# Combined approaches to map protein-ligand interactions using NMR and X-ray crystallography

Erin Shamini Anthonyrajah

PhD Biochemistry 2018

**A thesis re-submitted to the University of Kent for the degree of PhD in Biochemistry at the School of Biosciences, Faculty of Science, Technology and Medicine.**



# DECLARATION

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or Institute of learning.

Erin Shamini Anthonyrajah

November 2018

# A. Acknowledgements

Firstly, I would like to thank my supervisor Professor David G. Brown for all his help, guidance, support, time and patience throughout my PhD and thesis writing. I would like to express my sincere words of gratitude to Dr. Gary Thompson for helping me to understand the complexities of NMR and being very supportive in the analyses of my data. He was an excellent teacher and has supported me so well in the process of re-submission.

I would also like to thank Diamond (Oxfordshire) for all the technical help and support provided during crystallographic data collection, as well as Mr. Kevin Howland of the protein research facility for help with mass spectrometry with fluorescence analysis.

Special thanks go to past and present members of the Brown group, especially Dr. Susanne Schroeder and Dr. Abhimanyu Singh, for continuous support, encouragement and endless science- related conversations. I also wish to thank the members from Charles River Early Discovery, particularly Richard Bazin, Colin Robinson and Dr Steve Irving, for all the valuable help and advice.

All my love and gratitude go to my parents for their continued support and guidance throughout my life and studies. I would not have come this far in life without them! Finally, my deepest thanks go to my beloved husband Deny and my son Andrew for all the love, motivation and emotional support that have kept me going over the years, to whom I dedicate this work.

# B. Contents

**A. Acknowledgements .....3**

**B. Contents .....4**

**C. Abbreviations .....13**

**D. Abstract .....18**

**Chapter 1: Introduction ..... 20**

**1.1 Proteins**

1.1.1 Structure is related to function ..... 22

**1.2 Protein binding sites and interactions with ligands**

**1.3 Protein and Ligand Binding Affinity Studies**

1.3.1 Techniques used for determining binding affinities ..... 29

1.3.2 The use of protein binding sites into drug design ..... 30

**1.4 The determination of protein structure using X-ray crystallography**

1.4.1 The process of protein X-ray crystallography ..... 32

1.4.2 Crystal Definitions ..... 37

1.4.3 X-ray diffraction of protein crystal ..... 39

1.4.4 X-ray data acquisition and processing ..... 41

1.4.5 Molecular Replacement and Refinement ..... 44

1.4.6 The limitations of X-ray crystallography ..... 46

**1.5 Nuclear Magnetic Resonance**

1.5.1 Principle ..... 47

1.5.2 Saturation-Transfer Difference NMR ..... 48

1.5.3 Advanced NMR Screening Methods ..... 52

1.5.4 The development of fragment libraries ..... 52

**1.6 Other biophysical techniques in protein structure determination**

1.6.1 Biological Electron Microscopy ..... 54

1.6.2 Mass Spectrometry and Protein Mass Mapping ..... 55

1.6.3 Small Angle X-ray Scattering (SAXS) ..... 55

**1.7 Aims and Hypotheses**

**Chapter 2: Materials and Methods ..... 57**

**2.1 Materials**

2.1.1 Chemicals ..... 58

2.1.2 Bacterial Plasmids ..... 59

2.1.3 Media and solutions for bacterial work ..... 60

2.1.4 Media and solutions for DNA work ..... 63

2.1.5 Solutions for protein work ..... 65

2.1.5.1 Solutions for Nickel Affinity Chromatography ..... 65

2.1.5.2 Solutions for Desalting ..... 67

2.1.5.3 Solutions for ion-exchange chromatography ..... 67

2.1.5.4 Solutions for gel filtration chromatography ..... 68

2.1.5.5 Solutions for SDS gels ..... 69

2.1.6 Media and solutions for X-ray crystallography work ..... 70

2.1.7 Ligands ..... 72

2.1.7.1 Ligands for Cyclophilin A ..... 72

2.1.7.2 Ligands for Phosphodiesterase B1 ..... 73

2.1.7.3 Ligands for Bromodomain 4 ..... 74

## 2.2 Microbiological Methodology

2.2.1 Sterilisation Methods ..... 76

2.2.2 Preparation of *E. coli* Competent Cells ..... 76

2.2.3 Transformation of *E. coli* Competent Cells ..... 77

2.2.4 Preparation of glycerol stocks ..... 77

2.2.5 Inoculation of LB starter cultures from glycerol stocks ..... 77

2.2.6 Isolation of plasmid DNA ..... 78

## 2.3 Recombinant Protein Overproduction

2.3.1 Production of recombinant protein of a pET in T7 express competent cells in enriched culture medium ..... 78

2.3.2 Production of recombinant protein of a pET in T7 express competent cells in minimal medium for the purpose of isotopic labelling ..... 79

## 2.4 Molecular Biology Methodology

2.4.1 Design of primers ..... 80

2.4.2 Polymerase-Chain Reaction (PCR) ..... 81

2.4.3 DNA Gel Electrophoresis ..... 82

2.4.3.1 DNA Agarose Gel ..... 82

2.4.3.2 Visualisation of DNA ..... 82

2.4.4 Extraction and purification of DNA fragments from agarose gels ..... 82

2.4.5 Preparation of restriction digests ..... 83

2.4.6 DNA Ligation ..... 83

2.4.7 Mini Prep ..... 84

2.4.8 Identification of the colony containing both vector and insert DNA ..... 85

2.4.8.1 Digestion by Insert ..... 85

2.4.8.2 Sequencing ..... 86

## 2.5 Biochemical Methodology

2.5.1 Protein Purification Methods ..... 86

2.5.1.1 Immobilised Nickel Affinity Chromatography (IMAC) ..... 87

2.5.1.2 Buffer-Exchange columns ..... 87

2.5.1.3 Ion-Exchange Chromatography (IEC) ..... 88

2.5.1.4 Heparin Chromatography ..... 89

2.5.1.5 Gel Filtration Chromatography ..... 89

2.5.1.6 Affinity tag removal ..... 90

2.5.2 Purity of protein samples ..... 90

2.5.3 Crystallisation ..... 92

2.5.3.1 Hanging Drop Crystallisation ..... 92

2.5.3.2 Sitting Drop Crystallisation ..... 93

2.5.3.3 Soaking of protein crystals with ligands ..... 93

2.5.3.4 Co-crystallisation of protein with ligands ..... 94

2.5.3.5 Crystal Archiving ..... 94

2.5.3.6 Data Collection and Processing ..... 95

2.5.4 Nuclear Magnetic Resonance (NMR) ..... 95

2.5.5 Mass Spectrometry (MS) and Circular Dichroism (CD) ..... 96

**Chapter 3: The Determination of the Structure of CypA and Ligand Binding Sites using X-ray Crystallography and NMR ..... 97**

## 3.1 Introduction about Cyclophilin A (CypA)

3.1.1 Functional Studies of CypA ..... 99

3.1.2 Structure of CypA ..... 100

3.1.3 CypA as drug target ..... 101

**3.2 Recombinant expression and purification of CypA from *E.coli* and crystallisation**

3.2.1 Recombinant expression and purification ..... 104

3.2.2 Protein Crystallisation ..... 107

**3.3 Crystallographic Solutions of CypA**

3.3.1 Addition of Ligands ..... 108

3.3.2 Data collection and processing ..... 109

3.3.3 Crystallographic Data, Refinement and Ramachandran Plot Statistics with ligands ..... 110

**3.4 Protein Crystallisation Optimisation**

3.4.1 Crystallisation Conditions ..... 116

3.4.2 Micro-seeding experiments ..... 117

3.4.3 Additive Screen Experiments ..... 117

3.4.4 Concentration of the protein sample ..... 118

**3.5 Mutagenesis of Cyclophilin A**

3.5.1 Sequence of wild-type CypA ..... 119

3.5.2 Directed mutagenesis of wild-type CypA ..... 120

3.5.3 Recombinant Expression and Purification of the mutant CypA ..... 120

3.5.4 Mass spectrometry (MS) of mutant CypA ..... 124

3.5.5 Crystallisation of the mutant CypA ..... 125

**3.6 NMR screening methods on protein-ligand binding with CypA**

3.6.1 NMR Sample Preparation of CypA/3′ aminoacetophenone and CypA/4′ sulfamoylbenzoic acid complexes ..... 127

3.6.2 NMR Sample STD Data Acquisition ..... 127

3.6.3 An Introduction to Waterlogsy ..... 130

3.6.3.1 NMR Sample Waterlogsy Data Acquisition ..... 130

3.6.4 An Introduction to Carr-Purcell- Meiboom Gill (CPMG) ..... 133

3.6.4.1 NMR Sample Carr-Purcell- Meiboom Gill (CPMG) Data Acquisition ..... 133

**3.7 Characterisation of CypA by NMR Spectroscopy**

3.7.1 An Introduction to $^{15}N/^{1}H$ heteronuclear NMR experiments ..... 136

3.7.2 NMR Sample Preparation of $^{15}N/^{1}H$ CypA ..... 137

3.7.3 $^{15}N/^{1}H$ HSQC NMR Data Acquisition with CypA ..... 137

**3.8 Summary and Discussion .....**

**Chapter 4: The Determination of the Structure of TbrPDEB1 and Ligand Binding Sites using X-ray Crystallography and NMR……..141**

**4.1 Introduction about Phosphodiesterase B1 in *Trypanosoma brucei* (TbrPDEB1)**

4.1.1 Functional Studies of TbrPDEB1 ..... 142

4.1.2 Structure of TbrPDEB1 ..... 145

**4.2 Recombinant expression and purification of TbrPDEB1 from *E.coli* and crystallisation**

4.2.1 Recombinant expression and purification ..... 147

4.2.2 Protein Crystallisation ..... 150

**4.3 Crystallographic Solutions of TbrPDEB1**

4.3.1 Addition of Ligands ..... 152

4.3.2 Data collection and processing ..... 153

4.3.3 Crystallographic Data, Refinement and Ramachandran Plot Statistics with ligands ..... 155

4.3.4 TbrPDEB1 with ligands structures……158

4.3.5 Analysis of the TbrPDEB1 active site ..... 160

**4.4 Characterisation of TbrPDEB1 by NMR Spectroscopy**

4.4.1 NMR Sample Preparation of $^{15}N/^{1}H$ TbrPDEB1 ..... 163

4.4.2 $^{15}N/^{1}H$ HSQC NMR Data Acquisition with BRD4-BD1 ..... 165

4.4.3 NMR experiments with temperature variation ..... 166

**4.5 Summary and Discussion**

**Chapter 5: Crystal and NMR Structures of BRD4-BD1 in complex with 7 ligands……170**

**5.1 Introduction**

5.1.1 Functional Studies of BET BRD4 ..... 174

5.1.2 BET BRD4 as drug target ..... 176

5.1.3 Structural studies highlighting key residues and interactions in BET BRD4 protein ..... 178

**5.2 Recombinant expression and purification of BRD4-BD1 from *E.coli* and crystallisation**

5.2.1 Recombinant expression and purification ..... 180

5.2.2 Protein Crystallisation ..... 183

5.2.3 Addition of Ligands ..... 183

**5.3 Crystallographic Solutions of BRD4-BD1**

5.3.1 BRD4-BD1 with ligand structures ..... 185

5.3.2 Analysis of the BRD4-BD1 binding site ..... 190

**5.4 Summary and Discussion**

**Chapter 6: Ligand binding studies of BRD4-BD1 using NMR……196**

**6.1 Ligand-observed NMR screening methods of BRD4-BD1 with BTB 07004**

6.1.1 NMR Sample Preparation of BRD4-BD1/BTB 07004 complex ..... 198

6.1.2 NMR Sample STD Data Acquisition ..... 198

6.1.3 NMR Sample Carr-Purcell- Meiboom Gill (CPMG) Data Acquisition ..... 200

6.1.4 An Introduction to $^{19}$F CPMG ..... 202

6.1.4.1 NMR Sample $^{19}$F CPMG Data Acquisition..... 202

**6.2 Protein-observed NMR screening methods of BRD4-BD1 with BTB 07004**

6.2.1 NMR Sample Preparation of $^{15}$N/$^{1}$H BRD4-BD1 ..... 205

6.2.2 $^{15}$N/$^{1}$H HSQC NMR Data Acquisition with BRD4-BD1 ..... 206

**6.3 Characterisation of BRD4-BD1 by NMR Spectroscopy**

6.3.1 Introduction ..... 208

6.3.2 The Assignment of BRD4-BD1 ..... 211

6.3.2.1 NMR Sample Preparation of $^{15}$N/$^{13}$C/$^{1}$H BRD4-BD1 ..... 211

6.3.2.2 NMR Data Acquisition of CBCANH and CBCA(CO)NH experiments…….. 212

6.3.2.3 Amino Acid Sequential Backbone Resonance Assignment ..... 213

6.3.2.4 Sequential Backbone Assignment of BRD4-BD1..... 215

6.3.2.5 Secondary Structure Prediction from Chemical Shifts using TALOS ..... 219

**6.4 The Chemical Shift Map of BRD4-BD1 complexed with BTB 07004**

6.4.1 The Theory of Chemical Shift Mapping ..... 223

6.4.2 Chemical Shift Mapping of BRD4-BD1 with BTB 07004 ..... 225

6.4.2.1 NMR Sample Preparation of $^{1}$H/$^{15}$N BRD4-BD1 with BTB 07004..... 225

6.4.2.2 NMR Data Acquisition of $^{15}$N BRD4-BD1 in the presence and absence of BTB 07004..... 225

6.4.2.3 Chemical Shift Mapping Results of BRD4-BD1 in the presence of Ethanol……..231

**6.5 $^{19}$Fluoroindole-labelled ($^{19}$F) BRD4-BD1 Spectrum**

6.5.1 Introduction ..... 236

6.5.2 The expression and purification of $^{19}$F BRD4-BD1 ..... 236

6.5.3 MS experiments of $^{19}$F BRD4-BD1 ..... 239

6.5.4 NMR Data Acquisition of $^{19}$F BRD4-BD1 ..... 240

**6.6 Summary and Discussion**

**Chapter 7: Discussion….. …… 244**

**7.1 General Summary**

7.1.1 CypA Conclusions..... 247

7.1.2 TbrPDEB1 Conclusions ..... 248

7.1.3 BRD4-BD1 Conclusions ..... 249

**7.2 Future Studies: The development of Group Epitope Mapping**

**7.3 Conclusions**

**AppendixContents…….....254**

**Appendices.....255**

**References……276**

# C. Abbreviations

**1D** 1-Dimensional

**2D** 2-dimensional

**Amp** Ampicillin

**BET** Bromodomain and Extra-Terminal

**bp** base pairs

**BRD2** Bromodomain-2

**BRD3** Bromodomain-3

**BRD4-BD1** Bromodomain-4 BD1

**BRDT** Bromodomain-Testis Specific

**BSA** Bovine Serum Albumin

**C** Coil

**Carb** Carbenicillin

**CBP** CREB-Binding Protein

**CD** Circular Dichroism

**CPMG** Carr-Purcell- Meiboom Gill

**CsA** Cyclosporin A

**CV** Column volume (s)

**CypA** Cyclophilin A

**CypD** Cyclophilin D

**DLS** Dynamic Light Scattering

**DMSO** Dimethyl sulfoxide

**dNTP** Deoxynucleotide

**DSF** Differential Scanning Fluorometry

**DTT** Dithiothreitol

**EDTA** Ethylenediaminetetraacetic acid

**ELISA** Enzyme-Linked Immunosorbent Assays

**FCS** Fluorescence Correlation Spectroscopy

**FT** Flow-Through

**FTS** Fluorescence-based Thermal Shift

**GEM** Group-Epitope Mapping

**GF** Gel Filtration

**GFP** Green-Fluorescent Protein

**H3 / H4** Histone 3 / Histone 4

**HAT** Human African Trypanomiasis

**HCV** Hepatitis C Virus

**HEPES** 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

**HIF** Hypoxia-Inducible Factor

**HIV** Human Immunodeficiency Virus

**HR** Hampton Research

**HSQC** Heteronuclear Single Quantum Coherence Spectroscopy

**IEC** Ion-exchange Chromatography

**IEX** Ion-Exchange

**IMAC** Immobilised Nickel Affinity Chromatography

**IPTG** Isopropyl β-D-1-thiogalactopyranoside

**ITC** Isothermal Titration Calorimetry

**Kan** Kanamycin

**Kpsi** Kilopound per square inch

**LB** Luria-Bertani

**LCP** left circularly polarised

**LDS** Loading Dye Sample

**Lmj** *Leishmania major*

**LYS** Lysate

**MES** 2-(N-morpholino)ethanesulfonic acid

**MP** Marker Protein

**MR** Mean Residue

**mRNA** messenger Ribonucleic Acid

**MS** Mass Spectrometry

**MW** Molecular Weight

**MX** Macromolecular Crystallography

**NF-AT** Nuclear factor of activated T cells

**NF-κB** Nuclear Factor κB

**Ni** Nickel Eluant

**NIM811** N- methyl-4-isoleucine cyclosporine

**NiNTA** Nitrilotriacetic acid

**NMC** NUT Midline Carcinoma

**NMR** Nuclear Magnetic Resonance

**NOE** Nuclear Overhauser Effect

**NOESY** Nuclear Overhauser effect spectroscopy

**NUT** Nuclear Protein in Testis

**P** protein

**P-pocket** Parasitic pocket

**PBS** Phosphate Buffered Saline

**PCR** Polymerase Chain Reaction

**PDB** Protein Data Bank

**PDEs** Phosphodiesterases

**PEG** Polyethylene Glycol

**pET** Plasmid

**pI** Isolectric point

**pKa** Dissociation constant

**PPIA** Peptide Prolyl Isomerase A gene

**PPIA** Peptidyl-propyl isomerase A

**PPIase** Peptide Prolyl cis-trans Isomerase

**RCP** right circularly polarised

**RelA** Reticulo-Endotheliosis viral oncogene homolog A

**RNAi** RNA interference

**SAR** Structure-Activity-Relationship

**SAXS** Small-Angle X-ray Scattering

**SDS** Sodium Dodecyl Sulfate

**SOB** Super Optimal Broth

**SOC** Super Optimal Medium

**SPR** Surface Plasmon Resonance

**STD** Saturation-Transfer Difference

**TALOS** Torsion Angle Likelihood Obtained from Shift and sequence similarity

**TBE** tris/Boric Acid/EDTA

**Tbr** *Trypanosoma brucei*

**TBrPDEB1** Phosphodiesterase B1 expressed from *Trypanosome brucei*

**Tc** *Trypanosome cruzi*

**TEV** Tobacco Etch Virus nuclear-inclusion-a endopeptidase

**TOCSY** Total Correlation Spectroscopy

**TROSY** Transverse relaxation optimized spectroscopy

**v/v** volume per volume

**Ve** Elution Volume

**Vo** Void Volume

**VHL** Von Hippel-Lindau Tumor Suppressor Protein

**VSMC** Vascular Smooth Muscle Cells

**w/v** weight per volume

**WaterLOGSY** Water-Ligand Observed via Gradient Spectroscopy

**YT** Yeast Extract Tryptone

# D. Abstract

Understanding protein interactions with modulators, signalling molecules or inhibitors can give information about ligand binding sites, ligand interaction topology and allosteric regulation at an atomic level, which is key for the design of novel inhibitors and protein molecules. Many techniques in-vitro have been developed and utilised including X-ray crystallography, nuclear magnetic resonance (NMR), electron microscopy, mass spectrometry etc… Further advances in these biophysical techniques and combining the information from the different techniques are still required to provide detailed insights into the structure and function of proteins and relate these to the physiological environment.

