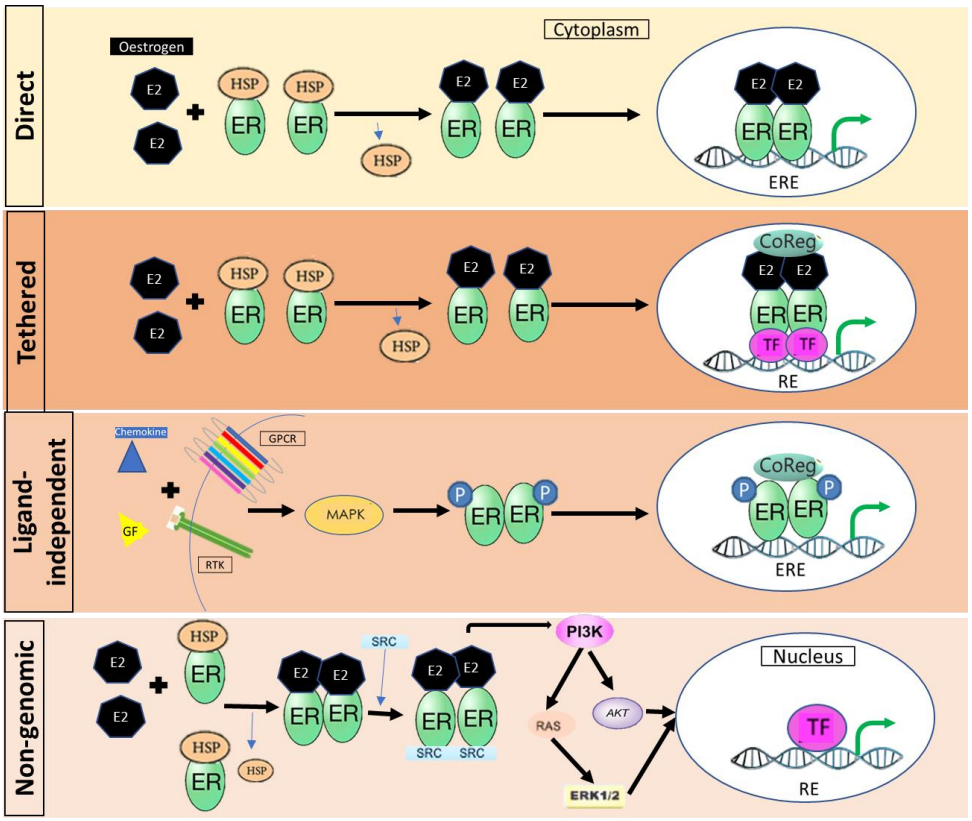


Figure 1.6: Oestrogen signalling pathways. Oestrogen receptor mediates transcriptional activity using 3 pathways. The genomic pathway includes both the direct and tethered interaction between oestrogen receptor and DNA. In the direct or classical pathway, the oestrogen receptor first dimerises upon interaction with oestrogen (E2) and thereafter, migrates to the nucleus and binds directly to the oestrogen response element. In the tethered pathway, the oestrogen receptor dimer interacts with other transcription factors and indirectly binds to DNA. A second pathway involves oestrogen-independent activation of the oestrogen receptor, which in this instance is activated by growth-factor mediated phosphorylation of the receptor. The third pathway is the non-genomic pathway, which is associated with the activation of various protein kinase cascades. As depicted here, the non-genomic pathway indirectly influences gene expression through signalling pathways and secondary messengers to ultimately act on transcription factors. Akt: protein kinase B, coReg: coregulators, ER: oestrogen receptor, ERE: oestrogen receptor elements, ERK: extracellular receptor kinase, GF: growth factor, HSP: heat shock proteins, MAPK: mitogen-activated protein kinase, P: phosphorylation, PI3K: phosphoinositide 3-kinases, RE: receptor elements, RTK: receptor tyrosine kinase, SRC: steroid receptor coactivator TF: transcription factors (Adapted from: Le Romancer *et al.*, 2011).

Notably, ER1 can also be activated independently of ligand-binding through post-translational modifications (Ignar-Trowbridge *et al.*, 1992). The activation of this pathway is primarily triggered by phosphorylation of specific residues of the ER or their association with coregulators (described below). This neurotransmitter or growth hormone-mediated ER1 activation is required in both reproductive and non-reproductive organs with aberrant unliganded activation of ER1 being associated with resistance to endocrine therapy in ER-positive cancer cells (Maggi, 2011). A small fraction of ER1 found in the plasma membrane can also elicit non-genomic responses that can rapidly modulate specific signalling pathways.

1.4.3) Structural composition

ER1 is composed of some highly conserved domain structures that are common among the nuclear receptor superfamily (Hewitt and Korach, 2018). As aforementioned, this includes the N-terminal domains A and B, which contains the AF-1 surface required for the ligand-independent activation of ER1. Domain C encompasses the DBD that allows for recognition and binding to specific DNA sequences, namely the oestrogen response elements, and aids in the dimerisation of the receptor. Domain D comprises the hinge region that interconnects the DBD and LBD and codes for the nuclear localisation signal. Domain E at the C-terminus constitutes the LBD that consists of the AF-2 surface. Transcriptional activity of ERs requires the synergistic action of both AF-1 and AF-2 (Dahlman-Wright *et al.*, 2006; Hewitt and Korach 2018). ER1 also consists of the highly variable carboxyl-terminal F domain, a feature unique to the ERs. The functionality of the F-domain is largely unknown, but it is proposed to have roles in ER intramolecular interaction and stability (Arao and Korach, 2018; Hewitt and Korach 2018).

1.4.3.1) Ligand-binding domain

The LBD is comprised of a ligand-binding site, a dimerisation interface, and a coactivator and corepressor interaction site. The LBD of ER1 is arranged in a three-layer, antiparallel α-helical fold that comprises of 12 helices. The central core of helices (which includes helices 4, 5, 6, 8 and 9) is sandwiched between helices 1 and 3 on one side, and helices 7, 10, and 11 on the other. This fold forms a hydrophobic ligand-binding pocket, as shown in Figure 1.7. Oestrogen and other ligands are buried in this hydrophobic cavity so that it is shielded from the environment. The helical fold is surrounded by a two-stranded anti-parallel β-sheet and three mobile helices. Notably, the LBD can form dimers with other agonist- and antagonist- bound LBDs. This dimerisation

interface is formed by helices 8, 9, 10 and 11, but the predominant contact surface is with the residues of helix 10 via a leucine zipper-like interaction zone, and hydrogen bonds (Tanenbaum *et al.*, 1998; Ruff *et al.*, 2000).

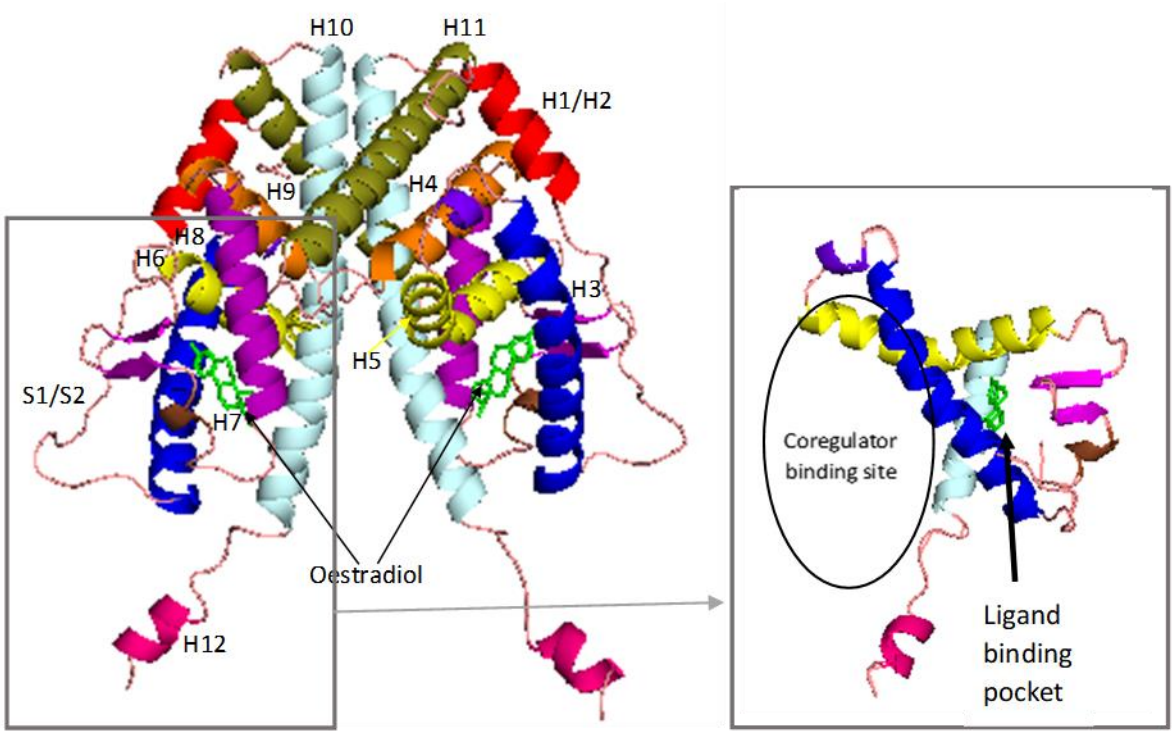


Figure 1.7: Structure of ER1 LBD. Cartoon representation of the dimeric ligand-binding domain of oestrogen receptor α. Oestrogen (green) is bound in a hydrophobic groove formed by helices 4, 5, 6, 8 and 9. Oestrogen binding changes the conformation of helix 12, such that it forms a coregulator binding site with helices 3, 4, and 5, as shown on the right (PDB ID:1A52; Tanenbaum *et al.*, 1998). Images rendered using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC.

Helix 12 is the most dynamic region of the LBD and undergoes a conformational change upon ligand binding, as shown in Figure 1.8. In the unliganded structure, helix 12 is flipped outwards, such that it allows oestrogens to enter the hydrophobic ligand-binding cavity; whereas in the oestrogen-bound structure, the conformation of helix 12 is repositioned to close the pocket and accommodate the ligand (L'Horset *et al.*, 1996; Farooq, 2015). Following the binding of a ligand, the entire LBD undergoes a conformational rearrangement resulting in the formation of a specific binding site for nuclear coregulators (Ruff *et al.*, 2000). This binding site is a hydrophobic groove formed by residues of helices 3, 4, 5 and 12, and the turn between helices 3 and 4, as depicted in

Figure 1.8. This constitutes the AF- 2 region. Ligand-binding facilitates the recruitment of coactivators to the AF-2 which may be required by the ER to regulate target gene transcription.

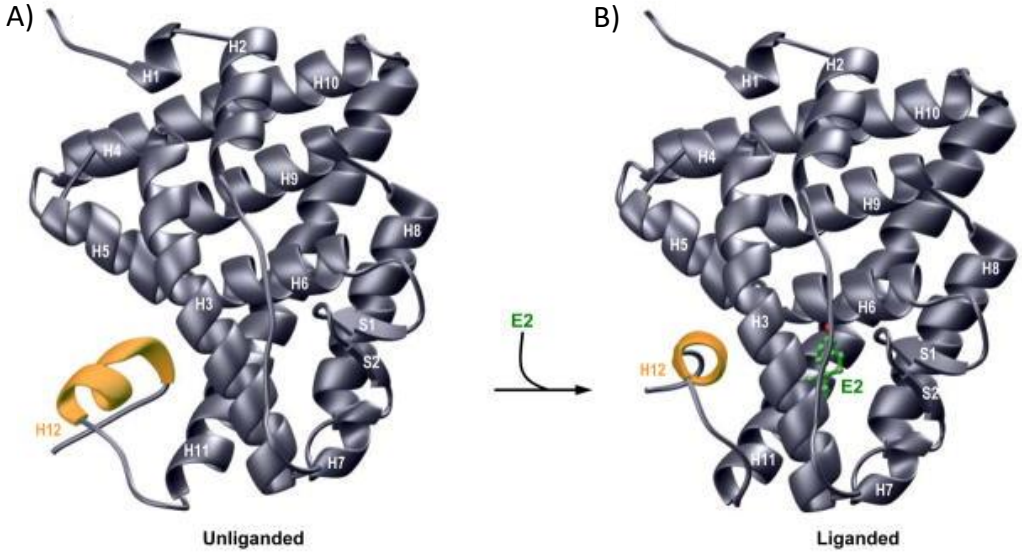


Figure 1.8: Cartoon representation of the ER1 ligand-binding domain in A) unliganded and B) liganded form. The monomeric LBD (grey) largely adopts a similar structure when bound to E2 (green) as it does when unbound, apart from the C-terminal H12 (yellow). H12 adopts a more structured conformation in the liganded than the unliganded structure as it turns and rotates by approximately 180° along the H11 axis. E2: oestradiol, H: helix (Retrieved from Farooq, 2015)

1.4.3.1.1) AF-2 and the LXXLL motif

Numerous coregulators, including BRCA1 and p160 proteins, have been shown to bind to the hydrophobic pocket of the AF-2 region of the LBD (Kawai *et al.*, 2002; Mak *et al.*, 1999). A common feature of these proteins is a signature alpha-helical LXXLL motif called the NR-box motif, with L denoting leucine residues and X indicating any amino acid. This motif sufficiently allows for an interaction between ER1 and its coregulators (Wärnmark *et al.*, 2002).

Additionally, co-crystallisation of ER1 LBD with a peptide derived from the NR-box II region of the coactivator GRIP1 protein showed that the NR-box fits into the AF-2 pocket of the ER1 LBD, as shown in Figure 1.9A (Shiau *et al.*, 1998). The NR-box peptide forms a short amphipathic α-helix that is recognised by the complementary AF-2 groove on the surface of the ER1 LBD. The peptide interacts with side chains of eleven AF-2 residues as shown in Figure 1.9B. These residues include: Leu354, Val355, Ile358, Ala361, and Lys362 from helix 3; Phe367 and Val368 from helix 4; Leu372 from the turn between helices 3 and 4; Gln375, Val376, Leu379, and Glu380 from helix

5 and Asp538, Leu539, Glu542, and Met543 from helix 12 (Shiau *et al.*, 1998). Notably, helix 12 of the LBD, specifically, is essential for this interaction between AF-2 and the NR-box of coactivators, as evidenced by the antagonist-bound ER1 in Figure 1.9. As aforementioned, AF-2 activity is blocked when ER1 is bound to antagonists. This is because when the LBD is bound to an antagonist, Leu539, Glu542, and Met543 are in the incorrect orientation, while other residues from the static regions of AF-2 are bound to helix 12 and thus are unable to interact with the coactivators. Hence, interaction with the antagonist alters the conformation of helix 12 such that it the AF-2 surface is incomplete and consequently unable to recognise coactivators (Shiau *et al.*, 1998).

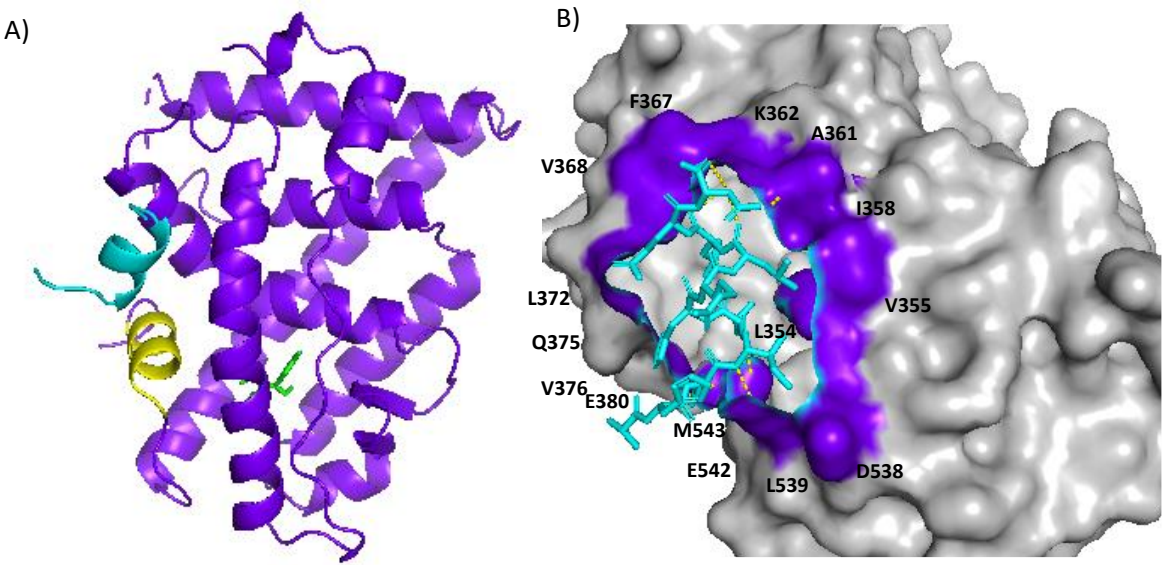


Figure 1.9: Interaction of the ER1 ligand-binding domain with a coactivator. A) Cartoon representation of ER1 LBD with an agonist and a coactivator (PDB ID:3ERD). Oestrogen receptor 1 (purple) interacts with a peptide containing the NR-box motif with the LXXLL sequence shown in cyan (where X denotes any residue) in the presence of diethylstilboestrol (green). Helix 12 is shown in yellow. B) A close view of the interaction between the LBD and coactivator (PDB ID:3ERD). The coactivator interacts with the ligand-binding domain by interacting with side chains of eleven residues, which are labelled and shown in purple. Images are rendered using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC. (Shiau *et al.*, 1998)

1.5) Probable Association Between FOXP2 and ER1

ER1 associates with several FOX proteins including FOXA1 (Carroll and Brown, 2006; Hurtado *et al.*, 2011), FOXK1 (Liu *et al.*, 2015), FOXM1 (Sanders *et al.*, 2013), FOXO3A (Zou *et al.*, 2008) and FOXP1 (Shigekawa *et al.*, 2011), yet the association between FOXP2 and ER1 has

neither been considered nor investigated. We postulate that there could be a probable association between FOXP2 and ER1 based primarily on 1) the potential role of oestrogen and ER1 on sex differentiation in language and speech, which are traits associated with FOXP2 and 2) the regulatory loop between various FOX TFs and ER1 in cancer, which could suggest a regulatory association between FOXP2 and ER1 as well. These two points have been mentioned above but have not been discussed in relation to both FOXP2 and ER1.

1.5.1) A role for FOXP2 and ER1: sexual differentiation of language and speech

As mentioned in section 1.3.2.1, speech and language and related disorders have characteristics which are differentiated based on sex. These differences in traits can putatively and at least partly be attributed to different endogenous sex steroids, such as oestrogen, which have functional roles in neurodevelopment and sexual differentiation (Gillies and McArthur, 2010; Maekawa *et al.*, 2014). The brains of males and females develop under different hormonal environments - α-fetoprotein is present in the female brain and sequesters circulating oestrogens in early life to protect the brain from masculinisation and defeminisation (Bakker *et al.*, 2006; Maekawa *et al.*, 2014).

With regards to language and speech associated traits in particular, oestrogenic modulation of dopaminergic function has been suggested to be the underlying reason for the sex differences that contribute to women outperforming men at speech articulation and fine motor skills (Jennings *et al.,* 1998). Improved motor performance was associated with oestrogen levels in women's blood. Furthermore, studies show that elevated testosterone levels during pregnancy and decreased oestrogen levels and decreased expression of the oestrogen receptor correlates with development of autism (Crider and Pillai, 2017). Another finding shows that low oestrogen levels elevate the risk of developing schizophrenic symptoms (Bergemann *et al.,* 2007). Additionally, oestrogen receptors are either decreased or found to be mutated in schizophrenic individuals (Crider and Pillai, 2017). Therefore, the activity of oestrogen is decreased by both lowered oestrogen levels and the altered attenuation and response to it in schizophrenia.

Hence, ERs can be key factors in mediating sexual dimorphism of certain traits. These sex differences could manifest functionally as sex-specific behaviours and traits. Hence, this commonality between ER activity in sexual differentiation and sexually differentiated nature of FOXP2-associated function and diseases has led us to speculate on whether there could be a link

between FOXP2 and ER1 which could mediate these sex differences and hormonal effects in FOXP2-linked functions and diseases. This hypothesis is given further credence by the existence of known interactions between ER and various FOX TFs.

1.5.2) ER1 and FOXP2 in breast cancer: inference from known ER1-FOXP TF interactions

FOXP2 as aforementioned also has a role in tumorigenesis in a variety of cancers, including breast cancer. The mechanism for its tumour-suppressive effect in cancer is not well understood nor has it been the subject of intensive research. However, given the relationship between ER1 and other FOX family members in breast cancer as well as the similarities in the structure of FOX proteins, FOXP2, which is downregulated in breast cancer, could also potentially mediate its role in breast cancer development and progression through association with ER1 in a manner similar to the other FOX proteins. This might be especially relevant considering FOXP1 associates with ER1 in breast cancer (Shigekawa *et al.,* 2011).

Both, FOXP1 and FOXP2 co-operate to regulate non-neural developmental processes (Shu *et al.,* 2007). Their expression also overlaps in the thalamus, striatum, inferior olive, and superior colliculus in the mature mouse brain (Ferland *et al.,* 2003). Additionally, *FOXP2* shares 64 % of its sequence with *FOXP1* while the FHD in particular is even more conserved with a sequence similarity score of 89 %. Both the proteins also contain an α-helical nuclear receptor box motif (with the sequence LXXLL) at the N-terminus that could directly mediate its interaction with ER1 as shown in Figure 1.10 This motif is remarkably similar in both proteins with only a single variation within the LXXLL motif itself and one variation within the flanking sequence. Therefore, the overlapping structure and function of FOXP1 and FOXP2 could suggest that FOXP2 could have some association with ER1, in a manner exemplified by FOXP1. Moreover, oestrogen and androgen receptors are found to colocalise with FOXP2 in the mouse amygdala, and rat brain (Bowers *et al.,* 2014; Lischinsky *et al.,* 2017). Hence, FOXP2 and ER1 interaction could be physiologically probable.

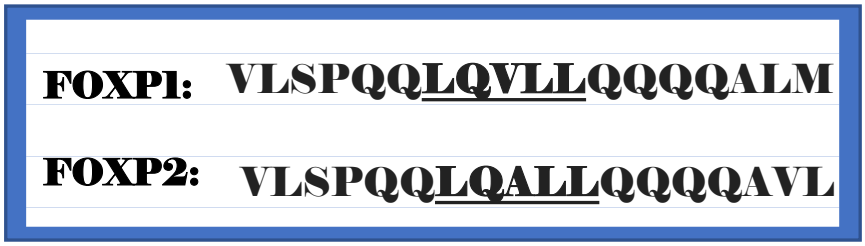


Figure 1.10: Nuclear receptor motif sequence of the FOXP1 and FOXP2 transcription factors. The LQALL motif that is sufficient for interaction with the ligand-binding domain is underlined. Sequence alignment performed using EMBOSS Needle (Needleman and Wunsch, 1970).

2. RATIONALE

Protein-protein interactions perform essential functions in various cellular processes including cell cycle progression, signal transduction, and metabolic and neurodevelopmental pathways (Keskin *et al.,* 2016). Given the role of PPIs in maintaining homeostasis, it is not surprising that perturbation of protein-protein interactions is the cause of various diseases. Hence, the study of PPIs is instrumental in providing a better understanding of cellular biochemistry and physiology and these interactions are the target of numerous therapeutic strategies. Considering that the tight regulation of gene expression is the consequence of complex and intricate interactions between transcription factors, coregulators, chromatin modifiers and other proteins, the transcriptional activity of FOXP2 is most likely regulated by a variety of these interactions.

FOXP2 transcription factors regulate the expression of various genes and are associated with language and speech acquisition and development, neurodevelopmental disorders, and cancers (Vernes *et al.,* 2008; Graham and Fisher, 2013). Notably, speech and language are traits that are differentiated in males and females (Wallentin, 2009). This sex differentiation can be mediated at least partly by sex hormones, including oestrogens, which mediate most of their effects through ER1 (Gillies and McArthur, 2010; Maekawa *et al.,* 2014).

Since FOXP2 has been found to be colocalised with ERs in some areas of the brain (Bowers *et al.,* 2014), an association between FOXP2 and ER1 could potentially mediate the engendered language and speech features. Moreover, both FOXP2 and ER1 have been implicated in breast cancer where they mediate their tumorigenic effect through common signalling pathways (Ito *et al.,* 2010; Chen *et al.,* 2018). It is plausible that a FOXP2 and ER1 interaction could potentially mediate oncogenic or tumour suppressor events in breast cancer by interacting with each other, as exemplified by the association of ER1 with various other members of the FOX family.

It is therefore of interest to investigate a potential association between ER1 and FOXP2. This can be accomplished through the use of recombinantly expressed proteins, namely the N-terminal region and forkhead domain of FOXP2 and the ligand-binding domain of ER1. By working with the isolated domains rather than the full-length protein, we are able to pinpoint which segment of the proteins is likely to mediate any specific interactions that may occur.

Furthermore, if there is indeed an interaction, the thermodynamics, regulation and functionality of FOXP2 could be explored further in the context of its interaction with ER1 as this will provide a comprehensive understanding of the mechanism of this protein-protein interaction, its occurrence and nature. Should the DNA-binding activity of FOXP2 be affected by the interaction, the protein-protein interaction could be important in regulating FOXP2's transcriptional activity in the body and maintaining homeostasis. Hence, this investigation serves to provide a solid foundation of information regarding the interaction so that a probable mechanism of action can be deduced, and any physiological relevance can also be identified. Consequently, this aids in elucidating whether the interaction could serve as promising target for mitigating relevant health problems.

Two domains of FOXP2 were chosen for this investigation, namely the forkhead domain (FHD) and nuclear receptor box - containing region at the N-terminal domain (NT). FOXP2 FHD, the DNA – binding domain, was selected as it could be involved in PPIs such that it would affect its's functional activity and consequently the functional activity of FOXP2. The FOXP2 NT domain was chosen as it contains the nuclear receptor box residues that can facilitate interaction with nuclear receptors such as ER1. Regarding ER1, the ligand-binding domain (LBD) was selected as it has an activation function region that can facilitate interaction with regulatory proteins. As domains have distinct fold and functions, their isolated use over full-length protein may be advantageous in obtaining more specified information such as identifying the independent role of and/or co-operativity and synergy between domains in mediating an interaction, the residues involved as well as the associated structural changes.

Investigating the association between FOXP2 and ER1 may help in understanding the link between language, speech, and sex, and could be important in enhancing the current understanding of speech and language acquisition and development. This could also have clinical relevance for speech and language-associated disorders, including various neurodevelopmental disorders such as autism. Additionally, elucidation of this interaction could serve as the basis for understanding the mechanism of FOXP2 in breast cancer, which could ultimately lead to the use of FOXP2 in breast cancer-related diagnosis and therapies in the future.

3. AIMS

FOXP2 and ER1 are both transcription factors that have a role in neural development and cancer. This study aims to investigate whether the overlapping functions of FOXP2 and ER1 occur through direct interaction of the ER1 LBD with either FOXP2 FHD or FOXP2 NT. The secondary aim of the study is to determine whether the interaction, if found, affects the binding of FOXP2 FHD with its cognate DNA.

The objectives of this study are to:

1) Determine if the FOXP2 transcription factor interacts with ER1 in MCF-7 cells to elucidate possible physiological relevance.

2) Overexpress FOXP2 FHD, FOXP2 NT and ER1 LBD and subsequently purify the proteins using immobilised-metal ion affinity chromatography (IMAC) and size-exclusion chromatography (SEC).
3) Determine the structural integrity of the proteins using far-UV circular dichroism and fluorescence spectroscopy.
4) Investigate a possible interaction between ER1 LBD and FOXP2 (FHD or NT) using fluorescence anisotropy.
5) Determine the thermodynamics of the ER1 LBD-FOXP2 FHD interaction by performing isothermal titration calorimetry (ITC).
6) Determine the effect of salt concentration on FOXP2 FHD-ER1 LBD interaction using fluorescence anisotropy and ITC.
7) Investigate the effect of the functional activities of both the ER1 LBD and FOXP2 FHD on the formation of ER1 LBD-FOXP2 FHD using fluorescence anisotropy and electrophoretic mobility shift assays (EMSA).

4. METHODS

This study focused on investigating the interaction between FOXP2 and ER1. This was accomplished by first identifying the interaction in breast cancer cells followed by a thorough biophysical characterisation of the interaction using recombinantly expressed proteins.

4.1) Materials

Specific materials used in the study include: FoxP2 (D55H9) Rabbit mAb (Cell Signalling Technology, USA), Estrogen Receptor α (D6R2W) Rabbit mAb (Cell Signalling Technology, USA), Anti-Rabbit, IgG, HRP-linked antibody (Cell Signalling Technology, USA), WesternBright™ ECL western blotting detection kit (Advansta, USA), Protein A/G PLUS-Agarose: sc-2003 (Santa Cruz Biotechnology, USA), Plasmids (Genscript, USA), GeneJET Plasmid Miniprep Kit (Thermofisher, USA), 17 β-oestradiol *(*Sigma-Aldrich, USA), PageRuler™ Prestained Protein Ladder, 10 to 180 kDa (ThermoFisher, USA), AlexaFluor™ 555 C 2 maleimide (Thermofisher, USA), Pierce™ dye removal columns (ThermoFisher, USA)

DNA: Nelson DNA oligonucleotide with sequence: 5'-AGGTGTTTACTTTCATAG-3', was synthesised as a duplex for this work (Integrated DNA Technology, South Africa) (Nelson *et al.,* 2013).

Stock Preparations: DNA was diluted in MiliQ water or 50 mM Tris, pH 7.8 buffer to required stock concentrations. E2 was dissolved in 99.9 % ethanol to a final stock concentration of 10 mM. AlexaFluor™ 555 C 2 maleimide was dissolved in DMSO (Sigma-Aldrich, USA) to a final concentration of 20 mM.

4.2) Protein-protein interaction in cancer cells

ER1 and FOXP2 have both been implicated in various cancers, including the breast. To determine if the commonality in functions may be due to an association between the proteins and could have some biological relevance, a co-immunoprecipitation study was performed.

4.2.1) Co-immunoprecipitation of protein-protein complex

To confirm the presence of the ER1-FOXP2 interaction in the cells, co-IP was employed. This tool is widely employed in the detection of protein-protein and protein-nucleic acid interaction both *in vivo* and *in vitro*. It is based on the principal that proteins are present in their native form by

themselves or in complexes along with various cellular components in a whole cell extract and the protein as well the complexes that it forms can be "captured" with the use of a protein-specific antibody. The process includes preparation of cell lysate, incubation with antibody that interacts with the protein of interest, immobilisation of the antibody-bound protein complex to an inert matrix or bed of beads and analysis of the protein-protein complexes formed (Journet and Cascales, 2017).

Co-IP is most suitable for isolation of complexes that are stable but some interactions in the cell, including PPIs, occur transiently or are weak which will make them difficult to detect. To mitigate this disadvantage, crosslinkers may be used. Crosslinkers include cell-permeable esters like dithiobis succinimidyl propionate that can interact with the side of chain of lysine and the N-terminal of proteins and consequently form a stable amide bond between the two interacting proteins (Wang *et al.*, 2019). Formaldehyde could also be used to crosslink interacting proteins by applying formaldehyde to live cells for a short incubation period (Sutherland *et al.*, 2008).

4.2.1.1) Cell culture and harvest

MCF-7 and HEK293 cells were kindly provided by Prof Mavri-Damelin (University of the Witwatersrand) with the cells being cultured and split by Jemma Falkov (MCF-7) and Thokozile Makhanya (HEK293). The cells were grown in Fetal Bovine Serum (FBS) media for 48 hours. Once the cells attained a confluency of 70-80 %, they were washed once with PBS and exchanged with serum-free media. Thereafter, the cells were treated with 1 nM E2 or vehicle control (0.001 % ethanol) for 24 hours. The cells usually grew to 80-90 % confluency following treatment. Some MCF-7 cells were treated with 1 % formaldehyde and incubated for 10 minutes at RT with shaking to cross-link protein-protein complexes in the cells. Thereafter, formaldehyde was quenched by treating the plates with 125 mM glycine and further incubated for 5 minutes at RT, shaking. Cells exposed to cross-linking reagent and otherwise were subsequently lysed, and proteins were isolated from the remainder of the cell components.

For cell lysis, the media was first discarded. Subsequently, the cells were washed 3x with phosphate-buffered saline (PBS) (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4). The cells were scraped and centrifuged at 11,000 x*g* at 4 °C for 10 minutes. The supernatant was discarded. The cells were lysed using native buffer (25 mM Tris, 75 M NaCl, 0.5 % Triton X-100, pH 8) and sonicated 5 times with 45 second intervals on ice. Thereafter, the cells

were shaken at 4 °C for 1 hour to ensure complete lysis. Following centrifugation at 5,000 x*g* and 4 °C for 5 minutes, the protein was collected in the supernatant and frozen at -80 °C until required for quantification and downstream application.

4.2.1.2) Bradford assay

The total protein content was determined using Bradford assay. Firstly, the Whatman filter paper was prepared by soaking it in water for 20 minutes followed by a soak in 95 % ethanol, 100 % ethanol and 100 % acetone separately for 5 minutes each. Once dry, 1, 3, 6, 12, 16 and 20 µL of 2 mg/mL BSA (in native lysis buffer) was applied to the filter paper and served as the control. Thereafter, 2 µL of protein sample isolated from cells was applied. The protein was fixed by shaking the filter paper in 7.5 % trichloroacetic acid (TCA) for 45 minutes. The protein was stained and destained for an hour each at RT, shaking. Each circle representing each of the standards and samples was cut and placed for 2 hours to overnight in elution buffer in the dark.

The samples were placed in a 96 well plate in triplicate. Using the Multiskan GO plate reader (ThermoFisher, USA) and SkanIt RE for Multiskan GO 3.2 software (ThermoFisher, USA), the 96-well plate was shaken for 15 seconds prior to measurement of absorbance of each well at 595 nm. The absorbance of the standards was used to construct a linear curve of the known BSA concentration on the x-axis against the relative absorbance on the y-axis. Using the equation of the standard curve and the absorbance of the protein samples, the total protein concentration was determined. The elution buffer was used as the blank.

4.2.1.3) Co-Immunoprecipitation

ER1 and the complexes it forms were isolated using co-IP and evaluated using SDS-PAGE and western blots; the workflow has been depicted in Figure 4.1. To determine if there was an interaction between ER1 LBD and FOXP2 FHD, 60 -80 µg/mL of protein in IP dilution buffer (20 mM Tris, 150 mM NaCl, 0.5 % Triton X-100, pH 8) and protease inhibitors (0.1 µg/mL of leupeptin, 1 µg/mL PMSF and 1 µg/mL aprotinin) was incubated separately with either FOXP2 or ER1 rabbit mammalian antibody (mAb) (1:250 dilution) (Cell Signalling Technology, USA) overnight at 4 °C. Similarly, the samples were incubated with protein A/G plus-agarose beads (Cell Signalling Technology, USA) overnight. The A/G beads bind to the antibodies which are expected to be bound to the protein of interest. Thereafter, the samples were centrifuged at 5000 x*g* for 5 minutes at 4 °C to pellet the beads, which were thereafter washed twice with salt buffer (20 mM

Tris, 150 mM NaCl, 0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, pH 8). The cells were centrifuged at 5000 x*g* for 5 minutes in between each wash. Finally, the cells were washed with a high salt wash buffer (20 mM Tris, 500 mM NaCl, 0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, pH 8, centrifuged, and then mixed with sample-loading buffer for SDS-PAGE.

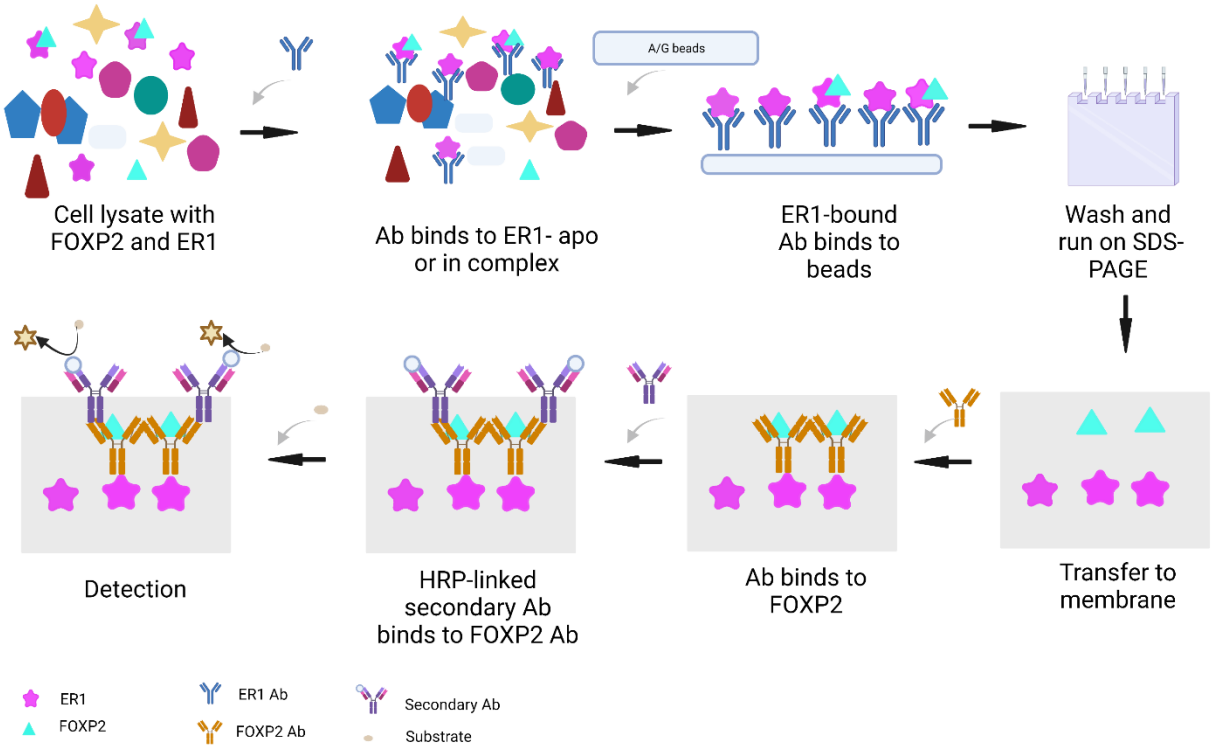


Figure 4.1: Protein-protein detection workflow. Mixture of proteins was first incubated with a primary antibody against ER1 and then incubated overnight with agarose beads that bind to the antibody. The sample was washed three times, denatured, and evaluated on an SDS-PAGE gel. The separated proteins were transferred onto a nitrocellulose membrane for western blot analysis, whereby the membrane was incubated firstly with primary antibody against FOXP2, secondly with an HRP-linked secondary antibody and finally with a substrate that underwent chemiluminescence by the secondary protein. A band would be detected only if ER1 formed a complex with FOXP2 and the complex was co-immunoprecipitated.

4.12.1.4) SDS-PAGE and Western blots

To confirm a potential interaction between ER1 and FOXP2, ER1 LBD bound to antibodies by itself or with interacting partners was denatured and resolved on a gel. A 10 % SDS-PAGE resolving (using 29:1 acrylamide/bisacrylamide solution, 20 mM Tris-HCl solution pH 8.8, 10 % SDS, APS and TEMED) and 4 % stacking (using 29:1 acrylamide/bisacrylamide solution, 20 mM

Tris-HCl pH 6.8, 10 % SDS, APS and TEMED) gel was made. The co-immunoprecipitated samples as well as total protein samples (40 - 60 µg/mL) were prepared by adding SDS sample buffer (0.25 M Tris-HCl pH 6.8, 8 % SDS, 10 % β-mercaptoethanol, 0.3 M dithiothreitol (DTT), 30 % glycerol) and heated for 5 minutes in a 90 °C water bath. The samples were loaded on to the gel along with a molecular weight marker and electrophoresed at 80 V for the first 30 minutes to allow stacking and thereafter 100 V until the dye front reached the bottom of the gel. The running buffer was comprised of 125 mM Tris, 1.25 M glycine and 3.5 M SDS.

Once the electrophoresis was completed, the gels were cut as per the size of the expected molecular weights of the proteins (FOXP2: 100 kDa, ER1: 66 kDa). The gel was placed on filter paper. Nitrocellulose membrane was placed on top on the gel and sandwiched with more filter paper into a cassette. Thereafter, a wet transfer electro-blot was performed using transfer buffer (25 mM Tris, 192 mM glycine, 20 % methanol) at 300 V for 2.5 hours at 4 °C. The nitrocellulose membrane was blocked with 5 % BSA in Tris-buffered saline with Tween (TBST) buffer (50 mM Tris -HCl, pH 7.5, 150 mM NaCl, 0.1 % tween) for 1 hour at RT. The membrane was washed once with TBST for 5 minutes and thereafter incubated overnight with ER1 or FOXP2 Rabbit mAb (Cell Signalling Technology, USA) at 4 °C, shaking. If antibody against FOXP2 was used in the co-immunoprecipitation step, the primary antibody against ER1 was used for western blot and vice versa. Thereafter, the cells were washed with TBST buffer 5 times for 5 minutes each by shaking the membrane in it at high speed at RT. The membrane was incubated with an anti-rabbit HRP-linked secondary antibody (1:1000 dilution in 5 % BSA solution prepared in TBST buffer) for an hour at RT in the dark. The membrane was washed in TBST five times as previously mentioned and then incubated with substrate solution (WesternBright^{TM} ECL western blotting detection kit) for 1 minute in the dark. The membrane was visualised using the HiSensitivity setting under blots using the Bio-Rad gel doc system. The molecular weight marker was visualised using the colorimetric setting under blots.