Both X-ray crystallography and NMR techniques were explored with a number of previously characterised protein models, including Cyclophilin A (CypA) (18 kDa), the catalytic domain of Phosphodiesterase B1 from *Trypanosoma brucei* (TbrPDEB1) (37 kDa) and the catalytic domain of Bromodomain 4 BD1 (BRD4-BD1) (15 kDa). The results were mixed for each protein system. CypA showed successful NMR spectra but X-ray structures were less amenable. TbrPDEB1 was a good crystallographic system, where four novel liganded crystal structures were solved, but was deemed unsuitable for NMR studies. BRD4-BD1 was a successful crystallographic system, and also suitable for ligand and protein-observed NMR utilising $^{1}H$, $^{15}N$ and $^{19}F$ nuclei.

Seven novel liganded BRD4-BD1 crystal structures and NMR assignments are presented in this thesis. The pair of triple resonance experiments HNCA, HN(CO)CA, HNCACB, HN(CO)CACB for BRD4-BD1 led to the sequential backbone assignment of $^{1}H_{N}$, $^{15}N$, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ nuclei. Major chemical shift differences ranging above 0.3 ppm were observed for the residues W81, V87, D88, A89, N93, L94, I146, A150, L158 and K160 with the ligand BTB 07004. This chemical shift mapping was in agreement with the binding site in the

BRD4-BD1 crystal structures, particularly for the residues W81, V87, L94, Y97, C136, Y139, N140 and I146.

The thesis confirms that combining X-ray crystallography and NMR leads to an advanced understanding of ligand interaction sites in proteins. This could be used for generating accurate topology maps of ligand-binding sites of any proteins with specific ligands, known as the “Group Epitope Mapping” (GEM) in the future and be extended to novel proteins.

# Chapter 1

## Introduction

## 1.1 Proteins

Proteins are fundamental molecules in all living cells and play a vital role in all various cellular processes [1] – [4]. They are defined as macromolecules due to their large size and high complexity [4]. Proteins are essential for developing and maintaining the structure, function and regulation of the body's tissues and organs [1] [2] [4].

The building blocks of a protein are amino acids [2] – [4]. Proteins are a polypeptide of amino acids linked with covalent peptide bonds. The polypeptide sequence can be illustrated as the "beads in a string". The polypeptide chain is formed on the ribosome during translation of *messenger* RNA from a specific gene resulting in protein synthesis in cells. All the amino acids consist of an amine group $NH_2$ at one end "amino" and of a carboxylic acid COOH group at another end "acid" [1] [3] [4]. The amine group and the carboxyl group are then linked together by a peptide bond. In most of the amino acids, both of the amino and carboxyl groups are attached to a central carbon atom. In addition, hydrogen atoms and another group of atoms (R) that varies with each amino acid type, are attached to the central carbon atom [1] [3] [4]. Consequently, the formula for an amino acid is $RCH(NH_2)COOH$; e.g. the simplest amino acid found is glycine, which has the following formula of $HCH(NH_2)COOH$ [1] [3] [4].

There are in total 21 different naturally-occurring amino acids, which can be arranged in several different ways to deliver different macromolecules, each one assigned with a particular function depending on its composition [4]. Each of these amino acids has its own and specific properties defined by the side-chain (R), and therefore give its unique characteristics in a protein structure. The propensity of the side-chain with polar solvent

(water) determines the classification of the amino acid: **hydrophobic/uncharged** (low propensity with water), **polar** or **hydrophilic/charged** (high propensity with water). The 21 amino acids are:

- **Hydrophilic/charged**: lysine (+), arginine (+), aspartic acid (-) and glutamic acid (-)
- **Polar**: serine, threonine, asparagine, glutamine, histidine and tyrosine
- **Hydrophobic/uncharged**: alanine, valine, leucine, isoleucine, proline, phenylalanine, tryptophan, cysteine, selenocysteine and methionine.

Glycine is an exception, as it does not have a side chain [4].

The sequence of amino acids of a polypeptide chain determines the three dimensional structure, function and properties of proteins. They are four levels of protein structure: primary, secondary, tertiary and quaternary [3] [4]. These four different structural levels inter-depend on each other, as together they make several interactions between hundreds to thousands of atoms in a highly complex manner [4].

### 1.1.1 Structure is related to function

Proteins are classified by their wide range of intra- and extracellular functions, e.g. structural roles in cytoskeleton, building proteins, catalytic enzymes, homeostasis, transport, storage, gene regulation, antibodies in immunity, assembly and hormones [4]. Each protein has a specific function in living organisms, such as the well-known, myoglobin, the first protein whose structure was revealed by X-ray crystallography [3]. Myoglobin is defined as a

cytoplasmic haemoprotein found in cardiac myocytes and oxidative skeletal muscle fibres, and binds $O_2$ using its heme residue (a porphyrin ring:iron ion complex) to provide oxygen to the working muscles [5].

Understanding the three-dimensional structure of proteins is important to deduce the associated functions at a molecular level in living cells for applications in protein engineering and drug design in pharmaceutical and biotechnological industries [6] [7]. Three-dimensional structure of a protein also informs about ligand-binding sites, ligand interactions and allosteric regulation [8]. It becomes useful particularly for targeted drug therapy in the design of novel inhibitors, but also design of protein molecules [8] (Figure 1.1). With the advent of Next Generation Sequencing, which predicts functions of new genes through sequence comparison with known protein folds is a key technique in understanding proteomics towards rational drug development [9]. Protein structure information can be utilised in the development of genomic medicines through the understanding of genetic variations in disease characterisation and diagnosis [9] [10].

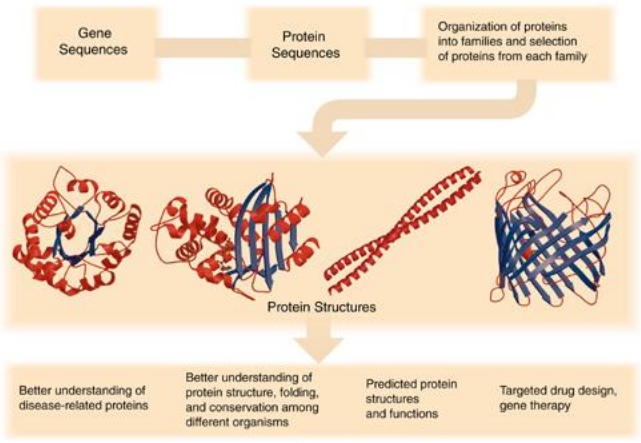



**Figure 1. 1. The structure-function relationship of a protein.** Understanding the intimate interaction between small molecule ligands and protein receptors' is key for drug design and therapeutic involvement [10]. *Adapted from*: https://www.nigms.nih.gov/about/budget/statements/archives/Pages/February25_1999.aspx

A large number of changes in a sequence may still lead to the same structure, therefore the structure is more conserved than a sequence [4], e.g. the anaerobic enzyme cobaltochelatase is involved in the synthesis of water-soluble vitamin $B_{12}$, whose function was well-known even before determination of structure. Cobaltochelatase is an enzyme that catalyses the insertion of cobalt into the corin ring of hydrogenobyrinic acid a,c-diamide [4]. Cobaltochelatase has a structural similarity with the enzyme ferrochelatase involved in heme biosynthesis (Figure 1.2). Although these two enzymes of cobaltochelatase and ferrochelatase only have 11 % sequence identity, their functions were discovered using their three-dimensional structure [4] [11] - [13].

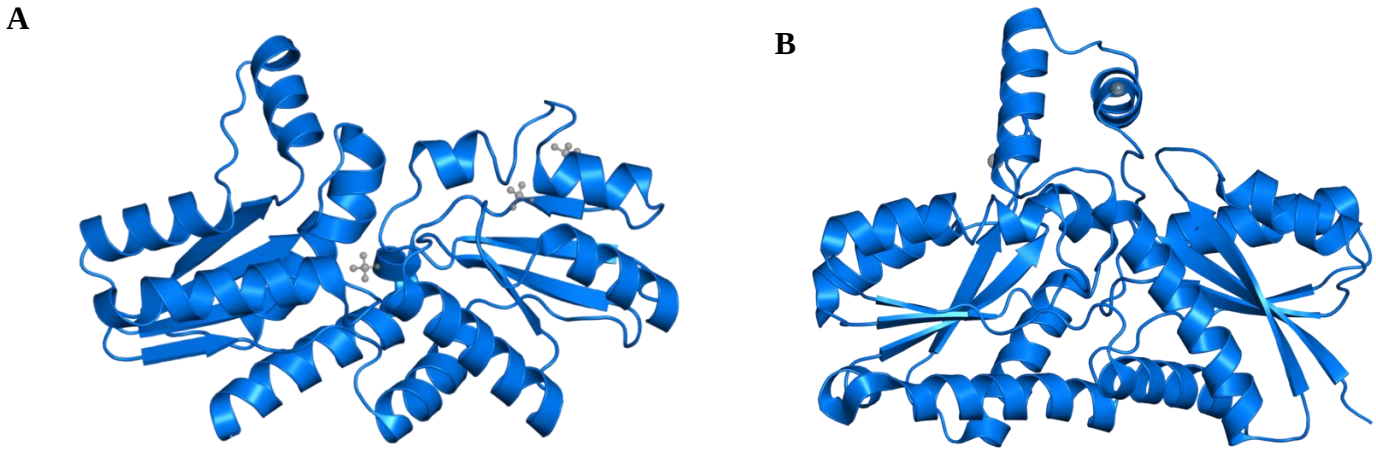



**Figure 1. 2. The structural similarity between cobaltochelatase and ferrochelatase** A) Crystal structure of cobaltochelatase (PDB: 1QGO) B) Crystal structure of ferrochelatase (PDB: 2QD4). Although cobaltochelatase and ferrochelatase are two different proteins with different functions, and only 11 % sequence similarity, they are structurally similar [12] [13].

However, a single sequential change in the protein's ligand binding site can change the ligand binding specificity of the protein, altering its associated function and even rendering the protein inactive [11].

Various analytical techniques have been developed to determine the three-dimensional structure of a protein, as proteins cannot be visualised by light microscopy [7] [14], such as X-ray crystallography, Nuclear Magnetic Resonance (NMR), electron microscopy, mass spectrometry and many more evolving techniques [7] [8] [15] – [21].

## 1.2 Protein binding sites and interactions with ligands

The function of a protein is fundamentally exerted through the binding of another molecule as a trigger. Majority of biological processes are dependent on protein-protein complex or protein-ligand complex [22] - [26]. Almost every existing biological reaction derives from a protein-ligand interaction, particularly in enzymology [27] [28]. High concentration of ligands could bind to multiple receptor pockets, known as allosteric regulation. Examples of receptors are enzymes, hormone receptors, cell signalling receptors, neurotransmitter receptors [25].

Proteins interact with other proteins and ligands through a complex array of intermolecular interactions (Figure 1.3) [29] [30]. These chemical interactions are controlled by polar specific interactions and thermodynamic forces (ionic bonds, hydrogen bonds, van der Waals' Forces and hydrophobic effects) within the binding site [30]. The interactions between the protein and the ligand molecule become complex with the flexibility of the protein binding sites and the protein structural rearrangements, following the binding of the ligand [30].

The folding of a protein is driven by reversible chemical processes, defined by an equilibrium constant Keq, which is the ratio of folded state over unfolded state [31]:

$$\mathbf{K} = \frac{folded\ state}{unfolded\ state}$$

So **K > 1** means predominantly **folded** and **K < 1** means predominantly **unfolded**.

At equilibrium and at standard rate, there is a chemical relationship between the equilibrium constant and the change in free energy (Gibbs energy):

$$\mathbf{\Delta G^{\ominus} = RTlnK}$$

where R is a constant, T is temperature in Kelvin and Keq is the equilibrium constant [31].

So $\mathbf{\Delta G^{\ominus} > 0}$ means predominantly **unfolded** and $\mathbf{\Delta G^{\ominus} < 0}$ means predominantly **folded**.

The change in $\Delta G^{\mathbf{o}}$ can be subdivided into the enthalpy change (ΔH) and the entropy change (ΔS):

$$\mathbf{\Delta G^{\ominus} = \Delta H - T\Delta S}$$

where ΔH is a measure of the heat released or taken up in a chemical reaction, and ΔS is a measure of disorder [31].

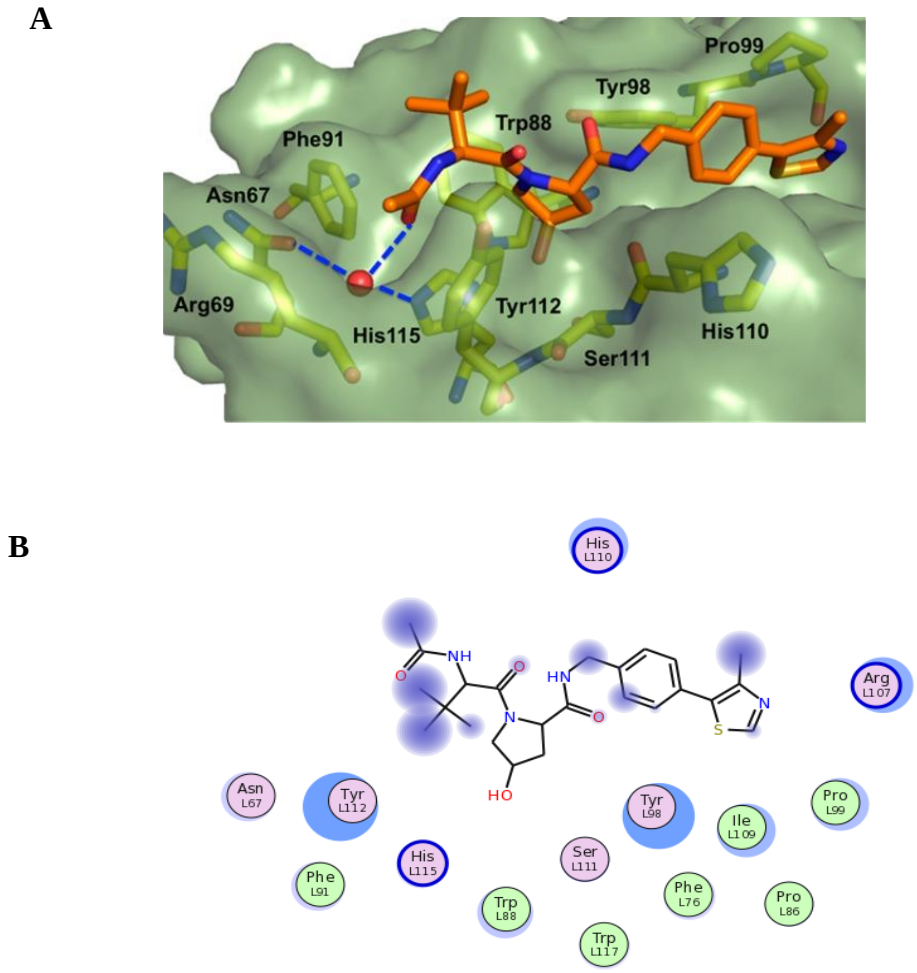



**Figure 1. 3. The chemical interactions between a protein and a ligand in a ligand binding pocket**

A) Crystal structure of VHL-HIF-alpha protein (PDB: 4W9H) in complex with the ligand (3JF) with nanomolar affinity. Blue-dashed lines represent hydrogen bonds with the structural water (red sphere). VHL-HIF-alpha protein is represented as a pale-green surface and residues from VHL-HIF-alpha protein forming the binding site as yellow sticks. B) The different 3JF ligand interactions with the protein VHL-HIF-alpha protein, represented in WinCoot FLEV. *Adapted from*: Galdeano C, Gadd MS, Soares P, Scaffidi S, Van Molle I, Birced I, Hewitt S, Dias DM, Ciulli A. (2014). Structure-guided design and optimization of small molecules targeting the protein-protein interaction between the von Hippel-Lindau (VHL) E3 ubiquitin ligase and the hypoxia inducible factor (HIF) alpha subunit with in vitro nanomolar affinities. *J Med Chem*. 57(20): 8657-63.

The function of a protein is dependent on the formation of highly specific binding pockets onto which small-molecule ligands would bind with the required affinities tailored to meet the needs of cellular biological function [24] [25].

It is crucial to understand precisely the binding sites within a protein and the interactions between receptors and ligands to reveal many biological processes [22] [32] [33], but also to apply that knowledge into drug discovery and drug design, particularly on target proteins with medical relevance [22] [26] [32] [33]. The prediction of protein binding sites helps with the geometry of protein-protein complexes as well as protein-ligand complexes. Examples of protein-protein complex are observed in aggregation, dimerisation or oligomerisation, causative of disease states such as Alzheimer's, Parkinson's and prions disease [22] [25]. Examples of protein-ligand complex are observed in membrane-bound G-protein coupled receptors, which are activated in response to a ligand, leading onto various downstream signal transduction pathways [23].

## 1.3 Protein and Ligand Binding Affinity Studies

The binding of ligands to proteins is studied using $K_D$, the dissociation constant of the reversible binding process. $K_D$ informs about the concentration of ligand required to activate the receptor, yielding a biological response or desensitisation of the receptor [23] [27].

Understanding $K_D$ is key for the knowledge of ligand-protein binding affinity, as it informs about the potency and efficacy of the ligand. $K_D$ provides crucial mechanistic insights into signalling pathways [23]. $K_D$ of protein-ligand interactions are usually found in the

nano-molar (nM) range, representing a strong binding. $K_D$ can also range in much weaker (milli-molar mM or micro-molar µM) or stronger (pico-molar pM) affinities. For instance, catecholamine epinephrine binds to beta-adrenergic receptors at low affinity (2 µM), whereas biotin binds to avidin at a very exceptionally high affinity of 1 femto-molar (fM). 1 fM of biotin is enough to activate avidin, therefore a very potent ligand [23] [27].

The interacting forces are interpreted from equilibrium binding kinetic measurements. $K_D$ is calculated from a mathematical equation with molecular models, based on the stoichiometry of the binding equilibrium [23] [27]. This provides a quantification of chemical interactions between the ligand and the protein receptor. The quantification from the $K_D$ values determines the strength of interactions of protein-ligand as thermodynamic measure. However, the calculations become difficult to manipulate as the protein's size and complexity increase [23] [27].

### 1.3.1 Techniques used for determining binding affinities

Many biophysical techniques are used to determine the molecular interactions controlling biological processes, either high-throughput or low-throughput screening techniques. These could be used for rationally controlling the protein function by altering the interactions of the protein complex [23] [27].

Enzyme-Linked Immunosorbent Assays (ELISA) and Surface Plasmon Resonance (SPR) are the most common techniques used for getting $K_D$ values. These techniques are based on immobilising either the protein receptor or the ligand of the complex, and detecting binding

by a specific signal upon addition of the other partner of the complex, e.g. colorimetric change for ELISA and resonance change in SPR [23].

Radiometric ligand binding assay requires a radioactive isotope labelled ligand. It can identify the $K_D$ of a ligand to a receptor directly on living cells and/or tissues, by detecting the radioactivity of the ligand in relation to its concentration [23].

Isothermal Titration Calorimetry (ITC) is used to investigate the thermodynamics of protein-ligand interactions. ITC involves the detection of ligand binding to its receptor through heat change [23].

Fluorescence Correlation Spectroscopy (FCS) monitors the size-dependent diffusion times in solution. Recently, FCS was used in cell membrane complexes. A similar technique to FCS is micro-scale thermophoresis carried out in-vitro in solution [23].

NMR titration experiments monitor the chemical shift changes caused by gradual addition of ligands while maintaining the protein concentration constant. The chemical shift changes are plotted against the varying concentration of ligand and $K_D$ is determined from the plot [34].

### 1.3.2 The use of protein binding sites into drug design

Understanding ligand interaction sites is crucial for structural-based drug design, utilising the three-dimensional atomic structure of protein-ligand complexes by X-ray diffraction, cryo electron microscopy and NMR [35] - [37]. The design of new lead small-molecule

compounds for targeted disease states has also become within the last 20 – 30 years with the study of protein-ligand complexed structures at the atomic scale [24] [38].

Since the localisation of functional sites on proteins has become an important area of interest, several computational methods have been developed to identify protein-protein interactions or protein-ligand interactions with low molecular weight drug-like compounds [22] [33] [37] [39]. The prediction of ligand binding-sites identifies large pockets involved in the binding of ligands [37]. The developed methods are based on sequence conservation, geometric and physico-chemical surface properties. The existing web-accessible methods for determining protein and ligand binding sites have been optimised with proteins with and without the presence of ligands [22]. Recent studies use phylogenetic analyses to identify key areas in the protein with a potential for high binding affinity by comparing with other proteins of known function with structural similarity [37].

The recent advances in structure-based drug design and improved knowledge about the energetics and dynamics of protein binding interactions have led to the implementation of protein-protein or protein-ligand interaction modulators [33]. These include:

i) Development of computational techniques to identify "hot spot" single residues energetically contributing for protein/ligand binding.

ii) Hypothesizing possible binding sites from non-ligand bound structures.

iii) Identification of allosteric binding sites as alternative sites to main binding pockets.

iv) Docking studies focusing on protein flexibility and external solvent effects on the electrostatic binding interactions.

v) Development of small fragment library to screen for modulators interfering with protein-protein interactions [33].

## 1.4 The determination of protein structure using X-ray crystallography

One of the main reasons for the success in computational structure guided drug design is the well-established methodology of protein X-ray crystallography. X-ray crystal structure informs about the key residues in ligand binding sites and the solvent interactions of proteins with ligands at atomic resolution. X-ray crystallography has helped in the synthesis and design of new biomolecules, e.g. Protein engineering of subtilisins to improve stability in detergent [40], and ligands through structure-based design. Historically, the first crystal protein structure was solved after 22 years in 1957. To date, the Protein Data Bank (PDB) consists of more than 90,000 entries, with 90% of structures solved by X-ray crystallography [11] [15] - [17] [20] [39] [41] [42].

### 1.4.1 The process of protein X-ray crystallography

X-ray crystallography is critically built on the principles of physical chemistry and thermodynamics. The formation of crystal nuclei and the growth of protein crystal are illustrated in Figure 1.4 [7]. The concentration space is divided into the undersaturated and supersaturated zones by the solubility curve. Protein will not crystallise in the undersaturated zone. The supersaturation zone is sub-divided into three zones based on the level of supersaturation: precipitation zone, nucleation zone and metastable zone. The precipitation

zone is the area where amorphous aggregates appear due to too many nuclei forming rapidly. The nucleation zone is the area where crystallisation occurs with the formation of few nuclei. The metastable zone is the region where crystallisation is induced with mechanical shock or seed crystal, as the protein solution has not entered the nucleation zone for a long time [7] [43].

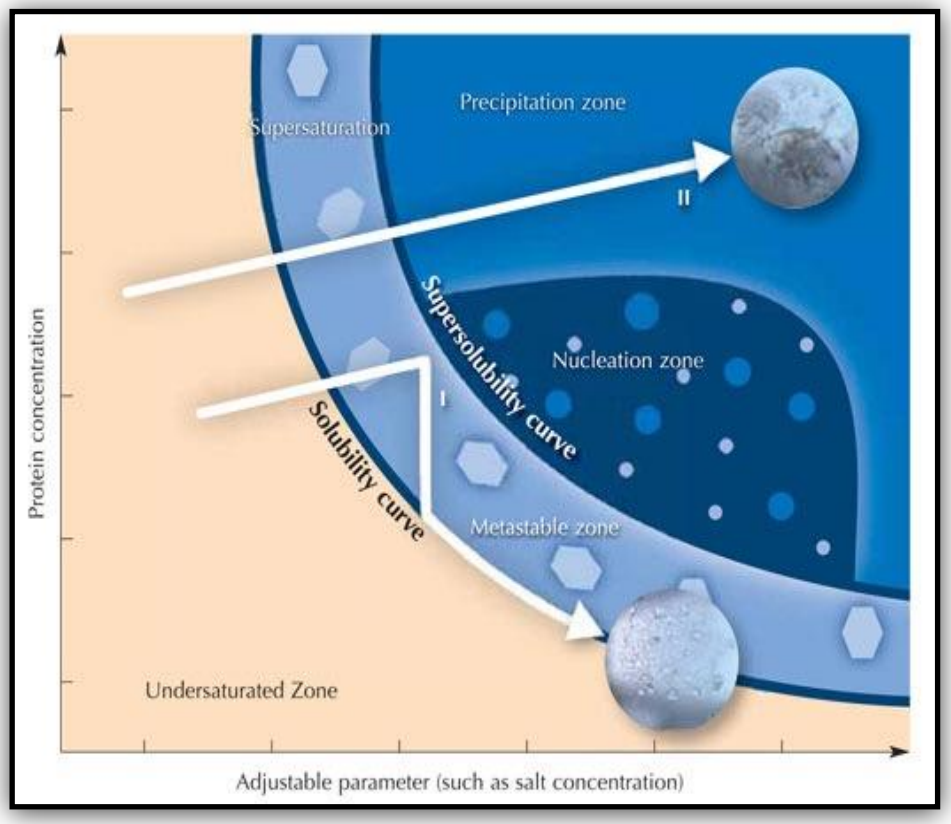



**Figure 1. 4. Two-dimensional crystallisation phase diagram.** The diagram shows the variation of protein concentration against precipitating agent concentration. The supersaturation of a protein solution results in crystallisation. Crystallisation can occur in two different zones: nucleation zone or metastable zone. (I) Crystallisation and/or (II) Precipitation (white arrows) decrease the protein concentration in the supersaturated state [7] [43]. *Adapted from*: Beat Blattmann and Patrick Sticher (2009) Growing crystals from protein. *Science in School Spring* 11: 30-36

Supersaturation is achieved by gradual decrease in protein solubility with precipitates, such as salts, organic solvents, polyethylene glycol (PEG) or ammonium sulfate. Precipitates compete with protein for hydration, resulting in an increased protein concentration due to low water availability to the protein [7] [43]. As the concentration of precipitant slowly increases, the amount of water solvent available for the protein ultimately decreases, resulting in protein precipitation and protein crystallisation if the conditions are optimal [4] [7] [43]. The most common method to grow protein crystals is vapour-diffusion through hanging-drop or sitting-drop (Figure 1.5), although other methods such as dialysis exist [7] [14]. The solubility of the protein is dependent on many factors, including temperature, concentration of precipitant, concentration of protein, pH of the buffer, additives (effectors, ligands, coenzyme substrates, inhibitors), ionic strength, organism source of protein, reducing/oxidising environment, metal ions, rate of equilibration and so on [4] [7] [14] [43].

The initial protein concentration within the drop is dependent on the protein solubility, i.e.: the higher the protein solubility, the higher the required protein concentration. Crystallisation is started with a protein concentration ranging between 5-10 mg/mL for most proteins [4] [7] [14].

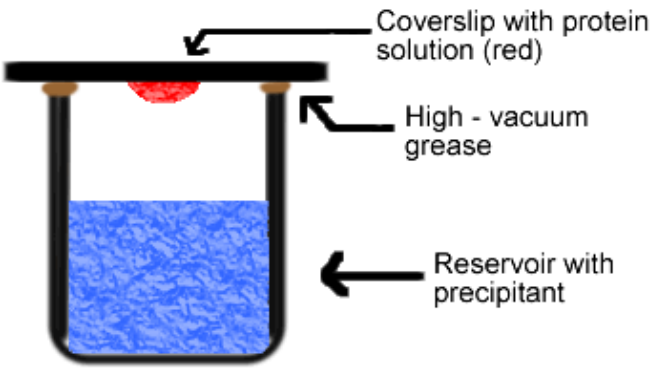



**Figure 1.5. The vapour-diffusion method.** The concentration of the precipitant in the reservoir solution is initially higher than in the protein drop. As the concentration of precipitant increases in the drop, the water solvent will evaporate towards the reservoir, increasing the protein concentration and reducing the volume of the drop. During the gradual increase of protein concentration, the protein will attain supersaturation and crystallise, if all the other factors (pH etc.) are optimal [44]. *Adapted from*: www.bio.davidson.edu/molecular/MolStudents/spring2003/Kogoy/protein.html

The crystallisation screens are a number of pre-made solutions with different precipitants, buffers, salt concentrations, which identify the "hit" condition producing crystals using the hanging-drop method in a 24-well crystallisation plate. The "hit" condition is optimised to obtain large single well-ordered crystals for X-ray diffraction experiments. Crystals are grown in a timescale ranging from hours to weeks (Figure 1.6) [4] [7] [14] [43] [44]. Common commercial screens used are Hampton Research, Molecular Dimensions or Qiagen. In cases of extensive screening, 96-well plates also exist, using 15 μL of protein sample for 96 different crystallisation conditions screening [4] [7] [14] [43] [44]. Special robotics connected to computers is nowadays available in industries to rapidly perform the screening experiments, requiring only a couple of minutes for the whole process to complete [4] [7] [14] [43] [44].

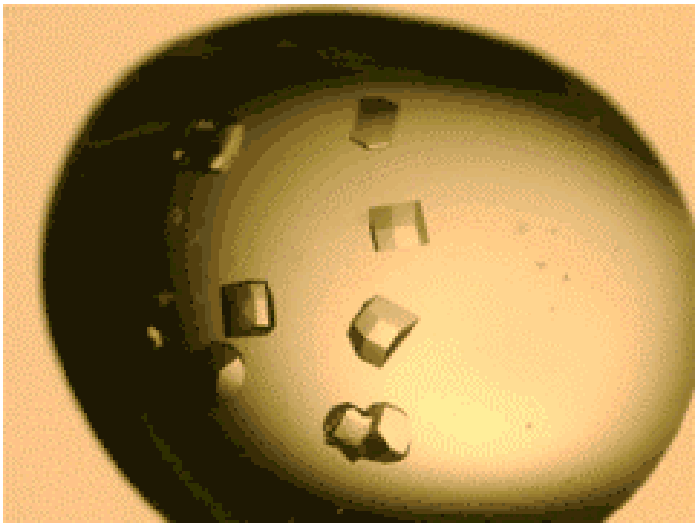


**Figure 1. 6. Example of lysozyme protein crystals in a hanging-drop experiment [45].** *From:* www.proteinstructures.com/Experimental/Experimental/crystallization-tools.html

The protein is recombinantly expressed and purified in relatively large quantities (approximately 10 mg/mL) in high purity and homogeneity [7]. The sample quality is assessed by Dynamic Light Scattering (DLS) or Small-Angle X-ray Scattering (SAXS) experiments to identify poly-dispersity from oligomerisation or aggregation in the protein solution, which can prohibit crystallisation [4]. The protein stability in different buffers, different ligands or additives can be determined by Differential Scanning Fluorometry (DSF) [4], which uses conventional real-time Polymerase Chain Reaction (PCR) instrument to identify low-molecular weight ligands that stabilise proteins [46]. Circular Dichroism (CD) Spectroscopy and analytical ultracentrifugation are other types of biophysical characterisation experiments used with X-ray crystallography for further protein structural characterisation. These techniques determine the correct folding and activity of the protein. CD is a common spectroscopic technique that measures the CD of proteins over a range of wavelengths. It is used to study the chirality of biological molecules. It also determines the different factors,

such as temperature or pH that can affect the formation of secondary structure with or without ligand molecules [47] [48]. Analytical ultracentrifugation is used for quantitatively analysing protein molecules in solution. It combines three optical systems, such as absorbance, interference and fluorescence, which determine precise sedimentation in real-time [49].

### 1.4.2 Crystal Definitions

A protein crystal consists of a regular packing model, known as the lattice model, where all the lattice points are in the same environment [50]. A crystallographic unit cell is formed by the joining of all lattice points with the highest symmetry, and a protein crystal is formed by the association of adjoining three-dimensional unit cells. The parameters of unit cells are defined by three vectors (a, b, c) and angles (α, β, γ). The space group of a crystal is the set of symmetrical elements of molecules arranged in a unit cell [50] [51]. Different space groups or crystal systems from the crystal lattice symmetry exist: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic. Several types of lattice centring are formed with the translational symmetry in three dimensions [50] [51]. These seven space groups and lattice centring lead to fourteen types of lattices: “Bravais lattices” (Figure 1.7). Thirty-two point groups symmetry operations used on a protein molecule include rotation axes, inversion axes and mirror planes, which are dependent to the external symmetry of a crystal. In addition, translation components result in screw axes and glide planes, totalling 230 possible space groups in a crystal from a Bravais lattice [52]. Protein chirality reduces

the number of possible space groups to sixty-five, as it does not permit mirror planes, glide planes or inversions [50] – [53].