Co-immunoprecipitation was followed by biophysical characterisation of the interaction. Therefore, all subsequent studies were performed using recombinantly expressed proteins. Due to difficulty in obtaining soluble and sufficient yields of full-length proteins, isolated domains were selected for all downstream applications.

4.3) Transformation

For recombinant protein production, bacteria need to be transformed with the plasmid containing the gene of interest. A vector for each of the protein constructs was constructed by incorporating the specific codon-optimised gene sequence into a pET-11a plasmid with an ampicillin-resistance gene insert as shown in Figure 4.2 (Genscript, USA). An N-terminal hexa-Histidine tag and a thrombin cleavage site (LVPRGS) were selectively inserted to aid in the purification, and the cleavage of the His-tag, respectively. The plasmid was under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG) - inducible T7 promoter derived from the T7 bacteriophage. RNA-polymerase has a high affinity for this promoter and as a result, it has a high frequency for initiating transcription.

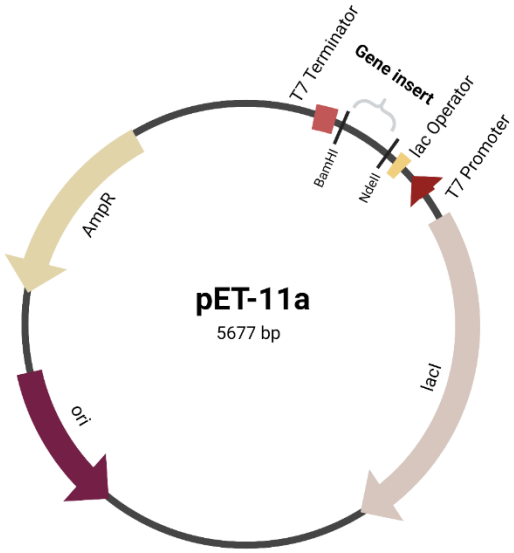


Figure 4.2: Schematic representation of the pET-11a plasmid. This is a bacterial vector with a T7 Promoter lacI repressor, pBR322 origin (ori), ampicillin resistance marker and multiple cloning sites. The gene sequence of the protein of interest was inserted between the BamHI and NdeI cloning site.

Competent T7 cells (New England Biolabs, USA) were transformed with the pET-11a plasmid containing the specific protein insert. Transformation was performed by incubating the T7 cells with the plasmid for 30 minutes followed by a 45 second heat shock treatment at 42 °C. Thereafter, the cells were cultured for 1 hour in super optimal broth with catabolite repression (SOC) media (2 % (w/v) tryptone, 0.5 % (w/v) yeast, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 20 mM glucose) at 37 °C shaking. This was followed by a 37 °C overnight incubation where the cells were

grown on agar plates containing 0.1 mg/mL ampicillin for selection. An isolated and well-formed colony was thereafter selected and grown overnight in Luria-Bertani (LB) broth (1 % tryptone, 0.5 % yeast extract, 1 % NaCl) at 37 °C. Subsequently, the culture was used to make up 1:1 glycerol stocks of the cell culture and 80 % glycerol. These glycerol stocks were used for all subsequent overexpression that were performed. Additionally, plasmids were isolated using the GeneJET plasmid miniprep kit and sent for sequencing (Inqaba Biotec, RSA) to ensure accuracy of the gene sequence.

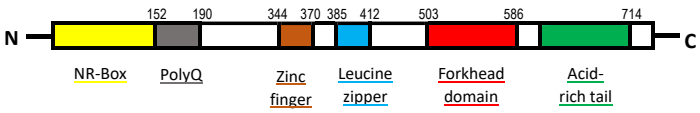
4.4) Overexpression of proteins

Following transformation, the proteins were expressed in bacterial cells. The expression of FOXP2 FHD has been previously optimised (Morris and Fanucchi, 2016). Conversely, expression trials had to be performed for FOXP2 NT and ER1 LBD.

4.4.1) FOXP2 forkhead domain

FOXP2 FHD was constructed with residues ranging between 503 and 586 of the FOXP2 protein as shown in Figure 4.3. Transformation of bacterial cells were performed as aforementioned. (Section 4.3).

A)



B)

NR-Box: MGSS HHHHHH SSG LVPR | GS MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSG DTSSEVSTVELLHLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVITPQQMQQ ILQQQVLSPQQLQALLQQQQAVMLQQQQLQEFYKKQQEQLHLQLL

FOXP2 FHD: MASMTGGQQMGM HHHHHH SSG LVPR | GS VRPPFTYATLIRQAIMESSDRQLTLNEI YSWFTRTFAYFRRNAATWKNAVRHNLSLHKCFVRVENVKGAVWTVDEVEYQKRRSQ

C)

	Molecular Weight (Da)	*pI*	*Extinction Co-efficient (M^{-1}. cm^{-1})*	*Number of residues*	
				Trp	**Tyr**
FOXP2 NT	18 912.28	7.11	1 490	0	1
FOXP2 FHD	12 972.75	10.41	22 460	3	4

Figure 4.3: FOXP2 structural organisation and constructs used for recombinant protein expression. A) Schematic representation of the modular organisation of FOXP2. FOXP2 is composed an N -terminal NR- box region (containing the LXXLL motif; X denotes any residue),

a poly-glutamine (PolyQ) region, zinc finger and leucine zipper motifs, the forkhead domain and an acid-rich tail at the C-terminus. B) Sequences of the constructs used for recombinant expression of the FOXP2 NR-box and FHD. The His-tag (purple), thrombin cleavage site (cyan; symbol (⏐) indicates where the thrombin cleaves) and the respective sequence of the protein (NR-box – yellow, FHD- red) are indicated. C) Physiochemical characteristics of the FOXP2 constructs. The molecular weight and isoelectric point were computed using the ExPASy Compute pI/Mw tool (Gasteiger *et al.,* 2005). The theoretical extinction co-efficient was calculated using the ExPASy ProtParam tool (Gasteiger *et al.,* 2005).

FOXP2 FHD was overexpressed using the protocol outline by Morris and Fanucchi (2016). An overnight culture of 100 mL 2xYT media and 0.1 mg/mL of ampicillin was inoculated with 2 mL glycerol stock and grown overnight at 37 °C. The overnight culture with a 1 in 50 dilution was used to inoculate flasks with 400 mL of 2xYT media and 0.1 mg/mL ampicillin. Cells were grown at 37 °C with shaking until an optical density (OD) of 0.6 was attained. The OD was measured at 600 nm. Following cold shock, the cells were induced with 0.5 mM IPTG and overexpressed at 20 °C for 22 hours at 200 rpm.

Subsequently, the cells were harvested by centrifugation at 5,000 x*g* for 25 minutes at 4 °C. The pelleted cells were resuspended in 45 mL of equilibration buffer (20 mM Tris, 500 mM NaCl, 30 mM imidazole, pH 7.8) for 3 L of culture. The cells were lysed with 0.1 mg/mL of lysozyme which hydrolyses the peptidoglycan of the bacterial cell walls. The lysate was then frozen at – 20 °C. Cells were lysed by initially leaving them to thaw at 20 °C. Subsequently, the lysate was sonicated 5 times at 18x for 45 seconds with 30 second intervals on ice. Thereafter, 0.01 mg/mL of DNase was added to the lysate to remove the viscosity of the lysate caused by DNA presence and improve extraction efficiency. The lysate was then left to incubate at room temperature (RT) for 30 minutes and thereafter centrifuged at 23,000 x*g* for 25 minutes at 4 °C to separate out the soluble fractions from the insoluble. The supernatant was collected and purified.

4.4.2) FOXP2 N – Terminal region

The FOXP2 NT was constructed using the region of FOXP2 spanning 1 – 151 amino acids, as shown in Figure 4.3. This construct was expected to have greater solubility than other longer constructs containing the poly-glutamine rich region, as determined by protein-sol software (https://protein-sol.manchester.ac.uk/). This construct was inserted into a pET-11a plasmid with a His-tag, ampicillin-selection gene, and a thrombin cleavage site. The plasmid was used to transform T7 express and BL21(DE3) *E. coli* cells through heat shock treatment. Cells were grown

and used to prepare glycerol stocks as stated above in section 4.3. The plasmid DNA was isolated and sent for gene sequencing (Inqaba Biotec, RSA).

The transformed *E. coli* cells were used to overexpress the FOXP2 NT. As this region of FOXP2 had never been overexpressed previously, the conditions for optimal expression and protein production were determined by performing expression trials. The conditions investigated are summarised in Table 4.1. Trials were performed by inoculating 100 mL flask of culture media containing 0.1 mg/mL of ampicillin. The cells were grown overnight at 37 °C with vigorous shaking. The cultures were subsequently used to inoculate fresh culture media flasks containing 0.1 mg/mL ampicillin at a 1 in 50 dilution. The flasks were incubated at 37 °C until the OD_{600} was 0.6. Thereafter, the culture was placed on ice to cool to 20 °C. The OD was measured using a JASCO V-630 spectrophotometer. The cultures were inoculated with different concentrations of IPTG, and protein was expressed at specific temperatures (noted in Table 4.1) with shaking. At different time intervals, 1 mL of samples were collected for analysis.

Table 4.1: Conditions investigated for optimal FOXP2 NT protein overexpression

VARIABLES UNDER STUDY	RANGE INVESTIGATED
Bacterial Cell Type	T7 Express, BL21(DE3)
IPTG Concentrations (mM)	0, 0.1, 0.3, 0.5, 0.7, 1
Media	LB (1 % (w/v) tryptone, 1 % (w/v) NaCl, 0.5 % (w/v) yeast extract), LB + sucrose 2xYT (1.6 % (w/v) tryptone, 0.5 % (w/v) NaCl, 1 % (w/v) yeast extract),
Temperature (°C)	15, 20, 25, 30, 37
Overexpression Time (h)	0 – 24
Additives	Glucose / Sorbitol (0-10 %), $MgCl_2$ (10 mM)

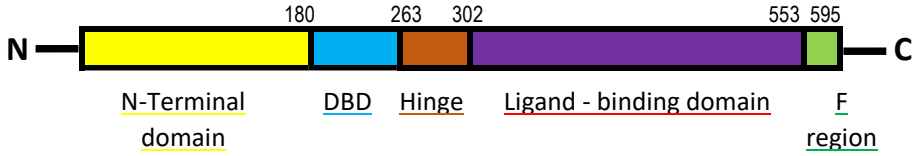
To harvest the cells, the samples were centrifuged at 12,000 x*g* for 10 minutes. The pellet was resuspended in 100 µL equilibration buffer (20 mM Tris, 500 mM NaCl, 30 mM imidazole, pH 7.5) and the supernatant was discarded. The cells were lysed by sonicating the samples at 3x speed on ice and centrifuged at 13,000 x*g* for 10 minutes at 4 °C. The pellet was resuspended in 100 µL of equilibration buffer and both the pellet and supernatant were resolved on a 16 % tricine SDS-

PAGE gel by adding reducing sample buffer (250 mM Tris/HCl, 30 % glycerol, 12 % SDS, 0.05 % Coomassie blue G-250, pH 7) to each of the samples (Schägger, 2006). Based on the results of the expression trials, the optimised condition for FOXP2 NT expressions was deduced to be as follows: FOXP2 NT transformed in BL21 (DE3) cells need to be grown to an OD_{600} of 0.6 in LB media supplemented with 5 % sucrose. Following cold shock, the cells require induction using 0.5 mM IPTG and should be expressed at 25 °C, 200 rpm for 6 hours.

4.4.3) Oestrogen receptor α ligand-binding domain

The ER1 LBD sequence between residues 302 and 552, along with a His-tag was inserted into a pET-11a plasmid as shown in Figure 4.4. The plasmid was used to transform either T7 express, BL21 (DE3) or BL21 (DE3) PLysS *E. coli* cells using heat shock treatment. Thereafter, using the aforementioned protocol (section 4.3), glycerol stocks of ER1 LBD-transformed *E. coli* cells were prepared.

A)



B)

Ligand-binding domain: MGSS HHHHHH KRS KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLD RNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDT LIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRLHAPTS

C)

	Quaternary structure	*Molecular Weight (Da)*	*pI*	*Extinction Co-efficient ($M^{-1}.cm^{-1}$)*	*Number of residues*	
					Trp	**Tyr**
ER1 LBD	Dimer	30 408.66 (monomer)	6.83	24 200	3	5

Figure 4.4: Structural information of the ER1 LBD. A) Schematic representation of the structural organisation of ER1 namely, the N-terminal domain, DNA-binding domain (DBD), hinge region, ligand-binding domain, and a poorly understood C-terminal tail (F region). B) Sequence of the construct that will be used to express recombinant ER1 LBD. The His-tag (pink) and the ligand-binding domain sequence (purple) are indicated. C) Physiochemical characteristics of ER1 LBD. The molecular weight and isoelectric point were computed using the ExPASy Compute pI/Mw tool (Gasteiger *et al.*, 2005). The theoretical co-efficient was calculated using the ExPASy ProtParam tool (Gasteiger *et al.*, 2005).

ER1 LBD has been successfully overexpressed and purified previously (Seielstad *et al.,* 1995; Eiler *et al.,* 2001). Using the protocol outlined by Eiler *et al.* (2001) as a basis, ER1 LBD overexpression trials were performed. The conditions outlined in Table 4.2 were investigated.

Table 4.2: Conditions investigated for optimal ER1 LBD overexpression

VARIABLES UNDER STUDY	RANGE INVESTIGATED
Bacterial Cell Type	T7 Express, BL21(DE3)
IPTG Concentrations (mM)	0, 0.1, 0.3, 0.5, 0.7, 1
Media	LB (1 % (w/v) tryptone, 1 % (w/v) NaCl, 0.5 % (w/v) yeast extract), LB + sucrose 2xYT (1.6 % (w/v) tryptone, 0.5 % (w/v) NaCl, 1 % (w/v) yeast extract), TB (2.0 % (w/v) tryptone, 2.4 % (w/v) yeast extract, 0.4 % (v/v) glycerol, potassium phosphate buffer (0.017M KH_2PO_4, 0.072 M K_2HPO_4)
OD_{600}	0.3-1.4
Temperature (°C)	15, 20, 25, 30, 37
Shaking Speed (Rpm)	110,150, 170, 200
Overexpression Time (h)	0 – 24
Additives	Glucose / Sorbitol (0-10 %), Glycerol (0 – 15 %), $MgCl_2$ (10 mM), Ethanol (1-3 %), 17 β-oestradiol (10 – 30 mM)
Cell Disruption Buffers	1) 20 mM Tris, 500 mM NaCl, 30 mM Imidazole at pH 7.5/8/9 Additives used: 25 - 200 mM Mannitol/ Sorbitol/ Sucrose, 0.1– 2% Triton x100) 2) 20 mM Tris, 300/600 mM NaCl, pH 7.5 3) 20 mM Tris, 500 mM KCl, pH 7.5 4) 50 mM Tris-HCl, pH 7.8, 50 mM NaCl, Additives: 0.1 - 2 M non- detergent sulfobetaine (NDSB), 1mM dithiothreitol (DTT)

Following expression trials, the optimal conditions were determined for the expression of soluble ER1 LBD. Cells were harvested by centrifugation and washed in several different types of ice-cold disruption buffers that are outlined in Table 4.2. NDSB is a zwitterionic compound with hydrophilic sulfobetaine head group. It aids in solubilising proteins without denaturing them. The cells were treated with 0.1 mM phenylmethanesulfonylfluoride (PMSF) and a protease inhibitor cocktail of leupeptin, aprotinin, antipain and chymostatin (1.25 µg/mL) before, during and following sonication. Protease inhibitors prevent unwanted proteolysis of ER1 LBD, Thereafter, the extract was centrifuged and the expression of the protein present in the supernatant was evaluated using SDS-PAGE (Eiler *et al.,* 2001).

4.5) Purification of proteins

FOXP2 FHD, FOXP2 NT and ER1 LBD were purified using various chromatography techniques.

4.5.1) Immobilised-metal ion affinity chromatography

Immobilised-metal ion affinity chromatography (IMAC) is a versatile, rapid and an inexpensive purification method employed on the basis of entrapment of affinity ligands. FOXP2 FHD was purified using a 2-step protocol as optimised by Blane and Fanucchi (2015). IMAC was performed using a HisTrap column (Cytiva, USA) charged with Ni^{2+}. The column was equilibrated with equilibration buffer (20 mM Tris, 500 mM NaCl and 30 mM imidazole at pH 7.8) prior to loading the cell lysate. The protein buffers used in this study are listed in Table 4.3. The His-tagged protein would bind to the immobilised Ni^{2+} ions while the contaminants were collected as flow-through. Thereafter, the column was washed with a high salt buffer (20 mM Tris, 1.5 M NaCl and 30 mM imidazole, pH 7.8) which is effective at removing non-specifically bound proteins. Imidazole was included in the wash buffer to increase the stringency of the wash so that non-specifically bound proteins can be effectively eluted. The column was re-equilibrated with equilibration buffer and eluted using elution buffer (20 mM Tris, 500 mM NaCl and 500 mM imidazole at pH 7.8). High concentration of the histidine analogue imidazole out competes the bound poly-histidine residues for binding to the immobilised metal ions and thus results in the elution of the protein of interest.

Table 4.3: Buffers used in the study	
Experiments	**Buffers**
IMAC	Equilibration: 20 mM Tris, 500 mM NaCl, 30 mM Imidazole, pH 7.8 Salt wash: 20 mM Tris, 1.5 M NaCl, 30 mM Imidazole, pH 7.8 Elution: 20 mM Tris, 500 mM NaCl, 500 mM Imidazole, pH 7.8
SEC	50 mM Tris, 300 mM NaCl, pH 7.8
Thrombin Cleavage	20 mM Tris, 2 mM KCl, 100 mM NaCl, pH 8
Quantification	20 mM Tris, 100 mM NaCl, pH 7.8
Fluorescence experiments/ EMSA	50 mM Tris, 100 mM NaCl, 1mM TCEP pH 7.8 (NaCl concentration was varied from 50-300 mM)
CD	5 mM Sodium phosphate, 1 mM TCEP pH 7.8

FOXP2 NT and ER1 LBD were purified similarly to the FOXP2 FHD with minor alterations. Instead of Ni^{2+} ions, FOXP2 NT was purified using an IMAC column immobilised with Co^{2+} ions as the protein appeared with fewer contaminating bands. The protein samples following IMAC purification were evaluated using tricine SDS-PAGE and all three proteins required a second purification step to attain complete sample purity.

4.5.2) Size-exclusion chromatography

Size-exclusion chromatography (SEC) purification is based on the separation of molecules by size through the use of a matrix of beads that serve as a sieve. Following IMAC, FOXP2 FHD, FOXP2 NT and ER1 LBD were purified using SEC (Blane and Fanucchi, 2015). SEC was performed by initially conducting an overnight equilibration of a Superdex^{TM} 75 prep grade column (Cytiva, USA) with SEC buffer (50 mM Tris, 300 mM NaCl, pH 7.8). Thereafter, the sample was injected on to the SEC column. It is optimal to inject less than 3 % of the column volume for SEC to avoid poor resolution. Hence, 11 ml of protein at maximum was injected at once. Fractions corresponding to peaks in UV reading were collected. The purity of the collected protein was evaluated using SDS-PAGE. The clean fractions were pooled and dialysed against storage buffer

(50 mM Tris, 100 mM NaCl at pH 7.8) using snakeskin dialysis tubing with 3.5 K (FOXP2 FHD and FOXP2 NT) or 10 K (ER1 LBD) cut-offs. This ensured buffer homogeneity between the proteins and the complete removal of imidazole and other contaminants from the sample.

4.5.3) Thrombin Cleavage of FOXP2

One of the downstream applications in this study required the removal of the His-tag of both the FOXP2 proteins. Hence, for both FOXP2 FHD and NT, the eluted protein fractions were pooled together following IMAC, and dialysed against cleavage buffer (20 mM Tris, 2 mM KCl, 100 mM NaCl, pH 8) overnight. Thereafter, the protein was incubated with thrombin for a period of 6 hours at RT. Cleaved FOXP2 was separated from the His-tag and uncleaved protein by connecting a 1 mL HiTrap benzamidine FF column (GE Healthcare, USA) in tandem to a Ni^{2+}/Co^{2+} charged IMAC column as described by Blane and Fanucchi (2018). Similar to IMAC, the columns were equilibrated with equilibration buffer, however the cleaved protein was collected in the flow-through. Thereafter, the benzamidine column was detached and elution buffer was passed through the IMAC column only to elute all uncleaved protein. The samples were evaluated on a 16% tricine SDS-PAGE gel.

4.6) Purity and molecular weight determination

The purity of the protein samples was evaluated following each purification step using sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE). Additionally, the molecular weights were approximated following the final purification step to ensure the purified fractions collected were indeed the proteins of interest.

4.6.1) Sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE)

Electrophoresis refers to the separation of macromolecules in an electric field. PAGE utilises polyacrylamide gels as the support medium as polyacrylamides are a polymer of acrylamide monomers and can form a porous gel matrix. Tricine SDS-PAGE is a modified version of the traditional Laemmli system which is not suitable for separation of proteins that are less than 15 kDa as the gels have poor resolution due to the comigration of SDS and smaller proteins (Laemmli, 1970; Schägger and Von Jagow, 1987). Gradient gels or gels with a high acrylamide concentration could improve the resolution. However, this is not preferred due to difficulties in gel casting, fragility, and irreproducibility. In contrast, tricine SDS-PAGE utilises tricine as trailing ions as opposed to glycine. The differences in pKa of tricine (8.15) and glycine (9.6) affects the stacking

of proteins. This is because tricine migrates much faster than glycine at pH values between 6.8 and 8.8 as a greater proportion of tricine is in the ionic form. Thus, the stacking limit is shifted to a low-molecular weight range as the ratio of the mobility of the trailing ions relative to the mobility of the proteins is increased (Schägger and Von Jagow, 1987). Tricine SDS-PAGE can be conducted using low acrylamide concentrations and in the absence of urea and thus is a simple and efficient method for separating low molecular weight proteins.

The molecular weight and purity of FOXP2 FHD, FOXP2 NT and ER1 LBD samples, prior to and following each purification step, were determined using tricine SDS-PAGE (Schägger, 2006). A 16 % separating gel and 4 % stacking gel were prepared with an acrylamide-bis-acrylamide stock (49.5% T and 3% C), gel buffer (3 mM Tris, 1 mM HCl and 0.3 % SDS at pH 8.45), ammonium persulfate (APS) and tetramethyl ethylenediamine (TEMED). APS aids in the polymerisation of acrylamide by producing free radicals while TEMED accelerates this process by catalysing the radical formation.

Tricine SDS-PAGE was performed by first adding 16 µL of reducing sample buffer (12 % (w/v) SDS, 6 % (w/v) β-mercaptoethanol, 30 % (w/v) glycerol, 0.005 % Coomassie blue G-250 and 150 mM Tris-HCl, pH 7.5) to 50 µL of sample. The samples were thereafter boiled at 95 °C for 5 minutes, followed by a brief vortex and centrifugation at 10,000 x*g* to ensure that the samples were properly mixed at the bottom of the tubes. Subsequently, 10 µL of samples were loaded and electrophoresed at 50 V until stacking had occurred before the voltage was increased to 140 V. The anode buffer was composed of 1 M Tris, adjusted to a pH of 8.0 with HCl and the cathode buffers of 1 M Tris, 1 M tricine and 1 % SDS at an unadjusted pH of approximately 8.25. The gel was stained for 2 hours using Coomassie blue and thereafter de-stained until the bands on the gel were clearly visible. The bands were compared against the molecular weight marker to approximate the size of the proteins.

4.7) Protein concentration determination

4.7.1) Bradford Assay

The Bradford assay is an inexpensive and sensitive protein assay based on the binding of certain protein residues to the Coomassie brilliant blue G-250 dye (Bradford, 1976). Bradford assay was used to quantify FOXP2 FHD, FOXP2 NT and ER1 LBD. A 2 mg/mL bovine serum albumin (BSA) stock sample was prepared in buffer (20 mM Tris, 100 mM NaCl, pH 7.8). BSA samples

with concentration ranging from 0 to 0.5 mg/mL were prepared and incubated with Bradford reagent for 30 minutes. Similarly, protein samples were prepared with Bradford reagent and incubated. The absorbance was measured at 595 nm on the JASCO V-630 spectrophotometer with buffer and Bradford reagent used as the blank. The concentrations of protein samples were determined by plotting a standard curve of the BSA concentrations against their respective absorbances (Olson and Markwell, 2007). The resultant concentrations were converted to molar concentrations using the Beer-Lambert law (Equation 1).

Equation 1: $A_{280} = \varepsilon.c.l$

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

4.7.2) Measurement of UV absorbance

Accurate quantitation of protein concentration is required for biochemical analysis of proteins. The simplest method for determining the protein concentration in solution is to measure the absorbance at 280 nm (A_{280}) as UV light is maximally absorbed by the side chains of aromatic amino acids and cystine residues in protein solution (Olson and Markwell, 2007).

4.7.2.1) Absorbance at 280 nm

Highly concentrated protein samples were usually at too low a volume to use the Bradford assay. As a result, the concentrations of purified FOXP2 FHD and ER1 LBD were determined by measuring the absorbance at 280 nm on the NanoDrop One spectrophotometer (ThermoFisher, USA) or the JASCO V-630 spectrophotometer. The absorbance values were corrected for the light scattering contributions made by buffers and aggregates (Equation 2)

Equation 2: $A_{280\text{-protein}} = A_{280\text{-protein}} - A_{280\text{-buffer}} - [(A_{340\text{-protein}} - A_{340\text{-buffer}})\ (A_{280\text{-buffer}} / A_{340\text{-buffer}})]$

and concentrations were calculated using the Beer-Lambert law (Equation 1). The theoretical extinction coefficients of FOXP2 FHD and ER1 LBD are 22 460 $M^{-1}. cm^{-1}$, and 24 200 $M^{-1}. cm^{-1}$, respectively, as tabulated in Figure 4.2 and 4.3 (Gasteiger *et al.*, 2005). Notably, when proteins are at high concentrations, the absorption co-efficient is influenced by electrostatic interactions between molecules at close proximity leading to inaccuracies in protein concentration calculations. Therefore, to keep within the region of linearity of the Beer-Lambert law, highly concentrated samples were diluted 10 to 50-fold. Additionally, the presence of nucleic acid contamination in

the protein samples was determined using the A_{260}/A_{280} ratio; this is indicative of the amount of contaminating DNA relative to the amount of protein. The protein sample were considered relatively pure or free from nucleic acid contamination if the ratio was approximately 0.6.

4.7.2.2) Absorbance at 205 nm

As FOXP2 NT does not contain any tryptophan residues and only a single tyrosine residue, it has an exceptionally low extinction co-efficient, which is calculated with a 10 % inaccuracy rate on ProtParam (Gasteiger *et al.,* 2005). In this instance, the concentration of FOXP2 NT was determined by measuring the absorbance of a dilute protein sample at 205 nm. At 205 nm, the absorbance of the peptide backbone is measured. Many proteins at 1 mg/mL have extinction coefficients ranging from 30 to 35 M^{-1}. cm^{-1} at 205 nm. More specifically, the absorbance of proteins with little to no tryptophan and tyrosine residues are assumed to have an extinction co-efficient of 31 mL. $mg^{-1}cm^{-1}$ (Scopes, 1974; Anthis and Clore, 2013).

In order to determine FOXP2 NT concentration, the protein sample was diluted in a quantification buffer (20 mM Tris, 100 mM NaCl, pH 7.8) to ensure that buffer components are kept to low concentrations that do not interfere with the absorbance reading at 205 nm. The absorbance was measured using the Nanodrop One spectrophotometer (ThermoFisher, USA) thrice and averaged. Concentrations were determined using the Beer Lamber law (equation 1) with the assumption that the protein extinction co-efficient is 31 mL. $mg^{-1}cm^{-1}$.

4.8) Structural Characterisation

The secondary structural content and quaternary structures of FOXP2 FHD, FOXP2 NT and ER1 LBD were evaluated using far-UV circular dichroism and size exclusion chromatography (SEC), respectively.

4.8.1) Far-UV circular dichroism

CD spectroscopy exploits the chirality of amino acid residues to provide structural information of proteins in solution. Far-UV CD spectroscopy was used to confirm the structural integrity of FOXP2 FHD, FOXP2 NT and ER1 LBD by characterising their secondary structure. CD was conducted by first dialysing the samples against CD buffer (5 mM sodium phosphate, pH 7.8). While this buffer is different to the one used consistently for all other applications in this study (50 mM Tris, pH 7.8), it was the only buffer in which all three proteins were stable without

compromising the signal-to-noise ratio of the spectra. The ellipticity was measured using a Jasco J1500 CD spectrometer at wavelengths ranging between 190−250 nm using 10 μM of FOXP2 proteins and 8 μM of ER1 LBD in a quartz cuvette. CD was performed in triplicate and normalised by converting the values from ellipticity (m.deg) to mean residue ellipticity (θ_{MRE}), with the units, deg.cm^2.dmol^{-1} (Equation 3) (Blane and Fanucchi, 2015). The data was deconvoluted using DichroWeb using the *k2d* neural network, an artificial intelligence-based algorithm that uses the CD spectra of proteins with known tertiary structure to quantitatively predict the secondary structural content of other proteins from their CD spectra (Andrade *et al.*, 1993; Whitmore and Wallace, 2004). A paired t-test was performed using SigmaPlot 15.0 to compare the significance of the variance between the apo- and bound-forms or the native and denatured forms of the proteins.

Equation 3: $\theta_{MRE} = \frac{(\theta \times ellipticity\ (m.\mathrm{deg}) \times 10^6)}{path\ length\ (mm) \times [protein](\mu M) \times No.of\ peptides}$

4.8.2) Size exclusion chromatography

SEC, as mentioned in section 4.5.2, was used to determine oligomeric states of the proteins. Size exclusion standards were loaded onto the Superdex^{TM} 75prep grade column (Cytiva, USA) with SEC buffer (50 mM Tris, 300 mM NaCl, pH 7.8). The standards included blue dextran (250,000 kDa), conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (15.7 kDa) and aprotinin (6.5 kDa). The size of the eluted protein was approximated by constructing a standard curve with the log of the known molecular weights of the standards against their respective ratio of the eluted volume against the void volume. The elution volume of blue dextran was considered the void volume as the glucose polymer is exceptionally large and thus excluded from the SEC column. A linear regression line was fitted to the standard curve using SigmaPlot 15.0 and the resulting equation was used to calculate the sizes of FOXP2 FHD, FOXP2 NT and ER1 LBD associated with the elution volume of the peaks observed when each of the proteins were loaded onto the SEC column.

4.9) Fluorescence anisotropy

Fluorescence anisotropy exploits specific properties of fluorescence to provide more information regarding the structure of proteins. Fluorescence spectroscopy is based on the fluorescence phenomenon where linearly polarised light of a specific wavelength is absorbed by fluorophores

and re-emitted at a longer wavelength (Lakowicz, 2006). The fluorescent molecules have an excitation as well as an emission dipole that are spatially aligned to the electric vector of the excitation and emission light, respectively. However, in the period between the excitation of an electron and fluorescence, the fluorophores in space are continuously changing orientations due to diffusional rotation. This, in turn dictates the polarisation of the emitted light (Lakowicz, 2006). The extent of polarisation is dependent on the speed of rotation of the fluorophores and is referred to as anisotropy. Hence, fluorescence anisotropy is based on the principle that fluorophores exhibit partially polarised emission when excited by polarised light. This can be mathematically stated as follows when molecules are excited by vertically polarised light (Equation 4).

Equation 4: $Anisotropy = \frac{I_{VV} - GI_{VH}}{I_{VV} - 2GI_{VH}}$; $G = \frac{I_{HV}}{I_{HH}}$

(I_{VV} indicates the fluorescence intensities of the vertically polarised emission, I_{VH} represents the intensities of horizontal emissions and G is the grating factor which corrects for the differential sensitivity of the instrument to vertically and horizontally polarised light.)

4.9.1) DNA interaction

FOXP2 FHD is a DNA-binding domain. To confirm the protein's functionality, fluorescence anisotropy was performed using RhodamineX (ROX)-labelled Nelson DNA. This contains the consensus sequence (5'-TGTTTAC-3') known to bind tightly to the FOXP2 FHD. ROX-labelled Nelson DNA was diluted to 0.2 µM in 50 mM Tris pH 7.8 buffer with either 50 mM or 100 mM NaCl and used to calculate the G-factor. FOXP2 FHD in the same buffer was the titrant. The experiment was performed using the PerkinElmer LS-50 fluorescence spectrophotometer equipped with an anisotropy filter. Excitation and emission wavelengths were set at 580 and 605 nm respectively with 5 nm excitation and emission slit. The experiment was repeated in triplicate for both salt concentrations. To eliminate any doubt of ER1 LBD's DNA-binding activity and it being a possible confounding factor for downstream applications, the experiment was also repeated using ER1 LBD as the titrant.

4.9.2) Identification of Protein-protein interaction

Fluorescence anisotropy can be utilised to study PPIs as the rate of rotation is dependent on the molecular size of the rotating molecule (Lakowicz, 2006). However, to successfully perform anisotropy to elucidate PPI, one of the interacting proteins needs to be fluorescently labelled.

4.9.2.1) Protein Labelling

The three types of fluorophores used to study protein interactions include the intrinsic tryptophan residues of the protein, extrinsically attached fluorophores and fluorescence fused partners. Prior to performing fluorescence anisotropy, the ER1 LBD was labelled with AlexaFluor™ C 2 maleimide 555. This was accomplished by taking 10-50 µM protein in SEC buffer and incubating it with three times the excess dye and shaking it for 1 hour at RT. Thereafter, the sample was left to incubate overnight. Excess dye was removed using Pierce™ dye removal spin columns (ThermoFisher, USA). The dye conjugate-to-protein ratio was determined by measuring the absorbance at 280 nm and 556 nm, the absorption maxima of the protein and the AlexaFluor dye, respectively. This was done using the NanoDrop one spectrometer (ThermoFisher. USA). The ratio was calculated using Equation 5. ER1 LBD labelled with a dye/protein ratio of 2 or higher was used for fluorescence anisotropy.

Equation 5: $Dye{:}\,Protein = \frac{A_{556}\times\varepsilon_{protein}}{[A_{280}\times 0.08]\times 158000}$

(Where A_{556} and A_{280} are the absorbance maxima of AlexaFluor 555 C 2 maleimide and protein, respectively; $\varepsilon_{protein}$ is the molar extinction co-efficient of the protein; 158 000 is the molar extinction co-efficient of the AlexaFluor dye; 0.08 is the correction factor for the absorbance of the fluorophore at 280 nm.)

In addition, ER1 LBD was also labelled using NTA-Atto 550 (Merck, USA) using the afore-mentioned protocol. Notably, this was not the preferred choice as the dye binds non-covalently to the His-tag of the protein and has the added of risk of dissociating from the protein when the equilibrium is disturbed following titrations in the fluorescence anisotropy experiments. To calculate the dye to protein ratio using equation 5, a correction factor of 0.12 and an extinction co-efficient of 120 000 were used.

4.9.2.2) Protein-protein interaction assays

Once the protein was labelled efficiently, it was diluted to 0.3 µM in 50 mM Tris, 100 mM NaCl, buffer of pH 7.8. This was placed in a quartz cuvette and used to measure the G-factor at excitation. FOXP2 FHD or FOXP2 NT dialysed against the appropriate buffer (50 mM Tris, 100 mM NaCl, at pH 7.8.) was titrated in continuously. Following each titration, the sample was incubated for 5 minutes in the dark. Thereafter the average of the first 10 anisotropy values was taken into

consideration. The experiment was performed in triplicate using the PerkinElmer LS-50 fluorescence spectrophotometer equipped with an anisotropy filter. The excitation wavelength was set at 556 nm and emission at 572 nm. The excitation and emission slit width was kept at 5 mm. The experiments were performed in triplicate. Thereafter, the data was exported, averaged, and used to form a single saturation ligand - binding non-linear regression curve using SigmaPlot 15.0 with the concentration of FOXP2 protein on the x-axis and the relative fluorescence anisotropy on the y-axis. The single saturation curve fitted best to the datapoints collected which was the reason for its selection. The equation of the fit of the fluorescence anisotropy curves used to deduce the K_d is noted in equation 6.

Equation 6: $y = \frac{Bmax(X)}{Kd+X}$

(Where Bmax denotes the maximum number of binding sites, X is the concentration of free ligand, K_d is the concentration of ligand required to reach half-maximal binding

4.9.3) Regulation of protein-protein interaction

The interaction between FOXP2 FHD and ER1 LBD was characterised further by investigating the interaction at different salt concentrations and in the presence of DNA and ligand. The anisotropy experiments were performed using the same protocol mentioned in section 4.8.2.2 with alterations in either buffer condition, labelled sample, or titrant.

Effect of salt: The anisotropy experiments were performed with labelled ER1 LBD and FOXP2 FHD in 50 mM Tris, pH 7.8 buffer with the following salt concentrations: 50 mM, 200 mM, and 300 mM. A one-way Analysis of Variance (ANOVA) test was performed on SigmaPlot 15.0 to ascertain whether the difference in the mean anisotropy readings at the different salt concentration following the titration of the same concentration of FOXP2 FHD was significant ($p < 0.05$). This test was selected as the data comprised of a single quantitative dependent (anisotropy) variable and an independent variable made up of four categories (different salt concentrations). In addition, the data passed both the normality and equal variance distribution test (Lee and Lee, 2018). Following ANOVA, the Bonferroni method was used to perform pairwise comparisons between each of the different salt concentrations.

Effect of ligand-binding: To determine the effect of ligand-binding on the ability of ER1 LBD to bind to FOXP2 FHD, anisotropy was performed using labelled ER1 LBD (0.3 µM) incubated with

excess E2 ligand (50 µM) overnight. The ER1 LBD was thereafter titrated with FOXP2 FHD (500 µM). The buffer for this experiment was 50 mM Tris, 100 mM NaCl, 0.25 % ethanol pH 7.8 to account for the ethanol-based solubilisation of E2. Functional characterisation of ER1 LBD's interaction with E2 has been described in the subsequent section.

Effect of DNA-binding: To determine if FOXP2 FHD's interaction with DNA influenced its ability to interact with ER1 LBD, FOXP2 FHD (450 µM) was incubated overnight with DNA (450 µM). This FOXP2 FHD was titrated stepwise into 0.3 µM fluorescently labelled ER1 LBD in 50 mM Tris, 100 mM NaCl, pH 7.8 buffer. Anisotropy was measured and analysed as stated in section 4.9.2.

Effect of protein-protein interaction on DNA-binding activity of FOXP2 FHD: Finally, ER1 LBD (0.3 µM) was incubated with saturating concentrations of FOXP2 FHD determined from previous anisotropy experiments in 50 mM Tris, pH 7.8 buffer with either 50 mM or 200 mM NaCl concentrations. The final concentration of FOXP2 FHD was 32 or 68 µM in buffer containing 50 mM or 200 mM NaCl, respectively. This ER1 LBD-FOXP2 FHD sample was then titrated with increasing concentrations of Nelson DNA ranging from 0-60 µM. The aim of performing this particular anisotropy experiment was to deduce if FOXP2 FHD in complex with ER1 LBD retains its ability to bind to Nelson DNA. The fluorescence anisotropy experiments performed have been summarised in Table 4.4. The results were fitted to a one-site saturation ligand-binding curve (equation 6) on SigmaPlot 15.0 where applicable. This has been noted in Table 4.4.

Table 4.4: Fluorescence anisotropy experiments to investigate protein-DNA and protein-protein interactions.

	Labelled molecule	Titrant	Buffer	Additional notes
4.7.1	ROX-labelled Nelson DNA (0.2 µM)	FOXP2 FHD (150 µM)	50 mM Tris, 50/100 mM NaCl- pH 7.8	
4.7.1	ROX-labelled Nelson DNA (0.2 µM)	ER1 LBD (50 µM)	50 mM Tris, 100 mM NaCl, 1mM TCEP - pH 7.8	
4.7.2	AF555-labelled ER1 LBD (0.3 µM)	FOXP2 FHD (700 -1200 µM)	50 mM Tris, 100 mM NaCl, 1mM TCEP - pH 7.8	
4.7.2	AF555-labelled ER1 LBD (0.3 µM)	FOXP2 NT (300 µM)	50 mM Tris, 100 mM NaCl, 1mM TCEP - pH 7.8	Poor fit with all binding curves, therefore not fitted.
4.7.3	AF555-labelled ER1 LBD (0.3 µM)	FOXP2 FHD (400-1000 µM)	50 mM Tris, 1mM TCEP - pH 7.8 NaCl concentrations: 50 mM, 200 mM, 300 mM	
4.7.3	AF555-labelled ER1 LBD (0.3 µM)	FOXP2 FHD (500 µM) + Nelson DNA (500 µM)	50 mM Tris, 100 mM NaCl, 1mM TCEP- pH 7.8	FOXP2 FHD incubated with Nelson DNA overnight at 4 °C before anisotropy experiment. Poor fit with all binding curves, therefore not fitted.
4.7.3	AF555-labelled ER1 LBD (0.3 µM) + E2 (50 µM)	FOXP2 FHD (700-800 µM)	50 mM Tris, 100 mM NaCl, 1mM TCEP - pH 7.8	ER1 LBD incubated with E2 overnight at 4 °C before anisotropy experiment.
4.7.3	AF555-labelled ER1 LBD (0.3 µM) + FOXP2 FHD (300-700 µM)	Nelson DNA (200-700 µM)	50 mM Tris, 50/200 mM NaCl, 1mM TCEP - pH 7.8	Labelled ER1 was used to calculate G-factor. FOXP2 FHD added to saturating concentration and sample incubated for 30 minutes at RT. Fluorescence reading taken and thereafter titrations with Nelson DNA.

AF555 -AlexaFluor™ 555 C 2 maleimide dye, E2 - 17β-oestradiol

4.10) Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a quantitative biochemical tool used in the characterisation of intermolecular interactions. Protein interactions, be they with nucleic acid, small molecule or proteins, result in changes in thermodynamic properties. ITC is able to precisely determine the enthalpy, entropy, Gibbs free energy and heat capacity changes associated with

binding. The binding affinity and enthalpy is determined directly while Gibbs free energy and entropy can be calculated using equations 8 and 9, respectively (Ladbury and Chowdhry, 1996).

Equation 8: $\Delta G = RTln(Ka)$

(where ΔG refers to Gibbs free energy (kJ/mol), R the gas constant (8.314 J/K.mol), T the temperature (K) and K_a the association constant or binding affinity.)

Equation 9: $\Delta G = \Delta H - T\Delta S$

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

4.9.1) Functional characterisation of ER1 LBD's ligand binding activity

ER1 LBD can interact with several types of ligands through its hydrophobic ligand binding pocket. This includes its natural ligands including E2 or endocrine therapy drugs such as tamoxifen (Ruff *et al.,* 2000). In addition, ER1 LBD can also interact with phytoestrogens, which are plant-based molecules that can mimic the activity of oestrogens. The functional activity of ER1 LBD was characterised using ITC.

The ITC experiment was performed using the Nano isothermal titration calorimeter (TA Instruments, USA). Stock solution of 10 mM E2 was prepared with 99.9 % ethanol and diluted to 100 µM in 50 mM Tris, 100 mM NaCl, 1 mM Tris (2-carboxyethyl) phosphine (TCEP), pH 7.8 buffer. The final concentration of ethanol in the E2 sample was 1 % (v/v). Purified ER1 LBD was dialysed against the same buffer and concentrated to 10 µM. To ensure perfect buffer match with the ligand, 0.5 % ethanol was added to the protein sample and placed in the sample cell. Using the 250 µL syringe containing E2, ITC was performed at 20 °C, with 5 µL injection volumes, 150 rpm stirring rate and 300 s equilibration time following each injection. The data was fitted using a one-site independent model using the NanoAnalyze 4.12.5 software (TA Instruments, USA). This model, which is based on the interaction of 'n' ligands binding to a single site on a macromolecule (usually 1:1 interaction), was chosen as the best fit was obtained compared to all other models.

4.10.2) Analysis of protein-protein interaction

ITC can also be used to determine the binding affinity and thermodynamic parameters of PPI. Therefore, to confirm and supplement the PPI identified using fluorescence anisotropy, ITC was

performed using the Nano isothermal titration calorimeter (TA Instruments, USA). ER1 LBD (8.5 µM) was placed in the cell and titrated continuously with 5 uL injections of FOXP2 FHD (50 µM). The buffer used was 50 mM Tris, 100 mM NaCl, 1 mM TCEP, pH 7.8. The stirring rate, integration time and temperature were set at 150 rpm, 350 seconds and 20 °C, respectively. In addition, the experiment was performed in the same buffer but with 300 mM NaCl using 15 µM of ER1 LBD and 400 µM FOXP2 FHD under the same conditions. The resultant curves were fitted to an independent model on NanoAnalyze 4.12.5 software.

4.11) Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) is an important technique for investigating gene regulation by evaluating the interaction between protein and DNA.

4.11.1) Functional characterisation

As previously mentioned, FOXP2 FHD is a DNA-binding domain and is known to bind the following consensus sequence with high affinity: 5'-TGTTTAC-3' (Nelson *et al.,* 2013). The activity of FOXP2 FHD in its apo and ER1 LBD-bound forms were evaluated in section 4.7.3 using fluorescence anisotropy. To re-affirm and supplement those results, EMSA was performed.

Nelson DNA (5'-AGGTGTTTACTTTCATAG-3') was synthesised as a duplex for this work (Integrated DNA Technology, South Africa). Samples were prepared with 0.5 µM DNA and increasing concentrations of FOXP2 FHD ranging from 0 to 20 µM in EMSA binding buffer (10 mM HEPES, 100 mM KCl, 1 mM Ethylenediaminetetraacetic acid (EDTA), 1 mM $MgCl_2$, 0.1 mg/mL bovine serum albumin, 10 % (v/v) glycerol, pH 7.5) and MilliQ water (Blane *et al.,* 2018). The samples were incubated at 4 °C for 60 minutes and loaded onto a 10 % polyacrylamide gel, which was prepared using acrylamide: bisacrylamide monomer solution, Tris-Borate-EDTA (TBE) buffer (0.089 M Tris, 0.089 M Borate and 2 mM EDTA, pH 8.4) and 20 % triethylene glycol. The gel was electrophoresed at 160 V for 1 h 45 minutes at 4 °C in TBE buffer. The gel was stained using GelRed dye (1:1000 dilution in distilled water) and viewed on the Molecular Imager® Gel Doc™ XR System (Bio-Rad). The process was repeated with FOXP2 NT and ER1 LBD to evaluate if they too had similar DNA-binding activity to FOXP2 FHD.

4.11.2) Effect of protein-protein interaction on FOXP2 FHD's functional activity

The interaction between FOXP2 FHD and ER1 LBD could affect the transcriptional activity of the FOXP2 protein. Hence, the ability of the FOXP2 FHD-ER1 LBD complex to bind to the cognate DNA sequence of FOXP2 FHD was investigated using EMSA. In this instance, FOXP2 FHD (2.5 and 5 µM) was incubated overnight at 4 °C with increasing ER1 LBD concentration (0-34 µM). This ER1 LBD was either in its apo-form or in the presence of excess E2 (300 µM). E2 was prepared from a 10 mM stock of E2 dissolved in 99.9 % ethanol. The samples were evaluated on an 8 % EMSA gel (prepared as outlined in section 4.10.1) using the set up shown in Figure 4.5 with each lane containing 0.5 µM Nelson DNA and the specified concentration of either or both FOXP2 FHD and ER1 LBD. Following electrophoresis, staining, and viewing, the gels were quantified relative to the lane containing only free DNA in ImageLab 5.2.1, as shown in Figure 4.5. The change in percentage intensity of the lower band across all ER1 LBD concentrations was plotted as a bar graph using SigmaPlot 15.0.

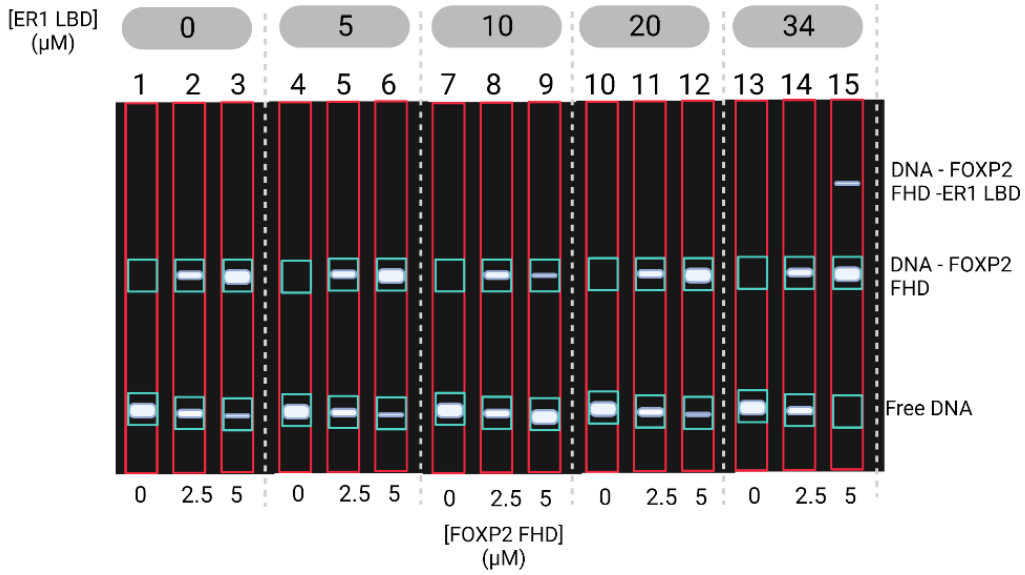


Figure 4.5: Set up and densitometric analysis of EMSA gels. Nelson DNA (0.5 µM) was incubated with different concentrations of FOXP2 FHD (0-5 µM). Each of these samples was replicated in the presence of increasing ER1 LBD concentrations (0-34 µM). Following electrophoresis, the gel was expected to have either one, two or three bands in each lane. To analyse the gel using ImageLab 5.2.1, lanes (red block) and bands (cyan block) were manually selected and thereafter, the bands correlating to free DNA and DNA-FOXP2 FHD complex in each lane were quantified using the first lane (containing only DNA) as the reference.

4.11) Molecular Docking

The interaction between FOXP2 FHD and ER1 LBD has been investigated in the presence of their other binding partners. In order to make sense of the results in terms of the structural properties of the proteins, molecular docking was employed. Three different docking tools were used based on protein sequence and structure. Additionally, there were two structure-based programmes used: one based on a rigid-docking protocol and the other on information-based flexible docking.

4.11.1) AlphaFold

AlphaFold is a sequence – based structure prediction tool. It incorporates the neural network architectures of proteins and has been trained to perform structure prediction taking the evolutionary, physical, and geometric constraints of protein structures into consideration (Jumper *et al.,* 2021). AlphaFold 2 is also capable of docking two or more protein structures when their sequences are uploaded as an input. It does this using a fold and dock approach whereby the proteins are folded and docked simultaneously, the accuracy is largely based on multiple sequence alignment accuracy (Bryant *et al.,* 2022).

AlphaFold docking was conducted using Google CoLab which employs a simplified version of AlphaFold v2.2.4: https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb (Mirdita *et al.,* 2022). Third party software and AlphaFold were first downloaded in the CoLab workbook. The amino acid sequence of ER1 LBD and FOXP2 FHD were used as input sequences. All prompts thereafter were followed. Briefly, the two input sequences were searched against various databases (UniRef90, MGnify and smallbfd) to generate multiple sequence alignment data. Thereafter, Bio sequence analysis was performed using Hidden Markov Models (HMM) followed by the execution of the AlphaFold run. The resultant PDB file of the predicted structure was downloaded and analysed.

4.11.2) ClusPro

ClusPro is a web-based server that performs direct docking of two interacting proteins. The direct method is based on thermodynamics and finding the structure conformation with minimum Gibbs free energy. It requires a model for free energy evaluation and an algorithm to perform energy minimisation (Kozakov *et al.,* 2004). The server first performs a rigid body docking by sampling billions of conformations using PIPER, a Fast Fourier Transform (FFT) approach. In this approach,

one protein called a receptor is placed at the origin of a coordinate system on a fixed grid. The second protein (denoted the ligand) is then placed on a movable grid and arranged with the receptor until all possible conformations have been exhausted. Thereafter, the server clusters 1000 structures with the lowest minimum energy - based on their root-mean-square deviation (RMSD). Finally, it refines the selected structure by performing energy minimisation.

To perform docking using ClusPro, the PDB structures of FOXP2 FHD (PDB ID: 2A07) and ER1 LBD (PDB ID: 1ERE) were uploaded (ClusPro 2.0: protein-protein docking) (Kozakov *et al.,* 2017; Desta *et al.,* 2020). ER1 LBD was denoted the receptor as it is larger than FOXP2 FHD, which was denoted the ligand. Prior to uploading, the two PDB files were prepared for docking using UCSF Chimera 1.16. FOXP2 FHD structure was modified to only include the FOXP2 FHD monomer. All other chains, including that of DNA were removed. DockPrep was performed to minimise the energy of the structure; charges were assigned to the Gasteiger residues. As no information is available on the protein-protein structure, the default settings were used to predict the docked structure.

4.11.3) HADDOCK

High ambiguity driven docking (HADDOCK) approach was used as a third docking tool. It is an information-based approach that makes use of biochemical and/or biophysical interaction data from NMR or X-ray crystallography. The information on the interacting residues is introduced as ambiguous interaction restraints (AIRs); this is divided into active and passive residues and drives the docking. Thereafter, the structures are ranked based on the sum of energy terms including electrostatic, van der Waals and AIR forces.

To perform protein-protein docking, the FOXP2 FHD (PDB ID: 2A07) and ER1 LBD (PDB ID: 1ERE) were first prepared in UCSF Chimera 1.16 by first removing all ligands (including Mg^{2+} ions) and all chains, leaving behind only a monomeric form of both proteins. Thereafter, DockPrep was employed. As HADDOCK cannot support multiple conformations of the same residue, the ER1 LBD PDB (ID: 1ERE) file was cleaned using the "selaltloc" setting on the PDBTOOLS webserver (Jiménez-García *et al.,* 2021). This tool selects the label with the highest occupancy and removes any other alternate location/conformation of the same atom. The resultant PDB files were used as input structures on the HADDOCK webserver (https://wenmr.science.uu.nl/haddock2.4/submit/1) (Van Zundert *et al.,* 2016, Honorato *et al.,*

2021). As no initial information is known on the protein-protein structure, *ab initio* protocol was employed using the Guru level access option. All other parameters were set to default, and the resultant complex structure was obtained as a PDB file and subjected to further analysis.

4.10.4) Complex Structure Analysis

The top-scored structure was obtained from each of the docking programmes and analysed. To compare the predicted structures, they were each aligned to each other using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC Molecular Graphics System, Version 2.5.5, Schrödinger, LLC. Furthermore, the predicted structures were aligned individually to the crystal structure of FOXP2 FHD bound to DNA (PDB ID: 2A07) as well as that of liganded ER1 LBD bound to a peptide (PDB ID:3Q95) (Stroud *et al.*, 2006). This analysis was performed to identify if the regions involved in the interaction with DNA, ligand and/or peptides were distinct from the regions involved in mediating the specific protein-protein association identified in this study.

5) RESULTS

This study is based on identifying and characterising a possible interaction between FOXP2 and ER1 proteins. In order to do so, the interaction was investigated in both the cellular and external environments. For the latter, the two chosen domains of FOXP2 (FHD and NT) and ER1 LBD needed to be successfully expressed and purified. The structure and functions of the proteins were evaluated to confirm that the proteins were correctly folded and retained their known function. Finally, the interaction between the proteins was investigated and characterised in the presence of different salt concentrations, ligand, and DNA.

5.1) Protein-protein interaction in MCF-7 cells

To determine if the FOXP2 FHD – ER1 LBD interaction occurs in cells, co-IP was performed. MCF-7 cells were treated with E2 or with 0.001 % ethanol. The latter served as a vehicle control. Following a 24-hour treatment, the cells were lysed, and proteins were separated from the remainder of the cell content. As shown in Figure 5.1, the total protein content was quantified using a Bradford assay. Once sufficient protein quantity was obtained, endogenous expression of ER1 and FOXP2 was assessed. MCF-7 cells expresses both FOXP2 and ER1, with E2-treated cells showing greater expression of ER1 as shown in Figure 5.1C. As HEK 293 cells did not show any expression of ER1, this cell line was selected as a negative control for co-IP studies.

Co-IP involves the capture of a specific protein using a primary antibody and beads that bind to that antibody. First, FOXP2 was captured on the beads; this FOXP2 could be in its apo-form or bound to other proteins. Secondly, to determine if the captured FOXP2 was involved in an interaction with ER1, western blot was performed using the antibody against ER1. Hence, if ER1 appeared on the membrane, this would indicate an endogenous interaction between FOXP2 and ER1. However, the western blot did not show any ER1 bands of 66 kDa. The band that appears slightly below the expected ER1 band indicates the heavy chains of the antibody (55 kDa) used in the co-IP step. To ensure that the PPI was not affected by the lysis and washing steps, the cells were treated with formaldehyde to crosslink any PPIs occurring in the cells. The cross-linkage too did not result in any ER1 band on the western blot and suggests that the FOXP2 protein does not form a complex with ER1 under the specific conditions.

However, this is not the case as different results appear when ER1 is captured instead of FOXP2. When ER1-bound antibody is captured onto beads and evaluated using western blots with a primary antibody targeting FOXP2, the co-immunoprecipitated proteins from the untreated, E2-treated and cross-linked cells all show a band corresponding to the size of FOXP2 on the membrane; HEK 293 cells do not seem to have this band. This indicates that ER1 does form a complex with FOXP2 in MCF-7 cells. Notably, the western blot has a lot of background noise when the antibody specific to FOXP2 is used in western blot analysis. More specifically, a band corresponding to the molecular weight of approximately 115 kDa appears with a high intensity not only in the Co-IP gel but also the gel evaluating FOXP2 expression (Figure 5.1C) and is consistent in the HEK 293 as well as the treated, untreated and cross-linked MCF-7 cells. It is likely that the region to which the FOXP2 antibody binds within the acid-rich tail may not be entirely specific to FOXP2 and thus contributes to the excess bands visible on the gel.

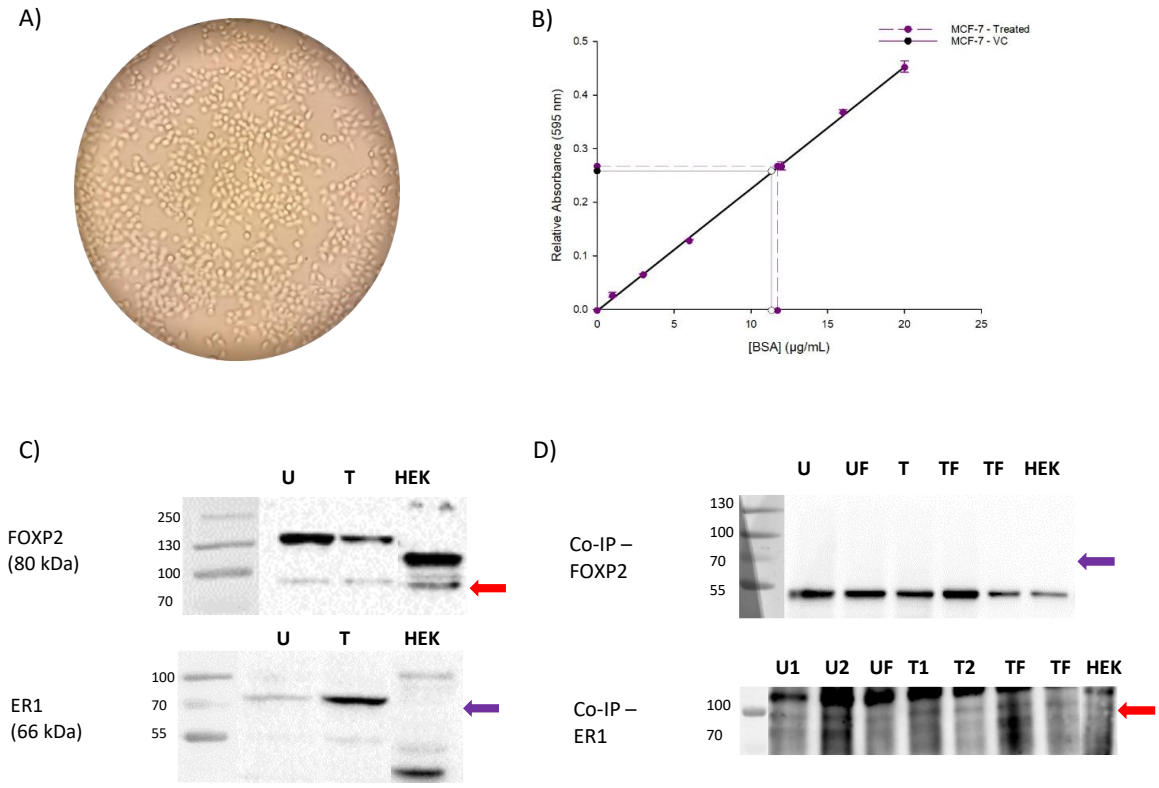


Figure 5.1: FOXP2-ER1 interaction in cells. A) Visualisation of MCF-7 cells. Cells with a confluency level of 80 to 90 % are used for downstream application. B) Quantification of total protein content. Bradford assay was performed using Bovine serum albumin (BSA) to construct a standard curve with the equation: y=0.0227x-0.0007 (r^2= 0.9992). The curve was used to approximate the total protein concentration in MCF-7 cells treated with 0.001 % ethanol (vehicle control (VC) AND 10 nM 17 β-oestradiol (E2). C) Protein expression in cells. To determine the endogenous expression of FOXP2 and ER1 in MCF-7 and HEK 293 cells, western blot was performed. Proteins were first separated based on molecular weight using a 10 % SDS-PAGE gel and transferred using wet-transfer onto a nitrocellulose membrane. The membrane was exposed to the respective primary antibody and thereafter an HRP-linked secondary antibody. Expression was visualised using the Molecular Imager® Gel Doc™ XR System (Bio-Rad). D) Co-immunoprecipitation of protein-protein complex and western blot analysis. Proteins were exposed to primary antibody against FOXP2 (top) and subsequently incubated with agarose beads. The antibody binds to apo- and complexed forms of FOXP2 whereas the beads capture the primary antibody. To assess if FOXP2 interacts specifically with ER1, western blot was performed of the captured FOXP2 and its associated complexes using a primary antibody against ER1. Additionally, the process was repeated using antibody against ER1 in the co-immunoprecipitation step (bottom). U: Untreated MCF-7 cells (VC); T: E2-treated MCF-7 cells; HEK: HEK 293 cells; UF: untreated and formaldehyde-crosslinked MCF-7 cells; TF: E2-Treated and formaldehyde-crosslinked MCF-7 cells.

5.2) Overexpression and purification of Proteins

The proteins used in this study were recombinantly expressed in bacteria and isolated from contaminants using chromatography - based purification techniques.

5.2.1) FOXP2 FHD

The FOXP2 FHD was overexpressed in T7 cells and thereafter purified using IMAC. All the lysate components that did not bind to the IMAC column eluted out first in the flow-through, as shown in Figure 5.2A. A salt wash was performed to ensure removal of any non-specifically bound contaminants. Following a change to a high imidazole buffer, the FOXP2 FHD that had initially bound to the resin via its N-terminal His-tag was released, eluted, and collected. The pellet and supernatant from the expression, the flow-through and the eluted peak from IMAC were all resolved on a tricine SDS-PAGE gel, as shown in Figure 5.2B. The FOXP2 FHD sample collected from IMAC was not entirely clean as can be seen by the contaminating bands in Figure 5.2B. Hence, the sample was subjected to a second step of purification, as shown in Figure 5.2C. The IMAC sample was loaded onto a preparative size exclusion column and the resultant peak was collected and resolved on a gel and is shown in Figure 5.2D. The presence of a lone band indicates sample purity. The molecular weight marker was used to construct a linear standard curve by plotting the relative migration of each of the bands against the log of their molecular weights.

FOXP2 FHD collected from SEC was approximated to be 13.53 kDa, which is close to the computed molecular weight of 12.97 kDa calculated using ProtParam.

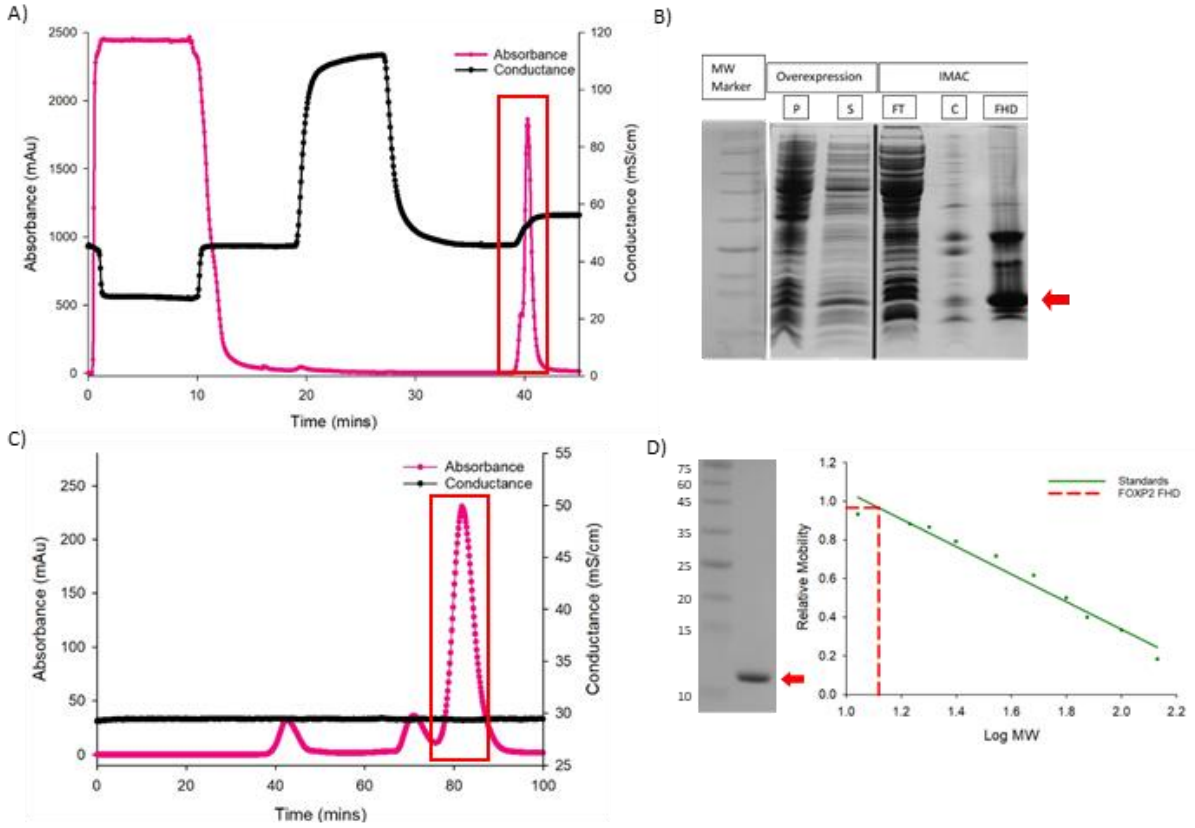


Figure 5.2: Purification profile of FOXP2 Forkhead Domain. A) Immobilised metal ion affinity chromatogram profile. Lysate from bacterial expression was centrifuged initially to pellet the insoluble fraction. The supernatant was loaded onto a nickel bound IMAC column. The flow-through and eluted sample (boxed in red) were collected. B) Gel analysis of IMAC samples. The differentiated pellet (P) and supernatant (S) samples from the bacterial expression were resolved on a 16 % tricine SDS-PAGE gel along with the flow-through (FT) and eluted fraction (FHD) from IMAC. The eluted fraction contained more than one band and was considered impure. C) Size exclusion profile of FOXP2 FHD. The elution fraction from IMAC was passed through a Superdex SEC column. The peak (boxed in red) was collected. D) Gel analysis of SEC sample. The sample collected from SEC was resolved on a 16 % tricine SDS-PAGE gel to confirm sample purity. The molecular weight marker was used to construct a standard curve (equation: $y = -0.71x+1.76$; $r^2=0.96$) to approximate the size of the collected FOXP2 FHD sample. The molecular weight is estimated to be 13.5 kDa.

5.2.2) FOXP2 NT

FOXP2 NT had not been previously expressed and therefore expression trials had to be conducted to determine the optimal conditions for its expression. The conditions examined were expression times and temperatures, IPTG concentrations, bacterial expression systems, and shaking speed during expression. Figure 5.3 shows the SDS-PAGE gels used to evaluate protein expression. Poor yields were obtained in the supernatant when T7 express cells were transformed with the FOXP2 NT plasmid in 2xYT media at 200 rpm. This was consistent over various IPTG concentrations, time periods and temperatures, although soluble protein yields were improved when expression was induced at higher ODs (Figure S1). Additionally, when BL21 (DE3) *E. coli* cells were transformed with FOXP2 NT plasmid, at least 50 % of the protein expressed was found in the soluble fraction under certain conditions, which included a 6-hour expression in 2xYT media at 25 °C and 200 rpm following induction with 0.5 and 1 mM IPTG, as shown in Figure 5.3B. Using this condition as a basis, the expression was evaluated in different media including LB, 2xYT and LB supplemented with sucrose. The latter gave the most soluble expression. Hence, the optimal condition for FOXP2 NT expression was deduced to be in BL21(DE3) cells, which are grown to OD_{600} of 0.6 and induced with 0.5 mM IPTG in LB media supplemented with 5 % sucrose. The expression time, temperature and shaking speed are 6 hours, 25 °C and 200 rpm respectively.

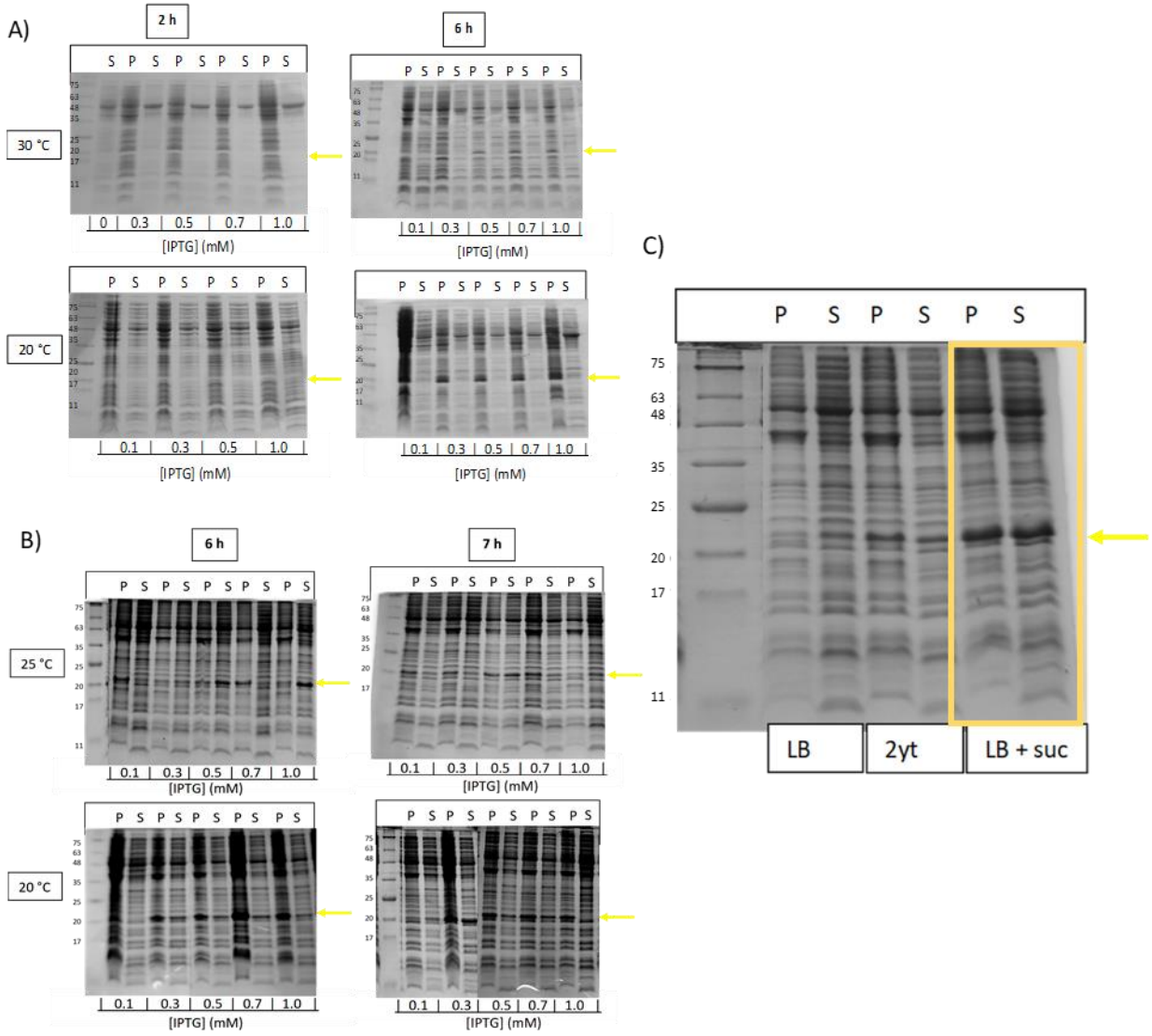


Figure 5.3: Evaluation of FOXP2 NT expression trials on SDS-PAGE gels. A) Expression in T7 Cells. Expression trials were performed at 30 °C and 20 °C in T7 competent cells by inducing at different IPTG concentrations. The gels with samples collected after 2 and 6 hours (h) show that the protein is predominantly found in the insoluble fraction in the pellet (P). B) Expression in BL21(DE3) cells. Expression trials were conducted at different temperatures and IPTG concentrations. The 6 and 7 h gels showed favourable soluble expression of FOXP2 NT at IPTG concentrations at and above 0.5 mM IPTG. C) Expression in different culture media. BL21(DE3) expression cells transformed with FOXP2 NT were expressed in LB, 2xYT and LB media supplemented with sucrose following induction with 0.5 mM IPTG at 25 °C. The latter condition enclosed in a yellow box provided the highest soluble yield of FOXP2 NT and was chosen as the optimal condition for its expression. The yellow arrow indicates the approximate location of the FOXP2 NT band at ~19 kDa. P: pellet, S: supernatant

Following the expression of FOXP2 NT, the protein was purified using IMAC as shown in Figure 5.4A. IMAC was performed by passing the lysate through a cobalt immobilised IMAC column. All the components of the lysate that did not have affinity to the resin, flowed out immediately. A salt wash was performed to remove any non-specifically bound proteins and imidazole was used to elute out FOXP2 NT. As FOXP2 NT does not have any tryptophan residues and only a lone tyrosine residue in its sequence, it only has a weak absorbance at 280 nm; however, a small but distinct peak is formed. This peak was collected, evaluated on a 16 % tricine SDS-PAGE gel and determined to be impure. Hence, a second step of purification was performed using SEC, as shown in Figure 5.4B. The protein's lack of aromatic residues presents a very noisy chromatogram, but the indicated peak was collected and evaluated on a gel. This sample was now free of contaminants. The gel was also used to approximate the molecular weight of the protein collected by constructing a standard curve of the relative migration of proteins with known molecular weight present in the molecular weight marker. Using the equation of the standard curve, the molecular weight was calculated to be 18.8 kDa which is as expected based on the ProtParam calculation (Gasteiger *et al.,* 2005).

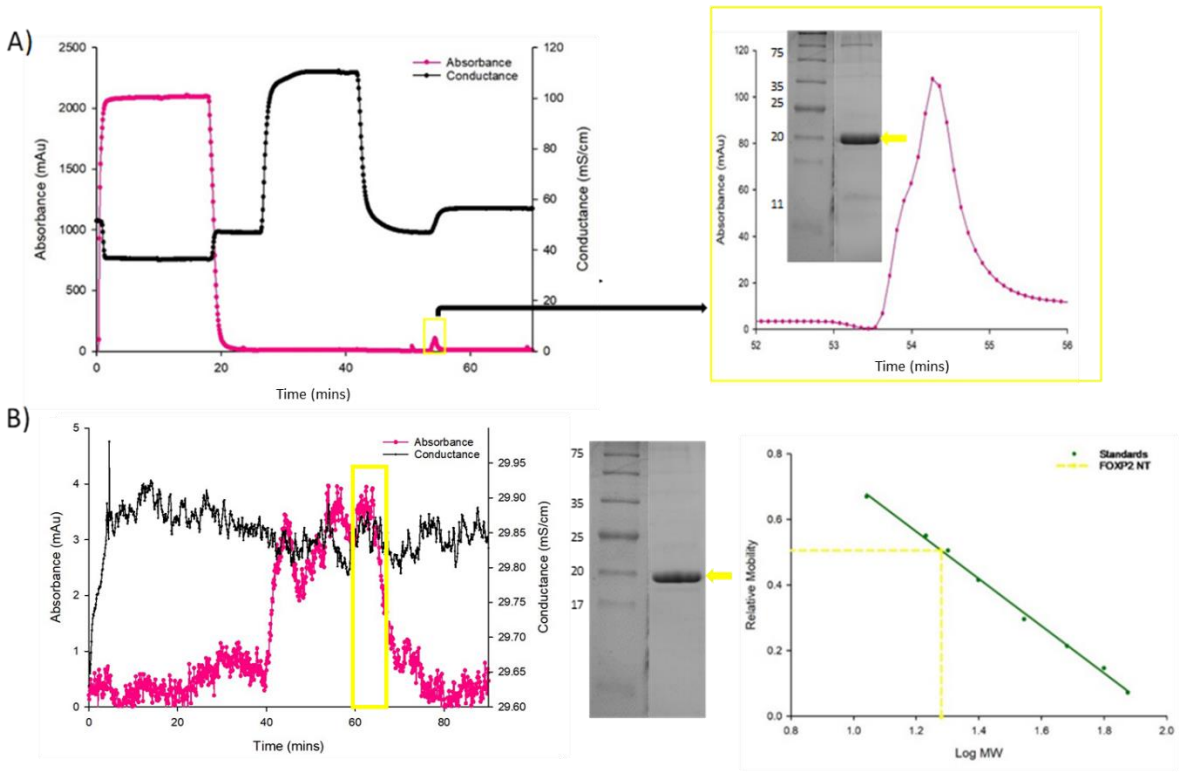


Figure 5.4: Purification of FOXP2 NT. A) IMAC profile and its evaluation on SDS-PAGE gels. FOXP2 NT was purified using cobalt-immobilised affinity chromatography. The initial increase in absorbance indicates the elution of all the components of the cell lysate that did not bind to the column. A salt wash visible by the increase in conductance was performed to remove any non-specifically bound contaminants. Following a salt wash, a high imidazole buffer was passed through the column to remove resin-bound proteins, including FOXP2 NT. This is observed as the small peak which is enlarged on the right. Alongside it is the tricine SDS-PAGE gel which used to evaluate the purity of the peak collected. The yellow arrow indicates the approximate location of the protein band. B) Size exclusion profile and gel of FOXP2 NT purification. FOXP2 NT was purified using size exclusion chromatography. The peak indicated with a yellow square was collected and ran on the gel alongside the chromatogram. The lone band indicates a pure sample with approximate molecular weight of the protein is 18.8 kDa. This was established by constructing a standard curve (equation: y= -0.72+1.43; r^2= 0.996) using the molecular weight marker, as shown on the right.

5.1.3) ER1 LBD

ER1 LBD was overexpressed for the first time in our lab and therefore induction trials were performed so as to establish optimal conditions for overexpression. (Figure 5.5A). Most of the expressed ER1 LBD is found in the pellet and this is consistent for the various conditions assessed in T7 cells in 2xYT media at 200 rpm. Based on this, various conditions were altered including

inducing at different ODs and using a slower shaking rate when expressing. However, none of these changes had a major influence, as most of the protein remained in the insoluble fraction. The ER1 LBD was then transformed into BL21(DE3) cells as depicted in figure 5.5B. While the majority of the protein remained insoluble, there was a greater quantity in the soluble fraction at both temperatures in comparison to the previous results.

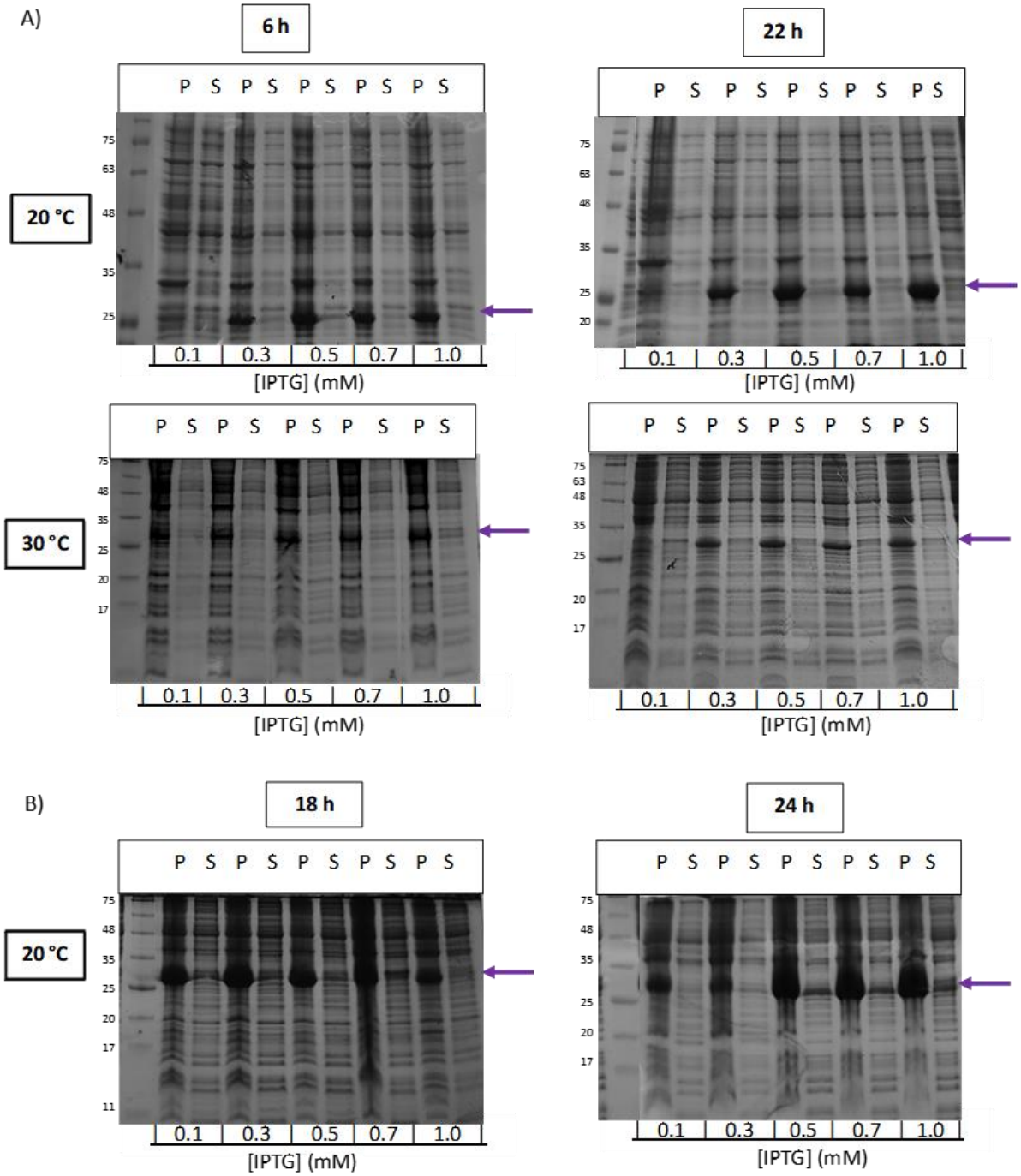


Figure 5.5: Expression trials of ER1 LBD. A) Expression in T7 express cells. Protein expression following induction at various IPTG concentrations and temperatures was evaluated after 6 and 22 hours. A high expression of ER1 LBD is visible; however, most of the protein is found in the pellet. B) Expression of ER1 LBD in BL21(DE3) cells. The protein is still predominantly in the

pellet but there is a small amount in the supernatant. P: pellet, S: supernatant. Purple arrow indicates the approximate position of ER1 LBD on the gel.