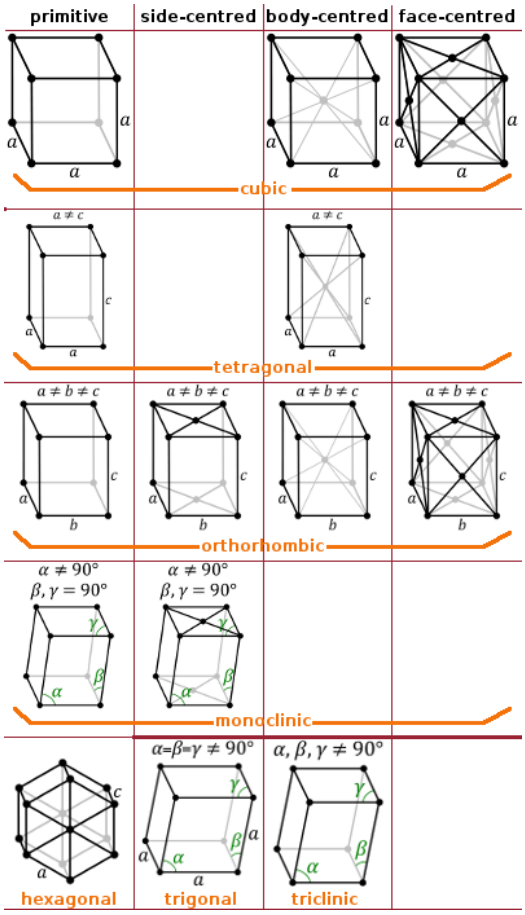



**Figure 1.7. Bravais Lattices in three dimensions.** Seven space groups from the crystal lattice symmetry lead to fourteen types of lattices: "Bravais lattices" [53]. *From:* [http://users.aber.ac.uk/ruw/teach/334/bravais.php](http://users.aber.ac.uk/ruw/teach/334/bravais.php)

### 1.4.3 X-ray diffraction of protein crystal

The crystals are cryocooled by plunging in liquid nitrogen to protect the crystals from radiation damage by the emission of very high intensity X-ray waves at the synchrotron. X-ray crystallographic data is collected at 100 degrees Kelvin [4] [7] [14]. Alternatively, in-situ crystallisation can be performed by placing the crystal plates directly into a lower intensity of X-ray beam. Prior to cryo-cooling, the crystals are briefly treated with a cryo-protectant solution, preventing crystalline ice formation in the drop around the crystal. Commonly used cryo-protectant solutions are ethylene glycol or glycerol [4] [7] [14] [43] [44].

In the beam, the crystal is exposed to monochromatic X-ray for a short period of time ranging from seconds to minutes depending on the intensity of the X-ray source beam-line, while the crystal is being rotated, changing its orientation and diffraction images collected typically from 0.1 to 1 degree a time, to collect a complete data set [4] [15] [41]. While most of the X-rays pass through the crystal, some X-rays interact with the electrons of the atoms. The electrons of the protein crystal's atoms diffract the X-ray beam into a pattern of scattered spots, known as the protein diffraction pattern recorded by the X-ray detector (Figure 1.8) [4] [15] [41]. The higher the numbers of electrons present in the atom, the greater the scattering power.

Regular arrangement of atoms in a protein crystal results in a constructive interference of the scattered X-ray waves and therefore an accurate diffraction pattern is formed [15] [41]. At least one thousand diffraction spots are needed for solving a protein structure from a well-ordered crystal, emphasising the need for growing high-quality protein crystals [4].

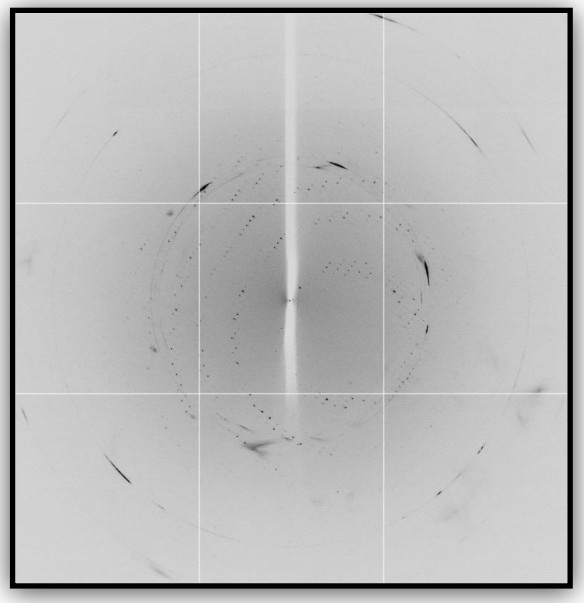


**Figure 1. 8. Example of X-ray diffraction pattern. Cyclophilin A protein has a space group P $3_2$ 2 1.** X-ray diffraction occurs as a result of X-ray wave hitting electrons of atoms within the crystal. The collision causes the wave to split and produce a scatter of waves by the electrons. Thus each atom contributes to each reflection through the electrons, hence each reflection informs about the atomic arrangements of the protein [4] [15] [41]. *Data generated from own experiments.*

In 1913, Sir William Bragg first investigated single crystal diffraction using monochromatic radiation in contrast to von Laue's earlier work on copper sulphate crystals where the X-ray radiation was polychromatic. Sir William Bragg described scattering as reflections from crystal "planes" (Figure 1.9) [15]. According to the Bragg's law, diffraction maxima are only observed when the path difference between the planes of the scattered waves is equal to the number of wavelengths [4] [15] [41].

$$n\lambda = 2dsin\theta$$

where λ is the radiation wavelength, d is the distance between the two waves, θ is the angle of reflections and n is an integer.

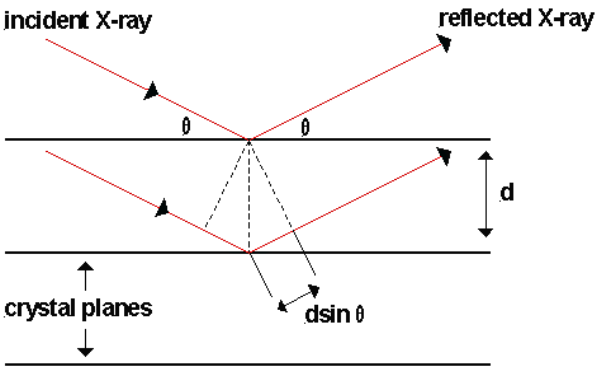



**Figure 1.9. Schematic diagram of Bragg's diffraction equation.** Reflected X-rays will be in the same phase as the incident X-rays if the path difference between them is equal to an integer (n) of the X-ray wavelength. Therefore X-rays will interfere constructively with the electrons from different atoms in the protein crystal [54]. *Adapted from*: http://www.academia.edu/16183346/MLE2101_Lab_Report

## 1.4.4 X-ray data acquisition and processing

The X-ray data is processed to obtain a reflection file (mtz) using processing softwares such as CCP4 package, XDS, *Xia2*, fast_dp and HKL-200, which extract the relative intensities of the diffracted X-ray beam [4].

*Xia2* automatically reduces the macromolecular crystallography (MX) diffraction data and determines the success of the diffraction dataset. It processes a complete MX dataset, containing multiple sequences of images at one or several wavelengths from images to structure with the help of the CCP4 Suite Package [54].

These steps are difficult to process during X-ray data collection with the increase in throughput of MX beamlines. Therefore, XDS (an independent software written by Wolgang Kabsch in Heidelberg), Scala and other tools from the CCP4 Suite Package (*Mosflm*) process the data automatically, while the user collects X-ray data. Synchrotron sources have started to have their own MX auto-solving and data processing softwares, e.g. *fast_dp* at Diamond Light Source [4]. "Conventional user interface" can also be implemented to manually input further information using natural language. *Xia2* process complex sweeps of data (a single sequence of images), in comparison to single sweeps of data processed by XDSME, *fast_dp* and RADP, which are scripts to automate the usage of XDS in a beamline [4]. Most of the MX datasets are formed of more than one wavelength, consisting of more than one sweep. Therefore, all of the data must be scaled and merged together in the same intensities [4].

XDS and Scala are series of scripts that can be run on any data processing servers, and they convert the diffraction images to mtz reflection files. The mtz files are used for the refinement of structure solution in the later stages [4]. The diffraction images are processed in the following way:

i) **Indexing the diffraction pattern**: positions of the observed reflection spots are analysed and the unit-cell vectors are determined. This is where the space group is identified [4].

ii) **Integration of the data**: the unit cell model is refined and the experimental geometry is integrated using profile-fitting methods. This step measures the intensities of all the reflections in the dataset [4].

iii) **Scaling**: all the raw integrated intensity data are not always on the same scale and are all placed on a common measurement to correct the systematic experimental

effects. Softwares such as aimless and scala aim to eliminate the systematic differences from the intensities by modelling the experiment. The scaling needs the dataset to have the right symmetry assigned and consistent definitions of the a, b and c unit cell vectors. In some instances, the shape of the protein could be misleading, and therefore could be a higher symmetry than the crystal symmetry itself. In these cases, it would be worth repeating the indexing and integration steps with the appropriate symmetry again. Unfortunately, the space group can only be confirmed and validated once the final structure is fully refined, which is quite inconvenient, as it required in the early stages of refinement. Scaling is dependent on replication of measurements based on crystal symmetry. So before scaling it is recommended to identify the point-group symmetry of the diffraction pattern, which can be determined using the Pointless software [4].

iv) **Merging**: all the scaled symmetry-related observations data are then averaged in each wavelength [4].

The resolution of the X-ray data is important in the processing of X-ray crystallographic data. Resolution in the context of X-ray crystallography can be defined as the total number of unique diffraction intensities collected for the given crystal lattice dimensions. High quality crystals diffract better, giving a better resolution. The quality of the electron density maps depends on the resolution of the X-ray data in a three-dimensional protein model [4].

## 1.4.5 Molecular Replacement and Refinement

Protein solution and solving is known as refinement, and is achieved with CCP4 package and WinCoot for the molecular graphics (Figure 1.10). The software compare experimental data with an existing model of the target protein structure through a process called molecular replacement to solve the phase problem [4]. The comparison is useful when solving flexible regions in the protein or missing atoms, resulting in negative/poor electron density map [4].

The model protein is first rotated and translated into the same position as the experimental target protein. A Fourier transformation of the model is used to calculate the structure. The comparison between the calculated structure or experimental data (Fcalc) and the model protein structure or existing data (Fobs) is measured by the refinement (R-factor), determining how closely the calculated data matches the observed data. R-factor indicates how well the refinement is incorporated within the protein structure [4].

$$R_{cryst} = \frac{\Sigma hkl \left| |F_{obs}| - k \, | F_{calc}| \right| \text{ x } 100\%}{\Sigma \; |F_{obs}|}$$

The refinement factor is described in percentage (%). The temperature factor (B-factor) measures the ordering of protein structure. The lower the temperature factor the higher the quality of the protein structure and well-ordered packing [4]. Other factors include the Ramachandran plot, bond distances and bond angles, which are monitored throughout the refinement process to ensure that the model geometry is in the expected range [4]. Isomorphous replacement and multiple wavelength anomalous dispersion are also used to determine phases and assign each reflection for a given phase in a diffraction experiment, when molecular replacement is unavailable [4].

**A** **B**

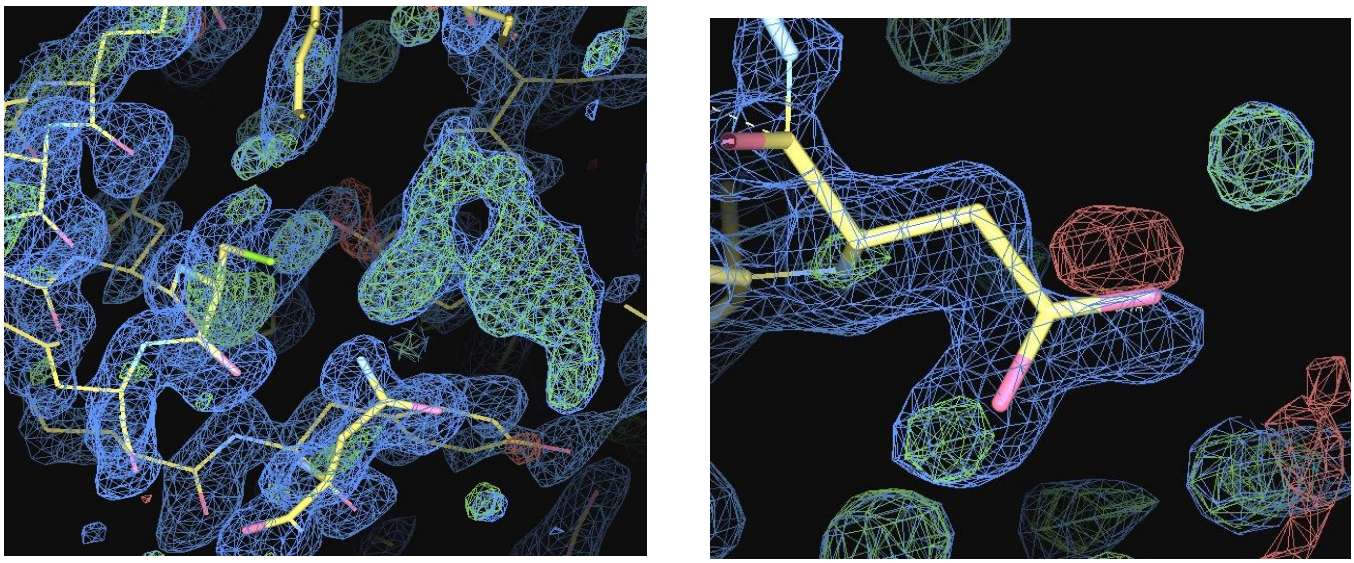


**Figure 1. 10. Positive and negative electron density of X-ray crystallography maps of TbrPDEB1 on WinCoot.**

A) Green positive densities, corresponding to water molecules, ligands (BTB 01148) or buffer molecules B) X-ray crystallographic data shows red negative densities, where atoms are missing or not geometrically correct. [4]. *Data generated from own experiments.*

The process of protein crystallisation, data acquisition, data refinement and processing are summarised in Figure 1.11.

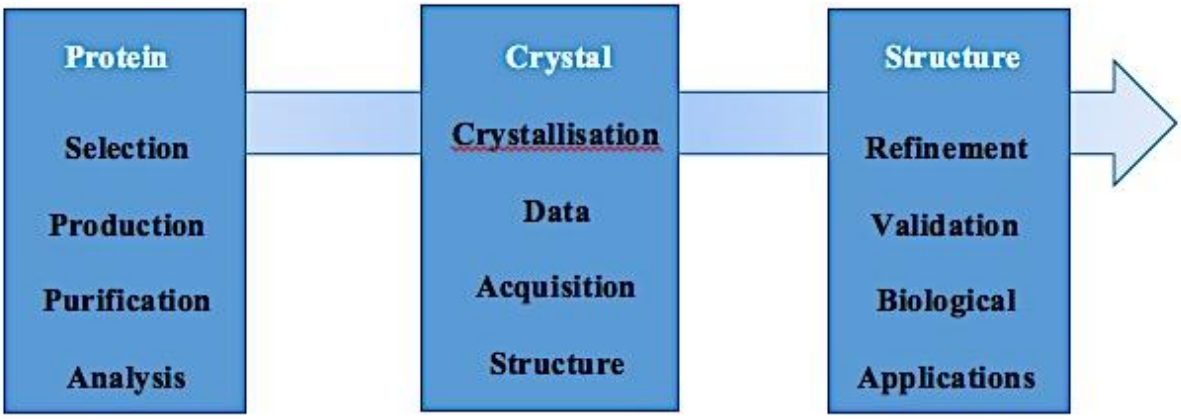



**Figure 1. 11. The workflow summary for protein structure determination using X-ray crystallography [7].** *Adapted from*: Beat Blattmann and Patrick Sticher (2009) Growing crystals from protein. *Science in School Spring* 11: 30-36

## 1.4.6 The limitations of X-ray crystallography

X-ray crystallography is dependent on the formation of good quality crystals: only well-ordered crystals have the potential to diffract to high resolution and produce a clear atomic pattern [41]. Technically, X-ray crystallography has been considered as a slow and difficult technique because of the task of growing high-quality crystals and optimisation of conditions [7]. Crystals are also at high risk of radiation damage, requiring the need for a cryoprotectant solution [41] [42]. Crystallisation artefacts could also pose a problem in X-ray crystallography. In this case, an artefact is defined as a different protein found in the crystal structure, other than the one expected. There are four reasons why artefacts could occur: (i) the wrong protein is expressed and purified (or co-purified) instead of the anticipated protein of interest. Generally, host proteins from *E. coli* could be purified during the process of

recombinant expression. (ii) the protein of interest is contaminated with an exogenous protein, such as DNAse during cell lysis or proteases/thrombin during his-tag cleavage. (iii) fusion protein is used to express a target protein, and only the fusion protein is purified and crystallised. (iv) human error leading to production of the wrong protein (mislabelling) [42].

## 1.5 Nuclear Magnetic Resonance

### 1.5.1 Principle

NMR Spectroscopy is used to determine the structure of biological proteins [8] [16] [19] - [21] [41]. In contrast to X-ray cystallography, NMR proteins are analysed in solution under physiological conditions [8] [41]. The protein solution is placed in a strong magnetic field and bombarded with radio waves. NMR resonances can be measured and they have a certain ppm value relative to a standard. The resonances can be defined as the energy of absorption and the intensity of the signal, which is proportional to the strength of the magnetic field. The NMR active nuclei absorb electromagnetic radiation at a frequency characteristic of the isotope. NMR produces complete spectra of signals, including set of distances between atomic nuclei [15] [21] [41], used to determine the model of the protein. Short inter-proton distances and torsion angles can also be measured [8]. Therefore, NMR can present various possibilities of structure and shape, instead of a single structure [15].

NMR also provides structural information of flexible regions and inherently flexible proteins [41]. Although it is a relatively rapid technique [19], it cannot provide structural details of large proteins (> 60 kDa), due to the presence of overlapping peaks in the spectra and signal broadening due to slower tumbling rates [41], and therefore it is only applicable to small to medium proteins for complete structural assignment.