To maximise the soluble yield of the protein, the conditions of Figure 5.5B were altered in different ways. Figure 5.6 shows protein expression in TB media at 170 rpm with varied OD_{600} of induction, IPTG concentrations and inclusion of additives including ethanol and E2. Ethanol has previously been shown to enhance the solubility of some recombinantly expressed proteins when included in small amounts in the culture media as it induces a shock response in *E. coli* cells that results in the production of chaperones (Bhatwa *et al.*, 2021). On the other hand, ligands such as E2 are included in culture media as they improve the stability of the protein to which they bind (Eiler *et al.*, 2001). In this instance, ethanol did not make a significant difference to the expression of ER1 LBD in the pellet or supernatant. In contrast, inclusion of relatively high concentrations of E2 (20-30 nM) seemed to have an adverse effect as the intensity of the ER1 LBD band decreased in the soluble fraction. Taking these findings into consideration, the optimal protocol for expression of ER1 LBD was identified to be as follows (Condition 2 in Figure 5.6): BL21(DE3) cells transformed with ER 1 LBD should be cultured in TB media and grown until an OD_{600} of 0.7 is attained. Following a 30 mins cold shock treatment at 4 °C, 0.3 mM IPTG should be added to the media to induce protein expression, which should be carried out at 20 °C for 24 hours with the shaking speed set at 170 rpm.

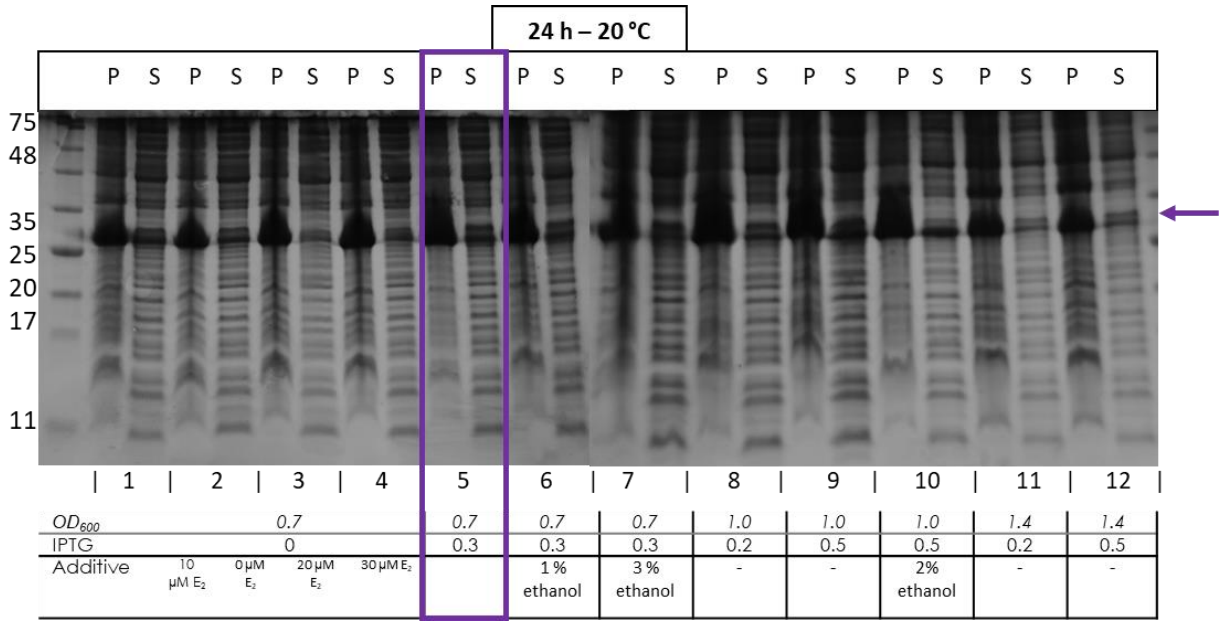


Figure 5.6: Expression trial of ER1 LBD in BL21(DE3) cells. The expression trial was conducted at 20 °C in TB media, 150 rpm and evaluated on a 16 % tricine SDS-PAGE gel. The bacterial culture was induced at different OD_{600} values and IPTG concentration and 17-β oestradiol (E2) and ethanol were also included in some instances. Conditions 1 and 4-9 showed relatively high protein expression in the supernatant; condition 5 was selected as the optimal as minimal additives were required while still acquiring a reasonable soluble yield. P: pellet, S: supernatant. Purple arrow indicates approximate location of ER1 LBD.

Following overexpression of soluble ER1 LBD, the protein was purified so as to isolate it from the rest of the contaminants. As with the other proteins used in the study, ER1 LBD plasmid was synthesised with a His-tag. Hence, the ER1 LBD was conveniently purified using a nickel bound IMAC resin. Figure 5.7 shows that the cell lysis buffer influences the amount of protein recovered from IMAC. Figure 5.7B indicates a greater yield with the inclusion of NDSB.Notably, it also appears more impure, but this may also be due to greater visibility of contaminants as the protein concentration is higher. The favourable effect of NDSB could be attributed to its ability to stabilise the ER1 LBD and prevent its aggregation even at the high concentrations released into the supernatant following cell lysis and present during the IMAC step (Eiler *et al.*, 2001). ER1 LBD requires further purification for use in downstream applications.

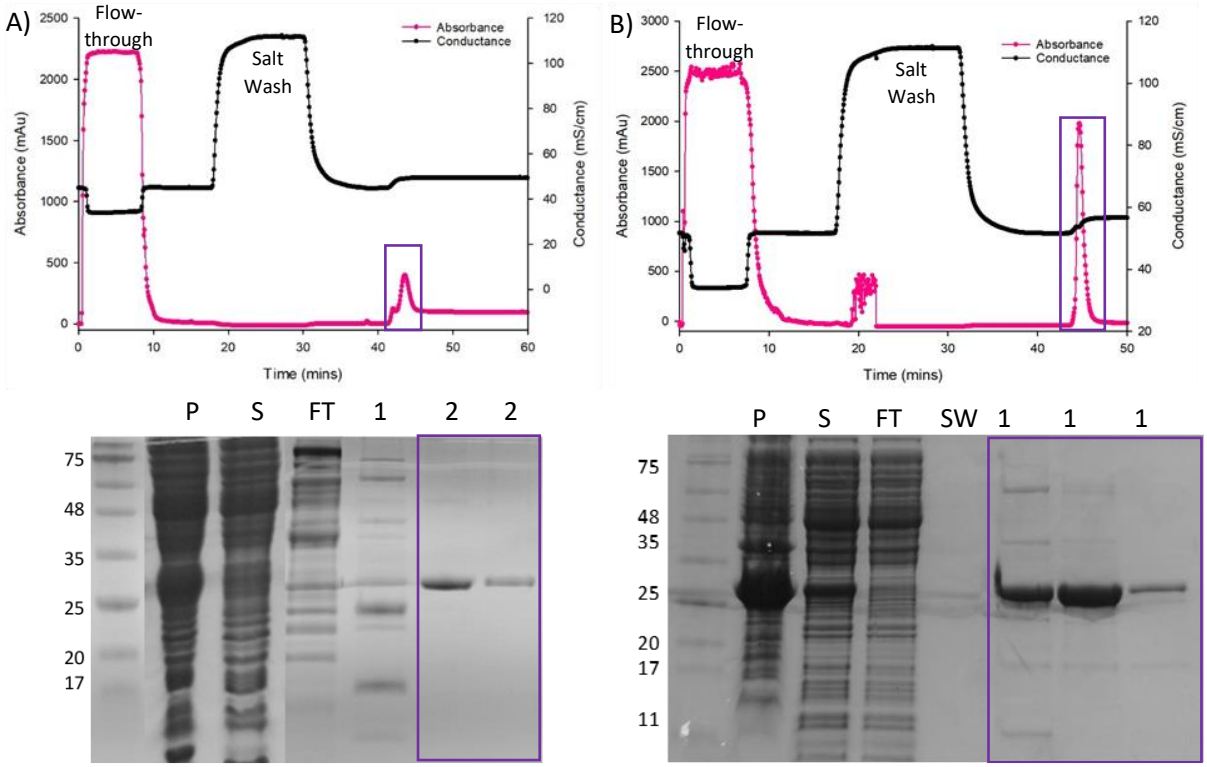


Figure 5.7: IMAC profiles of ER1 LBD lysed in disruption buffer A) without any additives and B) with NDSB. Protein expressed in BL21(DE3) cells was centrifuged to pellet the cells which were then resuspended in 20 mM Tris, 50 mM NaCl, 30 mM imidazole, pH 7.8 buffer with and without 1M NDSB. Following the separation of soluble and insoluble fractions, the protein was loaded onto an IMAC column pre-loaded with nickel ions. All the components that did not bind to the resin flowed out initially in the flow-through. Following a salt wash, re-equilibration and elution with 500 mM imidazole, the indicated peaks were collected and evaluated on the tricine SDS-PAGE gels shown below the respective chromatograms. In panel A, a single peak is observed with a shoulder preceding its elution. The SDS-PAGE gel shows that the shoulder (1) is predominantly comprised of contaminants whereas ER1 LBD is primarily eluted in the fractions corresponding to the peak (2). As the quantity of ER1 LBD obtained was poor, non-detergent sulfobetaine (NDSB) was included in the cell-disruption buffer prior to IMAC. Consequently, the chromatogram in panel B shows a markedly larger peak following elution when compared to that of panel A. This larger peak corresponds to a thicker protein band on the SDS-PAGE gel and suggest that inclusion of NDSB improved protein yield. P: pellet, S: supernatant, FT: flow-through, SW: Salt wash

Hence, the protein was loaded onto a preparative SEC column for further purification. As shown in Figure 5.8, SEC results in the elution of a single major peak that appears to be free of contaminants when evaluated on an SDS-PAGE gel. Subsequently, the migration of the bands of

the molecular weight marker were used to approximate the molecular weight of the collected protein. This was calculated to be 32.88 kDa and is as per expectation (Eiler *et al.*, 2001), providing confidence that the ER1 LBD had been successfully purified.

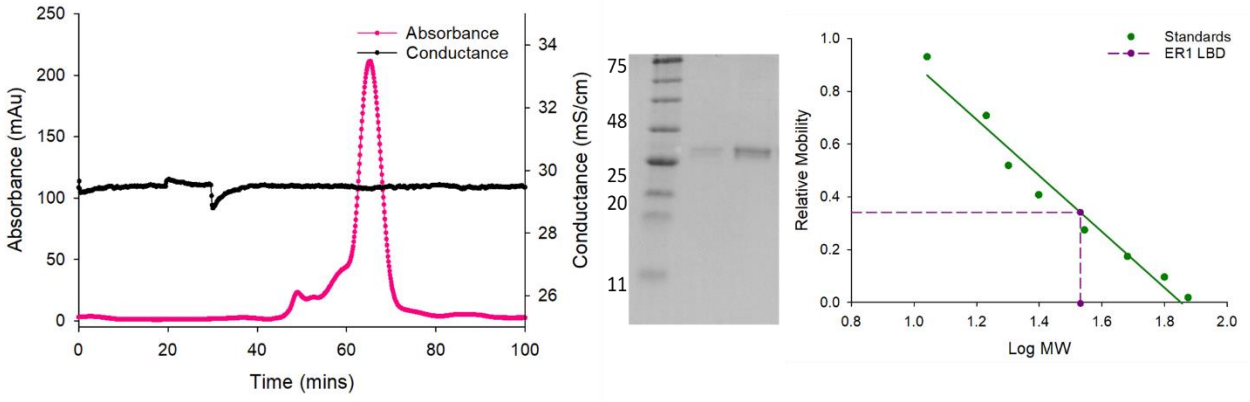


Figure 5.8: Size exclusion profile of ER1 LBD. Protein collected from IMAC was passed through a Sepharose SEC column along with 50 mM Tris, 300 mM NaCl, pH 7.8 buffer. One major peak was observed, collected, and evaluated on a tricine SDS-PAGE gel. Only one band is observed on the gel indicating sample purity. Additionally, the molecular weight marker was used to construct a standard curve (equation: y=-1.06+1.97, r^2=0.96) and the molecular weight of the protein was approximated to be 32.88 kDa based on the migration of the protein band indicating that ER1 LBD had been successfully purified.

5.3) Protein Quantification

Accurate protein quantification is required for the downstream applications. Hence, a combination of Bradford assay and UV Absorbance measurements were used to determine protein concentrations.

5.3.1) FOXP2 FHD and ER1 LBD

Following purification, the pure FOXP2 FHD and ER1 LBD proteins were pooled and dialysed against appropriate buffers and quantified using UV absorbance measurements. Figure 5.9 shows a typical absorbance spectrum of FOXP2 FHD and ER1 LBD. The peak observed at 280 nm with a shoulder at 295 nm indicates the absorbance of tyrosine and tryptophan residues, respectively. The lack of absorbance at 340 nm implies that there is no aggregation of protein within the sample. Additionally, the A280/A260 ratio was always calculated to be ~1.7 which indicates that both the proteins did not contain significant DNA contamination.

The absorbance at 280 nm was used to quantify protein concentration, which typically ranged close to the values tabulated in Figure 5.9 after the size exclusion purification step. Cumulatively, 10-15 mg of FOXP2 FHD was obtained per litre of expression. On the other hand, ER1 LBD expressed relatively well but only 5-8 mg protein was recovered per litre of culture following purification and dialysis. This was a result of aggregation which was usually minimalised by purifying and storing the protein at low concentrations (~2-8 µM).

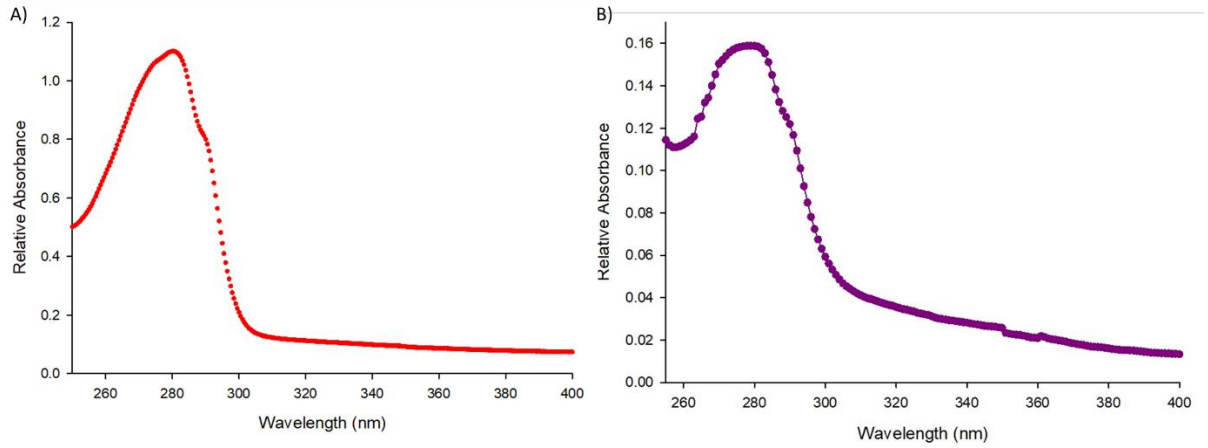
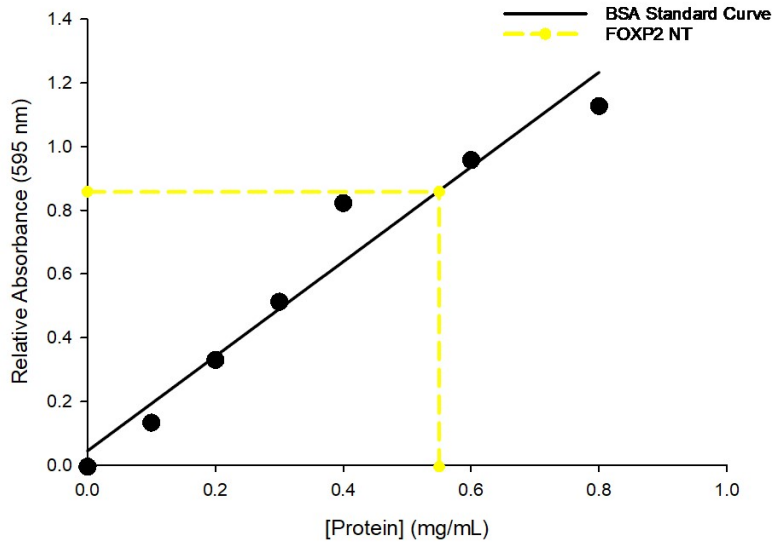


Figure 5.9: Absorbance spectrum of A) FOXP2 FHD and B) ER1 LBD. The absorbance at 280 nm less the absorbance at 340 nm is substituted into the Beer-Lambert equation and used to calculate the concentration of FOXP2 FHD and ER1 LBD which was 44 and 5 µM, respectively. These were the typical concentrations obtained for both proteins following size exclusion chromatography. The A280/A260 ratio, a measure of DNA contamination, was determined to be 1.7.

5.3.2) FOXP2 NT

Following purification, FOXP2 NT was dialysed against an appropriate buffer for downstream application. Any aggregation was removed by centrifugation before quantifying its concentration. As FOXP2 NT does not have any tryptophan residues and only a single tyrosine residue, any concentrations calculated using the absorbance reading at 280 nm would be inaccurate. Therefore, a Bradford assay was performed using BSA to construct a standard curve (Figure 5.10). The resultant fitted straight line equation was used in conjunction with the absorbance of the Bradford dye at 595 nm to calculate the concentration. As shown in Figure 5.10, the concentration of FOXP2 NT following purification was typically calculated to be 30 µM. This was compared with the absorbance reading at 205 nm obtained using the Nanodrop One spectrophotometer. The concentration of the same sample was estimated to be 29 µM. Hence, both the Bradford assay and

A205 readings produced a comparable concentration value and hence, in instances where the volume of the protein was limited, the NanoDrop One was used to quantify the FOXP2 NT concentration.



Average concentration based on Absorbance $_{205}$

mg/mL	A_{205}	A_{280}
0.578	17.85	0.03

Figure 5.10: Quantification of FOXP2 NT concentration. FOXP2 NT concentration was determined using a Bradford assay. A standard curve was constructed by measuring the absorbance of the Bradford dye bound to various BSA samples and plotting it against their known concentrations (equation: y=1.48x+0.05, r^2=0.94). The concentration of FOXP2 NT was calculated to be 0.588 mg/mL or 40 µM. This was equivalent to the concentration calculated using the absorbance at 205 nm (39 µM).

5.4) Quaternary structure of the proteins

To determine the quaternary structure of FOPX2 FHD, FOXP2 NT and ER1 LBD, the protein samples were resolved on a size exclusion column along with proteins of known molecular weights. The standards used for the construction of the plot were blue dextran (250 000 kDa), conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (15.7 kDa) and aprotinin (6.5 kDa). Blue dextran is larger than the fractionation range of the column and its elution is considered to represent the void volume.

When FOXP2 FHD was loaded onto the SEC column using an FPLC system (Figure 5.11A), three peaks were observed. The first peak eluted at 45 mL which corresponds to the elution volume of blue dextran and indicates the elution of large contaminants and higher order oligomers. The other two peaks appear at elution volumes of 74 mL and 86 mL and correspond to molecular weights of

25.7 kDa and 12.8 kDa, respectively. This is similar to the predicted size of the dimeric and monomeric forms of FOXP2 FHD as calculated using ExPASy ProtParam (Gasteiger *et al.*, 2005). As both the peaks appeared as a single band on an SDS-PAGE gel corresponding to the size of FOXP2 FHD, the size exclusion result suggests that FOXP2 FHD forms both a monomer and a homodimer. Taking the intensity of the peaks into consideration, FOXP2 FHD exists predominantly in its monomeric form in solution.

SEC was also used to investigate the oligomeric states of FOXP2 NT. As shown in Figure 5.11B, two peaks were identified which correspond to molecular weights of 60 kDa and 40 kDa, respectively. When evaluated on a gel, the samples from each peak appeared as a single band approximated to be ~20 kDa which is close to the molecular weight that was theoretically calculated for FOXP2 NT (18.9 kDa) using ProtParam. Hence, FOXP2 NT exists as a homodimer or homo-trimer. However, it is also important to consider the noisiness of the chromatogram owing to the lack of aromatic residues present in the protein. Therefore, it is difficult to confidently remark on all the oligomeric states the protein can form as the low absorbance values could obscure assemblies other than the two identified in the above figure; other assemblies may just be lower in proportion but still significant.

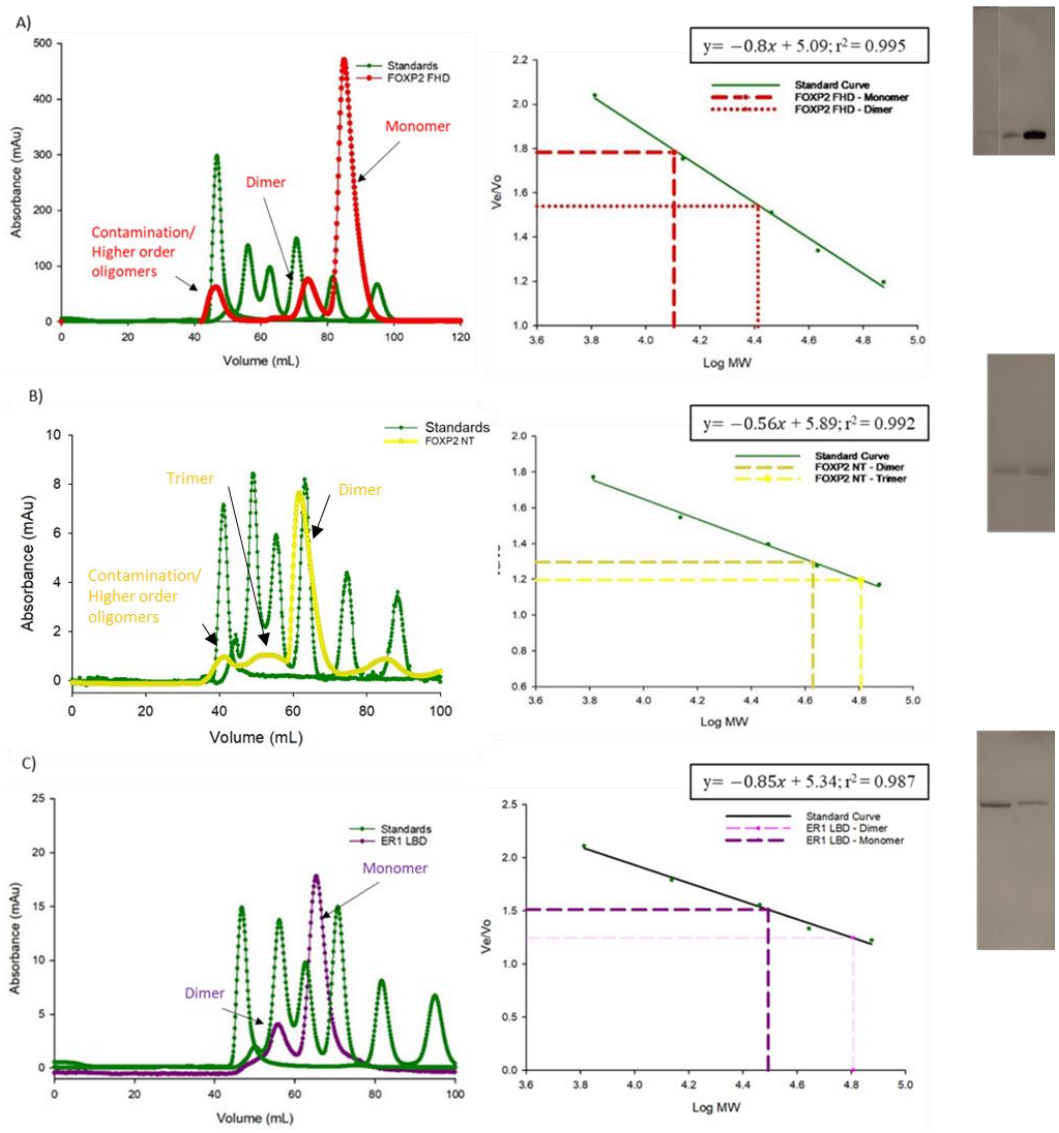


Figure 5.11: Quaternary structure of A) FOXP2 FHD, B) FOXP2 NT and C) ER1 LBD. Cytiva standards (green) were resolved on the HiLoad^{TM} 16/60 SuperdexTM 75 prep grade column comprised of blue dextran (250000 kDa), conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), ribonuclease A (15.7 kDa) and aprotinin (6.5 kDa). Blue dextran, which elutes at approximately 42 mL, is completely excluded from the column due to its high molecular weight and this volume is recognised as the void volume (Vo). The molecular weights of the eluted protein corresponding to each peak was determined by plotting the Ve/Vo of the standards against the log of their respective molecular weights and subsequently fitting the data to linear regression models using SigmaPlot 15.0. Sample from each distinct peak was also collected and evaluated using tricine SDS-PAGE gels.

Similarly, ER1 LBD seems to exist in two forms as well with a distinct peak observed at 52 mL and 60 mL in Figure 5.11C. From the standard curve, the molecular weight corresponding to each

peak was determined to be 65.1 and 30.9 kDa, which in turn corresponds to the size of the dimeric and monomeric forms of ER1 LBD, respectively. However, contrary to the FOXP2 proteins that either do not contain any cysteine residues (FOXP2 NT) or have a singly buried one (FOXP2 FHD), ER1 LBD contains three cysteine residues that are relatively exposed and could participate in formation of disulfide linkages. Therefore, size exclusion chromatography was performed in the absence and presence of the reducing agent DTT as shown in Figure 5.12A. The resultant chromatogram exhibits only a single peak in the presence of the reducing agent suggesting that the dimer was indeed disulfide-linked. This was confirmed on a gel where a non-reduced sample exhibited 2 bands – one at approximately 30 kDa and the other at 62 kDa, whereas the reduced sample comprised of only a single band at approximately 30 kDa (Figure 5.12B).

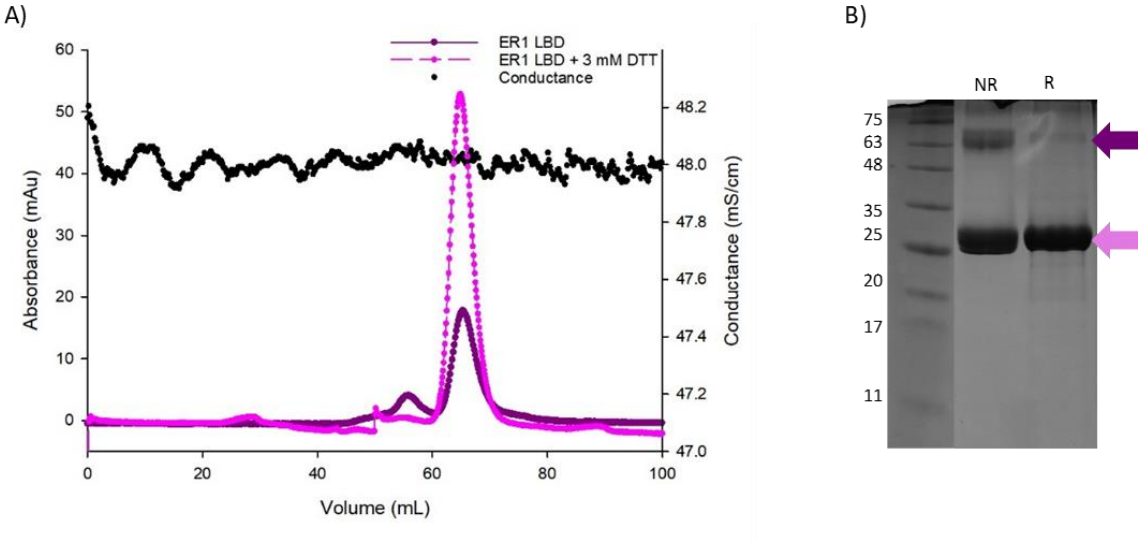


Figure 5.12: Role of disulfide bonds in determining ER1 LBD quaternary structure. A) Size exclusion profile in the presence of reducing agent. ER1 LBD does not form dimers when SEC is performed in the presence of 3 mM DTT. This is confirmed on a tricine SDS-PAGE gel, shown in panel B). The non-reduced sample (NR) does not have β-mercaptoethanol and thus has two major bands corresponding to the dimeric and monomeric forms of the protein as indicated by the purple and magenta arrows, respectively. In contrast, only the lower band appears intense in the reduced sample containing β-mercaptoethanol.

5.5) Functional characterisation of the proteins

FOXP2 FHD is the DNA - binding domain of FOXP2. Therefore, confirmation of FOXP2 FHD DNA-binding activity indicates that the protein is folded and functional. While ER1 is also a transcription factor, the isolated LBD that we work with in this study does not have DNA-binding

activity. The LBD binds to its ligand, E2, and it is this ligand binding activity that will be used to confirm functionality of ER1 LBD.

5.5.1) FOXP2 FHD

Since FOXP2 FHD is a DNA-binding domain, the functional activity of FOXP2 FHD was evaluated using EMSA. Nelson DNA was titrated against increasing concentrations of FOXP2 ranging from 0 to 12.5 µM. From Figure 5.13A, it is clear that FOXP2 FHD forms a complex with Nelson DNA when its concentration is at least 3 times greater than that of its interacting partner. This shows that FOXP2 FHD is functional and can interact specifically with its cognate DNA.

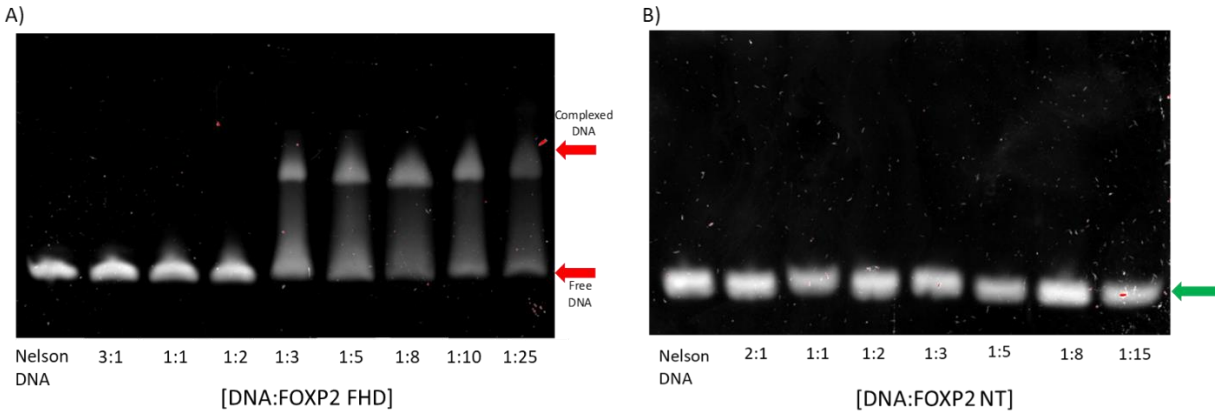


Figure 5.13: EMSA evaluating the DNA binding activity of A) FOXP2 FHD and B) FOXP2 NT. FOXP2 FHD and NT in 50 mM Tris, 100 mM NaCl, pH 7.8 are titrated at increasing concentrations (0 – 30 µM) against 0.5 µM Nelson DNA and resolved on a 10 % and 8 % PAGE gel, respecitvely. The FOXP2 FHD forms a complex with Nelson DNA as indicated by the shifted band on the gel when FOXP2 FHD concentration is at least 3x greater than that of DNA. Nelson DNA does not seem to form a complex with FOXP2 NT as only free DNA bands (green arrow) are visible irrespective of the concentration of NT in the sample.

5.5.2) FOXP2 NT

FOXP2 NT does not have any reported function as no prior studies have been conducted on this region alone. Therefore, to elucidate if FOXP2 NT has DNA-binding activity reminiscent of FOXP2 FHD, an EMSA was performed. At the concentration of FOXP2 NT titrated, the protein did not seem to bind to Nelson DNA as Figure 5.13B does not show any protein-DNA complex formation.

5.5.3) ER1 LBD

ER1 LBD is a ligand-binding domain that can interact with oestrogen among other ligands. To functionally characterise the ER1 LBD, isothermal titration calorimetry was performed by continuously injecting 100 µM E2 ligand into 10 µM ER1 LBD. Figure 5.14 shows the data of one of the replicates fitted to an independent binding curve on NanoAnalyze software. The K_d of the reaction is 0.3 µM and the interaction is spontaneous as indicated by the negative change in free Gibbs energy (ΔG). The spontaneity of the interaction arises due to both the enthalpic, and entropic changes associated with ligand-binding being favourable.

The reaction is exothermic and can be rationalised by considering the formation of hydrogen and van der Waals interactions between the E2 and ER1 LBD upon binding (Tanenbaum *et al.,* 1998). On the other hand, the interaction between ER1 LBD and E2 should result in a decrease in the conformational freedom of the protein. However, as the ΔS is positive, the solvent entropy associated with the release of water molecules from the hydrophobic ligand-binding pocket of ER1 LBD is large enough to overcompensate for the unfavourable decrease in conformational degree of freedom of the protein once bound to a ligand (Sasson and Notides, 1983, Wallerstein *et al.*, 2021). These thermodynamic parameters are in agreement with previous findings (Sasson and Notides, 1983). However, the disassociation constant obtained using this method varies from that obtained previously using radioactive ligand-binding assays (Eiler *et al.,* 2001). Hence, it must be noted that while the ITC qualitatively confirms the functionality of the protein, the lower dissociation constant obtained here may suggest that the ER1 LBD used in this study may not have bound oestrogen as strongly and effectively as expected. This may be due to proteins in the sample adopting alternative folds or aggregation of the protein during the ITC run.

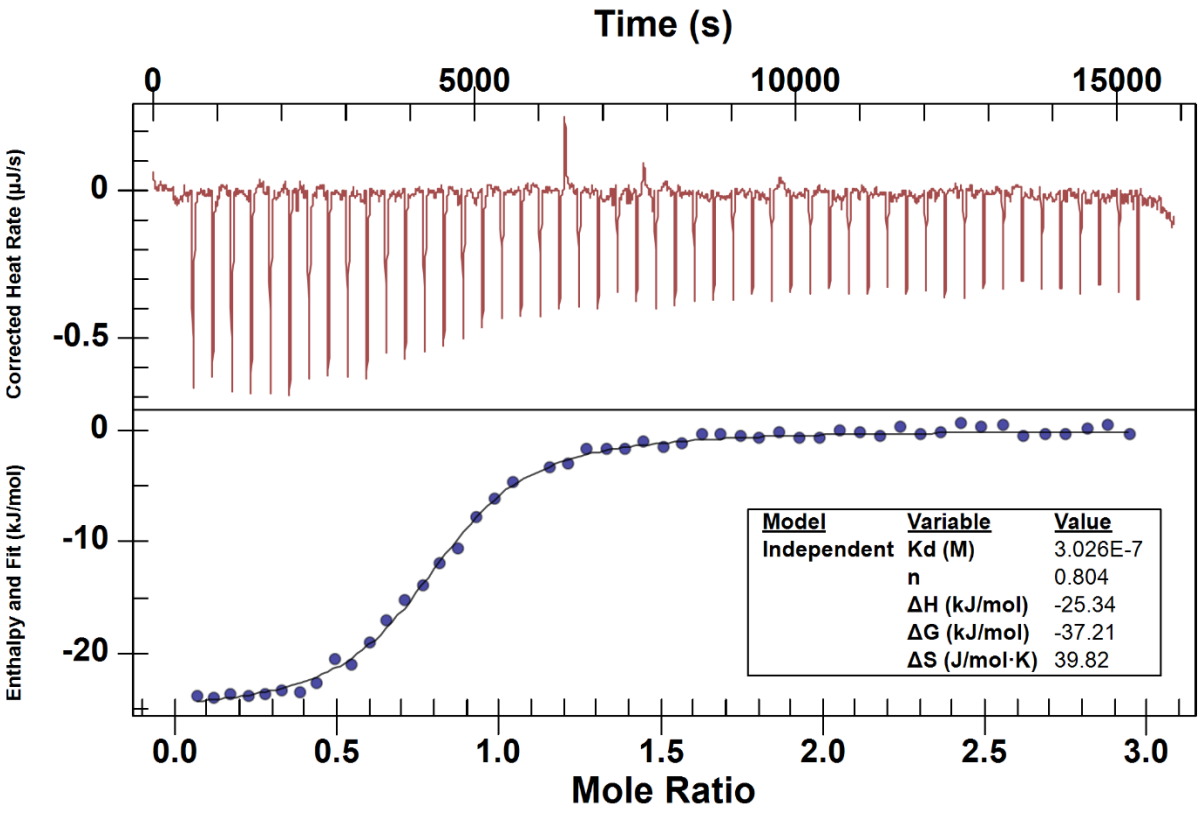


Figure 5.14: Interaction between ER1 LBD and E2. Isothermal titration calorimetry was performed by continuous 5 uL injection of 100 µM E2 into 10 µM ER1 LBD. The protein and ligand were kept in a 50 µM Tris, 100 mM NaCl buffer containing 1 mM TCEP and 1 % ethanol at pH 7.8. The Gibbs free energy change (ΔG) indicates that the reaction is spontaneous due to a favourable negative change in enthalpy (ΔH) and a favourable positive change in entropy (ΔS).

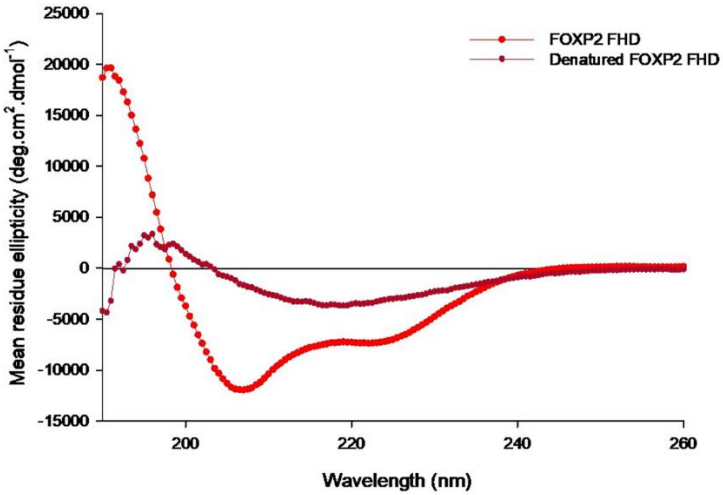
5.6) Structural Characterisation

The proteins were characterised according to their secondary and tertiary structures to deduce if they were indeed folded following recombinant expression in bacteria and purification. This was necessary prior to any further biophysical and biochemical studies on the proteins and their functions and interactions.

5.6.1) FOXP2 FHD

The secondary structure of apo FOXP2 FHD was characterised using far-UV CD. As shown Figure 5.15, FOXP2 FHD shows a CD spectrum with a trough at both 206 nm and 222 nm. Analysis of the spectra shows that the secondary structure of FOXP2 FHD is comprised of loops, alpha-helices as well as a β- sheeted component which is in agreement with previously obtained CD spectra and the crystal structure of the FOXP2 FHD (PDB ID: 2A07) (Whitmore and Wallace, 2004; Stroud

et al., 2006; Thulo *et al.,* 2021). FOXP2 FHD heated to 95 °C completely loses the alpha-helical structure. This confirms that the protein was folded originally and loses its structure when exposed to extreme heat.



	Alpha Helix	Beta Sheet	Random Coil
FOXP2 FHD	0.30	0.23	0.47

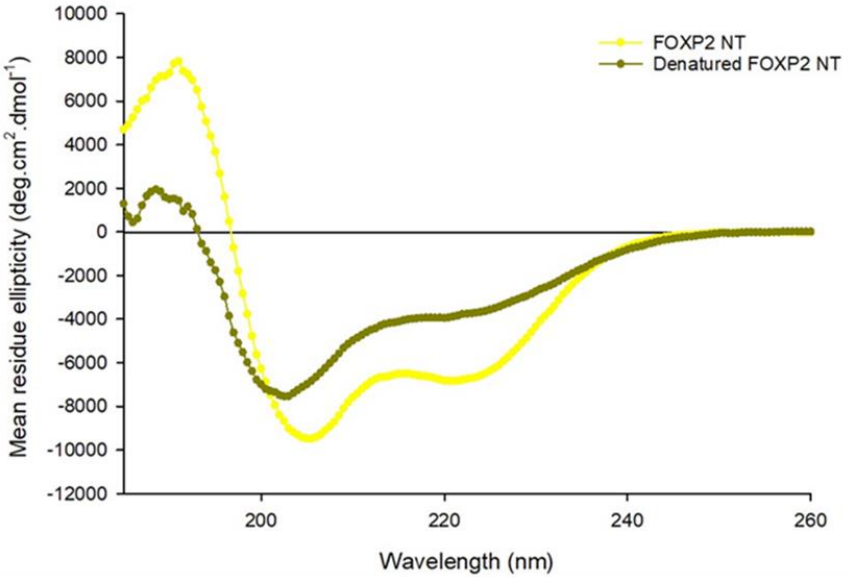
Figure 5.15: Secondary structure characterisation of FOXP2 FHD. A) Far-UV circular dichroism of FOXP2 FHD. CD was performed on the Jasco 1500 spectrophotometer using 10 µM FOXP2 FHD dialysed against 5 mM sodium phosphate buffer, pH 7.8 (red). CD was also performed with heat denatured FOXP2 FHD (maroon). The folded FOXP2 FHD exhibits two minima, typical of an alpha helical protein, whereas the denatured spectrum does not show this helicity. The CD data was analysed using DichroWeb and indicates that the native protein is comprised of both α-helices and β-sheets (Whitmore and Wallace, 2004).

5.6.2) FOXP2 NT

As with FOXP2 FHD, the secondary structure of FOXP2 NT was evaluated using far-UV circular dichroism. Tryptophan fluorescence could not be used to study the tertiary structure of FOXP2 NT because this protein has no tryptophan residues and only a single tyrosine residue which makes the signal too weak for any meaningful data to be collected. The resultant CD spectra depicted in Figure 5.16 show that the FOXP2 NT has two troughs, which are characteristic of an alpha helical structure. However, an all-alpha helical structure has two local minima of similar magnitude at 208 and 222 nm whereas FOXP2 NT deviates slightly with minima at 205.5 and 220.5 nm. Hence, the structure is likely to contain alpha helices among other structures. This was confirmed using

DichroWeb which shows that the protein is comprised of random coils, β-sheets and a small proportion of α-helices (Whitmore and Wallace, 2004).

Surprisingly, the FOXP2 NT retains some of its secondary structure even after being heated to 95 °C as the heated sample produced minima at 202.5 and 220 nm. There is, however, a marked difference between the folded and denatured states of the protein as depicted in Figure 5.16. Furthermore, the DichroWeb analysis indicates a reduction in both the alpha helical and beta sheeted content of the protein once heated (Whitmore and Wallace, 2004). Collectively, the secondary and quaternary structure analyses indicate that the FOXP2 NT is indeed folded. However, given the lack of previous information on this specific region and a known function, it is difficult to decipher whether the fold is indeed correct with complete certainty. For the purpose of this study, the structural evaluation was consistent across multiple overexpression and purification cycles and therefore, the FOXP2 NT was used downstream in fluorescence anisotropy experiments.



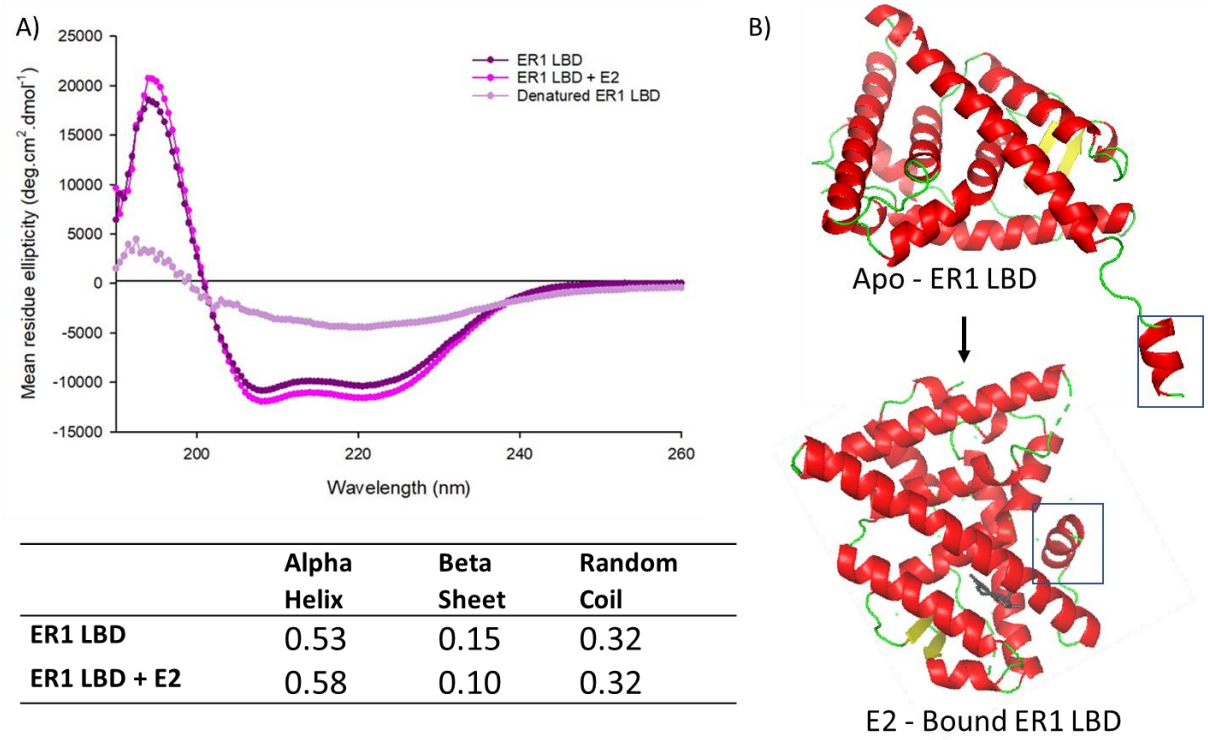
	Alpha Helix	Beta Sheet	Random Coil
FOXP2 NT	0.13	0.31	0.56
Denatured FOXP2 NT	0.08	0.29	0.63

Figure 5.16: Far-UV circular dichroism spectra of FOXP2 NT. FOXP2 NT was dialysed against 5 mM sodium phosphate, pH 7.8 and concentrated to 7.5 µM. The CD spectra were recorded between 180 and 260 nm in triplicates on a Jasco-J1500 spectrometer. The native FOXP2

NT (yellow) exhibits a trough with two minima at 205.5 and 220.5 nm with disproportionate ellipticity, whereas the FOXP2 NT heated to 95 °C (dark yellow) displays a minimum at 202.5 and 220 nm. The CD spectra were deconvoluted using DichroWeb, which shows that the FOXP2 NT has a high proportion of β-sheets in its folded and heat-denatured state (Whitmore and Wallace, 2004).

5.6.3) ER1 LBD

The secondary structure of ER1 LBD with and without its natural ligand (E2) was evaluated using far-UV CD. The spectra in Figure 5.17A exhibit a minimum at 208 nm and 222 nm which denotes a predominantly alpha-helical structure; this is in agreement with the CD spectra obtained by Ferrero *et al.* (2014). Furthermore, the protein's secondary structure remains unchanged despite the addition of E2 (Brandtand Vickery, 1997). This is confirmed by analysing the proportion of secondary structures in the apo- and E2-bound forms as DichroWeb analysis shows that the protein maintains its alpha-helical fold when bound to E2 (Whitmore and Wallace, 2004). As ligand-binding only alters the conformation of helix 12 and not much else as shown in Figure 5.17B, this result corresponds to the crystal structures of ER1 LBD and serves to validate the likelihood of the protein being folded in its native structure (Brzozowski *et al.*, 1997; Tanenbaum *et al.*, 1998). Conversely, denatured ER1 LBD no longer has a CD spectrum that shows two minima, reminiscent of an alpha-helical structure, indicating a significant difference between the native and denatured structures which is to be expected.



	Alpha Helix	Beta Sheet	Random Coil
ER1 LBD	0.53	0.15	0.32
ER1 LBD + E2	0.58	0.10	0.32

Figure 5.17: Far-UV CD structure of ER1 LBD. A) The secondary structural content of 7.5 µM ER1 LBD in its apo form (purple) and in the presence of excess E2 (pink). CD spectra were measured using the J1500 spectrometer and appear to be the same with two minima of similar ellipticity observed at 208 and 222 nm. DichroWeb analysis shows that the protein is primarily alpha-helical in its apo- and E2-bound forms (Whitmore and Wallace, 2004). The denatured ER1 LBD sample heated to 95 °C (lilac) does not resemble the other spectra and confirms a difference in native and denatured structures. The buffer (5 mM sodium phosphate, pH 7.8 and 1 mM TCEP), heated to 95 °C when required and otherwise, was used as a blank. B) The ER1 LBD crystal structure. The crystal structure shows that ER1 LBD is comprised of helices with one beta sheet and this secondary structural content remains unchanged in the apo- (PDB ID: 1A52) and E2-bound (PDB ID: 1ERE) forms as only the enclosed helix 12 changes conformation. Image rendered using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC (Brzozowski *et al.*, 1997; Tanenbaum *et al.*, 1998).

5.7) Identification of Protein – Protein Interaction

The potential interaction between ER1 LBD and either FOXP2 FHD or FOXP2 NT was evaluated using fluorescence anisotropy. ER1 LBD was labelled with NTA Atto 550 with a labelling efficiency of one dye molecule per protein. The labelled ER1 LBD was used to calculate the G-factor and was titrated with either FOXP2 FHD or FOXP2 NT. As shown in Figure 5.18A, the titration of up to approximately 20 µM of FOXP2 NT did not result in an increase in fluorescence

anisotropy and consequently, the data points could not be fitted to the non-linear regression binding model (Equation 6). This means that the molecular weight of the fluorophore-conjugated protein complex remained unchanged until the last couple titrations – tat is when the fluorescence anisotropy begins to increase. Given that the anisotropy did not increase despite the cumulative concentration of FOXP2 NT titrated in being almost a 100-fold in excess to that of the labelled ER1 LBD, the findings suggest that FOXP2 NT interacts only very weakly with ER1 LBD under the specific conditions used in the study.

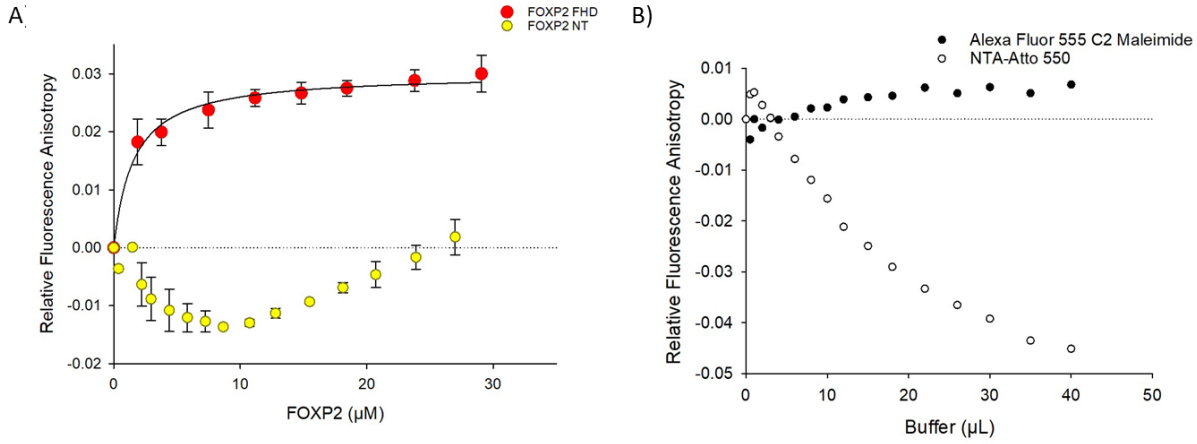
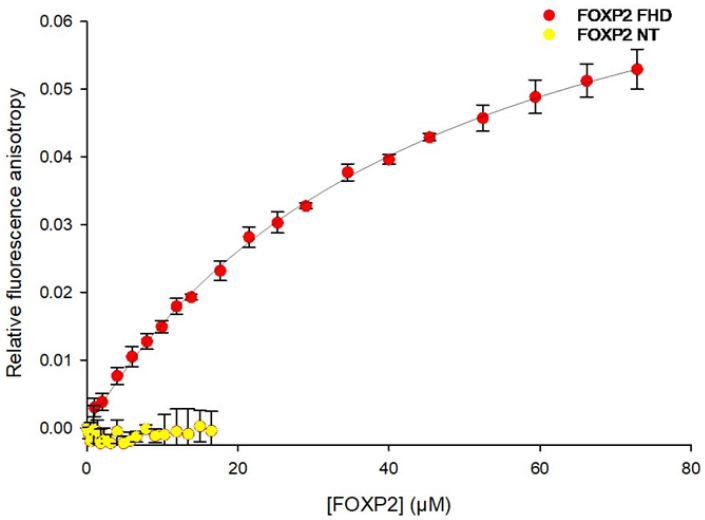


Figure 5.18: Interaction between ER1 LBD and FOXP2. ER1 LBD (0.3 μM) was labelled with NTA Atto 550 and titrated with either FOXP2 FHD (red) or FOXP2 NT (yellow). A) Fluorescence anisotropy of ER1 LBD and FOXP2 proteins. FOXP2 FHD clearly shows a hyperbolic curve when titrated against ER1 LBD, indicative of binding. The anisotropy curve was fitted to a one site saturation model which fitted well to the FOXP2 FHD data (r^2=0.98) and provided a K_d of 1.55 ± 0.25 μM and a Bmax of 0.03 ± 0.00 μM for the ER1 LBD-FOXP2 FHD interaction. On the other hand, the anisotropy following the increased titrations of FOXP2 NT only displays an increase following the titration of relatively high concentration of FOXP2 NT, illustrating that across the concentrations evaluated, FOXP2 NT shows only a weak ability to form a complex with ER1 LBD. B) Fluorescence anisotropy following titration of buffer. ER1 LBD labelled with either NTA-Atto 550 or Alexa Fluor 555 C2 maleimide was titrated with buffer (50 mM Tris, 100 mM NaCl, 1 mM TCEP, pH 7.8). The latter displays a decrease in anisotropy following each titration suggesting dissociation of the dye from the protein.

Conversely, titration of as little as 1 μM of FOXP2 FHD against the ER1 LBD resulted in a significant increase in fluorescence anisotropy with further titrations resulting in a hyperbolic curve. This indicates that there is a strong association between FOXP2 FHD and ER1 LBD. The

fit of this anisotropy data to a ligand saturation curve indicates that the ER1 LBD-FOXP2 FHD interaction has a K_d of 1.55 µM. However, as shown in Figure 5.18B, titrating buffer into NTA Atto-labelled protein has the unwanted effect of decreasing the relative fluorescence anisotropy. This is likely due to the dissociation of the dye from the protein following each titration and accounts for the negative values in anisotropy in Figure 5.18A. As a result, alternative labelling strategies were explored to mitigate these disadvantages.

One such strategy was to label the ER1 LBD with AlexaFluor 555 C2 maleimide with an efficiency of two dye molecules per protein. As shown in Figure 5.19A, the titration of up to approximately 18 µM of FOXP2 NT did not result in an increase in fluorescence anisotropy and is consistent with previous finding suggesting a weak interaction between FOXP2 NT and ER1 LBD. Similarly, the titration of FOXP2 FHD confirms that it does indeed interact with ER1 LBD. However, the K_d determined in this instance is 30-fold weaker at 46.3 µM. This suggests that labelling of the cysteines interferes with ER1 LBD's ability to bind to FOXP2. In spite of this flaw, later experiments were performed by labelling the cysteine residues as it continues to display trends similar to that of Atto-labelled protein without resulting in negative anisotropies.



	K_d (μM)	Bmax (μM)	R^2
ER1 LBD + FOXP2 FHD	46.28 ± 1.51	0.09 ± 0.00	0.99
ER1 LBD + FOXP2 NT	N/A	N/A	N/A

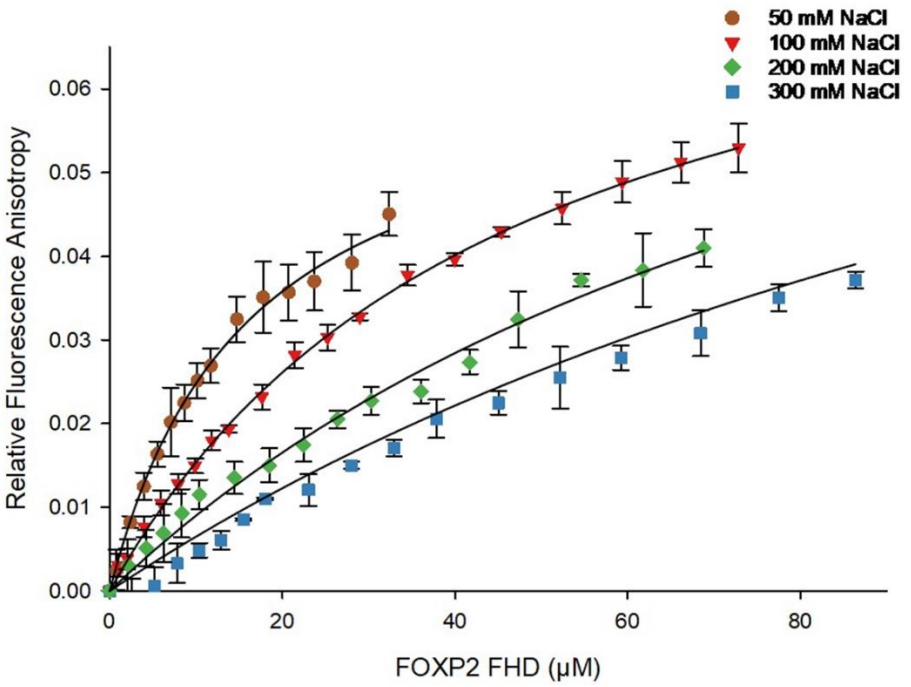
Figure 5.19: Interaction between ER1 LBD and FOXP2. ER1 LBD (0.3 μM) was labelled with AlexaFluor555 C2 maleimide and titrated with either FOXP2 FHD (red) or FOXP2 NT (yellow). Fluorescence anisotropy of apo-ER1 and FOXP2 proteins. FOXP2 FHD clearly shows a hyperbolic curve when titrated against ER1 LBD, indicative of binding. The anisotropy curves were measured in triplicate and globally fitted to a one site saturation model which fitted well to the FOXP2 FHD data and provided a K_d of 46.3 μM for the ER1 LBD-FOXP2 FHD interaction. On the other hand, the anisotropy following increased titrations of FOXP2 NT follows a mostly horizontally linear pattern, which is not well represented by the hyperbolic ligand binding model, illustrating that across the concentrations evaluated, FOXP2 NT does not show an ability to form a complex with ER1 LBD.

5.8) Effect of Salt on the Protein-Protein Interaction

Now that a PPI between ER1 LBD and FOXP2 FHD has been established, the interaction can be characterised further. Intermolecular interactions are governed by various types of bonds including van der Waals, hydrogen, and electrostatic interactions. Altering the salt concentrations of systems can be used to screen for electrostatic interactions to determine the forces dictating the interaction between proteins and to infer how the formation of such an association may be regulated physiologically (Glaser *et al.*, 2009; Zhang, 2012).

5.8.1) Fluorescence Anisotropy

To determine the effect of salt on the FOXP2 FHD-ER1 LBD interaction, fluorescence anisotropy was performed at different salt concentrating ranging from 50 to 300 mM NaCl. This range was selected to keep close to the physiologically relevant concentration of NaCl (150 mM) while also maintaining protein solubility (Leslie *et al.*, 2019). As shown in Figure 5.20, there is a difference in the strength of the interaction between FOXP2 FHD and ER1 LBD depending on the salt concentration. At 50 mM NaCl concentration, the fluorescence anisotropy curve reaches a plateau even though only ~32 µM of FOXP2 FHD has been titrated into the fluorophore labelled ER1 LBD. Thus, the K_d obtained (15.9 µM) is accurate. In contrast, the anisotropy curves at 200 and 300 mM NaCl concentrations follow a more linear progression and do not reach saturation despite the addition of ~80 µM of FOXP2 FHD. The K_d calculated at these concentrations may not be entirely accurate as the saturation point is estimated from the data points that have been measured. However, it is clear that the dissociation constant for the FOXP2 FHD-ER1 LBD interaction increases with the increase in buffer salt concentration and that there is an overall significant difference in the change in anisotropy following the addition of the same concentration of FOXP2 FHD at the different salt concentrations ($p = 0.002$).



[Salt] (mM)	K_d (μM)	Bmax	R^2
50	15.89 ± 1.25	0.06 ± 0.00	0.99
100	46.28 ± 1.51	0.09 ± 0.00	0.99
200	102.31 ± 17.70	0.10 ± 0.01	0.99
300	251.19 ± 43.92	0.11 ± 0.01	0.99

Figure 5.20: Effect of salt on the FOXP2 FHD-ER1 LBD interaction. ER1 LBD at a concentration of 0.3 μM and labelled with AlexaFluor 555 C 2 maleimide was titrated with increasing concentrations of FOXP2 FHD in various buffers containing different salt concentrations (in the range of 50-300 mM). The data for lower salt concentrations follow a more hyperbolic trend, typical of binding, whereas that for higher salt concentrations shows a more linear curve across the concentration of protein titrated (~70 – 80 μM) and do not reach saturation. The K_d values obtained from a fit to a single site binding model increase from 15.9 μM to 251.2 μM as the salt concentration increases. The average relative changes in anisotropy following the addition of the same concentration of FOXP2 FHD across the four salt concentrations are significantly different $p = 0.002$) and there is a clear trend of decreased affinity with increased salt concentration.

5.8.2) Isothermal Titration Calorimetry

To understand the molecular basis underlying the decrease in binding ability between the proteins with increasing salt concentrations, isothermal titration calorimetry was performed. FOXP2 FHD was titrated into a solution containing the ER1 LBD. The heat change that was observed is indicative of an interaction between the two proteins, confirming the anisotropy studies. The curve

fitted well to the independent model from the NanoAnalyze software and produced a K_d of 0.4 µM as shown in Figure 5.21A. The changes in enthalpy, entropy and Gibbs free energy were determined to be negative. This finding was consistent for all ITC runs performed at 100 mM NaCl.

However, there is an interesting change in the thermodynamic parameters of the FOXP2-ER1 interaction when the salt concentration is increased from 100 mM to 300 mM. Using the same concentrations of proteins at 300 mM NaCl as at 100 mM NaCl yielded no change in heat following the continuous addition of FOXP2 FHD into ER1 LBD. However, addition of 400 µM FOXP2 FHD into 15 µM ER1 LBD resulted in heat being released and indicated protein-protein interaction, as evidenced in Figure 5.22B. The K_d (50 µM) at the high salt is 125 times weaker compared to at 100 mM NaCl as clearly shown in Figure 5.22A. This is in agreement with the anisotropy findings of Figure 5.20, which showed a greater binding affinity between ER1 LBD and FOXP2 FHD at a lower salt concentration.

There is not only a difference in affinity at the two salt concentrations, the thermodynamic parameters obtained from the calorimetry also show interesting differences as summarised in Figure 5.22B. There is a change in sign of both the entropy and enthalpy terms at the two salt concentrations. At 100 mM NaCl, both ΔH and ΔS are negative whereas at 300 mM NaCl, these values are positive, suggesting an interesting difference in the forces that drive complex formation at the different salt concentrations The change in free energy upon complex formation is not significantly different at either salt concentration meaning that despite the difference in affinity and mechanism of interaction, the proteins will spontaneously associate at both salt concentrations tested.

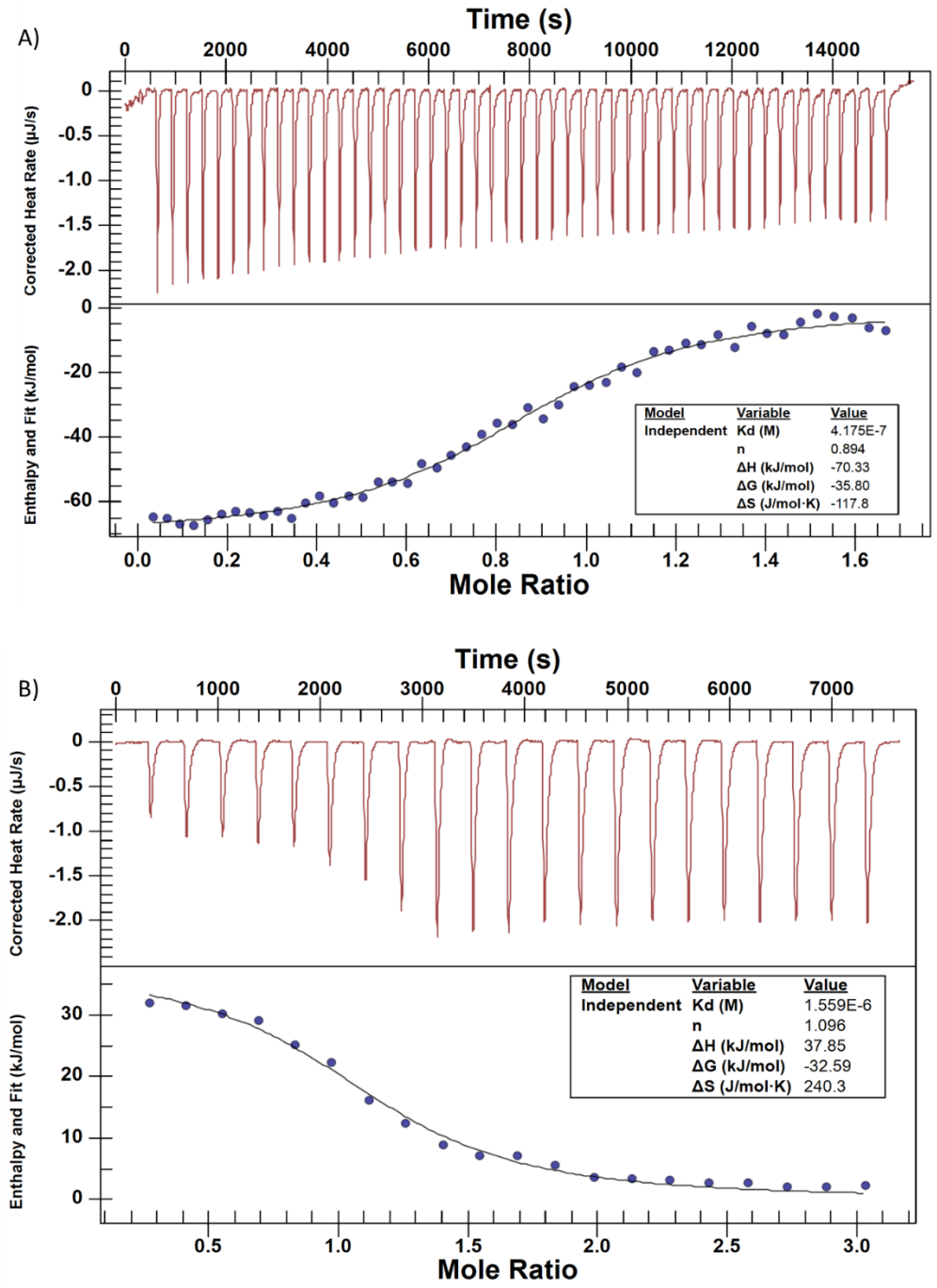


Figure 5.21: Thermodynamics of the FOXP2 FHD-ER1 LBD interaction. A) Protein-protein interaction at 100 mM NaCl. ITC was performed by titrating 8.5 µM of ER1 LBD with 50 µM FOXP2 FHD. The reaction occurred in 50 mM Tris buffer containing 100 mM NaCl and 1 mM TCEP, pH 7.8. B) The same Interaction at 300 mM NaCl. ER1 LBD (15 µM) in the cell was titrated with FOXP2 FHD (400 µM). The experiment was conducted in 50 mM Tris buffer containing 300 mM NaCl and 1 mM TCEP, pH 7.8. There is a change of sign in the entropy and enthalpy terms with an increase in salt and the K_d increases from 0.4 µM to 1.6 µM.

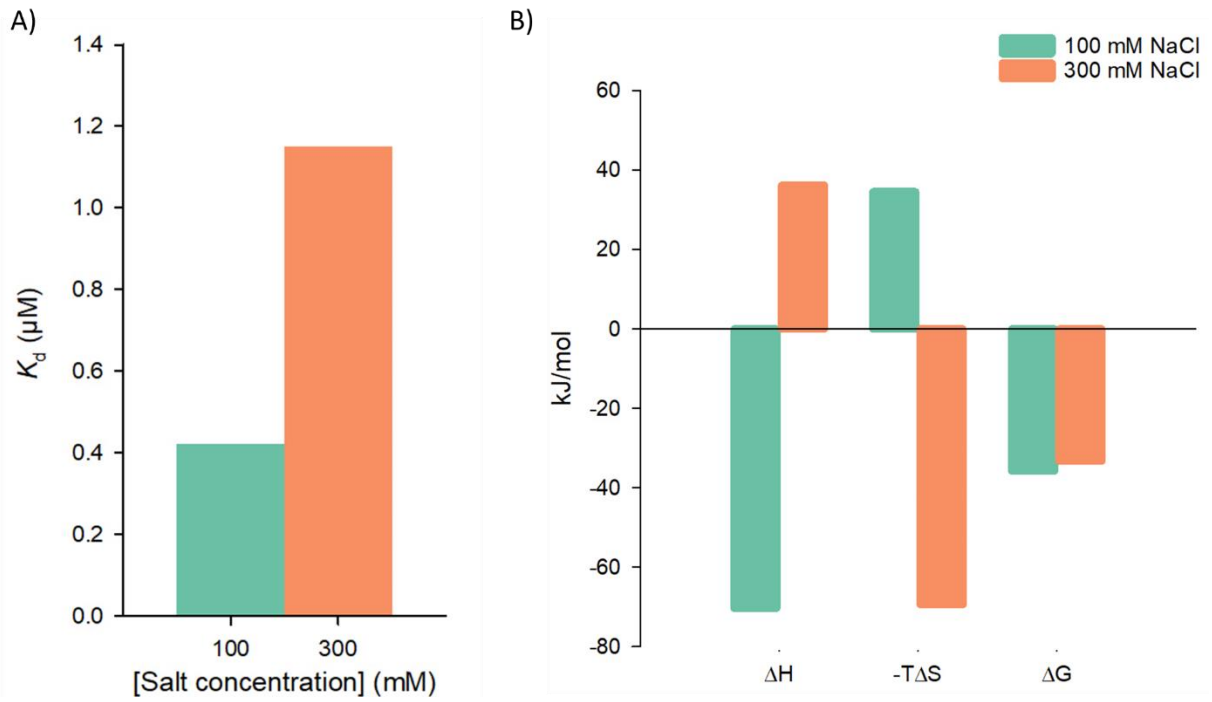


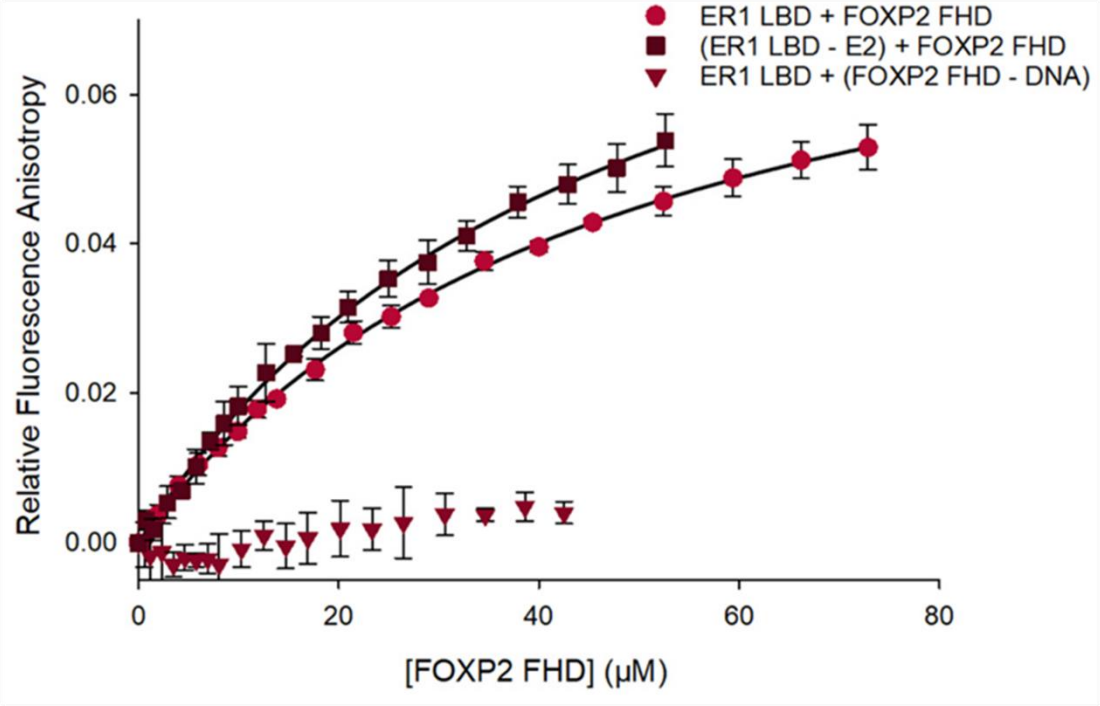
Figure 5.22: Comparative analysis of A) dissociation constants and B) thermodynamic parameters at different salt concentration. Isotherm was obtained from ITC by titrating FOXP2 FHD into ER1 LBD and fitted to an independent model on the NanoAnalyze software. In A, it is clear that binding affinity of the FOXP2 FHD-ER1 LBD interaction is salt dependent. In B, the change is free energy (Δ*G*) was negative at both 100 and 300 mM NaCl. However, changes in enthalpy (Δ*H*) and entropy (Δ*S*) differed considerably between the two-salt concentrations.

5.9) Effect of binding partners on the protein-protein interaction

Thus far, the interaction between FOXP2 FHD and ER1 LBD has been investigated with both the proteins in their apo-forms yet the functional activity of both proteins may be dependent on their interaction with their respective binding partners. Therefore, fluorescence anisotropy was performed to determines if and how the inclusion of the protein's binding partners could affect the interaction between FOXP2 FHD and ER1 LBD. ER1 LBD binds to E2 physiologically and *in vitro* (Tanenbaum *et al.*, 1998; Arao and Korach, 2018). Hence, ER1 LBD was successfully labelled with AlexaFluor 555 C2 maleimide and incubated with 50 μM E2 overnight in order to promote ligand binding. This experiment was performed on the assumption that the ER1 LBD interacted with E2 present in excess in the solution. This assumption was based on the ITC result shown in Figure 5.14 and previous literature (Eiler *et al.*, 2001). Following this, ER1 LBD was titrated with FOXP2 FHD in a similar manner to the apo experiments in section 5.6 (Figure 5.19). The resultant anisotropy curve (Figure 5.23) looked remarkably similar to the curve obtained

without incubating ER1 LBD with E2. The K_d too was almost identical. Since the curves in Figure 5.23 reached saturation, the dissociation constant could be accurately determined. The data suggest that E2 does not significantly affect the interaction between FOXP2 FHD and ER1 LBD.

Conversely, FOXP2 FHD, which is a DNA binding domain, was incubated with its cognate Nelson DNA in a 1:1 ratio overnight. EMSA was used to confirm that the protein formed a complex with the DNA. The FOXP2 FHD-Nelson DNA complex was then titrated into labelled ER1 LBD. There was no significant increase in the relative fluorescence anisotropy as the concentration of FOXP2 FHD-DNA was increased, as depicted in Figure 5.23. Moreover, the datapoints did not fit well to any particular binding model. This is considerably different to the anisotropy curve without the inclusion of DNA and implies that when FOXP2 FHD is bound to DNA, it is unable to full interact with ER1.



	K_d (μM)	Bmax (μM)	R^2
ER1 LBD + FOXP2 FHD	46.28 ± 1.51	0.09 ± 0.0	0.99
(ER1 LBD – E2) + FOXP2 FHD	45.95 ± 2.50	0.10 ± 0.0	0.99
ER1 LBD + (FOXP2 FHD – DNA)	N/A	N/A	N/A

Figure 5.23: Effect of the inclusion of binding partners (E2 or DNA) on the ER1 LBD-FOXP2 FHD interaction. ER1 LBD was labelled with AlexaFluor555 C2 maleimide and titrated with FOXP2 FHD as before, to obtain a hyperbolic anisotropy binding curve (red circle). This experiment was repeated with ER1 LBD incubated with E2 prior to FOXP2 FHD titration and a similar curve is obtained (brown square). These two plots were globally fitted to a ligand-binding nonlinear regression curve called the one site saturation on SigmaPlot 15.0. Conversely, when ER1 LBD is titrated with FOXP2 FHD pre-incubated with Nelson DNA overnight (1:1 ratio; pink triangle), there is no apparent increase in fluorescence anisotropy. This data is not fitted to any binding model due to poor correlation between the data points and the non-linear regression curves.

5.10) The effect of ER1 LBD-FHD interaction on FOXP2 FHD's DNA binding activity

The FHD is the DNA-binding domain of FOXP2 and its interaction with Nelson DNA has been evaluated in section 5.5.1. To see if this DNA-binding activity is affected by the interaction between FOXP2 FHD and ER1 LBD, EMSA and fluorescence anisotropy experiments were performed.

5.10.1) Electrophoretic mobility shift assay

The interaction between Nelson DNA and ER1 LBD was first evaluated as shown in Figure 5.24A to ensure that ER1 LBD does not indeed have any interaction with the DNA. This served as negative control and ensured that any subsequent results obtained are based only on the FOXP2-ER1 LBD and FOXP2-DNA interactions. As shown in Figure 5.24B, FOXP2 FHD at two different concentrations was incubated with increasing concentrations of ER1 LBD. This sample mix was then added to Nelson DNA. In the presence of no to relatively little ER1 LBD, the FOXP2 FHD is able to form a complex with Nelson DNA (red arrow). However, at higher concentrations of ER1 LBD at which FOXP2 FHD is likely to interact with the protein, the FOXP2 FHD – Nelson DNA complex formation seems to be reduced – the higher band is barely visible in the presence of 34 µM ER1 LBD. As shown in Figure 5.24C, this trend is consistent even when ER1 LBD is first incubated with E2 and then FOXP2 FHD.

Densitometric analysis was performed to quantitively assess free DNA content and protein-DNA complex formation. The intensity of the lower band indicated by the green arrow in Figure 5.24 was investigated. The lower green band represents free DNA and is assumed to be at 100 % intensity in lane one when only DNA is present in the sample. The intensity of this band in the remainder of the lanes was quantified relative to the intensity in the first lane. In the absence of ER1 LBD, as expected, increasing concentrations of FOXP2 FHD results in decreased free DNA due to increased FOXP2 FHD-DNA complex formation. However, when the ER1 LBD concentration in the sample is increased to 10 µM or higher, there is an increase in the density of the free DNA band as shown in Figure 5.24D. This is because there is a decrease in formation of the FOXP2 FHD-DNA complex. The result suggests that FOXP2's ability to bind to its DNA is hampered by its interaction with ER1 LBD

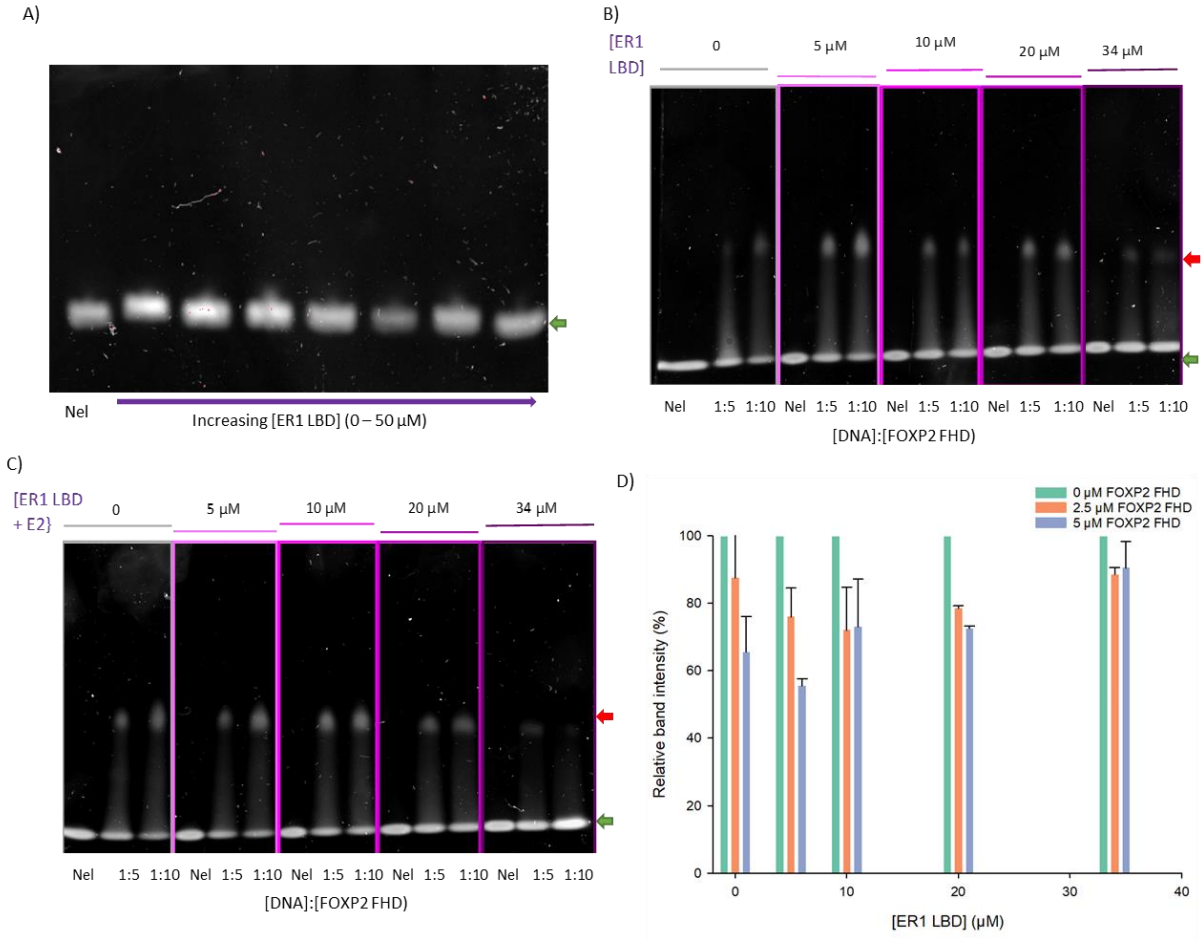


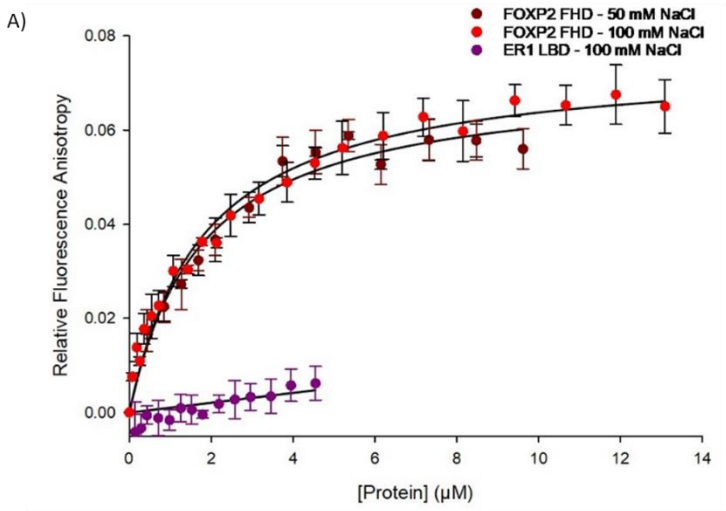
Figure 5.24: FOXP2 FHD's ability to bind to DNA following its incubation with ER1 LBD. A) The interaction between ER1 LBD and Nelson DNA. Nelson DNA (0.5 µM) was incubated with increasing concentrations of ER1 LBD and evaluated on an 8 % EMSA gel. The presence of only one band (green arrow) towards the bottom of the gel confirms that ER1 LBD does not interact with this DNA.FOXP2 FHD's functional activity in the presence of B) apo-ER1 LBD or C) E2-bound ER1 LBD was studied in a supershift EMSA. The free DNA (green arrow) and DNA bound to FOXP2 FHD (red arrow) were evaluated. FOXP2 FHD (2.5 and 5 µM) was incubated with 0 – 34 µM apo-ER1 LBD or ER1 LBD incubated with 100 µM E2 overnight. The FOXP2 FHD- ER1 LBD samples were then incubated with 0.5 µM Nelson DNA for 30 minutes and evaluated on an 8 % EMSA gel. D) Densitometric analysis of free DNA band. As ER1 LBD in its apo- and E2-bound form has a similar influence on FOXP2 FHD-DNA complex formation, the relative intensities of the lower band (green arrow) of gels in B) and C) were determined using ImageLab 5.2.1, averaged and plotted using SigmaPlot 15.0. The error bars represent standard deviation. The formation of the DNA-FOXP2 complex decreases with the increase in ER1 LBD concentrations.