### 1.5.2 Saturation-Transfer Difference NMR

NMR can determine structure and ligand-binding interactions between protein and ligands at an atomic level through Saturation-Transfer Difference (STD) NMR. This proton-based technique is often utilised as a screening tool to identify compounds that bind to a target protein. These "NMR hits" are then usually characterised by X-ray crystallography to give a complete picture of the ligand protein interactions at an atomic level. The closer the ligand is to the receptor, the stronger the STD signal will be; and therefore that will help to pinpoint the responsible areas in the receptor for binding through the Nuclear Overhauser Effect (NOE) transfer [55] [56]. The NOE is the process through which the polarisation of a nuclear spin is transferred from one nuclear spin population to another magnetic nucleus close in space through cross-relaxation using decoupling process or by a selective 180-degree pulse. It is commonly used to find out intra- and inter-molecular distances between a protein receptor and a ligand molecule [55] [56]. It causes the change in intensity of one nuclear resonance when the transitions of another nuclear resonance are affected by irradiation. Decoupling is the process by which an irradiation of a signal at a particular resonance frequency can affect the coupling of a nucleus to any others within the protein molecule [55] [56]. This effect can be accomplished within a timeframe of milliseconds, when the coupling will be all

deactivated. Saturation can be defined as the induction of the transitions between the α and β states of a proton being equilibrated following its irradiation (Figure 1.12). NOE can be quantitatively summarised by the following equation, determining the NOE signal at a given nucleus *i* when nucleus *S* is saturated [55] [56],

$$\eta i(S) = (I - I_0) / I_0$$

where $I_0$ corresponds to the normal intensity of a resonance; I corresponds to the intensity observed (i) as a result of irradiating another resonance (S) within the same protein molecule [57].

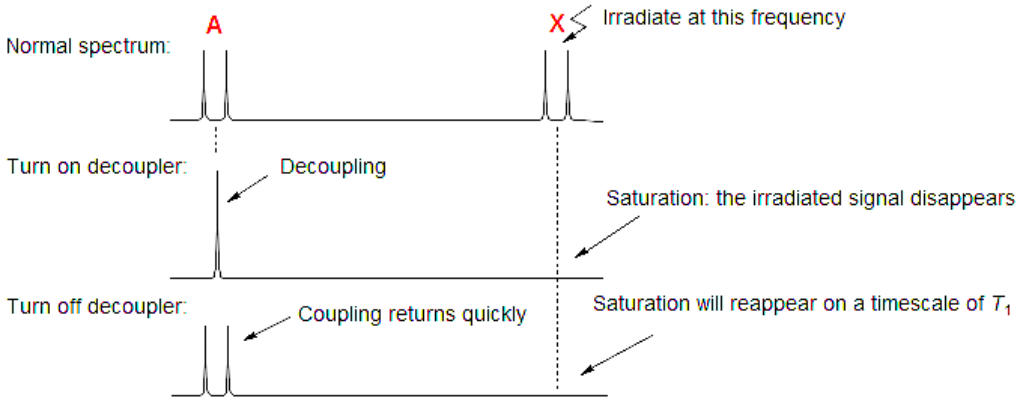



**Figure 1. 12. The Nuclear Overhauser Effect.** An irradiation will cause the spin population of a nucleus X to be altered and will be transferred to the magnetic nucleus A. Therefore, X will cause T1 relaxation of A by the interaction of their respective magnetic poles, known as dipole-dipole interaction. Dipole interaction is different to J coupling interaction, the latter being the polarised interaction between the bonding electrons of a molecule, and not through space-effect polarisation. T1 relaxation time can also be referred to spin-lattice relaxation time [55]. It measures the time at which the net magnetisation vector reaches its ground level of low energy state [55] [56]. When nuclei drop energy levels from high energy to low energy, it is often referred as the loss of energy to the surrounding nuclei. The term "spin-lattice" originates from the analyses of solids using NMR in the form of lattices [57] [58]. *From:* *http://www.usp.br/massa/2014/qfl2144/pdf/22_Maio_2014.pdf* [58]

STD-NMR is solely based on the signals of the ligand using small quantities of unlabelled proteins, and critically does not require any NMR information about the receptor [59]. This technique is used in ligand binding interactions occurring with a membrane-bound receptor, as membrane receptors can lose their structures and functionality when removed from their lipid environment [56]. This method can also be applied to protein-ligand and protein-peptide interactions [60], and can quickly produce data of even complex ligands and complex proteins with numerous post-translational modifications [61]. This method is popular in pharmaceutical industries, where thousands of compounds and fragments can be screened.

STD-NMR is the selective irradiation and saturation of the protein by intra-molecular spin-diffusion, achieved by a 1-Dimensional (1D) NMR experiment for 1-2 seconds, which details about the total number of ligand-binding sites in the protein [55] [56] [62]. Then, the ligand is added to the protein and the saturation will naturally spread onto the ligand by intermolecular saturation transfer. Based on the intensity of the signal in the STD-NMR spectrum of the individual protons between the ligand and the protein, the proximity of the ligand's protons to macromolecule's surface can be deduced according to the NOE [56]. All the non-specific binding of the ligand to the protein receptor will be determined. Finally, by subtracting the non-specific binding spectrum from the total binding spectrum, the difference spectrum will result, which is interpreted as the specific binding of the ligand to the macromolecule [55] [56] [62] (Figure 1.13).

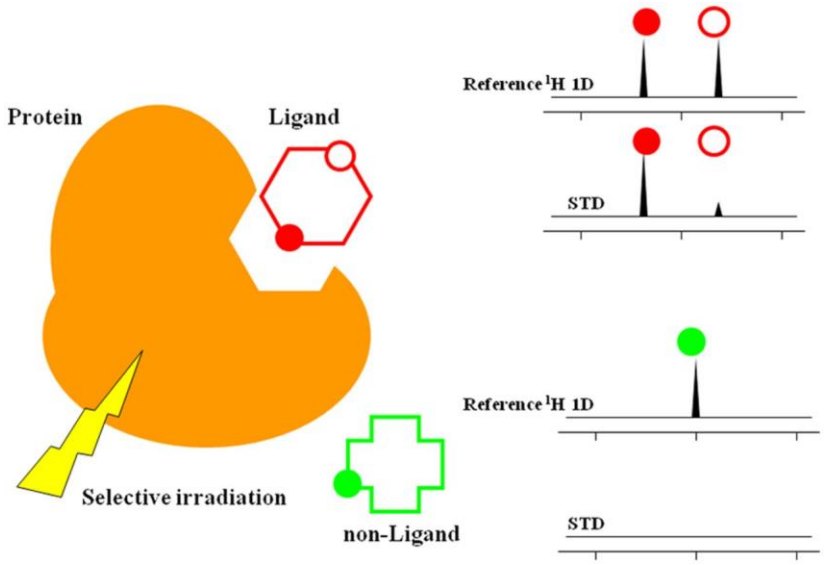



**Figure 1. 13. The concept of Saturation-Transfer Difference.** The red coloured sphere represents the area responsible for binding, as it is in close contact with the receptor and the signal stays strong. The clear sphere represents the area that is not involved in the binding process, and therefore its signal decreases as no STD signal takes place. The green coloured sphere represents no binding; as there is no STD signal occurring. However, tight binding can also show no signal [63]. *From: Hiraishi N, Tochio N, Kigawa T, Otsuki M, Tagami J (2013) Monomer-Collagen interactions studied by Saturation Transfer Difference NMR. Journal of Dental Research 92(3):284-8*

The techniques are dependent on the ligand's binding affinity range, which is related to the ligand-protein dissociation constant rates ($K_D$ and/or $IC_{50}$). Only ligands with low binding affinities (micro-molar µM range) to intermediate binding affinities can be detected using these NMR screening methods [63] [64] [65].

More involved NMR experiment helps one to understand the intimate interaction between small molecule ligands and proteins' receptors, and plot a map called "Group-Epitope Mapping" (GEM) [55] [56] [59] [61] [62]. This map would indicate the exact regions in a receptor responsible for binding to a particular type of ligands.

## 1.5.3 Advanced NMR Screening Methods

"Structure-Activity-Relationship" (SAR) by NMR technique is characterised as the best technique to be employed for proteins smaller than 30 kDa, because it provides detailed information about the localisation of the active site of a targeted protein [65] [66]. $^{15}N$-$^{1}H$ TROSY-based screening techniques can be used for larger protein targets; however, they require spectral assignment and are highly complex due to the spectral overlap [65] [66].

NMR can describe the protein ligand binding site using $^{1}H/^{15}N$ HSQC screening data, the conformation of the bound ligand using the NOE transfer through STD-NMR experiments, but also the accurate binding docking of the ligand to the protein's binding site using isotope-labelled NOESY experiments [65] [66]. NMR can yield even more detailed structural information when combined with other computational methods such as 3D database searching, virtual screening and structure-based ligand design [65] [66].

## 1.5.4 The development of fragment libraries

NMR is particularly popular technique in the field of fragment screening using ligand-observed NMR methods, identifying highly selective fragments, especially in cases of designing compounds targeting narrow binding pockets [64] [65]. Based on the identity of the selected compounds, novel scaffolds are developed in the practice of drug design and development [64] [65].

Introduction of cryo-cooled NMR probes have improved the screening of ligands to a particular protein target with increased sensitivity and reduced experiment times. This technology enables to screen more than 50 compounds for a same protein target at a time and

rapidly identifies positive binding ligands and negative non-binding ligands. Similarly to other NMR screening methods, this ligand-based NMR technology requires ligands with low to medium affinities [66] [67]. With the advancement of the technology, only small concentrations of protein and ligands are needed for a high quality data with increased resonance sensitivity, in comparison to the usual high concentrations around 10 mg/mL [66] [67]. However, mixtures containing up to 100 compounds for screening can lead to spectral overlap. In these cases, the data has to be separately de-convoluted using sophisticated software to yield the individual spectra of multiple bound ligands [66] [67].

The NMR screening technique remains an attractive high-throughput technique for its simplicity and its use in wide range of applications, such as protein-protein interactions, protein-ligand interactions, and as well as protein-nucleic acid interactions [64] [55]. This development is key in recognising a large number of low to medium binding affinity ligands to a target protein, assisting in the building of a pharmacore model of the protein's active site [65] [55] [66] - [69].

## 1.6 Other biophysical techniques in protein structure determination

There are several other techniques that can be used for the purpose of mapping protein structures and ligand-binding sites but they are much less information rich and provide only low resolution data. These include molecular electron microscopy, cryo-electron microscopy, mass spectrometry and small angle X-ray scattering [70].

### 1.6.1 Biological Electron Microscopy

It is used for the structural and chemical analysis of large proteins that are sometimes difficult to solve by X-ray crystallography or NMR [69] [70]. Electrons are used to visualise the shape of individual protein or biological protein complexes in a three-dimensional format [69] [70]. Electron diffraction can also be combined to electron microscopy for proteins embedded in membranes to obtain a high-resolution three-dimensional data and has the advantage of reducing beam damage to the specimen [41] [69]. Electron tomography obtains several views of different parts of a protein, which are then assembled to provide an overall three-dimensional shape [41]. Low resolution electron microscopy envelopes can be combined with X-ray crystallography or NMR to obtain more precise structural information about the protein and can help validating the findings by complementary approach [41] [69] - [71].

It is difficult to preserve the biological specimens, due to their aqueous and light elements composition, making the diffraction and visualisation very poor [71] [72]. The visualisation can slightly be improved with the negative stain technique, which is used to heighten the contrast in a sample. However, it still can cause modifications in the protein [71] [72].

Therefore, electron cryo-microscopy is used instead, where cryoprotectants are used to preserve the specimens, such as liquid-nitrogen [70]. Nevertheless, the samples must also be kept at low temperatures during the visualisation process on the electron microscopy and the beam of electrons must be reduced as much as possible in order to avoid structural damage to the protein [70].

### 1.6.2 Mass Spectrometry and Protein Mass Mapping

Mass spectrometry and tandem mass spectrometry are tools used to characterise the structure of proteins, and provide information about the elemental composition of protein complexes [73] - [77]. Proteins can either be observed directly or in the presence of sequence-specific proteolytic enzymes that would chemically modify and segment the protein into fragments. The resulting peptides are then analysed with mass spectrometry connected to a computing software, where the charged peptides are arranged according their mass [73] [75] [77]. Interestingly, protein mass mapping can also be used to determine the quaternary structures of proteins complexes, by highlighting the intra-protein structures [73] [76] [77]. Moreover, like the other techniques mentioned, it can provide information about protein-protein interactions and protein-ligand interactions, elucidating the individual residues and amino acids responsible for binding [73] [76] [77]. Electrospray ionisation [73] - [75] [77] and affinity chromatography can also be combined with mass spectrometry to monitor protein folding [73]. Ionisation has contributed a lot to the improvement of mass spectrometry, as it permits one to ionise peptides and small proteins directly, and informs about the molecular weight of proteins [74] [75].

### 1.6.3 Small Angle X-ray Scattering (SAXS)

The main problem of X-ray crystallography is the low like-hood of flexible proteins to form crystals. This can be overcome by SAXS, which is a technique used to characterise structures in the nanometer range. It is generally applied to the study of membrane proteins and is less

time-consuming than other structure-determining techniques. Low resolution techniques can be combined with higher resolution techniques to interpret the results [78] [79].

## 1.7 Aims and Hypotheses

Although many techniques have been developed and optimised over the past years, as described above, further research is still taking place to achieve better biophysical techniques [77]. Most structural details of proteins have been so far determined by single methods, although combined refinements have also been done in the past [80]. Thus, we also aim to combine both X-ray crystallography and NMR to understand about the general structure and ligand-binding sites. Our hypothesis is that better structural details and ligand-binding interaction topology will be obtained by combining both methods.

In order to test this methodology, a number of "model" proteins used were Cyclophilin A (CypA) (18 kDa), the catalytic domain of Phosphodiesterase B1 expressed from *Trypanosome brucei* (TBrPDEB1) (37 kDa) and the catalytic domain of human Bromodomain-4 BD1 (BRD4-BD1) (15 kDa), which are well characterised protein systems by X-ray crystallography. These proteins will be the starting proteins of the thesis.

This combined technique could serve as a tool for protein structure determination and ultimately drug discovery, as accurate information about the ligand- binding sites would be attained very precisely and quickly, even when the crystal structure is not available.

# Chapter 2

## Materials and Methods

# 2.1 Materials

## 2.1.1 Chemicals

Most of the chemicals and antibiotics used in this study were purchased from Sigma-Aldrich Ltd. Other materials were obtained from different other suppliers as follows: Agar, Agarose and IPTG from Melford Laboratories Ltd, NaCl, tris-HCl, 100% (v/v) glycerol and EDTA from Fisher Scientific Ltd, 0.2 μM Minisart single use syringe filters from Sartorial Epsom Ltd, Genejet Mini-Prep Kit, Genejet Gel Extraction Kit, NuPage SDS-PAGE gels, One Shot TOP10 chemically competent cells and SyBr Safe Stain from Life Technologies, $T_7$ express competent cells from New England Biolabs, Precision Plus Protein$^{TM}$ Prestained Marker and Coomassie Brilliant Blue R 250 from Bio-Rad, PCR Reagents (Q5 Reaction Buffer, dNTP mix, 100 bp DNA ladder and dye) from Promega, EDTA-protease free tablets from Roche, thrombin and TEV from Abcam, Hampton 24-well plates, micro-bridges, Polyethylene Glycol 3350 for X-ray crystallography from Hampton Research, X-ray crystallography coverslips from Molecular Dimensions, concentrators from Corning Ltd, NAP columns and DEAE Sepharose from Amersham Biosciences, HiTrap purification columns from GE Healthcare, $^{13}C$ glucose, $^{15}N$ ammonium sulfate, $D_2O$ and D6-DMSO from Goss Scientific, Nickel resin NiNTA from Qiagen. All the flammables were purchased from Fisher Scientific Ltd.

## 2.1.2 Bacterial Plasmids

| Construct | Parental Plasmid | Description | Source |
|---|---|---|---|
| **Cyclophilin A (CypA) Wild-Type** | pET21a | Bacterial expression and overproduction vector with T7lac promoter, $Amp^R$ | Charles River Early Discovery, Richard Bazin |
| **Cyclophilin A (CypA) Mutant** | pET21a | Bacterial expression and overproduction N-terminal $His_6$tag fusion protein with T7lac promoter, thrombin cleavage site, $Amp^R$ | Own construct |
| **Phosphodiesterase B1 from *Trypanosome brucei* (tbrPDEB1)** | pET28a | Bacterial expression vector, overproduction N-terminal $His_6$tag fusion protein with T7lac promoter, thrombin cleavage site, $Kan^R$ | Ke's Lab, Huanchen Wang |
| **Bromodomain-4 BD1 (BRD4)** | pET24a | Overproduction N-terminal $His_6$tag fusion protein with T7 promoter, $Kan^R$ | Charles River Early Discovery, Richard Bazin |

**Table 2. 1. Bacterial Plasmids**

Plasmids were provided by Charles River Early Discovery Laboratory, Ingram Building, Giles Lane, Canterbury, Kent, CT2 7NZ; and Ke's Lab, Huanchen Wang, The University of North Carolina.

## 2.1.3 Media and solutions for bacterial work

Luria-Bertani (LB) broth: - Tryptone 10 g

- Yeast extract 5 g
- NaCl 5g

Made up to 1 L with $dH_2O$ and autoclaved.

Luria-Bertani (LB) agar: - Tryptone 10 g

- Yeast extract 5 g
- NaCl 5 g
- Bacterial Agar 16 g

Made up to 1 L with $dH_2O$ and autoclaved. Antibiotics were added at a concentration of 100 μg/mL per plate.

2 xYT broth: - Tryptone 16 g

- Yeast extract 10 g
- NaCl 5 g
- Glucose 0.4 %

Made up to 1 L with $dH_2O$ and autoclaved.

2 xYT agar: - Tryptone 16 g

- Yeast extract 10 g
- NaCl 5 g
- Glucose 0.4 %
- Bacterial Agar 16 g

Made up to 1 L with $dH_2O$ and autoclaved. Antibiotics were added at a concentration of 100 μg/mL per plate.