5.10.2) Fluorescence anisotropy

Fluorescence anisotropy was used in addition to EMSA to further evaluate the interaction between FOXP2 FHD and its cognate DNA following its interaction with ER1 LBD. Anisotropy was performed with FOXP2 FHD and Nelson DNA first so as to determine the binding affinity of the protein-DNA interaction. Labelled DNA was titrated with increasing concentrations of FOXP2 FHD in buffers with different salt concentrations. The binding affinity of the FOXP2 FHD interaction with Nelson DNA is 1.7 and 1.8 µM at 50 and 100 mM NaCl concentrations, respectively as shown in Figure 5.25A. When ER1 LBD was titrated into labelled Nelson DNA there were no or very minute changes in fluorescence anisotropy which indicates that Nelson DNA does not interact with ER1 LBD confirming the EMSA result in Figure 5.24A.

In Figure 5.25B, labelled ER1 LBD was titrated against saturating concentrations of FOXP2 FHD (based on the findings from Figure 5.19). The ER1 LBD-FOXP2 FHD sample was then titrated with Nelson DNA to determine if the presence of DNA would affect the PPI between FOXP2 FHD and ER1 LBD – an increase in fluorescence intensity would indicate that Nelson DNA is able to bind to the FOXP2 FHD when it is in complex with ER1 LBD. This is evident at the higher salt concentration, which shows a slight increase in anisotropy.

In contrast, the opposite is observed at 50 mM NaCl as the addition of Nelson DNA results in a decrease in fluorescence anisotropy. Normally, this would indicate dissociation of the ER1 LBD-FOXP2 FHD complex but as the error bars are significantly large after the cumulative titration of approximately 40 µM Nelson DNA at which point the decreasing trend in anisotropy is observed, this change may be non-significant. Thus, at both the salt concentrations, there is only a very minute change in the relative anisotropy of ER1 LBD, and thus, the ER1 LBD-FOXP2 FHD complex remains largely unperturbed following the addition of Nelson DNA up to the concentration equivalent to that of FOXP2 FHD in the sample.



	K_d (μM)	Bmax	R^2
FOXP2 FHD (50 Mm NaCl)	1.7 ± 0.3	0.1 ± 0.0	0.97
FOXP2 FHD (100 mM NaCl)	1.8 ± 0.2	0.1 ± 0.0	0.98
ER1 LBD	N/A	N/A	0.54

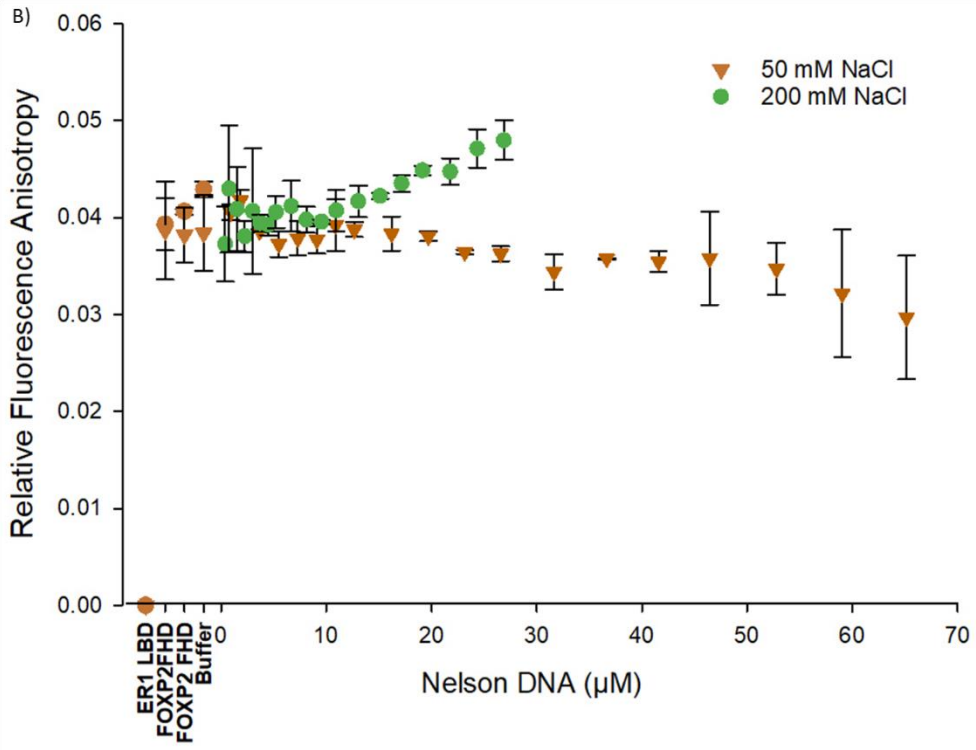


Figure 5.25: The effect of FOXP2's Interaction with Nelson DNA on the FOXP2 FHD -ER1 LBD interaction. A) FOXP2 FHD's interaction with Nelson DNA. Rox-labelled Nelson DNA (0.2 μM) was titrated with FOXP2 FHD in 50 mM Tris buffer containing either 50 or 100 mM NaCl. The resultant anisotropy plots were fitted to singe site saturation ligand-binding curves using SigmaPlot 15 to obtain a K_d of 1.7 and 1.8 μM at 50 and 100 mM NaCl, respectively. ER1 LBD

was titrated into the labelled Nelson DNA and exhibited no significant increase in anisotropy (purple). B) Addition of Nelson DNA into the FOXP2 FHD-ER1 LBD complex. Labelled ER1 LBD (0.2 µM) was incubated with saturating concentrations of FOXP2 FHD (32 and 68 µM in 50 mM and 200 mM buffer, respectively) for 30 minutes and thereafter titrated with Nelson DNA. There was a noisy increase in fluorescence anisotropy at the higher salt concentration and a decrease in anisotropy when the components were in the 50 mM NaCl buffer indicating that DNA may interfere with the PPI, but no curve could be fitted to this data.

5.11) Molecular Docking

The interaction between FOXP2 FHD and ER1 LBD was further analysed using molecular docking studies to visualise how the proteins might interact. The structural information may help in understanding the effect of the PPI on the structure as well as on either of the protein's functional activities observed *in vitro*.

5.11.1) Docked: Predicted ER1 LBD-FOXP2 FHD structures

Molecular docking can be used to predict the structure of the protein – protein interaction identified in this study. As molecular docking is merely a prediction tool with different tools utilizing slightly different techniques to do the docking, the FOXP2 FHD – ER1 LBD interaction has been docked using three different docking tools: AlphaFold (Mirdita *et al.,* 2022), ClusPro (Kozakov *et al.,* 2017; *Desta et al.,* 2020) and HADDOCK (Van Zundert *et al.,* 2016, Honorato *et al.,* 2021).

Firstly, the structure of the FOXP2 FHD-ER1 LBD complex was predicted using the sequence-based AlphaFold algorithm, which makes use of artificial-intelligence technology to combine information from homologous templates and multiple sequence alignment to generate structures (David *et al.,* 2022). The predicted structure of the complex is shown in Figure 5.26A. The confidence of the structure has been assessed using the predicted local distance difference test (pLDDT) score and the predicted aligned error (PAE). The pLDDT score suggests that most of the structure has been predicted with high confidence aside from the one helix of FOXP2 FHD and two helices of ER1 LBD which have a low confidence score. Regarding the PAE, a low value for a residue pair (x, y) from two separate domains suggests that they are positioned and oriented well in relation to the other.

In Figure 5.26B, the residues of the two proteins are separated by the red horizontal and vertical lines; the region spanning a larger number of residues is the ER1 LBD. If the y value of 150 and higher is selected, it corresponds to x values that are lighter in green in the FOXP2 FHD region

with a distance error of approximately 15 Å. A dark green shade indicates greater confidence in orientation, and exceptionally light shades or white indicate poor positioning and packing of the two domains. Hence, from the shading of the 2D PAE plot, the model can be considered with moderate confidence. The AlphaFold predicted ER1 LBD-FOXP2 FHD structure is depicted in Figure 5.27A and suggests that the DNA-binding helix (helix 3) of FOXP2 FHD interacts with ER1.

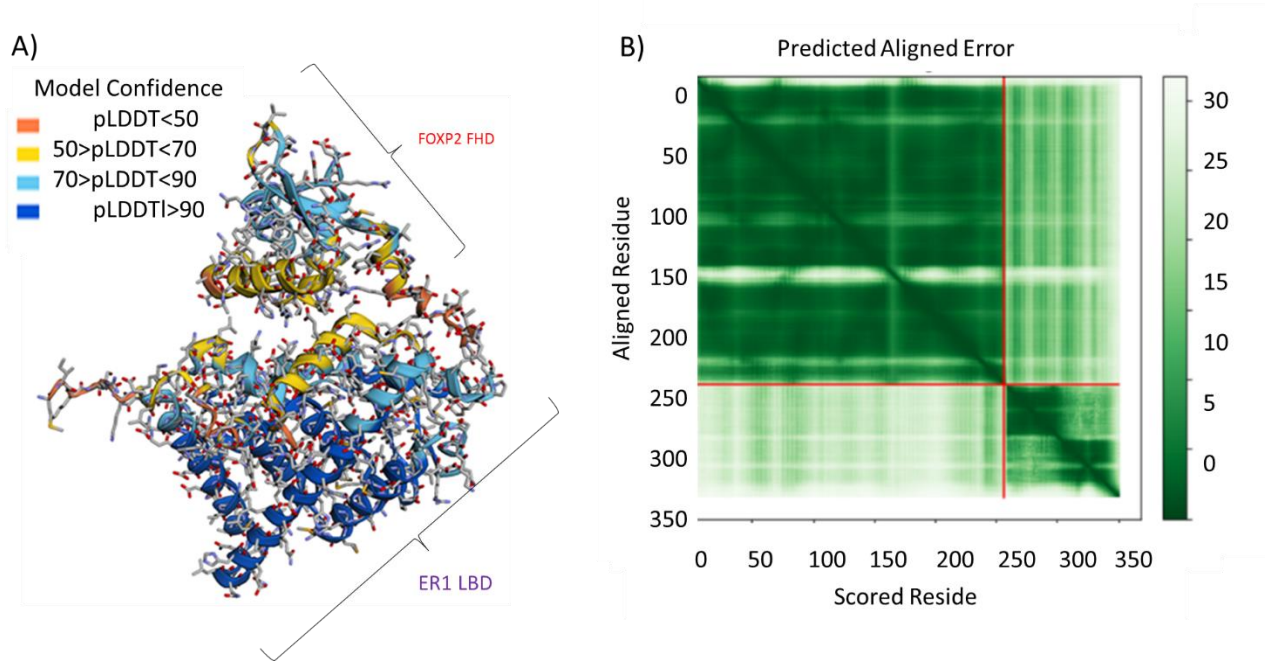


Figure 5.26: AlphaFold structure prediction of the FOXP2 FHD: ER1 LBD complex. A) AlphaFold predicted structure coloured in accordance with the per residue confidence score. The predicted local distance difference test (pLDDT) ranges from 1-100 and indicates if the residues are placed correctly in their local environment. The overall structure has been predicted with a relatively high confidence score; however, there are regions that have been predicted with a low and even very low confidence scores. B) A 2D predicted aligned error (PAE) plot of the structure. The PAE provides an estimate of positional error at a specific point (x) when the structure is aligned on a specific residue on the y-axis. It ranges from 0.-35 Å with a darker shade indicative of greater confidence. The red vertical and horizontal lines separate the residues of ER1 LBD (1-256) and FOXP2 FHD (257-341). Image generated using Google CoLab (Mirdita *et al.*, 2022).

Thereafter, the proteins were docked using structure-based programmes: ClusPro and HADDOCK. ClusPro uses fast Fourier transform correlation techniques to perform rigid docking

of protein structures. The predicted ER1 LBD-FOXP2 FHD structure is shown in Figure 5.27B and is based on a cluster of 78 members with highly negative energies being calculated for the interaction (Kozakov *et al.,* 2017; Desta *et al.,* 2020).

HADDOCK, on the other hand, is an information-based flexible docking programme. However, it can be employed for *ab initio* docking when no structural information is available (Van Zundert *et al.,* 2016, Honorato *et al.,* 2021). The *ab initio* docked structure has been shown as a cartoon structure in Figure 5.27C. The size of the cluster was 2 and resulted in a HADDOCK score of -49.6 +/- 6.4 and a Z-score of -1.1. The former is a weighted sum of various energy terms including van der Waals, desolvation, electrostatic and bond restraint violation energies and is used to determine the best pose for the PPI. The Z-score, on the other hand, is a measure of the number of standard deviations from where this cluster is located (Van Zundert *et al.,* 2016). The negative scores indicate more energetically stable and reliable structure prediction (Razzaghi-Asl and Ebadi, 2021).

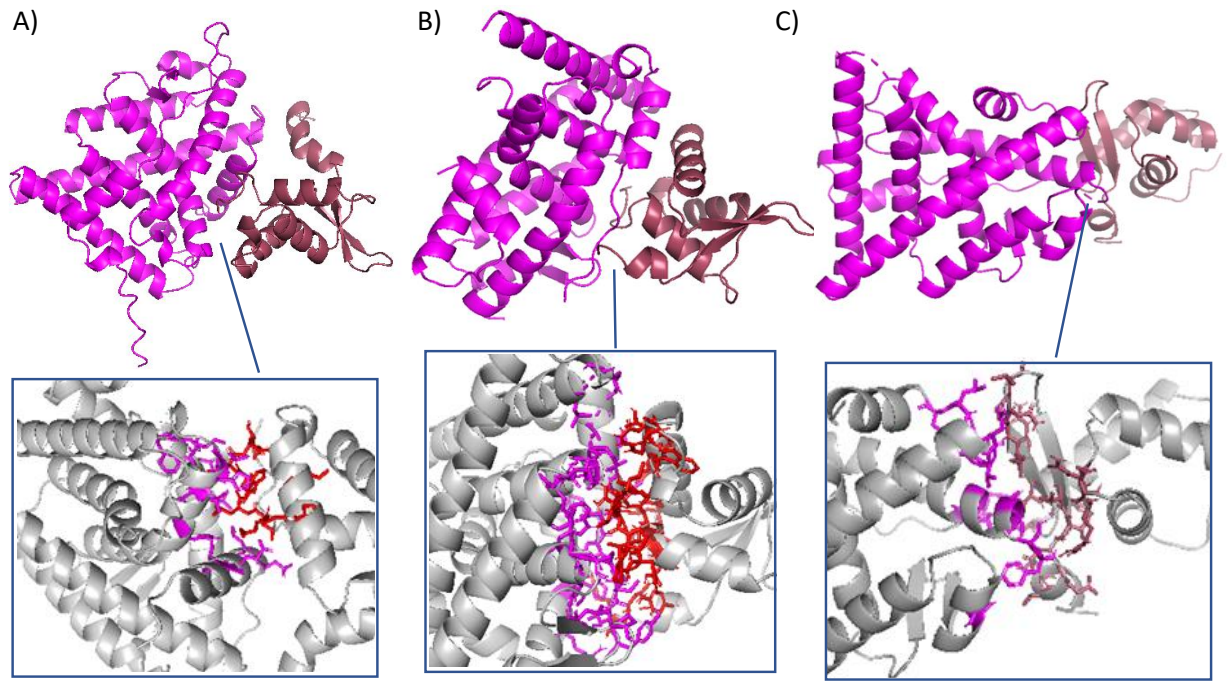


Figure 5.27: Structure-based docking of ER1 LBD and FOXP2 FHD. Docking was performed using the crystal structure of ER1 LBD (magenta; PDB ID: 1ERE) and FOXP2 FHD (raspberry red; PDB ID: 2A07) (Tanenbaum *et al.*, 1998; Stroud *et al.*, 2006). The interface residues of ER1 LBD (magenta) and FOXP2 FHD (red) have been identified and depicted as sticks using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC. A) Sequence-based AlphaFold prediction. The sequences of FOXP2 FHD and ER1 LBD were used to generated multiple sequence alignment, which was used to predict the structure. (Mirdita *et al.*, 2022). B) Rigid docking using ClusPro. The docking was based on a cluster of 78 members and resulted in a structure with the central and lowest energy scores of -802.7 and -975.2 J, respectively (Kozakov *et al.*, 2017; Desta *et al.*, 2020). C) Flexible docking using HADDOCK. The structure was obtained following an ab initio protocol. The cluster size was 2 with the HADDOCK score and z- score being – 49.6+/- 6.4 and of -1.1, respectively. The energetic contribution for the interaction comprised of Van der Waals (-24.5 +/- 5.5), Electrostatic (-130.3 +/- 10.6) and desolvation (-0.0 +/- 1.3) energies. The buried surface area was determined to be 998.4 +/- 112.4 (Van Zundert *et al.*, 2016, Honorato *et al.*, 2021).

The structures of the protein-protein complex obtained from AlphaFold, ClusPro and HADDOCK were super-imposed to identify any differences in the prediction. As shown in Figure 5.28, each program has predicted a different position and conformation of FOXP2 FHD relative to ER1 LBD. AlphaFold predicts that helix 1 of FOXP2 FHD and helix 12 of ER1 LBD predominantly participate in the formation of polar contacts between the proteins. Conversely, ClusPro shows that all helices of FOXP2 are involved in polar contacts with helices 1,2,3,5 and 8 of ER1 LBD.

HADDOCK shows minimal polar contacts restricted mostly to the winged region of FOXP2 FHD and helices 1 and 2 of ER1 LBD.

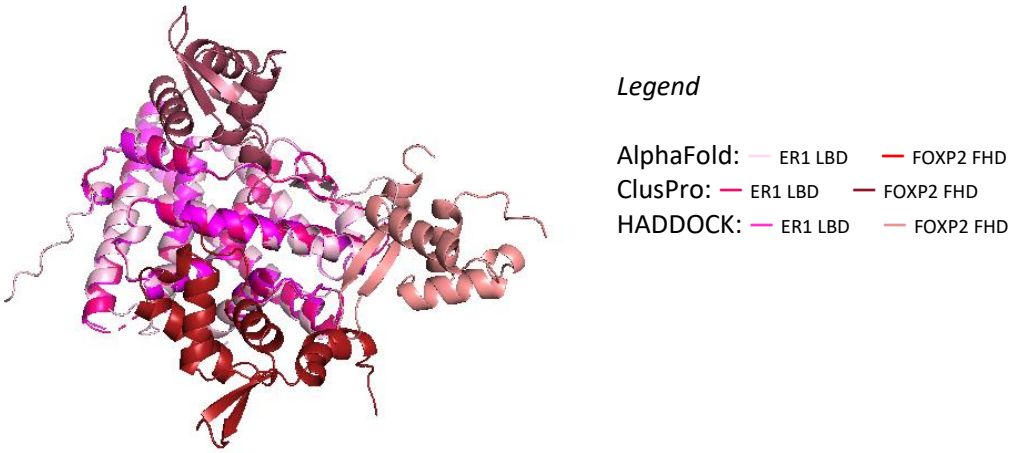


Figure 5.28: Alignment of ER1 LBD – FOXP2 FHD structure predicted using different docking programmes. The AlphaFold, ClusPro and HADDOCK structures are aligned to each other as cartoon structures in The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC. Each program identifies different binding interface.

The specific residues involved in polar contact between the proteins were identified using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC and have been summarised in Table 5.1. The interaction is predominantly mediated by polar residues based on the complexes predicted by all three docking programs. Moreover, it is the positively charged residues of FOXP2 FHD including arginine and lysine that are involved in the interaction as opposed to ER1 LBD, which is predicted to interact with FOXP2 FHD primarily through aspartic acid and glutamic acid, the negatively charged residues.

Table 5.1: Residues involved in ER1 LBD-FOXP2 FHD interaction obtained using different docking programs

	FOXP2 FHD	ER1 LBD
AlphaFold	Y509 (H1)	N536 (H12)
	A510 (H1)	D542 (H12)
	T511 (H1)	E546 (H12)
	E585 (W2)	
ClusPro	I502 (H1)	D313 (H1)
	V503 (H1)	D321 (H2)
	R504 (H1)	E323 (H2)
	P505 (H1)	P324 (H2)
	T508 (H1)	I326 (H2)
	Y509 (H1)	R352 (H3)
	Y540 (H2)	N359 (H3)
	R543 (H2)	W393 (H5)
	N544 (H3)	N439 (H8)
	T547 (H3)	Q441 (H8)
	H554 (H3)	E443 (H8)
	Y580 (W2)	
	R583 (W2)	
	R584 (W2)	
HADDOCK	R523 (H1)	E330 (H2)
	K560 (W1)	F337 (H3)
	V575 (W2)	

***+ charged residues, - charged residues, polar uncharged residues, non-polar residues,**

H: helix, W: wing

5.11.2) ER1 LBD-FOXP2 FHD complex in the presence of oestrogen

The findings in this study suggest that the interaction between FOXP2 FHD and ER1 LBD was not mediated by the LXXLL-containing FOXP2 NT region nor was it regulated by ER1 LBD's interaction with E2. To understand how this relates to the structure of the protein-protein complex and where FOXP2 FHD binds on ER1 LBD, the predicted protein-protein complex was super-imposed with the crystal structure of ER1 LBD bound to an agonist called estriol and a peptide containing the LXXLL motif (Rajan *et al.*, 2012). The estriol binds to the hydrophobic pocket whereas the peptide binds to the AF-2 region that mediates ligand-regulated binding coregulatory proteins. The docked structures from Figure 5.27 were aligned to this crystal structure to determine if they bound to the same region as the peptide. As shown in Figure 5.29, the interface residues of ER1 LBD that are involved in interaction with the peptide and FOXP2 FHD are found in two distinctly separate regions for the ClusPro and HADDOCK - predicted structures. Conversely, AlphaFold predicts the binding of FOXP2 FHD to ER1 LBD to be in the same region as that of the peptide.

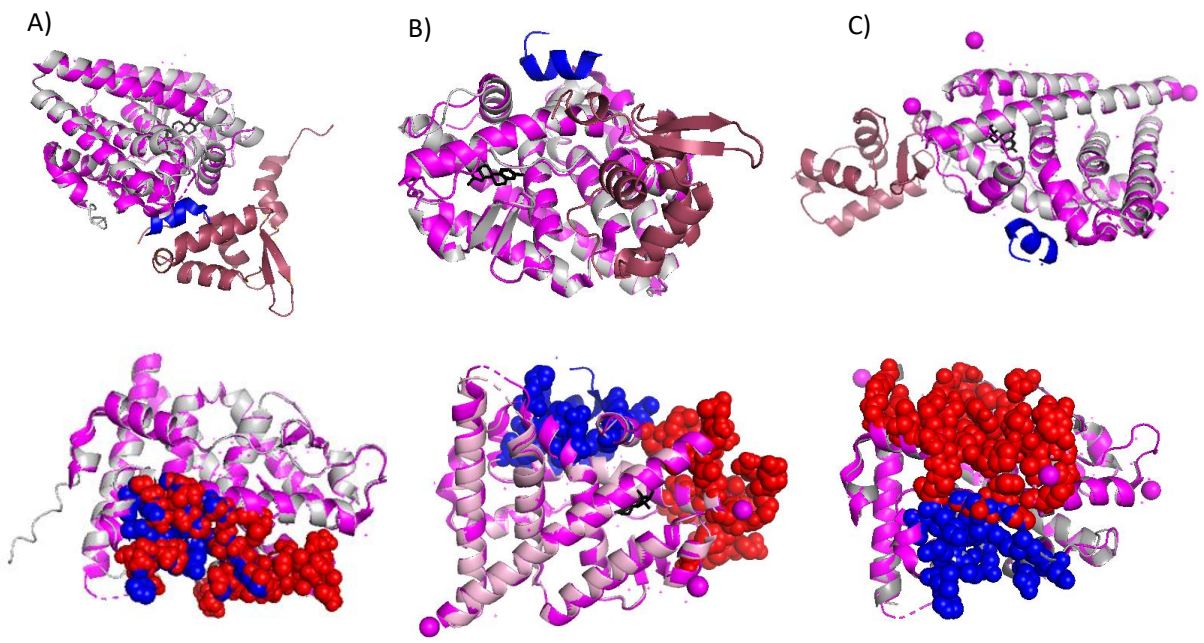


Figure 5.29: ER1 LBD crystal structure in alignment with predicted complex structure from A) AlphaFold, B) ClusPro and C) HADDOCK. The docked structure of ER1 LBD (Grey) and FOXP2 FHD (red) complex was aligned to the ER1 LBD crystal structure (magenta) bound to estriol (black) and an LXXLL-containing peptide (blue) (PDB ID: 3Q95). The corresponding structure below shows the interface residues of ER1 LBD in interaction with the peptide (blue) and FOXP2 FHD (red) as spheres. Aside from the AlphaFold structure, the other two docked complexes do not show FOXP2 binding to the same region as the peptide. Structures rendered using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC. (Van Zundert *et al.*, 2016, Kozakov *et al.*, 2017; Desta *et al.*, 2020; Honorato *et al.*, 2021, Mirdita *et al.*, 2022)

5.11.3) ER1 LBD-FOXP2 FHD complex in the presence of DNA

To put the protein – protein interaction into context with FOXP2 FHD's DNA binding activity, the FOXP2 FHD interaction with DNA is shown in Figure 5.30. The FOXP2 FHD has been crystallised in the presence of a specific DNA sequence (Wang *et al.*, 2003; Stroud *et al.*, 2006). This structure has been analysed and polar contacts between the protein and DNA have been shown.

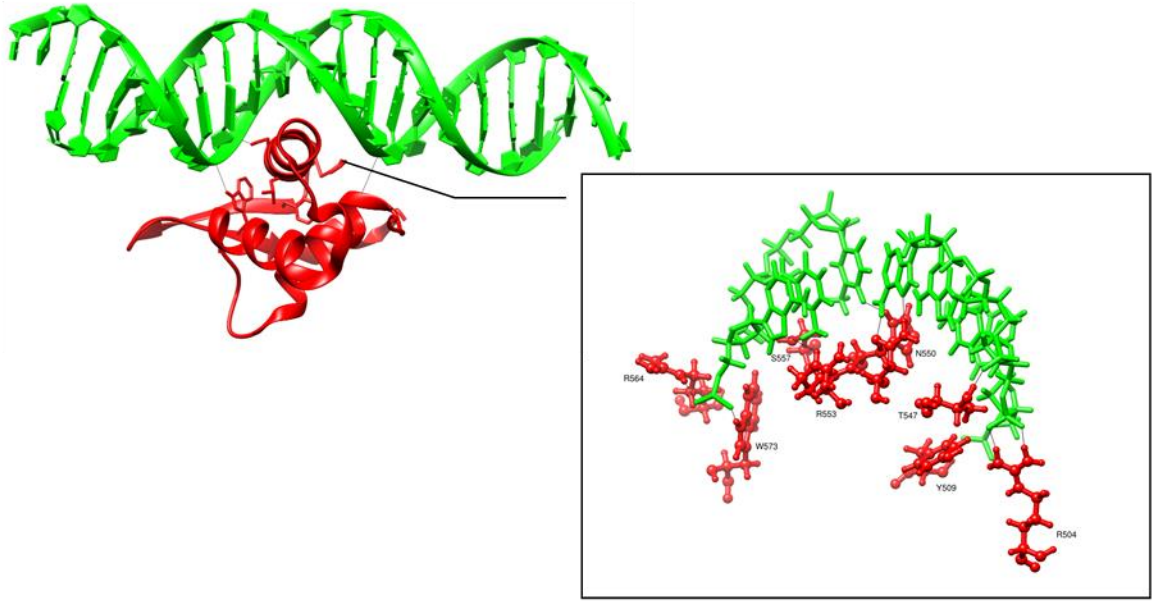


Figure 5.30: Interaction between FOXP2 FHD and Nelson DNA. FOXP2 FHD (red) in the presence of Wang DNA (green) has been shown with the black lines indicating Hydrogen bonds. Image rendered using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC (PDB ID: 2A07; Stroud *et al.*, 2006). The residues involved in polar contacts have been indicated as a ball and stick model and labelled. Image rendered using Chimera 1.16.

Fluorescence anisotropy showed that DNA-binding affected the ability of FOXP2 FHD to bind to ER1 LBD and vice versa. To determine how this would correlate with the structure of the protein-protein complex in relation to FOXP2's binding with DNA, the docked structures were super-imposed with the crystal structure of FOXP2 FHD with DNA (Stroud *et al.*, 2006). The images depicted in Figure 5.31 make it abundantly clear that some of the FOXP2 FHD residues involved in DNA binding are in close proximity to the ones involved in PPI with ER1 LBD. This is because helix 3 of FOXP2 FHD is involved in the interaction with DNA and as apparent for the docked structure from all three programmes, it also forms part of the interface in interaction with ER1 LBD

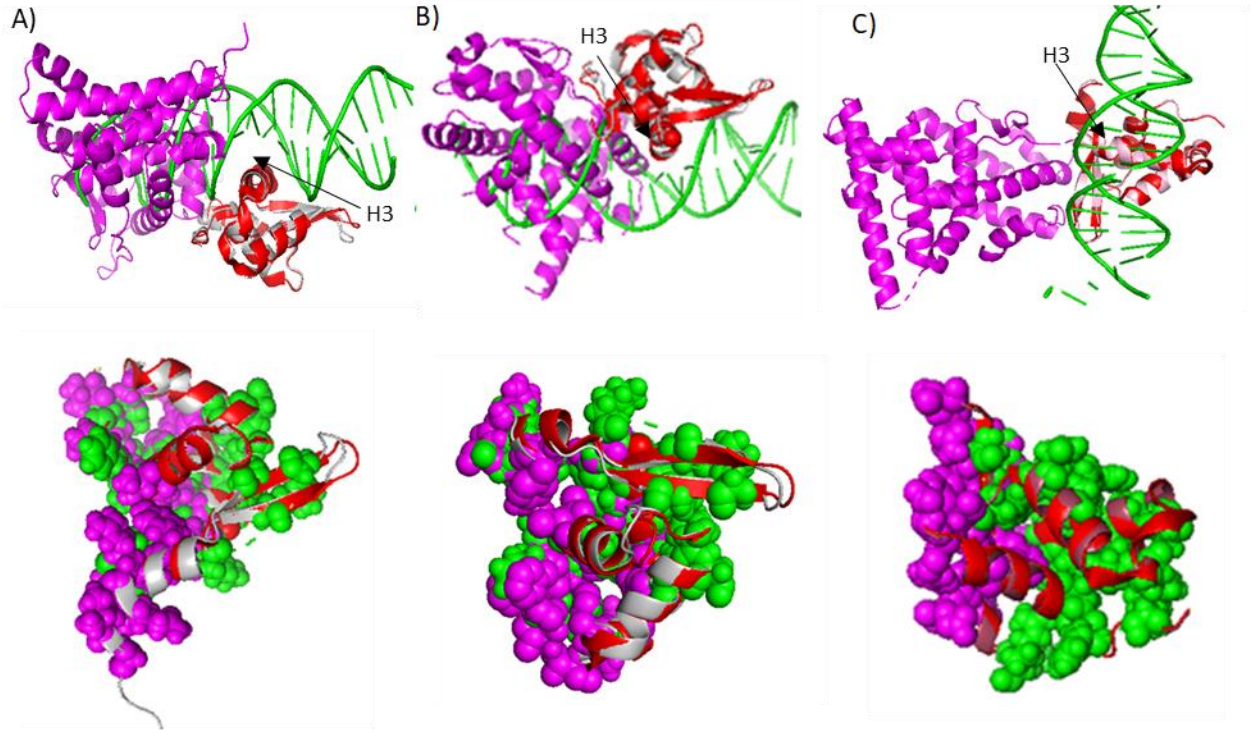


Figure 5.31 Analysis of the FOXP2 FHD – ER1 LBD interaction in the presence of DNA using the predicted structures from A) AlphaFold B) ClusPro and C) HADDOCK. ER1 LBD (magenta) is shown interacting with FOXP2 FHD (grey). The ER1 LBD-FOXP2 FHD complex was super-imposed with the structure of FOXP2 FHD (red) in complex with DNA (green) (PDB ID: 2A07; Stroud *et al.,* 2006). In the corresponding image below, the FOXP2 FHD residues involved in interaction with DNA (green) and ER1 LBD (magenta) are shown as spheres while the remainder of the protein has been indicated as a cartoon structure in The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC. (Van Zundert *et al.,* 2016, Kozakov *et al.,* 2017; Desta *et al.,* 2020; Honorato *et al.,* 2021, Mirdita *et al.,* 2022)

6) DISCUSSION

FOXP2 and ER1 are both transcription factors that have a role in neural development and cancer. FOXP2 has a role in language and speech development and acquisition, features that are conceived to be sexually dimorphic. Sexually differentiated traits are at least in part associated with sex hormones including oestrogen that mediates its effect through ERs. Furthermore, numerous FOX proteins have been shown to have a relationship with ER1 in breast cancer. FOXP2 too has a function is breast cancer and a structure that is relatively similar to other FOX proteins. Therefore, this research was focussed on investigating the potential interaction between FOXP2 and ER1 transcription factors to deduce whether their functionality may be dependent on or regulated by the PPI in an effort to expand the current understanding of sexual differentiation of speech and language as well the mechanism of FOXP2's tumorigenic activity in breast cancer.

6.1) ER1-FOXP2 association in MCF-7 cells

ER1 and FOXP2 have been found to interact in MCF-7 cells. Interestingly, the interaction is only identified when ER1 is immunoprecipitated using its specific antibody. One possible reason for this finding could be that FOXP2's interaction with ER1 impedes its ability to interact with its specific antibody and thus the antibody is unable to bind to FOXP2 bound to ER1. The FOXP2 antibody used in this study was produced by immunising animals with a synthetic peptide corresponding to residues surrounding L623 of the human FOXP2 protein. Therefore, the binding of ER1 has to be in proximity to L623 and surrounding residues to influence FOXP2's ability to bind to this antibody. This residue occurs at the C-terminal end of FOXP2 past the FHD that spans residues 503 to 586. As only the FOXP2 FHD structure has been resolved and the far C-terminal end is predicted with poor confidence using AlphaFold, it is very difficult to confer whether it is indeed the FOXP2 FHD that is involved in the interaction with ER1 such that it affects the contacts formed by L623 and surrounding residues with the FOXP2 antibody. Therefore, the cell studies credibly indicate an interaction between FOXP2 and ER1 that is most likely mediated through FOXP2's C-terminal end but do not confirm that it is specifically the FHD that is involved.

Hence, to investigate the specific of the interaction, two domains of FOXP2 and one of ER1 was recombinantly expressed. This included the DNA-binding forkhead domain and the nuclear receptor-box containing N-terminal region of FOXP2. The former was specifically selected to determine if the transcriptional activity of FOXP2 could be directly affected by a potential

interaction with ER1 whereas the latter contains the conserved residue that is known to mediate interaction with nuclear receptors. With regards to ER1, its ligand-binding domain was selected as this region facilitates both the interaction with ligands and coregulatory proteins. In order to conduct the study, FOXP2 FHD and NT as well as ER1 LBD were first expressed in *E. coli* cells and purified using chromatography techniques.

6.2) Structure and functionality of expressed proteins

The helical structure and DNA-binding function of FOXP2 FHD is well characterised whereas FOXP2 NT has not been studied at all, be it in isolation or through inclusion in larger constructs (Stroud *et al.,* 2006, Blane and Fanucchi, 2015, Morris and Fanucchi, 2016, Blane *et al.,* 2018). This is largely due to the high content of glutamine residues that make the region prone to aggregation (Häußermann *et al.,* 2019). Given its propensity for self-assembly, it makes sense that the FOXP2 NT does not exist as a monomer, but rather a dimer and trimer in solution (Figure 5.11). Notably, FOXP2 NT seems to be comprised of helices and coils, similar to the secondary structure of other proteins with poly-glutamine regions near their vicinity (Figure 5.16) (Totzeck *et al.,* 2017). As these sequences are usually found on the surface of proteins, it is likely that the NT lies on the surface of the full FOXP2 protein, modulating its folding (Davies *et al.,* 2008; Totzeck *et al.,* 2017). However, it's contribution to folding may not be significant as FOXP2's functional activity identified thus far remains unaffected in its absence (Vernes *et al.,* 2006; Estruch *et al,* 2016). In line with that, FOXP2 NT also did not interact with Nelson DNA and is unlikely to possess any DNA-binding activity (Figure 5.13).

On the other hand, ER1 structure has been well-characterised. The ER1 LBD is comprised of 12 helices which correlate with the CD spectrum obtained in this (Figure 5.11) and previous studies (Ferrero *et al.,* 2014). Binding of oestrogen is mediated by two direct hydrogen bonds as well as water mediated hydrogen bonds and van der Waals forces (Tanenbaum *et al.,* 1998). The configurational change in H12 along with the expulsion of water from the binding pocket likely contributes to the increase in disorder whereas the multitude of non-covalent interactions contribute to the favourable enthalpy of the reaction. Therefore, the reaction is expected to be spontaneous at all temperatures.

Notably, the binding affinity of the E2-ER1 LBD interaction is much lower than expected (Figure 5.18) as ligand-binding assays with radio-labelled E2 have been used previously to obtain a K_d of

0.2 nM (Eiler *et al.,* 2001). While slight variations in binding affinities obtained using different methods are expected, the binding affinity obtained here is too high for it to be accurate (Kišonaitė *et al.,* 2014). Given that the protein-ligand binding is expected to be extremely tight, the ER1 LBD concentration required to obtain a reasonable *c*-value is below the limit of detection of the calorimeter and contributes to the over-estimation of the K_d. The effect is probably compounded by the instability of the protein in the absence of its ligand as maintaining the 450 $Å^3$ hydrophobic cavity would require major adjustments (Brzozowski *et al.,* 1997). Tanenbaum *et al.* (1998) had even proposed that the specific contact between the ligand and protein allows the ligand to act as a scaffold for the folding of the lower part of ER1 LBD. Therefore, larger concentrations of protein were prone to aggregation the absence of E2 or NDSB and were very much avoided in the absence of the ligand where possible. However, the ITC was useful in qualitatively indicating that ER1 LBD is functional and interacts with E2. Consequently, ER1 LBD was labelled to assess PPI with FOXP2.

6.3) FOXP2 NT does not interact with ER1 LBD

FXOP2 NT did not show any association with ER1 LBD in this study. However, this does not necessarily mean that the N-terminal region of FOXP2 does not interact with ER1 LBD at all. Firstly, it may interact very weakly with ER1 LBD which would only be apparent if a much higher concentration of FOXP2 NT was titrated in. This was not possible in this study as the FOXP2 NT had a tendency to bind to the membrane when concentrating the protein. Therefore, obtaining higher concentrations of the protein presented a challenge. This was the case when concentrating with both the Amicon stirred cell and ultra-centrifugal filters (Merck, USA). This may be due to self-assembly of the protein or the protein having non-specific adsorbance to the membrane due to exposure of its hydrophobic residues. Hydrophobic proteins generally have low expression, which was indeed the case with FOXP2 NT as the yields were much lower than what was obtained for FOXP2 FHD and ER1 LBD (van Gils *et al.,* 2022).