Minimal Media broth:

| Component (stock concentration) | Stock concentration | Working concentration | Sterilisation method |
|---|---|---|---|
| **$(^{15}NH_4)_2SO_4$ (x 50)** | 30 g/L | 0.6 g/L | Filter |
| **$PO_4$ / NaCl (x 10)** | 68 g/L $Na_2HPO_4$<br>30 g/L $KH_2PO_4$<br>5 g/L NaCl | 6.8 g/L $Na_2HPO_4$<br>3.0 g/L $KH_2PO_4$<br>0.5 g/L NaCl | Autoclave |
| **$Na_2SO_4$ (x 1000)** | 46 g/L | 46 mg/L | Autoclave |
| **EDTA Trace (x 100)** | 10 g/L EDTA<br>3.2 g/L $MnCl_2$<br>1 g/L $FeCl_3$<br>0.1 g/L $ZnCl_2$<br>20 mg/L $CuCl_2$<br>20 mg/L $CoCl_2$<br>20 mg/L $H_3BO_3$ | 0.1 g/L EDTA<br>32 mg/L $MnCl_2$<br>10 mg/L $FeCl_3$<br>1 mg/L $ZnCl_2$<br>0.2 mg/L $CuCl_2$<br>0.2 mg/L $CoCl_2$<br>0.2 mg/L $H_3BO_3$ | Autoclave |
| **$MgSO_4$ (x 1000)** | 246 g/L | 0.246 g/L | Filter |
| **$Cacl_2$ (x 1000)** | 44.1 g/L | 44.1 mg/L | Filter |
| **Biotin (x 1000)** | 1 g/L | 1 mg/L | Filter |
| **Thiamine (x 1000)** | 1 g/L | 1 mg/L | Filter |
| **Glucose / $^{13}$C Glucose (x 50)** | 150 g/L | 3 g/L | Filter |
| **MilliQ $H_2O$** | - | - | Autoclave |
| **Antibiotic** | As required | As required | Filter |

**Table 2. 2. Minimal Media Components**

Super Optimal Broth (SOB): - Tryptone 10 g

- Yeast extract 5 g
- NaCl 0.25g
- 250 mM KCl 5 mL

Made up to 500 mL with $dH_2O$ and autoclaved.

Before using the SOB, 5 mL of 2 M sterile $MgCl_2$ were added. SOB was usually stored without the addition of $MgCl_2$.

Super Optimal Medium (SOC): - Tryptone 20 g

- Yeast extract 5 g
- NaCl 0.5g
- 250 mM KCl 0.186 g

Made up to 1 L with $dH_2O$ and autoclaved.

Just before using the SOC, 10 mL of 1 M sterile $MgCl_2$ and 20 mL of 1 M filter-sterilised glucose were added.

Antibiotics:

| Antibiotic | Stock concentration | Working concentration |
|---|---|---|
| Ampicillin | 100 mg/mL | 150 µg/mL |
| Carbenicillin | 100 mg/L | 150 µg/mL |
| Kanamycin | 30 mg/mL | 30 µg/mL |

**Table 2. 3. List of Antibiotics**

Additives:

| Additive | Stock concentration | Working concentration |
|---|---|---|
| Isopropyl β-D-1-thiogalactopyranoside (IPTG) | 1 M | 1 mM |
| $MgCl_2$ | 2 M | 20 mM |
| Glucose | 1 M | 20 mM |

**Table 2. 4. List of additives**

## 2.1.4 Media and solutions for DNA work

1 x TBE Buffer: - tris Base 10.8 g

- Boric Acid 5.5 g

- 0.5 M EDTA pH 8.0 4 mL

Made up to 1 L with $dH_2O$.

1 % (w/v) Agarose: - Agarose 1 g

Made up to 100 mL with 1 x TBE buffer, and dissolved by heating in the microwave.

100 bp DNA Hyperladder (Promega):

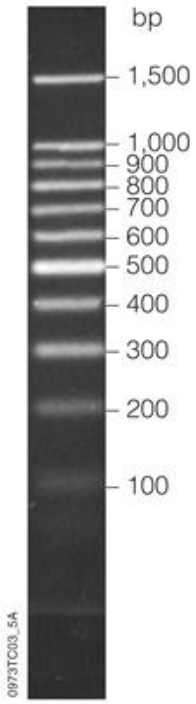



**Figure 2. 1. 10 µL Hyperladder separated on a 1 % (w/v) agarose gel**

6 x DNA Loading Buffer: - Sucrose 4 g

- Bromophenol Blue 25 mg

Made up to 10 mL with $dH_2O$.

## 2.1.5 Solutions for protein work

### 2.1.5.1 Solutions for Nickel Affinity Chromatography

His Charge Buffer: - 0.1 M $NiSO_4$ 0.15 g

Made up to 10 mL with $dH_2O$.

Nickel Strip Buffer: - 20 mM $NaH_2PO_4$ pH 7.4

- 0.5 M NaCl
- 50 mM EDTA

Nickel Chromatography Buffer A (CypA): - 50 mM tris pH 7.8

- 300 mM NaCl
- 30 mM imidazole
- 1 mM beta-mercaptoethanol

Nickel Chromatography Elution Buffer B (CypA): - 50 mM tris pH 7.8

- 300 mM NaCl
- 530 mM imidazole
- 1 mM beta-mercaptoethanol

Nickel Chromatography Buffer A (tbrPDEB1): - 20 mM tris pH 8.0

- 300 mM NaCl
- 10 mM imidazole
- 10 % (v/v) glycerol
- 1 mM beta-mercaptoethanol

Nickel Chromatography Elution Buffer B (tbrPDEB1): - 20 mM tris pH 8.0

- 300 mM NaCl
- 260 mM imidazole
- 10 % (v/v) glycerol
- 1 mM beta-mercaptoethanol

Nickel Chromatography Buffer A (BRD4): - 50 mM Hepes pH 7.5

- 500 mM NaCl
- 10 mM imidazole
- 5 % (v/v) glycerol

Nickel Chromatography Elution Buffer B (BRD4): - 50 mM Hepes pH 7.5

- 500 mM NaCl
- 250 mM imidazole
- 5 % (v/v) glycerol

### 2.1.5.2 Solutions for Desalting

Desalt Buffer (CypA): - 25 mM $NaH_2PO_4$ pH 6.5

- 0.5 mM DTT

Desalt Buffer (tbrPDEB1): - 20 mM tris pH 7.5

- 100 mM NaCl

- 5 % (v/v) glycerol

- 1 mM beta-mercaptoethanol

### 2.1.5.3 Solutions for ion-exchange chromatography

IEX Buffer A (CypA): - 50 mM Hepes pH 6.8

- 1 mM DTT

- 2.5 mM EDTA

- 100 µM phenylmethanesulfonyl fluoride

IEX Elution Buffer B (CypA): - 50 mM Hepes pH 6.8

- 1 mM DTT

- 2.5 mM EDTA

- 400 mM NaCl

- 100 µM phenylmethanesulfonyl fluoride

IEX Buffer A (tbrPDEB1): - 20 mM tris pH 7.5

- 100 mM NaCl
- 5 % (v/v) glycerol
- 1 mM beta-mercaptoethanol

IEX Elution Buffer B (tbrPDEB1): - 20 mM tris pH 7.5

- 1 M NaCl
- 5 % (v/v) glycerol
- 1 mM beta-mercaptoethanol

### 2.1.5.4 Solutions for gel filtration chromatography

GF Buffer (CypA): - 25 mM tris pH 7.5

- 100 mM NaCl
- 0.5 mM DTT
- 0.5 mM EDTA

GF Buffer (TbrPDEB1): - 20 mM tris pH 7.5

- 50 mM NaCl
- 5 % (v/v) glycerol
- 1 mM beta-mercaptoethanol

GF Buffer (BRD4): - 10 mM Hepes pH 7.5

- 50 mM NaCl

- 5 % (v/v) glycerol

For NMR experiments, the protein was always eluted in the following buffer:

GF Buffer: - 25 mM $NaH_2PO_4$ pH 6.5

- 100 mM NaCl

- 1 mM DTT

### 2.1.5.5 Solutions for SDS gels

Coomassie Stain: - Coomassie Brilliant Blue R250 1 g

- Methanol 50 % (v/v)

- Glacial Acetic Acid 10 %

Made up to 1 L with $dH_2O$. The solution was stirred for 3-4 hours and then filter through Watman filter paper.

Destain Buffer: - Methanol 50 % (v/v)

- Glacial Acetic Acid 40 %

Made up to 1 L with $dH_2O$.

Precision Plus Protein All Blue Stain (Promega):

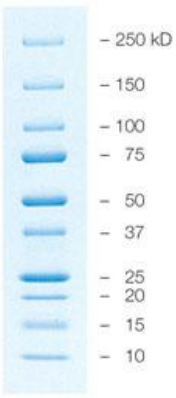



**Figure 2. 2. 10 µL Precision Plus Protein All Blue Stain applied to SDS-Page 4-12% Bis-tris gel from Promega**

## 2.1.6 Media and solutions for X-ray crystallography work

All solutions, if not purchased, were filtered through a 0.2 µM filter unit prior to use.

Reservoir solutions were prepared within the wells of the crystal plates.

| | CypA | TbrPDEB1 | BRD4 |
|---|---|---|---|
| **Crystallisation Condition** | 35% - 41% Ammonium Sulfate (stock of 66%), 0.1 M Tris-HCl pH 7.2 - 8.5 (stock of 1 M). | 14% - 24% PEG 3350 (stock of 50%), 0.4 M Sodium Formate (stock of 5 M), 0.3 M Guanidine Chloride (stock of 5 M), 0.1 M MES pH 6.5 (stock of 1 M). | 3.5 M - 5 M Sodium Formate (stock of 5 M). |
| **Reservoir Volume** | 500 µL | 1,000 µL | 500 µL |
| **Drop Volume** | 2 µL | 2 µL | 2 µL |
| **Protein Concentration** | 16 mg/mL | 10 mg/mL | 10 mg/mL |
| **Incubation Temperature** | 20 °C | 4 °C | 4 °C |
| **Cryosolvents** | 40 % Ammonium Sulfate, 0.1 M tris-HCl pH 7.2 - 8.5, 15 % glycerol | 16 % PEG 3350, 0.4 M Sodium Formate, 0.3 M Guanidine Chloride, 0.1 M MES pH 6.5, 20% Ethylene Glycol | 4 M Sodium Formate, 20% Ethylene Glycol |

**Table 2. 5. The crystallisation conditions for each protein**

Cyclophilin A: Necessary information for determining the protein concentration using nanodrop: Extinction coefficient (ε) is 8,610 $M^{-1}cm^{-1}$ ; MW=18,000 Da ;

A = ε x concentration x path length

Phosphodiesterase B1: Necessary information for determining the protein concentration using nanodrop: Extinction coefficient (ε) is 14,150 $M^{-1}cm^{-1}$ ; MW=37,000 Da

Bromodomain 4: Necessary information for determining the protein concentration using nanodrop: Extinction coefficient (ε) is 26,270 $M^{-1}cm^{-1}$; MW=15,000 Da

## 2.1.7 Ligands

All the ligands used in this study have been provided by different suppliers, which are summarised in the following sub-sections.

### 2.1.7.1 Ligands for Cyclophilin A

Ligands for Cyclophilin have been originally obtained from Sigma-Aldrich Ltd, but have been kindly provided by Dr Michelle Rowe. All the compounds were prepared at a stock concentration of 30 mM and stored at room temperature. All ligands were dissolved and prepared as stock solutions in D6-DMSO.

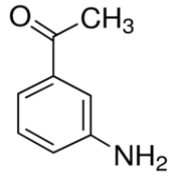
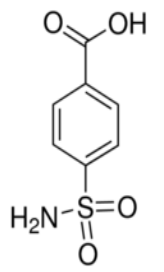
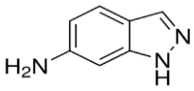
| Ligand Name | Smiles | Structure |
|---|---|---|
| **3′ Aminoacetophenone** | C1(CC(CC1)N)C(=O)C |  |
| **4′ Sulfamoylbenzoic Acid** | O=S(=O)(N)C1CCC(C(=O)O)CC1 |  |
| **6′ Aminoindazole** | [nH]1nCC2CCC(CC12)N |  |

**Table 2. 6. List of CypA ligands**

## 2.1.7.2 Ligands for Phosphodiesterase B1

Ligands for TbrPDEB1 have been provided by VU Free University of Amsterdam as part of their collaboration. Stocks have been prepared at variable concentrations. All ligands were dissolved and prepared as stock solutions in D6-DMSO and stored at room temperature in the dark.

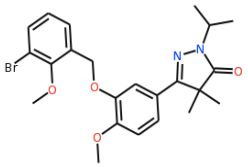
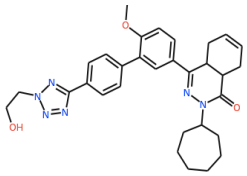
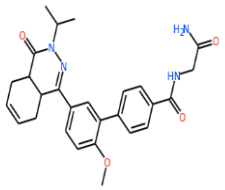
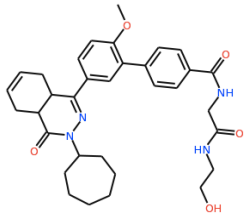
| Ligand Name | Smiles | Structure | Stock |
|---|---|---|---|
| **VUF 14450** | COC1=C(OCC2=CC=CC(Br)=C2OC)C=C(C2=NN(C(C)C)C(=O)C2(C)C)C=C1 |  | 5 mM |
| **VUF 14234** | COC1=CC=C(C2=NN(C3CCCCCC3)C(=O)C3CC=CCC23)C=C1C1=CC=C(C2=NN(CCO)N=N2)C=C1 |  | 50 mM |
| **PPS 59083** | COC1=CC=C(C2=NN(C(C)C)C(=O)C3CC=CCC23)C=C1C1=CC=C(C(=O)NCC(N)=O)C=C1 |  | 50 mM |
| **PPS 60036** | COC1=C(C2=CC=C(C(=O)NCC(=O)NCCO)C=C2)C=C(C2=NN(C3CCCCCC3)C(=O)C3CC=CCC23)C=C1 |  | 50 mM |

**Table 2. 7. List of TbrPDEB1 ligands**

## 2.1.7.3 Ligands for Bromodomain 4

Ligands for Bromodomain 4 have been provided by Maybridge Ltd (Cambridge) as part of their collaboration. All the ligands have been prepared at a stock concentration of 100 mM and stored at room temperature. All ligands were dissolved and prepared as stock solutions in ethanol.

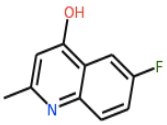
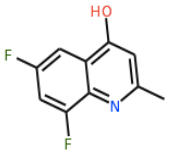
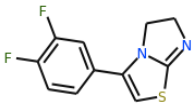
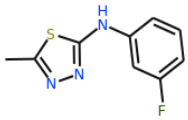
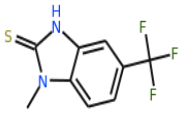
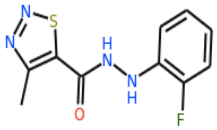
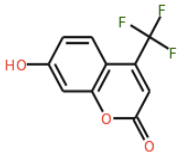
| Ligand Name | Smiles | Structure |
|---|---|---|
| **BTB01148** | C12=C(N=C(C=C1O)C)C=CC(=C2)F |  |
| **BTB06033** | C1=2C(N=C(C=C1O)C)=C(F)C=C(C2)F |  |
| **BTB07004** | N12C(SC=C1C=1C=C(F)C(=CC1)F)=NCC2.Br |  |
| **RJF00210** | C=1(SC(C)=NN1)NC=1C=C(F)C=CC1 |  |
| **TG00013** | C1(N(C2=CC=C(C(F)(F)F)C=C2N1)C)=S |  |
| **HTS05027** | C1(C(NNC=2C(F)=CC=CC2)=O)=C(N=NS1)C |  |
| **RJF00002** | C=1(C(F)(F)F)C2=C(C=C(C=C2)O)OC(C1)=O |  |

**Table 2. 8. List of BRD4-BD1 ligands**

## 2.2 Microbiological Methodology

### 2.2.1 Sterilisation Methods

Unless stated otherwise, all the media and buffers used for microbiological purposes were sterilised by autoclaving at 121 °C, 1 bar pressure for 15 minutes. Substances that were sensitive to high temperatures were sterilised using 0.2 μM Minisart single use syringe filters.

### 2.2.2 Preparation of *E. coli* Competent Cells

A sterile loop was used to scrape off a portion from a commercial aliquot of T7 express competent cells and were streaked onto an LB agar plate without any antibiotic resistance, and incubated at 37 °C overnight. The next day, a 50 mL SOB media starter culture was inoculated from a single colony without antibiotic. The culture was incubated at 37 °C with ~ 200 rpm shaking until an $OD_{600}$ of 0.5. The cells were harvested at 3,000 rpm (The Beckman SX4750A Centrifuge with an Allegra X-15R rotor) for 15 minutes at 4 °C, and the pellet was resuspended in 25 mL of ice-cold 50mM $CaCl_2$. The cells were left on ice for 30 minutes and centrifuged again at 3,000 rpm (The Beckman SX4750A Centrifuge with an Allegra X-15R rotor) for 15 minutes at 4 °C. The pellet was resuspended in 2mL of ice-cold 50 mM $CaCl_2$ containing 15% (v/v) sterile glycerol. The cells were rapidly aliquoted into 50 μL samples and stored at -80 °C.

### 2.2.3 Transformation of *E. coli* Competent Cells

One aliquot of competent cells was defrosted on ice for 10 minutes and 1 μL of plasmid DNA was added, and the cells were incubated on ice for a further 30 minutes, heat-shock was performed in the water-bath at 42 °C for 1 minute. The cells mixture was then kept on ice again for 2 minutes before the addition of 300 μL of SOC media. Incubation at 37 °C in a dry-block for 60 minutes was done and the cells were then plated on LB-agar containing the appropriate antibiotics and incubated at 37 °C overnight.

### 2.2.4 Preparation of glycerol stocks

For the purpose of long-term storage of bacterial cells, glycerol stocks were prepared. Just before the induction step in the protein production bacterial expression protocol (c.f.: 2.3.1), 1 mL culture sample was collected. The 1mL pre-induction culture volume was transferred into 100 μL aliquots in cryotubes. Ice-cold sterile glycerol was added to the aliquots to a final concentration of 10% (v/v). The aliquots were stored in ice for 30 minutes and frozen at -80 °C.

### 2.2.5 Inoculation of LB starter cultures from glycerol stocks

100 μL of fully thawed glycerol stock was added to an LB medium containing the appropriate antibiotic. The culture was incubated overnight at 37 °C shaking at ~ 200 rpm.

### 2.2.6 Isolation of plasmid DNA

Plasmids were isolated from 1-5mL overnight *E.coli* culture in LB medium using a GeneJet Spin Mini-Prep Kit. The DNA plasmid was isolated following the manufacturer's instructions. The concentration of the nucleic acid was measured on the nano-drop (BioDrop Ltd) using the elution buffer as the nano drop blank solution. Plasmids were stored at -20 °C.

## 2.3 Recombinant Protein Overproduction

### 2.3.1 Production of recombinant protein of a pET in T7 express competent cells in enriched culture medium

A starter culture was prepared by inoculating 25 mL LB medium supplemented with appropriate antibiotics with a single colony. The starter culture was incubated at 37°C shaking at 200 rpm for 2-3 hours and used to inoculate the main cultures of 1 L LB medium supplemented with appropriate antibiotics. The main culture was incubated at 37 °C shaking at 200 rpm until an $OD_{600}$ of ~ 0.6 was reached. The culture was moved at 20 °C with shaking at 200 rpm until $OD_{600}$ of 0.85 was reached. 500 μL of culture pre-induction sample was taken and stored at -20 °C until use. 1 mM of IPTG was added to induce recombinant protein overproduction at 20 °C at 160 rpm overnight.

On Day 2, another 500 μL of culture post-induction sample was taken and was stored at -20 °C until use. The main culture was harvested by centrifugation at 6,500 rpm (Beckman JLA-16.250) for 20 minutes. The supernatant was discarded; pellet was weighed and frozen at -80 °C until needed.

### 2.3.2 Production of recombinant protein of a pET in T7 express competent cells in minimal medium for the purpose of isotopic labelling

In order to produce isotopically enriched ($^{15}N$, $^{15}N/^{13}C$), recombinant fusion protein, the minimal medium recipe described in table 2.1 was used. Phosphate, $Na_2SO_4$ and EDTA trace elements were all added in an Erlenmeyer flask that was five times the culture volume required and sterilised by autoclaving. All other components were prepared separately as stock solutions and sterilised using 0.2 μM Minisart single use syringe filters and added aseptically to the flask prior to inoculation.

A single colony was picked from a fresh LB plate and used to inoculate 2 x 50 mL minimal medium starter cultures supplemented with the appropriate antibiotic.

The starter cultures were grown overnight at 37 °C shaking at 200 rpm overnight until $OD_{600}$ reached 0.6 - 1.0. The cells were sedimented in a benchtop-centrifuge at 3,500 rpm (The Beckman SX4750A Centrifuge with an Allegra X-15R rotor) for 15 mins and the cells pellet was resuspended in 4 mL of minimal medium without antibiotic. The 4 mL were used to inoculate 3 x 400 mL main cultures supplemented with the appropriate antibiotic. The cultures were grown at 37 °C shaking at 200 rpm until $OD_{600}$ of 0.6 was reached. From this

stage onwards, protein production in minimal medium was carried out the same as described for protein production in enriched medium (c.f.: 2.3.1).

## 2.4 Molecular Biology Methodology

### 2.4.1 Design of primers

A list of all primers designed for amplification purposes in this study can be found in Table 2.9. All primers were ordered from Sigma-Aldrich Ltd.

| Primer Name | Sequence | Enzyme Restriction Site |
|---|---|---|
| **CypA T7 forward primer** | 5′ TATCATATGGTCAACCCCACCGTGTTC 3′ | NdeI |
| **CypA T7 reverse primer** | 5′GTGCTCGAGTTAGTGATGGTGATGGTGATGGAG TTGTCCACAGTCAGCAATGGTGAT 3′ | XhoI |

**Table 2. 9. List of primers.** The restriction sites are underlined in each of the sequences. The segment of the sequence highlighted in red show the his-tag that has been used for the purpose of mutagenesis as substitution method.

## 2.4.2 Polymerase-Chain Reaction (PCR)

All PCR reactions were carried out in a Techne 3Prime Thermal Cycler using the Q5 High-Fidelity DNA Polymerase in sterile Eppendorf tubes.

The standard PCR reactions performed are described in Table 2.10. The reactions were set up on ice.

| | **µL** |
|---|---|
| **10 x Q5 Reaction Buffer** | 5 |
| **dsDNA template (clone)** | 0.2 |
| **oligonucleotide primer concentration, forward** | 1 |
| **oligonucleotide primer concentration, reverse** | 1 |
| **dNTP mix (10 mM)** | 1 |
| **nuclease-free water** | 41 |

**Table 2. 10. Standard PCR reaction**

| **Total reaction volume was 50 µL.Step** | **Temperature (°C)** | **Duration (minutes)** | **Number of cycles** | **Function** |
|---|---|---|---|---|
| **1** | 95 | 1 | 30 | Denaturation |
| **2** | 55 | 1 | 30 | Annealing |
| **3** | 72 | 1 | 30 | Elongation |
| **4** | 4 | HOLD | HOLD | HOLD |

**Table 2. 11. Standard temperature protocol of PCR reaction**

The final PCR product was analysed by agarose gel electrophoresis (c.f.: 2.4.3).

## 2.4.3 DNA Gel Electrophoresis

### 2.4.3.1 DNA Agarose Gel

The size of DNA fragments to be separated determines the percentage of agarose gel to be used [160]. In this study, 1% (w/v) agarose gel were used. An agarose gel was prepared in 1 x TBE buffer with the addition of SyBr Safe (10 μL). DNA loading dye buffer: sample was prepared on the basis of 1:10 ratio. A maximum of 20 μL sample was loaded in each well of the agarose gel, and the gel was run at 80 volts for ~ 1 hour.

### 2.4.3.2 Visualisation of DNA

DNA was stained with the SyBr Safe dye. It intercalates between the base pairs of DNA. SyBr Safe works by absorbing the UV blue light at 509 nm and emitting green light at 524 nm [161]. The DNA was observed after placing the stained DNA gel onto a transilluminator and photographed through a red filter.

## 2.4.4 Extraction and purification of DNA fragments from agarose gels

Once the DNA band at the desired size was identified, the DNA fragment was excised using a clean scalpel razor blade. The DNA was then extracted and purified from the excised gel

slice using the GeneJet Gel Extraction Kit according to the manufacturer's instructions. The concentration of DNA was assessed using nano drop measurements using the elution buffer as the nano drop blank solution.

## 2.4.5 Preparation of restriction digests

All reactions were carried out in sterile Eppendorf tubes and performed as outlined in Table 2.12. The corresponding buffers for the enzymes were used as directed by New England BioLabs information. The digested reactions were gently mixed by pipetting and centrifuged for 1 minute at low speed. The reactions were incubated at 37 °C for 2 hours.

| | **Insert Preparation (µL)** | **Vector Preparation (µL)** |
|---|---|---|
| **Mini-prepped DNA plasmid** | 20 | 5 |
| **10 x Buffer** | 5 | 3 |
| **N-terminal restriction enzyme** | 2 | 1 |
| **C-terminal restriction enzyme** | 2 | 1 |
| **nuclease-free water** | 21 | 20 |

**Table 2. 12. Standard Insert/Vector Digestion Protocol.** (Total reaction volume was 50 µL).

## 2.4.6 DNA Ligation

All DNA ligation reactions were carried out in sterile Eppendorf tubes following the standard ligation protocol as described in Table 2.13. The ligation reaction was gently mixed by pipetting and centrifuged for 1 minute at low speed. The reaction was incubated at 20 °C

overnight. The ligation mix was then incubated on the following day at 4 °C for three nights, instead of 3 hours (this was determined after several rounds of optimisation experiments). For the purpose of cloning the digested DNA insert into the vector, the ligation reaction was transformed by chemically competent cells One-Shot TOP10. 5 µL of ligation mix was added to 50 µL of TOP10 cells.

| | **µL** |
|---|---|
| **Insert DNA Digest** | 6 |
| **Vector DNA Digest** | 2 |
| **10 x T4 DNA Ligase Buffer** | 1 |
| **DNA Ligase** | 1 |

**Table 2. 13. Standard Ligation Protocol.** Total reaction volume was 10 µL.

## 2.4.7 Mini Prep

Few colonies deriving from the transformation using chemically competent cells were randomLy picked. It was crucial to carefully select colonies that were not close to each other on the LB plate. The selected colonies were individually streaked on four different LB agar plate with the same antibiotic resistance from which they were picked up in order to have replicates of the selected colonies as back-up, and immediately inoculated into 3 mL of LB medium with antibiotic. The four replica LB plates were incubated at 37°C overnight. The LB cultures were incubated at 37 °C with shaking at ~ 200 rpm overnight.

LB cultures were grown overnight and a mini-prep was performed to isolate the plasmids. The plasmids were generated using the exact same protocol as highlighted in section 2.2.7. The concentrations of the purified DNA plasmids were assessed by nano drop measurements using the elution buffer as the nano drop blank solution.

## 2.4.8 Identification of the colony containing both vector and insert DNA

### 2.4.8.1 Digestion by Insert

Each individual mini prepped plasmid was digested again with restriction enzymes. All the digestion reactions were undertaken in a sterile Eppendorf tube on ice as described in Table 2.14.

| | **µL** |
|---|---|
| **Mini-prepped DNA plasmid** | 20 |
| **10 x Buffer** | 5 |
| **N-terminal restriction enzyme** | 2.5 |
| **C-terminal restriction enzyme** | 2.5 |
| **nuclease-free water** | 20 |

**Table 2. 14. Standard Digestion by Insert Protocol.** Total reaction volume was 50 µL

Each of the reactions were gently mixed by pipetting and centrifuged for 1 minute at low speed. The reactions were incubated at 37 °C for 2 hours.

### 2.4.8.2 Sequencing

DNA sequences were verified by sequencing carried out by Beckman Coulter Genomics (Hope End, Takeley, Essex CM22 6TA) [151] with plasmids at a concentration of 100 ng/μL.

## 2.5 Biochemical Methodology

### 2.5.1 Protein Purification Methods

All the purifications were generally performed at 4 °C, but certain purification preparations were also done at room temperature at 20 °C. The purifications were undertaken on the Äkta protein purification system.

Pellets were defrosted and resuspended in the appropriate 10 x the weight of the wet-cell pellet binding buffer containing EDTA-free protease tablets. The cells were disrupted at 25 kpsi 4°C using the Constant Cell Disruption System. The lysate was centrifuged at 18,000 rpm (The Centrifuge Beckman with JA-25.50 rotor) at 4 °C for 30 minutes. The resulting pellets were discarded and the supernatant containing the protein. Buffers used varied according to the protein purified and are listed in Sections 2.1.4.

### 2.5.1.1 Immobilised Nickel Affinity Chromatography (IMAC)

IMAC is a form of affinity chromatography based on the interaction between divalent metal ions immobilised on a matrix ($Co^{2+}$, $Ni^{2+}$, $Cu^{2+}$, $Zn^{2+}$ and $Fe^{3+}$) and specific polyhistidine amino acid chains fused to the recombinant protein. This strong interaction is achieved by the donor groups on the histidine imidazole ring bonding with the immobilised metal matrix [81].

IMAC was either performed on an Äkta system using 5 mL HiTrap Chelating Columns (GE Healthcare) or by gravity flow chromatography with NiNTA matrix for very small volume preparations.

The column was equilibrated with 5-10 column volumes (CV) (Nickel Chromatography Buffer A) at 2 mL/min flow-rate and the supernatant resulting from section 2.5.1 was loaded onto the 5 mL HisTrap Chelating Column at 1 mL/min. After removal of non-specifically binding proteins with washes of 5 CV Buffer A at 2 mL/min, the recombinant protein of interest was eluted using a buffer containing of a high concentration of imidazole (Nickel Chromatography Buffer B) at 5 mL/min. The elution was monitored at $A_{280}$ nm.

### 2.5.1.2 Buffer-Exchange columns

The presence of high concentrations of imidazole or salt can destabilise the protein [82]. Therefore, proteins eluted from IMAC were desalted by buffer exchange through a 26/10 HiPrep Desalting Column into desalt buffers [82]. The column was equilibrated with 5 CV of mobile phase buffer at 5 mL/min flow-rate.

### 2.5.1.3 Ion-Exchange Chromatography (IEC)

Another common method used for purifying protein molecules is IEC, which purifies protein molecules on the basis of their ionic charge [83]. There are two different types of ion-exchange chromatography: cation-exchange chromatography consisting of a negatively-charged matrix and attracting positively charged molecules; anion-exchange chromatography containing a positively charged matrix and attracting negatively charged molecules [83]. The number and position of the ionic groups on the sample molecule and on the functional ionic group on the matrix determines the strength of the ionic interaction [83].

IEX was either performed on an Äkta system using 5 mL HiTrap Chelating Columns or gravity flow chromatography using DEAE Sepharose matrix for very small volumes preparations. The buffers used vary according to the protein's pI or the protein's acid dissociation constant (pKa). The pH of the mobile phase buffer should be between the pI or pKa of the charged molecule and the pKa of the charged group on the column matrix, and it generally must be with low salt concentrations, as salt has ionic strength. In this study, most of the ion-exchange chromatography was undertaken using weak anion-exchanger.

The column was equilibrated with 5-10 CV of IEX Buffer A at 2 mL/min. The protein sample was then loaded onto the column at 1 mL/min. After loading, the column was washed with 5-10 CV of IEX Buffer A until baseline stable at 2 mL/min flow-rate. The protein was eluted by increasing the salt concentration in the IEX Buffer B using a linear gradient at 5 mL/min. Molecules with weakest ionic interaction start to elute first, and the stronger interaction molecules come off later as the salt concentration gradually increases later in the gradient.

### 2.5.1.4 Heparin Chromatography

Heparins are known as linear polysaccharides that are negatively charged. They bind to a variety of biomolecules such as enzymes, serine protease inhibitors, growth factors, extracellular matrix proteins, DNA binding proteins and hormone receptors. Heparins are commonly used in chromatography as an ion-exchanger with high charge density [84]. For the purpose of heparin chromatography, 1 mL HiTrap Heparin Column was used. Heparin columns act the same way as cation exchanger and were used exactly in the same way, as described in section 2.5.1.3.

### 2.5.1.5 Gel Filtration Chromatography

Gel Filtration, also known as size-exclusion chromatography, is a type of chromatography separating proteins, peptides and oligonucleotides on the basis of size. The gel filtration column is made up of a bed of porous beads, through which the molecules pass at different speeds [85] [86]. Smaller molecules would diffuse into the pores of the beads, and would get retarded, whereas larger molecules enter less or completely avoid the beads and get eluted first. Gel filtration is usually the last step of any purification process, as it also serves as a method of buffer exchange [85] [86].

Gel filtration chromatography was undertaken using Superdex S75 16/60 and HiPrep Sephacryl S300 26/60 columns. The efficiency of the columns was determined by the calibration of the columns using Gel Filtration Markers Kit for Protein Molecular Weights

29,000 - 700,000 Da. The kit consists of Carbonic Anhydrase (29 kDa), Albumin (66 kDa), Alcohol Dehydrogenase (150 kDa), beta-amylase (200 kDa), Apoferritin (443 kDa), Thyroglobulin (669 kDa) and Blue Dextran (2,000 kDa). The calibration of the column was undertaken following the instructions on the kit's manual.

Columns were equilibrated with 1-2 CV of mobile phase buffer (Buffer A) at 2 mL/min. The protein sample was concentrated using 10 K MW cut-off Centricon to 2 mL and loaded onto the column. The elution was collected in 4 mL fractions at 1 mL/min.

#### 2.5.1.6 Affinity tag removal

For X-ray crystallography purposes, the polyhistidine tag was removed by incubating the eluted protein with thrombin or TEV (stock 2 units/μL) on the basis of 5 units/mL and 2 units/mL respectively at 4°C overnight. For NMR purposes, the tag was usually not cleaved.

### 2.5.2 Purity of protein samples

Purity was assessed by SDS PAGE gel. SDS-PAGE is a powerful technique that separates proteins on the basis of their size by electrophoresis by using a discontinuous polyacrylamide gel and sodium dodecyl sulfate (SDS) as a denaturation method for the proteins [87] [88].

4 x loading dye sample (LDS) was added to the protein sample. The samples were boiled at 95 °C for 15 minutes and 10 µL of the sample were loaded onto the NuPage SDS gel, along with a protein marker with defined molecular weights. The gel was run at 200 volts, 120 amps for 30 minutes. The SDS-PAGE gel was stained on Coomassie Blue for 1 hour and 30 minutes, and destained using destain buffer and distilled water until the bands were visible.