Secondly, maybe a larger construct of the FOXP2 N-terminal region is required for the protein to fold properly and engage in correct PPI with ER1 LBD. Only the acid residues prior to the long poly-glutamine rich tract were used to construct FOXP2 NT based on the assumption that inclusion of more glutamine residues would adversely affect protein solubility. However, this meant that the residues in this construct constitute only half of the one helix, as shown in Figure 6.1. It is in this

incomplete helix that the nuclear receptor box is found. As aforementioned, the nuclear receptor box can sufficiently mediate interaction of nuclear receptor LBDs with coregulatory proteins. Should the incompleteness of the helix hinder its proper folding, it may have hindered the construct's ability to interact with ER1 LBD. The absence of a poly-glutamine tract in androgen receptor N-terminal domain results in the protein having a less pronounced alpha-helical structure and an impaired ability to bind to other proteins (Davies *et al.*, 2008). Hence, it may be advisable to generate a construct of the N-terminal region with the poly-Glu tract included.

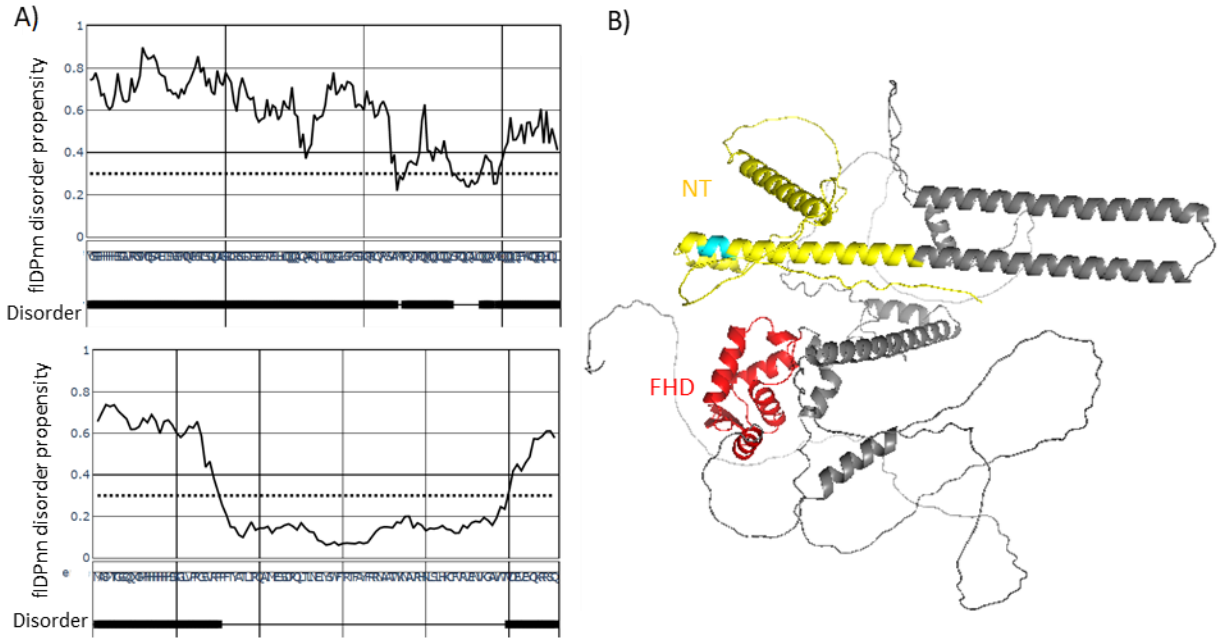


Figure 6.1: The disordered region of FOXP2. A) Predicted structure of FOXP2 NT (top) and FOXP2 FHD (bottom). FOXP2 NT is predicted to be predominantly a disordered region using flDPnn. The FOXP2 FHD is not predicted to have as much disorder, and this corresponds to its crystal structure which shows that the protein is predominantly alpha-helical. B) AlphaFold predicted structure of full length FOXP2. The forkhead domain is the only region with experimentally resolved structure. The three-dimensional structure of the rest of the protein including the NT has not been resolved hence the structure of the entire FOXP2 protein has been downloaded from the AlphaFold database. The FOXP2 NT (yellow) is comprised of 2.5 alpha-helices which were predicted with good accuracy. The bends or random coils which make up most of the structure is predicted with low confidence and lacks proper structure. The predicted structure of the FOXP2 FHD (red) corresponds to the crystal structure.

The AlphaFold predicted structure of FOXP2 also indicates that the NT region may be largely unstructured as it is represented by a considerable region that spans loops that have been predicted

with low confidence (Figure 6.1B). This correlates with the disorder predicted using the putative function- and linker-based Disorder Prediction using deep neural network (flDPnn) webserver. There is greater disorder found in the FOXP2 NT construct used in this study compared to the FOXP2 FHD and this is supported by the amino acid residue content of FOXP2 NT. It is comprised of mostly charged and polar residues and thus does not have sufficient bulky aromatic residues (FOXP2 NT comprises of only 2 H, 1 Y and 1 F) to mediate co-operative folding and so it forms an intrinsically disordered region (IDR). This region can contribute to PPI through exposure of linear motifs (Babu, 2016). IDRs can bind to several partners that may be structurally quite different to one another and have the propensity to form low affinity complexes that are important for transcriptional regulation and signalling pathways. Therefore, while it is unlikely that the FOXP2 NT construct used in this study (1 – 151 residues of the FOXP2 sequence) could be involved in the interaction with the ER1 LBD, the anisotropy findings do not completely negate the possibility of an interaction between the nuclear receptor box of FOXP2 and ER1 LBD. A larger construct may result in the FOXP2 NT assuming a different fold that could enable binding between its LXXLL sequence and ER1 LBD's AF-2 region.

6.4) FOXP2 FHD associates with ER1 LBD *in vitro*

Despite no interaction being detected between FOXP2 NT and ER1 LBD, this study shows clear evidence of a novel interaction between ER1 LBD and FOXP2 FHD with a binding affinity that can be considered to be moderate (Kastritis *et al.,* 2011). Notably, the K_d determined using ITC was 4-fold less than that obtained using anisotropy and Atto-labelled ER1 LBD. The data obtained from the isotherm are considered accurate as the *c* value calculated on the K_d obtained is 20. Hence, the slight difference in K_d may be associated with the techniques themselves – both techniques use different variables as a measure of interaction (i.e., heat change or anisotropy signal) which is bound to cause variations. However, the significantly higher K_d obtained by labelling the ER1 LBD using a maleimide confirms that the labelling of cysteine weakens ER1 LBD's ability to interact with FOXP2 FHD. ER1 LBD contains four cysteine residues, each at vastly different regions on the protein as shown in Figure 6.2. As labelling usually resulted in a dye to protein ratio of two, a couple of cysteines would be typically bound to the AlexaFluor dye for all the anisotropy experiments. This may explain the weaker affinity for FOXP2 in the anisotropy experiments. Notably, the three docked structures of ER1 LBD and FOXP2 show FOXP2 FHD binding to

distinct and different regions of ER1 LBD, and a cysteine is located on the interface for all three docked structures albeit at distinct locations.

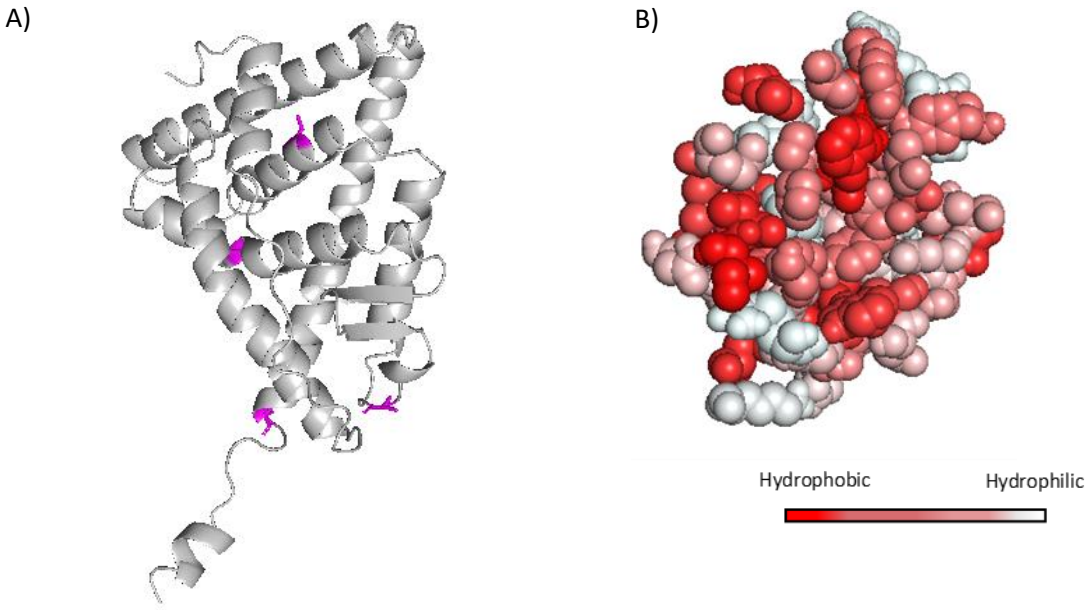


Figure 6.2: ER1 LBD structure. A) A focus on cysteine residues. ER1 LBD (Gray; PDB ID: 1A52) has been shown as a cartoon structure with the 4 cysteines (C381, C417, C447 and C530) indicated in magenta. Image rendered using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC. B) Hydrophobicity of the interface residues between ER1 LBD and FOXP2 FHD. The interface residues of the docked structure of ER1 LBD and FOXP2 FHD obtained from ClusPro has been depicted as spheres and coloured according to the hydrophobicity using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC (Kozakov *et al.*, 2017; Desta *et al.*, 2020).

6.5) ER1 LBD-FOXP2 FHD interaction is dependent on salt concentration

The affinity of the interaction between ER1 LBD and FOXP2 FHD shows a clear dependence on the salt concentration. The experiments were conducted at a pH of 7.8 and therefore ER1 LBD (pI= 6.8) and FOXP2 FHD (pI = 10.41) would have a net negative and positive charge, respectively. Therefore, one of the forces mediating the PPI will be electrostatic forces. It is likely that the increase in salt concentrations effectively shields the electrostatic forces on the surface of

the protein, thus hampering the ability of the proteins to interact (Ratanasumawong *et al.*, 2015). This effect can be explained using the law of matching water affinities.

The effect of salt on PPI is dependent on its ability to be hydrated which in turn is dependent on the strength of the interaction between the ion and water molecule compared to that between water molecules. An ion is considered strongly hydrated when it interacts more strongly with water than water interacting with itself and vice versa – this is considered the ion's water affinity. According to the law of matching water affinities, ions with opposing charges will form ionic pairs or an inner sphere when they have matching water affinities (Collins, 1997, 2006). Selected ions have been ranked on their water affinity in Figure 6.3. Monovalent anions like chlorine are weakly hydrated as opposed to monovalent cations which are strongly hydrated (Collins 2006). Sodium is relatively hydrated with the strength of its interaction with water almost equivalent to that between water molecules. On the protein surface, negatively-charged amino acids – Asp and Glu – are strongly hydrated carboxylates, while positively charged residues which are derivatives of ammonia are weakly hydrated. Polar and non-polar amino acids are also considered to be weakly hydrated (Collins, 2006). Therefore, the positively charged residues on the surface of FOXP2 FHD will interact strongly with the negatively charged chloride ions owing to their matching water affinities at pH 7.8 (Zhang, 2012). Consequently, increasing the NaCl concentration would increase the number of interactions between the chlorine ions and positively charged amino acid residues leading to neutralisation of the charge on the FOXP2 FHD. This will decrease the electrostatic interactions between the FOXP2 FHD, approaching a pseudo-neutral (or at least a decreased net positive) state and the net negatively charged ER1 LBD.

Conversely, there is a mismatch between the water affinity of sodium and the negatively charged residues on the surface of ER1 LBD, so it is likely that the decrease in binding affinity of the ER1 LBD-FOXP2 FHD interaction with increasing salt is driven more greatly by the increase in chloride ions in solution than the sodium ions. Furthermore, the chloride ions will be attracted to the weakly hydrated uncharged residues on the surface of both ER1 LBD and FOXP2 FHD and could increase the repulsive electrostatic interaction (Zhang, 2012). Ultimately, this would decrease the affinity of the PPI. This correlates with the molecular docking results obtained from all three programmes. From the polar contacts identified (Table 5.1), FOXP2 FHD comprises primarily of positively charged residues that interact with the negatively charged residues found

on ER1 LBD. Additionally, the HADDOCK program suggests that electrostatic interactions are pivotal for this specific-specific interaction as electrostatic energy is the largest energy contributor.

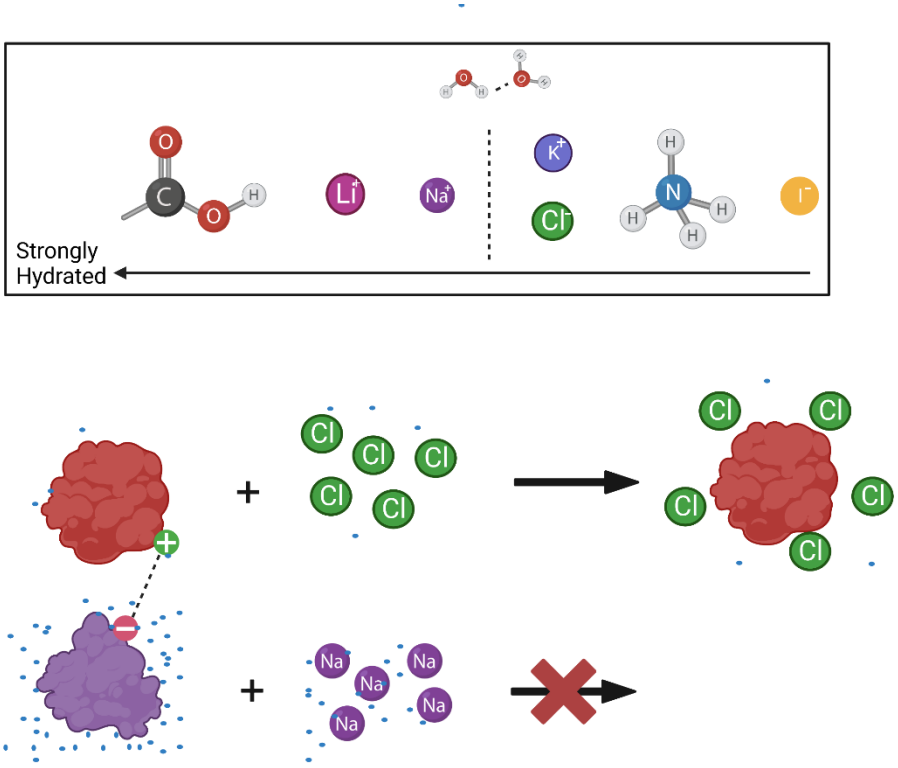


Figure 6.3: Water-matching affinities in regulation of protein-protein interaction in salt solutions. According to the matching water affinities, molecules of opposite charge with matching water affinities are likely to interact (Collins, 2006). At pH 7.8, positively charged residues on the surface of FOXP2 FHD interact with chloride ions in solution as they have similar affinities for hydration. Conversely, strongly hydrated negatively charged amino acids on ER1 LBD surface are unlikely to interact well with the relatively hydrated sodium ions. The change in ion concentration of the proteins' environment thus affects the attractive interaction between the positively and negatively charged protein.

Should the PPI exist under physiological conditions, the effect of NaCl could have an interesting role in regulating the interaction between FOXP2 AND ER1. Sodium ions are present in low concentrations (~5-15 mM) in the intracellular fluid relative to the intercellular fluid whereas chloride ions are found in greater proportion within the cell (~100 mM) (Allu and Tiriveedhi, 2021). Considering that ionic concentrations may vary by an order of magnitude or more between cell types and under different environments (including pH), the interaction could be specific to

certain cell types and physiological conditions. Interestingly, a three-fold increase in intracellular sodium levels has been noted in cancer cells as opposed to normal cells (Nagy *et al.,* 1981). On the other hand, increased chloride ions during brain maturation have been associated with autism (Fernell *et al.,* 2021). In addition, cellular environments are comprised of several other ions that could potentially shield the charges on the surface of the proteins and exemplify the effect of NaCl such as potassium ions which have altered levels in tumorigenic cells (Allu and Tiriveedhi, 2021). Therefore, such variations in ionic concentrations have the potential of either interfering with or attenuating the interaction between FOXP2 and ER1 if the *in vitro* results are to be taken into consideration. However, at present this is mere a postulation and would require substantial research on the interaction in the cells and *in vivo* to make an informed assertation.

6.6) ER1 LBD-FOXP2 FHD interaction is only enthalpically driven at low salt concentrations

The ER1 LBD-FOXP2 interaction is spontaneous at RT at both 100 and 300 mM NaCl concentration (Figure 5.21). In addition, the interaction is exothermic at the lower salt concentration indicating that the energy being released from the formation of bonds between ER1 LBD and FOXP2 FHD is greater than the energy required to break the original bond between either of the protein and solvent, solutes and solvent and solvent molecules. PPIs are mediated by several types of bonds including the aforementioned electrostatic as well as hydrogen, hydrophobic and van der Waals. Given that the K_d is quite low for the interaction (0.4 μM), there may be a relatively high number of intermolecular interactions mediating a relatively tight association between FOXP2 FHD and ER1 LBD.

Moreover, the ER1 LBD – FOXP2 FHD interaction is characterised by not only a negative enthalpy but also a largely negative entropy. PPIs are characterised by the release of water molecules from the reaction interface and into the solvent which would result in an increase in entropy of the solvent. The mechanism of the interaction can be deduced by considering the system studied by Tse *et al.* (2020), in which the protein-protein interface was primarily made of hydrophobic residues but also included some hydrophilic ones. Similarly, the interface between ER1 LBD and FOXP2 FHD is comprised of a mixture of hydrophobic and hydrophilic residues with the two hydrophobic patches surrounded by hydrophilic residues as shown in Figure 6.2. Therefore, the mechanism underlying the interaction is likely to be as follows: the approaching proteins cause water molecules to be expelled from the interface when they are a specific distance

apart from each other (distance less than the diameter of water-3.4 A˚) resulting in a large favourable change in solvent entropy (Tse *et al.,* 2020). However, as the proteins approach each other, their side chains lose a certain degree of freedom which decreases the conformational entropy of the proteins resulting in only a slight increase in the overall entropy. Thereafter, when distance between the proteins is at a minimum, the binding surfaces of the proteins adapt to maximise their interaction energy (including the bond formation between hydrophilic residues) causing a favourable decrease in enthalpy, which occurs at the expense of a large entropic loss associated with the constraints in movement of the interface residues of the proteins. It makes sense then that the tighter the binding between macromolecules, the greater the constraints in movement which would result in a largely negative entropy. The favourable enthalpy contributes to the negative change in free energy and makes the interaction spontaneous.

At the higher salt concentration, it is the enthalpic loss that is being compensated for by the gain in entropy. Burying of hydrophobic groups could be one of the causes of the unfavourable enthalpic change and favourable desolvation entropy (Sharp *et al.,* 1991). If the PPI is mediated through hydrophobic interactions, loss of hydrophobic patches deprives water molecules of orienting themselves on the protein surface such that they can no longer maintain their network of hydrogen bonds. Therefore, the decrease in hydrogen bonds with water accounts for the increase in enthalpy. Additionally, it would mean that the desolvation of polar groups requires greater energy contribution than the energy released from the interaction between ER1 LBD and FOXP2 FHD. It is then likely that the interaction at higher salt is formed by a lower number of bonds between the two proteins than the number of bonds mediating the interaction at the lower salt concentration (Choudhury and Pettitt, 2006).

The lower enthalpy of the PPI at high salt suggests that the binding may not be very tight, allowing for one or both proteins to have some conformational freedom and therefore a more positive conformational entropy than at lower salt concentrations (Freire, 2008). The disorder is compounded by the solvation entropy resulting from the release of water molecules from the protein surface during the PPI. This then means that increase in sodium and chloride ions in solution disrupts the hydrophilic interactions between proteins that would normally occur under a lower salt concentration ensuring the interaction is mediated largely by hydrophobic interactions. It may be likely that the chloride ions, which have matching water affinity to the positively charged

residues of FOXP2 FHD, interact with the charged polar residues and consequently prevent them from engaging in interaction with ER1 LBD. The salting-in effect of the increased salt concentration is consistent with the stability of FOXP2 FHD in solution as it is prone to aggregation due to self-assembly at lower salt concentrations (50 mM) than at higher salt concentrations (between 300 and 500 mM).

6.7) Not one or the other but both: effect of E2

Other than salt concentration, the interaction between the proteins can also be regulated by the presence of other interacting partners. ER1 LBD has a hydrophobic cavity to which oestrogenic compounds like E2 are able to bind. The difference in the structure of ER1 LBD when in its apo-form compared to its E2-bound form seems primarily localised to only one of the 12 helices comprising its structure (Brzozowski *et al.,* 1997*;* Tanenbaum *et al.,* 1998*).* The observation that inclusion of E2 does not seem to influence the PPI, suggests that the binding of FOXP2 FHD is impervious to the E2- induced conformational changes in helix 12 of ER1 LBD. This finding can be contextualised to the known mechanism of ER1-E2 signalling which is as follows: ER1 LBD binds to oestrogen or its other agonist, which induces a conformational change in Helix 12 and locks it in an "active" configuration that is now open to binding with co-regulatory proteins or even peptides (Hewitt and Korach, 2018). Additionally, all proteins found to interact with the AF-2 region thus far contain the LXXLL motif that FOXP2 FHD lacks (although it is found in FOXP2 NT). Thus, it is likely that the interaction between ER1 LBD and FOXP2 FHD is not mediated through LBD's AF-2 region.

However, the AlphaFold predicted structure suggests the opposite (Figure 5.29) as FOXP2 FHD is predicted to bind to the same region as the LXXLL peptide which binds to the AF-2 region. Should this structure be an accurate representation of the protein-protein complex, it would not be possible for the interaction to remain unaffected by the absence and presence of an ER1 ligand like E2. This is because some of the interface residues of the AlphaFold structure are the same as the ones involved in interaction with the LXXLL motif in a crystallised structure (Shiau *et al.,* 1998). This includes the following residues: Val355, Ile358, Lys362, Val368, Leu372, Val376, Leu539, Glu542 and Met 543. The last three residues are subject to change in orientation when binding to ligands such that when ER1 LBD is bound to an antagonist, Leu539, Glu542, and Met543 are in the incorrect orientation while other residues from the static regions of AF-2 are bound to helix

12. Consequently, the protein is unable to interact with its coactivators. This would be the case with FOXP2 FHD as well, should it bind to a similar region as the peptide.

Notably, ClusPro and HADDOCK predict a binding site for FHD to be distinct from the AF-2 region. Therefore, based on the experimental results from this work, it is likely that the ClusPro and HADDOCK structures predict the binding site of FOXP2 FHD on ER1 LBD more accurately than AlphaFold, although ClusPro and HADDOCK do suggest entirely different regions of binding. HADDOCK predicts the binding of FOXP2 FHD to be below H3, H7 and H11 of ER1 LBD, whereas ClusPro predicts the binding to be near the regions between H2, H3, H8, H9 and H10 as shown in Figure 6.4A. Given that the position of H12 would be positioned outwards in the absence of a ligand, this would make H12 lie directly next to the HADDOCK-predicted positioning of FOXP2 FHD and would influence the PPI. Hence, the ClusPro prediction is most consistent with the laboratory findings from the three docked structures.

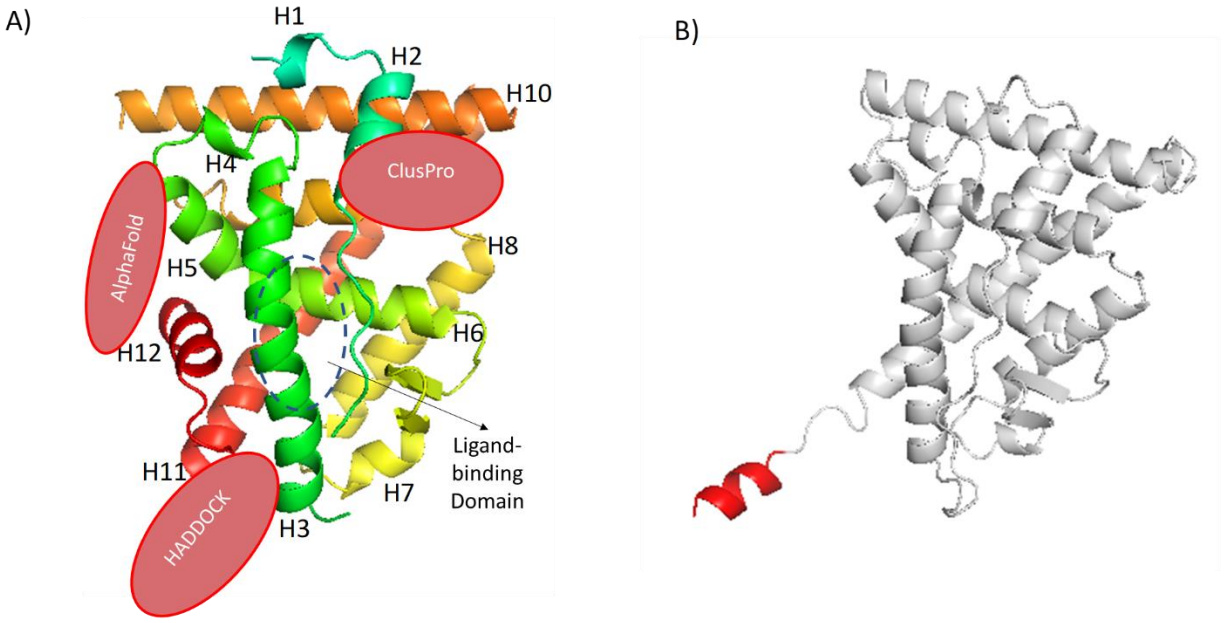


Figure 6.4: Cartoon structure of ER1 LBD. A) Liganded ER1 LBD (PDB ID: 1ERE) coloured by chains. The relative positioning of FOXP2 FHD based on the models generated by AlphaFold, ClusPro and HADDOCK are indicated. (Brzozowski *et al.*, 1997). B) Position of H12 when unliganded. ER1 LBD (PDB ID: 1A52) is shown in grey with helix 12 indicated in red. Imaged rendered using The PyMOL Molecular Graphics System, Version 2.5.5, Schrödinger, LLC (Tanenbaum *et al.*, 1998).

6.8) One or the other but not both: effect of DNA

While the ER1 ligand does not affect the PPI, inclusion of DNA in the study influences the ability of FOXP2 FHD to participate in interaction with ER1 LBD as has been summarised in Figure 6.4. The findings presented here (Figures 5.23 and 5.24) suggest that FOXP2 FHD can only bind to one of the macromolecules at a given time and not both simultaneously. It could also suggest that if FOXP2 binds to DNA first, it can no longer bind to ER1 LBD and vice versa. However, this theory would not make sense if the kinetics of the interactions are considered, as the protein-DNA and protein-protein complexes will associate and dissociate continuously in solution. If that happens, FOXP2 FHD will form a complex with its preferred binding partner.

Considering that FOXP2 FHD has greater affinity for DNA than for ER1 LBD, DNA would be its preferred partner. To confirm this hypothesis, labelled ER1 LBD was mixed with DNA prior to titration with FOXP2 FHD. The resultant anisotropy curve differed to the one obtained when ER1 LBD alone served as the titrand as it was no longer linear in appearance and initially indicated either little to no increase in fluorescence anisotropy (Figure S2). This would suggest that FOXP2 FHD does bind selectively to its DNA rather than to ER1 LBD. However, one has to consider that while the ER1 LBD and background DNA were the limited variables in the experiment, the concentration of labelled ER1 LBD was 0.3 µM and background DNA concentration was 7 µM. The latter could not be decreased further as too low a DNA concentration would make its inclusion obsolete; the concentration of FOXP2 FHD would supersede that of DNA within a couple titrations and would saturate the DNA present in the sample even prior to observing a change in fluorescence anisotropy of the ER1 LBD. Hence, it is more likely that the titrated FOXP2 FHD interacted with the DNA that was present in higher concentration rather than the more limited ER1 LBD. Should the fluorescence signal of ER1 LBD at a greater concentration not have had a fluorescence intensity beyond the limit of the fluorimeter, it would have been advantageous to perform anisotropy with equimolar concentration of labelled ER1 LBD and background DNA concentration and deduce which macromolecule FOXP2 is partial to.

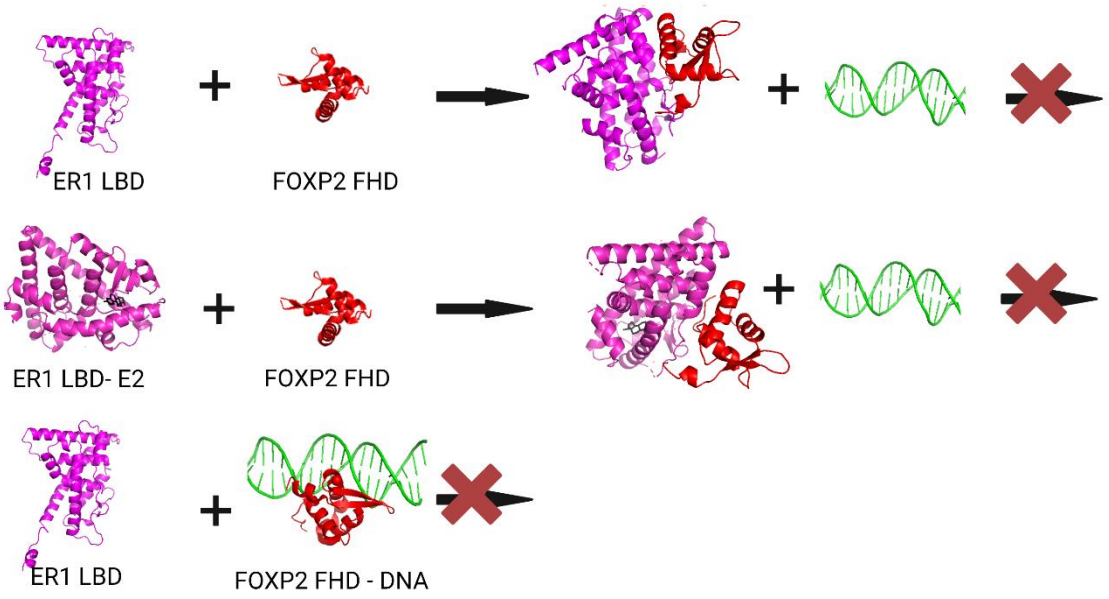


Figure 6.5: Interaction between ER1 LBD and FOXP2 FHD. ER1 LBD in its apo (PDB 1D: 1A52) and E2-bound forms (PDB ID:1ERE) associates with FOXP2 FHD (PDB ID:2A07) to form a protein-protein complex, hindering FOXP2 FHD's ability to bind to its cognate DNA. However, when FOXP2 FHD is already bound to DNA, it cannot associate with ER1 LBD to form a complex.

Similarly, ER1 LBD or DNA were in limited quantity in all the experiments performed with FOXP2 FHD, DNA and ER1 LBD in this study. ER1 LBD was the limited moiety in fluorescence anisotropy experiments, whereas the DNA was the limited entity in EMSA experiments. Based on these findings, FOXP2 FHD could bind to either the DNA or ER1 LBD at any given point in time but not to both of them simultaneously. Additionally, FOXP2 bound to the macromolecule that was present in higher concentration rather than the one that was limited. Hence, it is likely that the limited molecule is being outcompeted by the other molecule for binding with FOXP2 FHD.

Molecular docking was performed to correlate the finding of the PPI in the presence of DNA to the structure of the complex. Notably, docking approaches have shortcomings and may not always generate accurate models for the interaction (Vreven *et al.*, 2015; Lensink *et al.*, 2020). Although three different approaches were used to mitigate this disadvantage, there is no certainty that any of these models, even if similar, accurately represent the protein-protein complex *in vivo*. Hence, they are just used as a basis to explain some finding determined experimentally.

The predicted ER1 LBD-FOXP2 FHD structures obtained from the three different approaches were aligned with the FOXP2 FHD structure bound to DNA. All three structures show that the

interfaces of FOXP2 FHD in interaction with both DNA and ER1 LBD are not distinct and have considerable overlap (Figure 5.31). This is especially the case for the AlphaFold and ClusPro predicted structures but less so for the HADDOCK structure. The latter shows the DNA and ER1 LBD-binding interfaces to be adjacent to each other but there is an overlap on a small region which could still be significant. Therefore, all three structures could correlate with the experimentally derived information suggesting that FOXP2 FHD binds to either DNA or ER1 LBD at a given time and does not bind both simultaneously. Contextually, the results suggest that ER1 LBD could regulate the interaction between FOXP2 FHD and DNA. ERs have been found to regulate transcription without direct binding to DNA through a process known as transcriptional crosstalk (Bjornstrom and Sjoberg, 2005). In such instances, ER is tethered to another transcription factor to regulate gene expression.

6.9) Towards a role in oncogenesis

The interaction of ER1 and FOXP2 in non-malignant cancers taken together with the finding that the ER1 LBD-FOXP2 FHD interaction impedes FOXP2's ability to interact with its DNA suggests that the interaction could be physiologically relevant. Hence, the interaction could be studied further in cells to identify what significance the PPI could have, if any, on oncogenesis. ER1 is overexpressed in breast cancer and contributes to cancer progression by activating the PI3K/AKT signaling pathway (Leung *et al.,* 2015, Arun *et al.,* 2018, Liu *et al.,* 2020). However, ER1 is also associated with better prognosis as opposed to ER-negative breast cancers (Dou *et al.,* 2017, Liu *et al.,* 2020). In contrast, FOXP2 expression is downregulated in breast cancer and thus its tumour suppressive activity is suppressed (Chen *et al.,* 2018). However, both proteins are involved in two pivotal tumour-promoting pathways. This includes the TGF-β and PI3K/AKT signalling pathways. Regarding the former, ER1 blocks the pathway by partly being involved in ligand-independent cross-talks with various TGF-β signalling proteins whereas downregulation of FOXP2 is associated with elevated levels of various TGF-β pathway related proteins (Stope *et al.,* 2010; Chen *et al.,* 2018). More interestingly however are their roles in the PI3K/AKT pathway.

FOXP2 is linked to the PI3K/AKT signaling pathway through the X-linked ribosomal S6 kinase 4 (RSK4), a ribosomal protein that can act as a tumour suppressor by inhibiting the PI3K/AKT signaling pathway (Huo *et al.,* 2019, Yang *et al.,* 2022). RSK4 has decreased expression in breast cancer (Li *et al.,* 2014). Yang *et al.* (2022) have also found that FOXP2 binds directly to the

promotor region of RSK4 and modulates its expression level in thyroid cancer, a cancer in which FOXP2 has a tumour suppressor role. On the other hand, Huo *et al* (2019) showed that ER upregulation is related to the decreased expression of RSK4 in ER-positive breast cancer, which in turn is linked to endocrine therapy resistance to potent chemotherapeutic drugs, such as Doxorubicin (Mei *et al.*, 2020). As this study on recombinant proteins shows that ER1 LBD and DNA compete for binding with FOXP2 FHD, there is a possibility that the upregulation of ER1 in breast cancer could lead to elevated ER1-FOXP2 interaction which could inhibit FOXP2's transcriptional activity, preventing it from binding to the promoter region of RSK4 and ultimately inhibiting RSK4's tumour suppressor activity. Hence, further studies are paramount to unravelling the potential ramifications of the FOXP2-ER1 interaction in breast cancer cells.

6.10) Limitations of the study

In this study, the functional characterisation of ER1 LBD binding to E2 was qualitatively assessed using ITC as the binding affinity obtained was unreliable. In the absence of a suitable competitive ligand, displacement ITC was not performed but could be used in future to determine a more accurate measure of the interaction. Additionally, the association between ER1 LBD and FOXP2 FHD was investigated using only the ER1 LBD and thus it cannot be ascertained whether the same results will be obtained with larger constructs or even the full protein as contacts between the residues of the protein may influence the PPI. Penultimately, while the study illustrates that FOXP2's interaction with DNA is a regulator of and may even be regulated by FOXP2 FHD's interaction with ER1 LBD *in vitro*, more research into the nature and specificities of the competing interactions is required to determine the relevance on functional activity of either protein. Finally, the ER1- FOXP2 interaction was investigated in breast cancer cells to glean a possible physiological relevance. While the interaction was identified, it needs to be confirmed using a different antibody to eliminate the background noise so that the results are clear and indisputable. Both proteins also have roles in other cancers including prostate. Furthermore, the interaction may also have implications on sexual differentiation of language and speech acquisition development and diseases. Hence, there is a possibility that the interaction could affect other biological processes, the relevance of which should not be ignored in future studies.

7) CONCLUSION

FOXP2 and ER1 are transcription factors and have varied but pivotal roles in the body. FOXP2 has a role in language development, neuronal development and has been implicated in language, speech and related disorders and cancers. ER1 has an extensive role in various organ systems, including the cardiovascular and central nervous systems, is important for maturation of female reproductive organs, reproduction and is implicated in various cancers as well as resistance to endocrine therapies. To understand language and speech, including its development and differences between males and female sexes, an interaction between FOXP2 and ER1 was investigated using specific domains of both proteins.

The fluorescence anisotropy, ITC and EMSA results combinatorically suggest an interaction between FOXP2 FHD and ER1 LBD. This interaction is regulated by NaCl concentrations - increase in salt weakens the affinity between the proteins. This is likely due to the change in mechanisms of binding as evidenced by the converse effects of enthalpic and entropic changes driving the association. Molecular simulations may confirm the mechanism of the interaction at the different salt concentrations as postulated in this study.

Moreover, the binding of DNA and ER1 LBD to FOXP2 may be competitive and is in agreement with molecular docking studies that show considerable overlap in the FOXP2 FHD interface involved in association with DNA and the interface mediating interaction with ER1 LBD. Notably, the FOXP2 FHD -DNA interaction is not regulated by NaCl concentration but as the PPI between ER1 LBD and FOXP2 FHD is regulated by salt, salt (and electrostatic interactions) seemingly affects the interplay between ER1 LBD, FOXP2 FHD and DNA. Hence, the interaction between ER1 LBD and FOXP2 FHD may interfere with the functional activity of FOXP2 by hindering its ability to interact with DNA. As this finding is dependent on salt concentration and the relative concentration of ER1 and DNA, it may indicate that the PPI and consequently its effect on gene expression is cell and situation specific. As this interaction has not been previously investigated, it is unknown whether it occurs physiologically and under what conditions. This needs to be elucidated prior to seeking any physiological relevance.

In conclusion, protein-protein interactions such as these are essential for homeostasis and form an attractive yet complex avenue for targeted therapies. This research suggests a regulatory

mechanism for FOXP2's transcriptional activity *in vitro* and could potentially aid in elucidating the basis for sexual dimorphism of language and speech acquisition and development and how it may be exploited for the treatment of various neurodevelopmental disorders which are known to affect males disproportionately and more aggressively. Additionally, as both the transcription factors are pivotal in a variety of cancers, specifically breast cancers and even prostate cancer, this interaction could provide an alternate avenue for research that could have the potential of bearing fruitful results.

8. REFERENCES

Afek, A., Schipper, J.L., Horton, J., Gordân, R. and Lukatsky, D.B., 2014. Protein− DNA binding in the absence of specific base-pair recognition. *Proceedings of the National Academy of Sciences*, *111*(48), pp.17140-17145.

Allu, A.S. and Tiriveedhi, V., 2021. Cancer salt nostalgia. *Cells*, *10*(6), p.1285.

Andrade, M.A., Chacon, P., Merelo, J.J. and Morán, F., 1993. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Engineering, Design and Selection*, *6*(4), pp.383-390.

Anthis, N.J. and Clore, G.M., 2013. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Science*, *22*(6), pp.851-858.

Arao, Y. and Korach, K.S., 2018. The F domain of estrogen receptor α is involved in species-specific, tamoxifen-mediated transactivation. *Journal of Biological Chemistry*, *293*(22), pp.8495-8507.

Arun, A., Ansari, M.I., Popli, P., Jaiswal, S., Mishra, A.K., Dwivedi, A., Hajela, K. and Konwar, R., 2018. New piperidine derivative DTPEP acts as dual-acting anti-breast cancer agent by targeting ER α and downregulating PI 3K/Akt-PKC α leading to caspase-dependent apoptosis. *Cell proliferation*, *51*(6), p.e12501.

Auboeuf, D., Hönig, A., Berget, S.M. and O'Malley, B.W., 2002. Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science*, *298*(5592), pp.416-419.

Auwerx, J., Baulieu, E., Beato, M., Becker-Andre, M., Burbach, P.H., Camerino, G., Chambon, P., Cooney, A., Dejean, A., Dreyer, C. and Evans, R.M., 1999. A unified nomenclature system for the nuclear receptor superfamily. *Cell*, *97*(2), pp.161-163.

Bach, D.H., Long, N., Luu, T.T.T., Anh, N., Kwon, S. and Lee, S., 2018. The dominant role of Forkhead box proteins in cancer. *International journal of molecular sciences*, *19*(10), p.3279.

Babu, M.M., 2016. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions*, *44*(5), pp.1185-1200.

Bakker, J., De Mees, C., Douhard, Q., Balthazart, J., Gabant, P., Szpirer, J. and Szpirer, C., 2006. Alpha-fetoprotein protects the developing female mouse brain from masculinization and defeminization by estrogens. *Nature neuroscience*, *9*(2), pp.220-226.

Beato, M. and Sánchez-Pacheco, A., 1996. Interaction of steroid hormone receptors with the transcription initiation complex. *Endocrine Reviews*, *17*(6), pp.587-609.

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

Berggård, T., Linse, S. and James, P., 2007. Methods for the detection and analysis of protein–protein interactions. *Proteomics*, *7*(16), pp.2833-2842.

Bergemann, N., Parzer, P., Jaggy, S., Auler, B., Mundt, C. and Maier-Braunleder, S., 2007. Estrogen and comprehension of metaphoric speech in women suffering from schizophrenia: results of a double-blind, placebo-controlled trial. *Schizophrenia bulletin*, *34*(6), pp.1172-1181.

Bhatwa, A., Wang, W., Hassan, Y.I., Abraham, N., Li, X.Z. and Zhou, T., 2021. Challenges associated with the formation of recombinant protein inclusion bodies in Escherichia coli and strategies to address them for industrial applications. *Frontiers in Bioengineering and Biotechnology*, *9*, p.630551.

Bitan, T., Lifshitz, A., Breznitz, Z. and Booth, J.R., 2010. Bidirectional connectivity between hemispheres occurs at multiple levels in language processing but depends on sex. *Journal of Neuroscience*, *30*(35), pp.11576-11585.

Bjornstrom, L. and Sjoberg, M., 2005. Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. *Molecular endocrinology*, *19*(4), pp.833-842.

Blane, A., Dirr, H.W. and Fanucchi, S., 2018. A phosphomimetic study implicates Ser557 in regulation of FOXP2 DNA binding. *The Protein Journal*, *37*(4), pp.311-323.

Blane, A. and Fanucchi, S., 2015. Effect of pH on the Structure and DNA Binding of the FOXP2 Forkhead Domain. *Biochemistry*, *54*(25), pp.4001-4007.

Blainey, P.C., Luo, G., Kou, S.C., Mangel, W.F., Verdine, G.L., Bagchi, B. and Xie, X.S., 2009. Nonspecifically bound proteins spin while diffusing along DNA. *Nature structural & molecular biology*, *16*(12), p.1224.

Blaustein, J.D., 2004. Minireview: neuronal steroid hormone receptors: they're not just for hormones anymore. *Endocrinology*, *145*(3), pp.1075-1081.

Bogan, A.A. and Thorn, K.S., 1998. Anatomy of hot spots in protein interfaces. *Journal of molecular biology*, *280*(1), pp.1-9.

Bowers, J.M., Perez-Pouchoulen, M., Roby, C.R., Ryan, T.E. and McCarthy, M.M., 2014. Androgen modulation of Foxp1 and Foxp2 in the developing rat brain: impact on sex specific vocalization. *Endocrinology*, *155*(12), pp.4881-4894.

Brandt, M.E. and Vickery, L.E., 1997. Cooperativity and dimerization of recombinant human estrogen receptor hormone-binding domain. *Journal of Biological Chemistry*, *272*(8), pp.4843-4849.

Brzozowski, A.M., Pike, A.C., Dauter, Z., Hubbard, R.E., Bonn, T., Engström, O., Öhman, L., Greene, G.L., Gustafsson, J.Å. and Carlquist, M., 1997. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature*, *389*(6652), pp.753-758.

Carnesecchi, J., Forcet, C., Zhang, L., Tribollet, V., Barenton, B., Boudra, R., Cerutti, C., Billas, I.M., Sérandour, A.A., Carroll, J.S. and Beaudoin, C., 2017. ERRα induces H3K9 demethylation by LSD1 to promote cell invasion. *Proceedings of the National Academy of Sciences*, *114*(15), pp.3909-3914.

Carroll, J.S. and Brown, M., 2006. Estrogen receptor target gene: an evolving concept. *Molecular Endocrinology*, *20*(8), pp.1707-1714.

Chen, Y., Chen, C., Zhang, Z., Liu, C.C., Johnson, M.E., Espinoza, C.A., Edsall, L.E., Ren, B., Zhou, X.J., Grant, S.F. and Wells, A.D., 2015. DNA binding by FOXP3 domain-swapped dimer suggests mechanisms of long-range chromosomal interactions. *Nucleic acids research*, *43*(2), pp.1268-1282.

Chen, J., Sawyer, N. and Regan, L., 2013. Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Science*, *22*(4), pp.510-515.

Chen, M.T., Sun, H.F., Li, L.D., Zhao, Y., Yang, L.P., Gao, S.P. and Jin, W., 2018. Downregulation of FOXP2 promotes breast cancer migration and invasion through TGFβ/SMAD signaling pathway. *Oncology letters*, *15*(6), pp.8582-8588.

Choudhury, N. and Pettitt, B.M., 2006. Enthalpy− entropy contributions to the potential of mean force of nanoscopic hydrophobic solutes. *The Journal of Physical Chemistry B*, *110*(16), pp.8459-8463.

Chu, Y.P., Chang, C.H., Shiu, J.H., Chang, Y.T., Chen, C.Y. and Chuang, W.J., 2011. Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: Insight into its domain swapping and DNA binding. *Protein science*, *20*(5), pp.908-924.

Clark, K.L., Halay, E.D., Lai, E. and Burley, S.K., 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, *364*(6436), pp.412-420.

Collins, K.D., 1997. Charge density-dependent strength of hydration and biological structure. *Biophysical journal*, *72*(1), pp.65-76.

Collins, K.D., 2006. Ion hydration: Implications for cellular function, polyelectrolytes, and protein crystallization. *Biophysical chemistry*, *119*(3), pp.271-281.

Conte, L.L., Chothia, C. and Janin, J., 1999. The atomic structure of protein-protein recognition sites. *Journal of molecular biology*, *285*(5), pp.2177-2198.

Crespi, B., Read, S. and Hurd, P., 2017. Segregating polymorphisms of FOXP2 are associated with measures of inner speech, speech fluency and strength of handedness in a healthy population. *Brain and language*, *173*, pp.33-40.

Crider, A. and Pillai, A., 2017. Estrogen signaling as a therapeutic target in neurodevelopmental disorders. *Journal of Pharmacology and Experimental Therapeutics*, *360*(1), pp.48-58.

Cuiffo, B.G. and Karnoub, A.E., 2016. Silencing FOXP2 in breast cancer cells promotes cancer stem cell traits and metastasis. *Molecular & cellular oncology*, *3*(3), p.e1019022.

Dahlman-Wright, K., Cavailles, V., Fuqua, S.A., Jordan, V.C., Katzenellenbogen, J.A., Korach, K.S., Maggi, A., Muramatsu, M., Parker, M.G. and Gustafsson, J.Å., 2006. International union of pharmacology. LXIV. Estrogen receptors. *Pharmacological reviews*, *58*(4), pp.773-781.

David, A., Islam, S., Tankhilevich, E. and Sternberg, M.J., 2022. The AlphaFold database of protein structures: a biologist's guide. *Journal of molecular biology*, *434*(2), p.167336.

Davies, P., Watt, K., Kelly, S.M., Clark, C., Price, N.C. and McEwan, I.J., 2008. Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. *J Mol Endocrinol*, *41*(5), pp.301-14.

Desouza, J., Gadkar, S., Jagtap, D. and Sachdeva, G., 2019. Size, site, and signaling: Three attributes of estrogen receptors. *Biomedical Research Journal*, *6*(2), p.37.

Desta, I.T., Porter, K.A., Xia, B., Kozakov, D. and Vajda, S., 2020. Performance and its limits in rigid body protein-protein docking. *Structure*, *28*(9), pp.1071-1081.

Dill, K.A., 1990. Dominant forces in protein folding. *Biochemistry*, *29*(31), pp.7133-7155.

Dragan, A.I., Read, C.M. and Crane-Robinson, C., 2017. Enthalpy–entropy compensation: the role of solvation. *European Biophysics Journal*, *46*(4), pp.301-308.

Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M. and Uversky, V.N., 2005. Flexible nets: the roles of intrinsic disorder in protein interaction networks. *The FEBS journal*, *272*(20), pp.5129-5148.

Eiler, S., Gangloff, M., Duclaud, S., Moras, D. and Ruff, M., 2001. Overexpression, purification, and crystal structure of native ERα LBD. *Protein expression and purification*, *22*(2), pp.165-173.

Estruch, S.B., Graham, S.A., Chinnappa, S.M., Deriziotis, P. and Fisher, S.E., 2016. Functional characterization of rare FOXP2 variants in neurodevelopmental disorder. *Journal of neurodevelopmental disorders*, *8*(1), pp.1-19.

Evans, R.M., 1988. The steroid and thyroid hormone receptor superfamily. *Science*, *240*(4854), pp.889-895.

Farooq, A., 2015. Structural and functional diversity of estrogen receptor ligands. *Current topics in medicinal chemistry*, *15*(14), pp.1372-1384.

Featherstone, K., White, M.R.H. and Davis, J.R.E., 2012. The prolactin gene: a paradigm of tissue-specific gene regulation with complex temporal transcription dynamics. *Journal of neuroendocrinology*, *24*(7), pp.977-990.

Ferland, R.J., Cherry, T.J., Preware, P.O., Morrisey, E.E. and Walsh, C.A., 2003. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *Journal of comparative Neurology*, *460*(2), pp.266-279.

Fernández, A. and Scheraga, H.A., 2003. Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proceedings of the National Academy of Sciences*, *100*(1), pp.113-118.

Fernell, E., Gustafsson, P. and Gillberg, C., 2021. Bumetanide for autism: Open-label trial in six children. *Acta paediatrica*, *110*(5), pp.1548-1553.

Ferrero, V.E.V., Pedotti, M., Chiadò, A., Simonelli, L., Calzolai, L., Varani, L. and Lettieri, T., 2014. Rational modification of estrogen receptor by combination of computational and experimental analysis. *PLoS One*, *9*(7), p.e102658.

Freire, E., 2008. Do enthalpy and entropy distinguish first in class from best in class?. *Drug discovery today*, *13*(19-20), pp.869-874.

Fuentes, N. and Silveyra, P., 2019. Estrogen receptor signaling mechanisms. *Advances in protein chemistry and structural biology*, *116*, p.135.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Humana press.

Germain, P., Staels, B., Dacquet, C., Spedding, M. and Laudet, V., 2006. Overview of nomenclature of nuclear receptors. *Pharmacological reviews*, *58*(4), pp.685-704.

Gijsbers, A., Nishigaki, T. and Sánchez-Puig, N., 2016. Fluorescence anisotropy as a tool to study protein-protein interactions. *JoVE (Journal of Visualized Experiments)*, (116), p.e54640.

Gillies, G.E. and McArthur, S., 2010. Estrogen actions in the brain and the basis for differential action in men and women: a case for sex-specific medicines. *Pharmacological reviews*, *62*(2), pp.155-198.

Glaser, B.T., Bergendahl, V., Anthony, L.C., Olson, B. and Burgess, R.R., 2009. Studying the salt dependence of the binding of σ70 and σ32 to core RNA polymerase using luminescence resonance energy transfer. *PLoS One*, *4*(8), p.e6490.

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

Graham, S.A. and Fisher, S.E., 2013. Decoding the genetics of speech and language. *Current opinion in neurobiology*, *23*(1), pp.43-51.

Gronemeyer, H. and Moras, D., 1995. How to finger DNA. *Nature*, *375*(6528), pp.190-191.

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

Häußermann, K., Young, G., Kukura, P. and Dietz, H., 2019. Dissecting FOXP2 oligomerization and DNA binding. *Angewandte Chemie*, *131*(23), pp.7744-7749.

Herrero, M.J. and Gitton, Y., 2018. The untold stories of the speech gene, the FOXP2 cancer gene. *Genes & cancer*, *9*(1-2), p.11.

Hewitt, S.C. and Korach, K.S., 2018. Estrogen receptors: new directions in the new millennium. *Endocrine reviews*, *39*(5), pp.664-675.

Honorato, R.V., Koukos, P.I., Jiménez-García, B., Tsaregorodtsev, A., Verlato, M., Giachetti, A., Rosato, A. and Bonvin, A.M., 2021. Structural biology in the clouds: the WeNMR-EOSC ecosystem. *Frontiers in Molecular Biosciences*, p.708.

Huo, H., Ye, X., Yang, H., Li, Q. and Jiang, Y., 2019. RSK4 inhibits breast cancer cell proliferation and invasion in vitro, and is correlated with estrogen receptor upregulation in breast cancer. *Oncology Reports*, *42*(6), pp.2777-2787.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D. and Carroll, J.S., FOXA1 is a critical determinant of Estrogen Receptor function and endocrine response.

Ignar-Trowbridge, D.M., Nelson, K.G., Bidwell, M.C., Curtis, S.W., Washburn, T.F., McLachlan, J.A. and Korach, K.S., 1992. Coupling of dual signaling pathways: epidermal growth factor action involves the estrogen receptor. *Proceedings of the National Academy of Sciences*, *89*(10), pp.4658-4662.

Ito, I., Hanyu, A., Wayama, M., Goto, N., Katsuno, Y., Kawasaki, S., Nakajima, Y., Kajiro, M., Komatsu, Y., Fujimura, A. and Hirota, R., 2010. Estrogen inhibits transforming growth factor β signaling by promoting Smad2/3 degradation. *Journal of biological chemistry*, *285*(19), pp.14747-14755.

Jackson, B.C., Carpenter, C., Nebert, D.W. and Vasiliou, V., 2010. Update of human and mouse forkhead box (FOX) gene families. *Human genomics*, *4*(5), p.345.

Jansen, R., Greenbaum, D. and Gerstein, M., 2002. Relating whole-genome expression data with protein-protein interactions. *Genome research*, *12*(1), pp.37-46.

Jennings, P.J., Janowsky, J.S. and Orwoll, E., 1998. Estrogen and sequential movement. *Behavioral neuroscience*, *112*(1), p.154.

Jiménez-García, B., Teixeira, J.M., Trellet, M., Rodrigues, J.P. and Bonvin, A.M., 2021. PDB-tools web: A user-friendly interface for the manipulation of PDB files. *Proteins: Structure, Function, and Bioinformatics*, *89*(3), pp.330-335.

Journet, L. and Cascales, E. eds., 2017. *Bacterial Protein Secretion Systems: Methods and Protocols*. Humana Press.

Kastritis, P.L., Moal, I.H., Hwang, H., Weng, Z., Bates, P.A., Bonvin, A.M. and Janin, J., 2011. A structure-based benchmark for protein–protein binding affinity. *Protein Science*, *20*(3), pp.482-491.

Katzenellenbogen, B.S. and Katzenellenbogen, J.A., 2000. Estrogen receptor transcription and transactivation Estrogen receptor alpha and estrogen receptor beta: regulation by selective estrogen receptor modulators and importance in breast cancer. *Breast Cancer Research*, *2*(5), p.335.

Kawai, H., Li, H., Chun, P., Avraham, S. and Avraham, H.K., 2002. Direct interaction between BRCA1 and the estrogen receptor regulates vascular endothelial growth factor (VEGF) transcription and secretion in breast cancer cells. *Oncogene*, *21*(50), pp.7730-7739.

Keskin, O., Ma, B. and Nussinov, R., 2005. Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. *Journal of molecular biology*, *345*(5), pp.1281-1294.

Keskin, O., Tuncbag, N. and Gursoy, A., 2016. Predicting protein–protein interactions from the molecular to the proteome level. *Chemical reviews*, *116*(8), pp.4884-4909.

Kišonaitė, M., Zubrienė, A., Čapkauskaitė, E., Smirnov, A., Smirnovienė, J., Kairys, V., Michailovienė, V., Manakova, E., Gražulis, S. and Matulis, D., 2014. Intrinsic thermodynamics and structure correlation of benzenesulfonamides with a pyrimidine moiety binding to carbonic anhydrases I, II, VII, XII, and XIII. *PLoS One*, *9*(12), p.e114106.

Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D. and Vajda, S., 2017. The ClusPro web server for protein–protein docking. *Nature protocols*, *12*(2), pp.255-278.

Kulkarni, J., Gavrilidis, E., Worsley, R., Van Rheenen, T. and Hayes, E., 2013. The role of estrogen in the treatment of men with schizophrenia. *International journal of endocrinology and metabolism*, *11*(3), p.129.

L'Horset, F., Dauvois, S., Heery, D.M., Cavaillès, V. and Parker, M.G., 1996. RIP-140 interacts with multiple nuclear receptors by means of two distinct sites. *Molecular and Cellular Biology*, *16*(11), pp.6029-6036.

Ladbury, J.E. and Chowdhry, B.Z., 1996. Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chemistry & biology*, *3*(10), pp.791-801.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature*, *227*(5259), pp.680-685.

Lai, C.S., Gerrelli, D., Monaco, A.P., Fisher, S.E. and Copp, A.J., 2003. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain*, *126*(11), pp.2455-2462.

Lakowicz, J.R., 2006. Principles of Fluorescence Spectroscopy: Springer Science. *New York*, pp.158-204.

Le Romancer, M., Poulard, C., Cohen, P., Sentis, S., Renoir, J.M. and Corbo, L., 2011. Cracking the estrogen receptor's posttranslational code in breast tumors. *Endocrine reviews*, *32*(5), pp.597-622.

Lee, H.R., Kim, T.H. and Choi, K.C., 2012. Functions and physiological roles of two types of estrogen receptors, ERα and ERβ, identified by estrogen receptor knockout mouse. *Laboratory animal research*, *28*(2), pp.71-76.

Lee, S. and Lee, D.K., 2018. What is the proper way to apply the multiple comparison test?. *Korean journal of anesthesiology*, *71*(5), pp.353-360.

Lensink, M.F., Nadzirin, N., Velankar, S. and Wodak, S.J., 2020. Modeling protein-protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition. *Proteins: Structure, Function, and Bioinformatics*, *88*(8), pp.916-938.

Lepp, S., Anderson, A. and Konopka, G., 2013. Connecting signaling pathways underlying communication to ASD vulnerability. In *International review of neurobiology* (Vol. 113, pp. 97-133). Academic Press.

Leslie, T.K., James, A.D., Zaccagna, F., Grist, J.T., Deen, S., Kennerley, A., Riemer, F., Kaggie, J.D., Gallagher, F.A., Gilbert, F.J. and Brackenbury, W.J., 2019. Sodium homeostasis in the tumour microenvironment. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *1872*(2), p.188304.

Leung, H.W., Wang, Z., Yue, G.G.L., Zhao, S.M., Lee, J.K.M., Fung, K.P., Leung, P.C., Bik-San Lau, C. and Tan, N.H., 2015. Cyclopeptide RA-V inhibits cell adhesion and invasion in both estrogen receptor positive and negative breast cancer cells via PI3K/AKT and NF-κB signaling pathways. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1853*(8), pp.1827-1840.

Li, Q., Jiang, Y., Wei, W., Ji, Y., Gao, H. and Liu, J., 2014. Frequent epigenetic inactivation of RSK4 by promoter methylation in cancerous and non-cancerous tissues of breast cancer. *Medical oncology*, *31*(1), pp.1-8.

Li, H., Xue, Y., Ma, J., Shao, L., Wang, D., Zheng, J., Liu, X., Yang, C., He, Q., Ruan, X. and Li, Z., 2019. SNHG1 promotes malignant biological behaviors of glioma cells via microRNA-154-

5p/miR-376b-3p-FOXP2-KDM5B participating positive feedback loop. *Journal of Experimental & Clinical Cancer Research*, *38*(1), p.59.

Lischinsky, J.E., Sokolowski, K., Li, P., Esumi, S., Kamal, Y., Goodrich, M., Oboti, L., Hammond, T.R., Krishnamoorthy, M., Feldman, D. and Huntsman, M., 2017. Embryonic transcription factor expression in mice predicts medial amygdala neuronal identity and sex-specific responses to innate behavioral cues. *Elife*, *6*, p.e21012.

Liu, Y., Ao, X., Jia, Z., Bai, X.Y., Xu, Z., Hu, G., Jiang, X., Chen, M. and Wu, H., 2015. FOXK2 transcription factor suppresses ERα-positive breast cancer cell growth through down-regulating the stability of ERα via mechanism involving BRCA1/BARD1. *Scientific reports*, *5*(1), pp.1-12.

Liu, Y., Ma, H. and Yao, J., 2020. ERα, a key target for cancer therapy: A review. *OncoTargets and therapy*, *13*, p.2183.

Maekawa, F., Tsukahara, S., Kawashima, T., Nohara, K. and Ohki-Hamazaki, H., 2014. The mechanisms underlying sexual differentiation of behavior and physiology in mammals and birds: relative contributions of sex steroids and sex chromosomes. *Frontiers in neuroscience*, *8*, p.242.

Maggi, A., 2011. Liganded and unliganded activation of estrogen receptor and hormone replacement therapies. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1812*(8), pp.1054-1060.

Mak, H.Y., Hoare, S., Henttu, P.M. and Parker, M.G., 1999. Molecular determinants of the estrogen receptor-coactivator interface. *Molecular and Cellular Biology*, *19*(5), pp.3895-3903.

Marino, M., Galluzzo, P. and Ascenzi, P., 2006. Estrogen signaling multiple pathways to impact gene transcription. *Current genomics*, *7*(8), pp.497-508.

Mazaira, G.I., Zgajnar, N.R., Lotufo, C.M., Daneri-Becerra, C., Sivils, J.C., Soto, O.B., Cox, M.B. and Galigniana, M.D., 2018. The nuclear receptor field: A historical overview and future challenges. *Nuclear receptor research*, *5*.

McEwan, I.J., Claessens, F. and Gewirth, D.T., 2004. DNA recognition by nuclear receptors. *Essays in biochemistry*, *40*, pp.59-72.

Mei, Y., Liao, X., Zhu, L. and Yang, H., 2020. Overexpression of RSK4 reverses doxorubicin resistance in human breast cancer cells via PI3K/AKT signalling pathway. *The Journal of Biochemistry*, *167*(6), pp.603-611.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M., 2022. ColabFold: making protein folding accessible to all. *Nature Methods*, pp.1-4.

Moreira, I.S., Fernandes, P.A. and Ramos, M.J., 2007. Hot spots—A review of the protein–protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics*, *68*(4), pp.803-812.

Morris, G. and Fanucchi, S., 2016. A key evolutionary mutation enhances DNA binding of the FOXP2 forkhead domain. *Biochemistry*, *55*(13), pp.1959-1967.

Mozzi, A., Riva, V., Forni, D., Sironi, M., Marino, C., Molteni, M., Riva, S., Guerini, F.R., Clerici, M., Cagliani, R. and Mascheretti, S., 2017. A common genetic variant in FOXP2 is associated with language-based learning (dis) abilities: Evidence from two Italian independent samples. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *174*(5), pp.578-586.

Myatt, S.S. and Lam, E.W.F., 2007. The emerging roles of forkhead box (Fox) proteins in cancer. *Nature Reviews Cancer*, *7*(11), pp.847-859.

Nagy, I.Z., Lustyik, G., Nagy, V.Z., Zarandi, B. and Bertoni-Freddari, C., 1981. Intracellular Na+: K+ ratios in human cancer cells as revealed by energy dispersive x-ray microanalysis. *The Journal of cell biology*, *90*(3), pp.769-777.

Nakagawa, S., Gisselbrecht, S.S., Rogers, J.M., Hartl, D.L. and Bulyk, M.L., 2013. DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proceedings of the National Academy of Sciences*, *110*(30), pp.12349-12354.

Needleman, S.B. and Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, *48*(3), pp.443-453.

Nelson, C.S., Fuller, C.K., Fordyce, P.M., Greninger, A.L., Li, H. and DeRisi, J.L., 2013. Microfluidic affinity and ChIP-seq analyses converge on a conserved FOXP2-binding motif in

chimp and human, which enables the detection of evolutionarily novel targets. *Nucleic acids research*, *41*(12), pp.5991-6004.

Nooren, I.M. and Thornton, J.M., 2003. Diversity of protein–protein interactions. *The EMBO journal*, *22*(14), pp.3486-3492.

Olson, B.J. and Markwell, J., 2007. Assays for determination of protein concentration. *Current Protocols in Pharmacology*, *38*(1), pp.A-3A.

Oñate, S.A., Tsai, S.Y., Tsai, M.J. and O'Malley, B.W., 1995. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science*, *270*(5240), pp.1354-1357.

Oswald, F., Klöble, P., Ruland, A., Rosenkranz, D., Hinz, B., Butter, F., Ramljak, S., Zechner, U. and Herlyn, H., 2017. The FOXP2-driven network in developmental disorders and neurodegeneration. *Frontiers in cellular neuroscience*, *11*, p.212.

Patel, S., Homaei, A., Raju, A.B. and Meher, B.R., 2018. Estrogen: the necessary evil for human health, and ways to tame it. *Biomedicine & Pharmacotherapy*, *102*, pp.403-411.

Pinel, P., Fauchereau, F., Moreno, A., Barbot, A., Lathrop, M., Zelenika, D., Le Bihan, D., Poline, J.B., Bourgeron, T. and Dehaene, S., 2012. Genetic variants of FOXP2 and KIAA0319/TTRAP/THEM2 locus are associated with altered brain activation in distinct language-related regions. *Journal of Neuroscience*, *32*(3), pp.817-825.

Plevin, M.J., Mills, M.M. and Ikura, M., 2005. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends in biochemical sciences*, *30*(2), pp.66-69.

Pratt, W.B. and Toft, D.O., 1997. Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocrine reviews*, *18*(3), pp.306-360.

Rao, V.S., Srinivas, K., Sujini, G.N. and Kumar, G.N., 2014. Protein-protein interaction detection: methods and analysis. *International journal of proteomics*, *2014*.

Ratanasumawong, S., Hagiwara, T. and Sakiyama, T., 2015. Effect of sodium chloride on the adsorption of proteins from pink shrimp (Pandalus eous) onto stainless steel surfaces. *Food Science and Technology Research*, *21*(3), pp.327-331.

Razzaghi-Asl, N. and Ebadi, A., 2021. In silico design of peptide inhibitors of tubulin: amyloid-β as a lead compound. *Journal of Biomolecular Structure and Dynamics*, *39*(6), pp.2189-2198.

Record Jr, M.T., Lohman, T.M. and De Haseth, P., 1976. Ion effects on ligand-nucleic acid interactions. *Journal of molecular biology*, *107*(2), pp.145-158.

Reuter, M.S., Riess, A., Moog, U., Briggs, T.A., Chandler, K.E., Rauch, A., Stampfer, M., Steindl, K., Gläser, D., Joset, P. and Krumbiegel, M., 2017. FOXP2 variants in 14 individuals with developmental speech and language disorders broaden the mutational and clinical spectrum. *Journal of medical genetics*, *54*(1), pp.64-72.

Rhee, D.Y., Cho, D.Y., Zhai, B., Slattery, M., Ma, L., Mintseris, J., Wong, C.Y., White, K.P., Celniker, S.E., Przytycka, T.M. and Gygi, S.P., 2014. Transcription factor networks in Drosophila melanogaster. *Cell reports*, *8*(6), pp.2031-2043.

Rosenfeld, M.G., Lunyak, V.V. and Glass, C.K., 2006. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes & development*, *20*(11), pp.1405-1428.

Ruff, M., Gangloff, M., Wurtz, J.M. and Moras, D., 2000. Estrogen receptor transcription and transactivation structure-function relationship in DNA-and ligand-binding domains of estrogen receptors. *Breast Cancer Research*, *2*(5), p.353.

Sanders, D.A., Ross-Innes, C.S., Beraldi, D., Carroll, J.S. and Balasubramanian, S., 2013. Genome-wide mapping of FOXM1 binding reveals co-binding with estrogen receptor alpha in breast cancer cells. *Genome biology*, *14*(1), p.R6.

Sanjuán, J., Tolosa, A., González, J.C., Aguilar, E.J., Pérez-Tur, J., Nájera, C., Moltó, M.D. and de Frutos, R., 2006. Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations. *Psychiatric genetics*, *16*(2), pp.67-72.

Sarmady, M., Dampier, W. and Tozeren, A., 2011. HIV protein sequence hotspots for crosstalk with host hub proteins. *PLoS One*, *6*(8).

Sasson, S. and Notides, A.C., 1983. Estriol and estrone interaction with the estrogen receptor. II. Estriol and estrone-induced inhibition of the cooperative binding of [3H] estradiol to the estrogen receptor. *Journal of Biological Chemistry*, *258*(13), pp.8118-8122.

Scopes, R.K., 1974. Measurement of protein by spectrophotometry at 205 nm. *Analytical biochemistry*, *59*(1), pp.277-282.

Seielstad, D.A., Carlson, K.E., Katzenellenbogen, J.A., Kushner, P.J. and Greene, G.L., 1995. Molecular characterization by mass spectrometry of the human estrogen receptor ligand-binding domain expressed in Escherichia coli. *Molecular Endocrinology*, *9*(6), pp.647-658.

Sharp, K.A., Nicholls, A., Friedman, R. and Honig, B., 1991. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry*, *30*(40), pp.9686-9697.

Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A. and Greene, G.L., 1998. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*, *95*(7), pp.927-937.

Shigekawa, T., Ijichi, N., Ikeda, K., Horie-Inoue, K., Shimizu, C., Saji, S., Aogi, K., Tsuda, H., Osaki, A., Saeki, T. and Inoue, S., 2011. FOXP1, an estrogen-inducible transcription factor, modulates cell proliferation in breast cancer cells and 5-year recurrence-free survival of patients with tamoxifen-treated breast cancer. *Hormones and Cancer*, *2*(5), p.286.

Shu, W., Lu, M.M., Zhang, Y., Tucker, P.W., Zhou, D. and Morrisey, E.E., 2007. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development*, *134*(10), pp.1991-2000.

Shu, W., Yang, H., Zhang, L., Lu, M.M. and Morrisey, E.E., 2001. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. *Journal of Biological Chemistry*, *276*(29), pp.27488-27497.

Smutny, T., Mani, S. and Pavek, P., 2013. Post-translational and post-transcriptional modifications of pregnane X receptor (PXR) in regulation of the cytochrome P450 superfamily. *Current drug metabolism*, *14*(10), pp.1059-1069.

Stope, M.B., Popp, S.L., Knabbe, C. and Buck, M.B., 2010. Estrogen receptor α attenuates transforming growth factor-β signaling in breast cancer cells independent from agonistic and antagonistic ligands. *Breast cancer research and treatment*, *120*(2), pp.357-367.

Stroud, J.C., Wu, Y., Bates, D.L., Han, A., Nowick, K., Paabo, S., Tong, H. and Chen, L., 2006. Structure of the forkhead domain of FOXP2 bound to DNA. *Structure*, *14*(1), pp.159-166.

Sutherland, B.W., Toews, J. and Kast, J., 2008. Utility of formaldehyde cross-linking and mass spectrometry in the study of protein–protein interactions. *Journal of mass spectrometry*, *43*(6), pp.699-715.

Tanenbaum, D.M., Wang, Y., Williams, S.P. and Sigler, P.B., 1998. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proceedings of the National Academy of Sciences*, *95*(11), pp.5998-6003.

Tellinghuisen, J., 2008. Isothermal titration calorimetry at very low c. *Analytical biochemistry*, *373*(2), pp.395-397.

Thomas, M.C. and Chiang, C.M., 2006. The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology*, *41*(3), pp.105-178.

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

Totzeck, F., Andrade-Navarro, M.A. and Mier, P., 2017. The protein structure context of polyQ regions. *PloS one*, *12*(1), p.e0170801.

Tse, C., Wickstrom, L., Kvaratskhelia, M., Gallicchio, E., Levy, R. and Deng, N., 2020. Exploring the free-energy landscape and thermodynamics of protein-protein association. *Biophysical journal*, *119*(6), pp.1226-1238.

van Gils, J.H.M., Gogishvili, D., van Eck, J., Bouwmeester, R., van Dijk, E. and Abeln, S., 2022. How sticky are our proteins? Quantifying hydrophobicity of the human proteome. *Bioinformatics Advances*, *2*(1), p.vbac002.

Van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., De Vries, S.J. and Bonvin, A.M.J.J., 2016. The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. *Journal of molecular biology*, *428*(4), pp.720-725.

Vernes, S.C., Newbury, D.F., Abrahams, B.S., Winchester, L., Nicod, J., Groszer, M., Alarcón, M., Oliver, P.L., Davies, K.E., Geschwind, D.H. and Monaco, A.P., 2008. A functional genetic

link between distinct developmental language disorders. *New England Journal of Medicine*, *359*(22), pp.2337-2345.

Vernes, S.C., Nicod, J., Elahi, F.M., Coventry, J.A., Kenny, N., Coupe, A.M., Bird, L.E., Davies, K.E. and Fisher, S.E., 2006. Functional genetic analysis of mutations implicated in a human speech and language disorder. *Human molecular genetics*, *15*(21), pp.3154-3167.

Von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P., 2002. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, *417*(6887), pp.399-403.

Vreven, T., Moal, I.H., Vangone, A., Pierce, B.G., Kastritis, P.L., Torchala, M., Chaleil, R., Jiménez-García, B., Bates, P.A., Fernandez-Recio, J. and Bonvin, A.M., 2015. Updates to the integrated protein–protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. *Journal of molecular biology*, *427*(19), pp.3031-3041.

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. and Fisher, S.E., 2012. The DISC1 promoter: characterization and regulation by FOXP2. *Human molecular genetics*, *21*(13), pp.2862-2872.

Wallentin, M., 2009. Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and language*, *108*(3), pp.175-183.

Wallerstein, J., Ekberg, V., Ignjatović, M.M., Kumar, R., Caldararu, O., Peterson, K., Wernersson, S., Brath, U., Leffler, H., Oksanen, E. and Logan, D.T., 2021. Entropy–Entropy Compensation between the Protein, Ligand, and Solvent Degrees of Freedom Fine-Tunes Affinity in Ligand Binding to Galectin-3C. *JACS Au*, *1*(4), pp.484-500.

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

Wang, H., He, M., Willard, B. and Wu, Q., 2019. Cross-linking, immunoprecipitation and proteomic analysis to identify interacting proteins in cultured cells. *Bio-protocol*, *9*(11), pp.e3258-e3258.

Wärnmark, A., Treuter, E., Gustafsson, J.Å., Hubbard, R.E., Brzozowski, A.M. and Pike, A.C., 2002. Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor α. *Journal of Biological Chemistry*, *277*(24), pp.21862-21868.

Webb, H., Steeb, O., Blane, A., Rotherham, L., Aron, S., Machanick, P., Dirr, H. and Fanucchi, S., 2017. The FOXP2 forkhead domain binds to a variety of DNA sequences with different rates and affinities. *The Journal of Biochemistry*, *162*(1), pp.45-54.

Weigelt, J., Climent, I., Dahlman-Wright, K. and Wikström, M., 2001. Solution structure of the DNA binding domain of the human forkhead transcription factor AFX (FOXO4). *Biochemistry*, *40*(20), pp.5861-5869.

Weikum, E.R., Liu, X. and Ortlund, E.A., 2018. The nuclear receptor superfamily: A structural perspective. *Protein Science*, *27*(11), pp.1876-1892.

Whitmore, L. and Wallace, B.A., 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic acids research*, *32*(suppl_2), pp.W668-W673.

Wohlgemuth, S., Adam, I. and Scharff, C., 2014. FoxP2 in songbirds. *Current opinion in neurobiology*, *28*, pp.86-93.

Wurtz, J.M., Bourguet, W., Renaud, J.P., Vivat, V., Chambon, P., Moras, D. and Gronemeyer, H., 1996. A canonical structure for the ligand-binding domain of nuclear receptors. *Nature structural biology*, *3*(1), pp.87-94.

Yang, F., Xiao, Z. and Zhang, S., 2022. FOXP2 regulates thyroid cancer cell proliferation and apoptosis via transcriptional activation of RPS6KA6. *Experimental and Therapeutic Medicine*, *23*(6), pp.1-10.

Yaşar, P., Ayaz, G., User, S.D., Güpür, G. and Muyan, M., 2017. Molecular mechanism of estrogen–estrogen receptor signaling. *Reproductive medicine and biology*, *16*(1), pp.4-20.

Zhang, Y.L., Sun, F.T., Zhang, Z., Chen, X.X., Liu, A.X., Pan, J.J., Peng, F., Zhou, S. and Sun, L.J., 2014. Comprehensive expression analysis suggests functional overlapping of human FOX transcription factors in cancer. *Asian Pac J Cancer Prev*, *15*(23), pp.10475-81.

Zou, Y., Tsai, W.B., Cheng, C.J., Hsu, C., Chung, Y.M., Li, P.C., Lin, S.H. and Hu, M.C., 2008. Forkhead box transcription factor FOXO3a suppresses estrogen-dependent breast cancer cell proliferation and tumorigenesis. *Breast cancer research*, *10*(1), p.R21.

9) APPENDIX

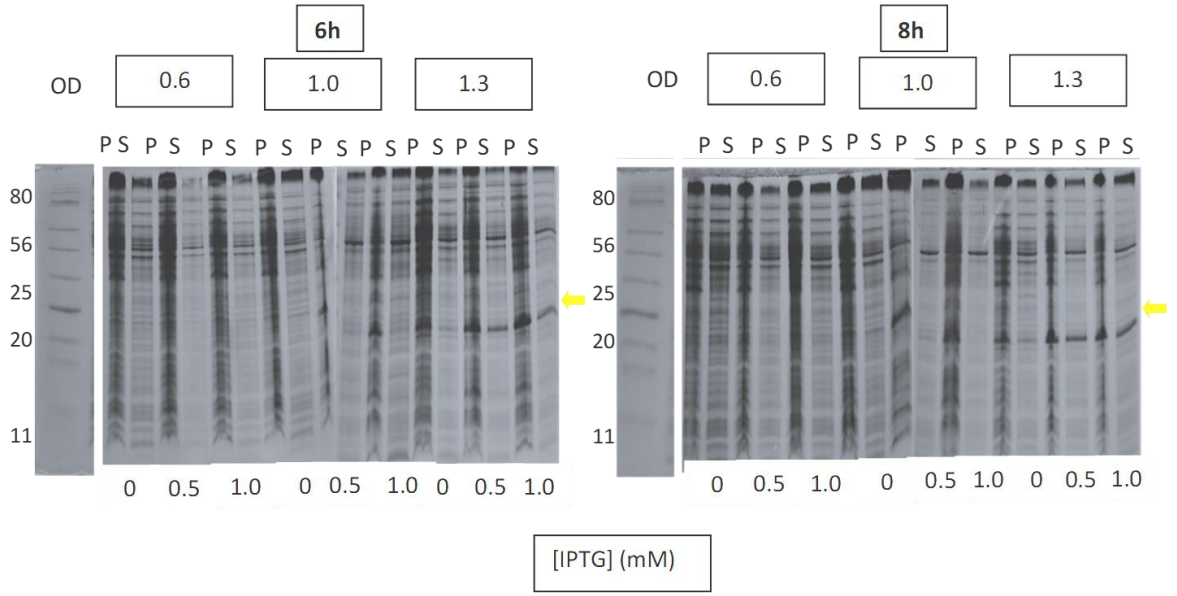


Figure S1: Induction trials of FOXP2 NT at different optical density. FOXP2 NT in T7 express cells were grown to different OD in 2xYT media. The OD was measured at 600 nm using the Jasco V-630 spectrophotometer. Thereafter, cells were induced with different IPTG concentrations (0-1 mM), and the protein was expressed at 20 °C, 150 rpm. Following cell harvest and lysis, and separation of pellet (P) and supernatant (S), the samples were evaluated on a 16 % tricine SDS-PAGE gel. Yellow arrow indicates position of the FOXP2 NT band.

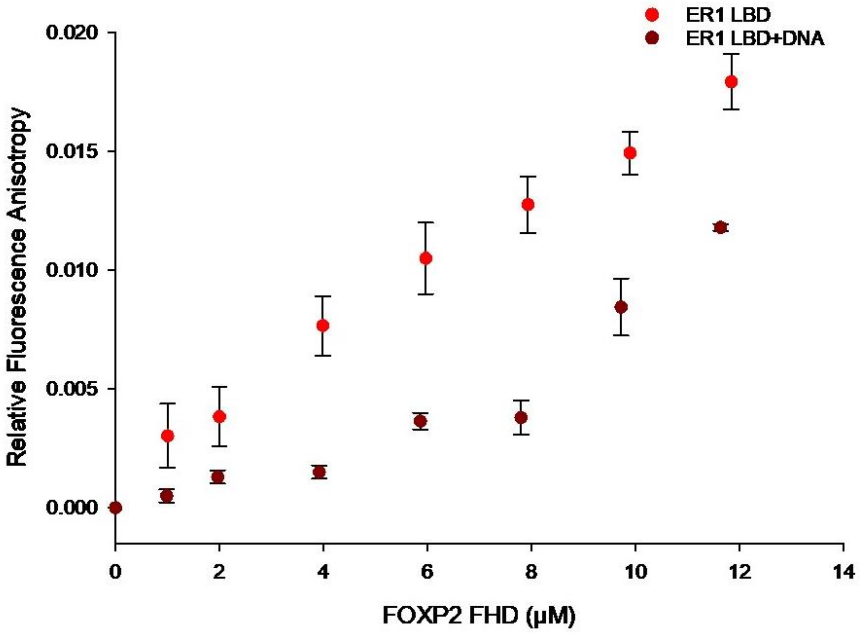


Figure S2: ER1-FOXP2 interaction in the absence and presence of DNA. Fluorescently labelled ER1 LBD (0.3 μM) samples were prepared in 50 mM Tris, 100 mM NaCl, 1 mM TCEP, pH7.8. Fluorescence anisotropy was performed by titrating labelled ER1 LBD by itself or in the presence of Nelson DNA (7 μM) with FOXP2 FHD. In the absence of background DNA, FOXP2 FHD binds to ER1 LBD in almost a linear fashion. Conversely, the fluorescence anisotropy of ER1 LBD with background Nelson DNA only begins to increase after at least 8 μM of FOXP2 FHD has been titrated. This may indicate that the FOXP2 FHD chooses to initially bind DNA and only begins to interact with ER1 LBD when the titrated FOXP2 FHD concentration is close to or exceeding the background DNA concentration.