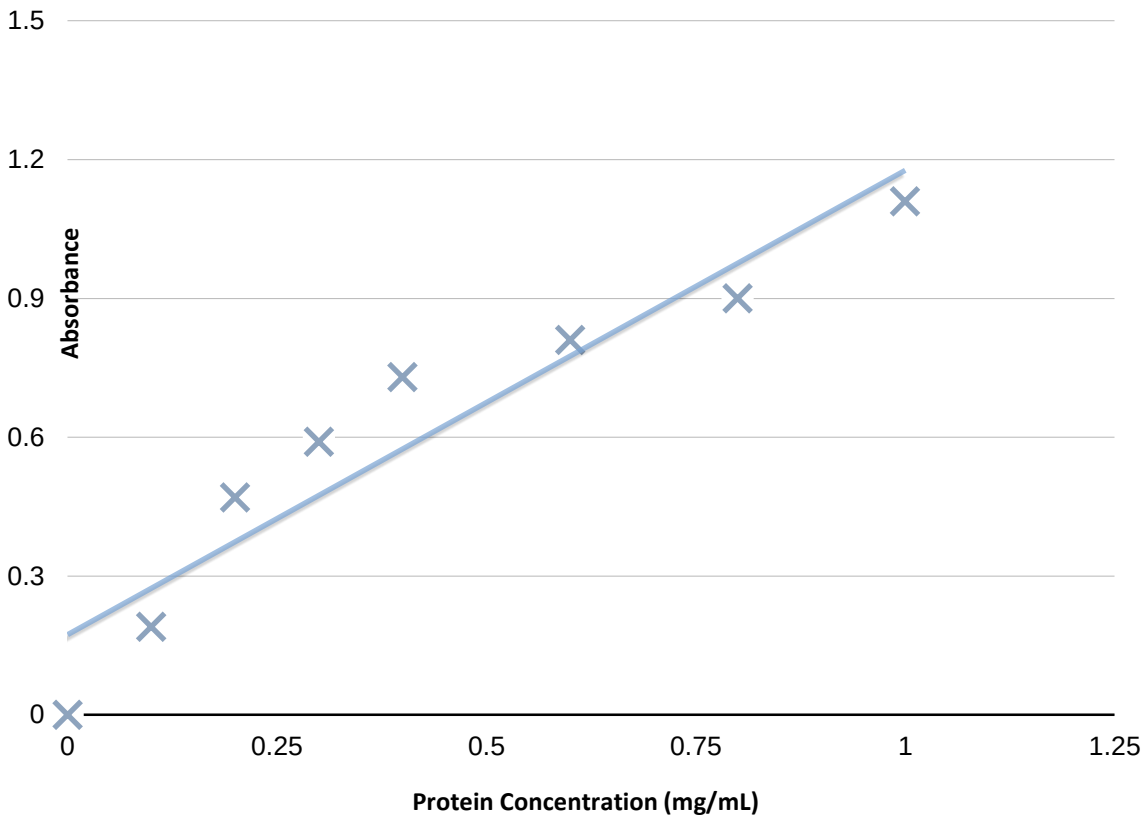



**Figure 2. 3. Chart of Bradford Assay**

Bradford Assay was performed to determine the presence of proteins eluting from gravity-flow purifications as well as determining the concentration of proteins in parallel to

nano-drop measurement. Dilution series were prepared with a standard of different concentrations of bovine serum albumin (BSA) and Bradford reagent, from which protein concentration of samples were determined.

### 2.5.3 Crystallisation

Crystallisation experiments were performed using the hanging drop or sitting drop vapour diffusion methods [144]. They were usually incubated at a constant temperature of either 4 °C or 20 °C in a crystal growth incubator (Molecular Dimensions) for a day or until growth of crystal was observed. Initial conditions were screened against additive screens Hampton Research I (HR2-138) and II (HR2-428) on the automated Mosquito system TTP LabTech. Crystallisation conditions are listed in Table 2.5.

#### 2.5.3.1 Hanging Drop Crystallisation

Hanging Drop experiments were carried out in 24 well Hampton plates. Each experiment involved siliconizing a cover slip and pipetting onto it a 2 μL drop consisting of an equal mixture of protein solution and reservoir precipitant solution. For optimisation purposes, different drop sizes, as well as protein to reservoir ratios were varied. Hanging drops were equilibrated against 500 μL or 1 mL of reservoir solution [144].

### 2.5.3.2 Sitting Drop Crystallisation

Sitting Drop experiments were carried out in 96 well Greiner plates. In each well of the plate, 2 μL drop consisting of an equal mixture of protein solution and reservoir precipitant solution were tried. For optimisation purposes, different drop size with different proportions were tried. Sitting drops were equilibrated against 320 μL of reservoir solution. The plate was sealed with plastic film (Hampton Research) [144].

### 2.5.3.3 Soaking of protein crystals with ligands

A common method of preparing a protein-ligand complex is to soak a protein crystal in reservoir solution containing an excess of ligand. Generally a ratio of 1:1 protein:ligand is initiated depending on the binding efficiency and affinity of the ligand. The ligand concentration should ideally be 10-1000 x the dissociation constant ($K_D$) of the ligand [162].

A separate crystallisation plate was prepared by adding micro-bridges to the wells that contained 1 mL of the reservoir solution. 20 μL were added to the micro-bridge and crystals were quickly transferred into the drop using loops. Ligand was then added to the micro-bridge drop, and then sealed with a coverslip. The drop was regularly checked to see if the ligand was well dissolved and the crystals were still in intact condition. Soaks were incubated for 2 days at either 4 °C or 20 °C.

### 2.5.3.4 Co-crystallisation of protein with ligands

In cases where protein-ligand complexes could not be obtained through soaking experiments, co-crystallisation experiments were set up. Here the protein is incubated with ligand before crystallisation [162] [163].

For co-crystallisation experiments, a ratio of 1:2 protein:ligand ratio was adopted. The protein sample was initially centrifuged at low speed for 5 minutes at 4 °C to remove any precipitants. Ligand was added to the protein and the mixture was incubated on ice for 30 minutes. The mixture was spun again briefly at low speed on the microfuge at 4 °C, and the crystallisation tray was prepared as described in section 2.5.3.1.

### 2.5.3.5 Crystal Archiving

For data collection purposes, the crystals were cryo-cooled in liquid nitrogen. The crystals were carefully picked up using loops, and placed on the cryoprotectant solution. The cryoprotectant solution prevents against cracking of the crystal due to the freezing conditions. The crystals were then picked up and inserted into the crystallography pucks, which were also being cooled in liquid nitrogen in a dewar [45].

### 2.5.3.6 Data Collection and Processing

All X-ray crystallography data was collected at Diamond Synchrotron Oxfordshire and processed through Xia2 software. The data were refined using Refmac5 which is part of the CCP4 package. The structures were visualised and corrected using WinCoot. The further refinement were performed with Buster program run on a Linux system [45] [138] [139].

## 2.5.4 Nuclear Magnetic Resonance (NMR)

All the NMR experiments were undertaken at the Protein NMR facility based at the University of Kent on a Bruker Avance III 600 MHz NMR spectrometer with a 4-channel and 5-amplifier configuration. The experiments were performed, processed and analysed by Dr Mark Howard, Dr Michelle Rowe and Dr Gary Thompson.

Protein samples were required at a concentration of 10 mg/mL concentration in deuterated GF Phosphate Buffer. The sample is deuterated with 5-10 % (v/v) $D_2O$ (Goss Scientific). Three different types of tubes were used: 3 mm tubes taking 200 μL sample, 5 mm Shigemi tube taking 300 μL sample, and 6 mm tube taking 600 μL sample [89]. The tubes were cleaned with distilled water and dried with compressed air after usage. The NMR data was processed and analysed using TopSpin and Kaleidograph softwares. The assignments were performed using the software Analysis 2.3.

## 2.5.5 Mass Spectrometry (MS)

MS is an analytical chemistry tool used to measure the molecular mass of a protein sample. It measures the amount and type of chemicals present in a sample [90] [91]. CD is a common spectroscopic technique that measures the chirality of proteins over a range of wavelengths [92] [93]. All the MS and CD experiments were undertaken at the Bimolecular Science Facility at the University of Kent. MS were performed using Electrospray Mass Spectrometer Bruker micrOTOF-Q and CD using Circular Dichroism Spectropolarimeter JASCO J-715. All the experiments were performed, processed and analysed by Mr Kevin Howland.

For MS experiments, protein samples with a concentration of 100 μg/mL were prepared in a volume of 100 μL.

# Chapter 3

## The Determination of the Structure of CypA and Ligand Binding Sites using X-ray Crystallography and NMR

This chapter's main emphasis is to understand the interactions of CypA with the ligands 3'aminoaetophenone, 4′ sulfamoylbenzoic Acid and 6′ aminoindazole using X-ray crystallography and various NMR methodologies, and to characterise the protein residues involved in ligand binding. The different NMR experiments are ligand-observed $^{1}$H STD, $^{1}$H WaterLOGSY and $^{1}$H CPMG binding studies, and protein-observed chemical shift perturbation using $^{1}$H/$^{15}$N HSQC.

## 3.1 Introduction about Cyclophilin A (CypA)

CypA is an 18 kDa archetypal protein that derives from the cyclophilin or immunophilin families, and is encoded by the gene Peptide Prolyl Isomerase A (PPIA) [94] [95]. It is ubiquitously distributed, structurally well-conserved in evolution [95], and present in both eukaryotes and prokaryotes, including plants, insects, fungi and bacteria [96] – [101]. The human cyclophilin family consists of 16 different cyclophilin proteins that are distinct from each other in terms of structural architecture. CypA is known as the most abundant family member within the cyclophilins, as it accounts to nearly 0.1 - 0.6 % of the total number of cytosolic proteins, and is the first identified member in mammals among the family. Unlike other cyclophilin members, CypA is found in all mammalian tissues [96] – [102].

### 3.1.1 Functional Studies of CypA

The protein's predominant function is involved in protein folding and trafficking through the activity of its peptide prolyl cis-trans isomerase (PPIase) [95]. PPIase enzymes catalyse the isomerisation of peptide bonds from trans-form to cis-form in proline residues, which contributes to the fidelity of folded proteins in the cell [103] - [105].

CypA was hypothesised to be an intracellular protein, but recent studies showed that inflammatory stimuli, such as hypoxia, infection, and oxidative stress, can also lead to the secretion of CypA. The secreted form of CypA can then act as an autocrine or paracrine factor that is able to mediate intercellular communication [106]. The growth of mouse embryonic stem cells is indeed mediated by the autocrine factor stimulation of the extracellular CypA on endothelial cells and vascular smooth muscle cells (VSMC) using pro-inflammatory signals [95] [106] [107]. Both the intracellular and extracellular forms of CypA can be involved in pathogenicity. Lower doses of extracellular CypA promote cell migration and proliferation, whereas higher doses lead to cytotoxicity, highlighting the biphasic effect of the protein [95] [108] [109].

CypA is also able to attract leukocytes, monocytes and lymphocytes using its potent chemotactic signals [95]. Apart from its protein folding and trafficking functions, CypA is also involved in many biological roles and is often defined as a protein with multi-functional properties, such as T-cell activation, T-cell assembly, immune-modulation, chaperoning and cell signalling. Helekar et al. confirmed that CypA is involved in the maturation of homo-oligomeric receptors by either acting directly or indirectly as a prolyl isomerase or as a molecular chaperone [95] [110] [111]. The protein is therefore involved in several pathological conditions, mostly inflammatory

disorders [95]. Examples include the role of CypA in the pathogenesis of abdominal aortic aneurysm, atherosclerosis and other cardiac diseases [95] [112]. CypA has been shown to be also involved in the pathogenesis of diabetes [113], virology [95] [114] – [116], cancer [117] – [120], neurodegenerative disorders (Alzheimer's Disease) [121], rheumatoid arthritis [122] – [125], sepsis [126], asthma [127], periodontitis [128] and ageing [129] [130].

## 3.1.2 Structure of CypA

Human CypA consists of two α-helices that surround an eight-stranded anti-parallel β-barrel structure, which itself encloses a hydrophobic core made up of seven aromatic rings and hydrophobic residues (18 kDa). This hydrophobic core region is considered as the active site of the protein, and is where most inhibitors bind (Figure 3.1) [95] [131] [132].

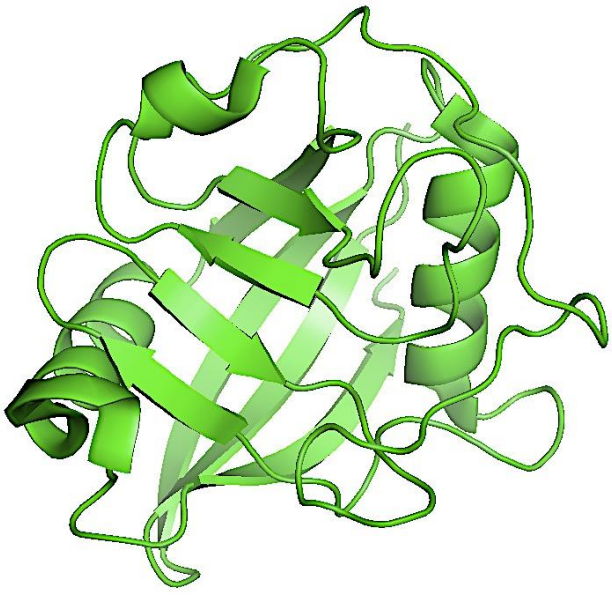


**Figure 3.1. CypA crystal structure (PDB: 3K0O)** Ribbon representation of CypA crystal structure (PDB: 3K0O) showing two α-helices surrounded by eight-stranded anti-parallel β-barrel structure.

### 3.1.3 CypA as drug target

The immunosuppressive drug Cyclosporine A (CsA) binds to CypA in a ternary complex with calcineurin to exert its inhibitory effects, and is the most well characterised and tightest binding inhibitor of CypA identified so far. CsA is able to inhibit both the cytosolic and secreted form of CypA, and exerts its effects by blocking the activity of the PPIase [133] [134]. CsA binds to the hydrophobic core region ranging from His 114 to Lys 121, and four β-strands ranging from β3-β6. The CsA-CypA complex then binds to the interface between the catalytic and regulatory subunits of calcineurin. A crystal structure of the CypA-CsA complex has been successfully resolved at 2.1 Å (Figure 3.2) [95] [133] [134].

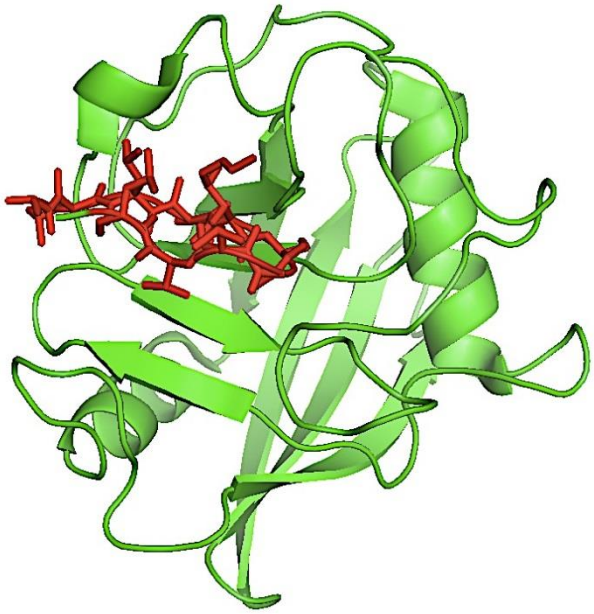


**Figure 3.2. Crystal structure of CypA complexed with Cyclosporine (CsA) (PDB: 1CWA)** CypA is shown as green ribbon and CsA as red sticks.

CsA inhibits the calcium-calmodulin activated serine/threonine-specific protein phosphatase calcineurin, preventing the biological function of calcineurin and the translocation of nuclear factors in activated T-cells (NF-AT) from the cytosol to the nucleus (Figure 3.3). This collectively results in the prevention of the transcription of genes responsible for encoding pro-inflammatory cytokines [133] [134]. The binding of the CypA-CsA complex to calcineurin has been shown to stabilise the complex itself, and the proteins in the complex show increased resistance to damage by proteolysis. When CsA binds to CypA, the charges and hydrophobicity of the complex become conjoined with the binding site of calcineurin [133] [134].

A number of inhibitors of CypA are known of which NIM811 & MM218 are the most important. NIM811 is a derivative of CsA and is known to act as an anti-inflammatory inhibitor of CypA. It is used in the treatment of inflammatory disorders, such as arthritis (Figure 3.3) [133].

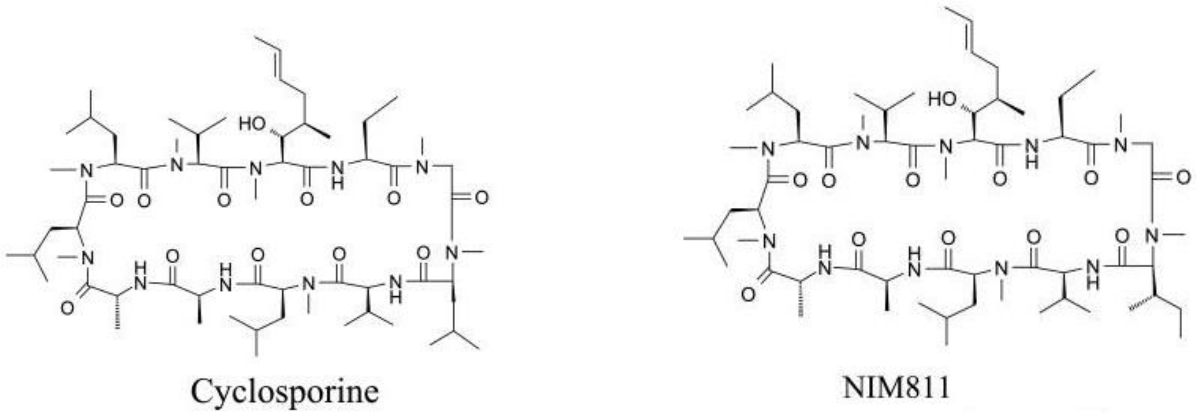



**Figure 3.3. Chemical structure of Cyclosporine (CsA) and its derivative NIM811.** *Adapted from*: Ma, S., Boerner, J. E., TiongYip, C., Weidmann, B., Ryder, N. S., Cooreman, M. P., & Lin, K. (2006). NIM811, a cyclophilin inhibitor, exhibits potent in vitro activity against hepatitis C virus alone or in combination with alpha interferon. *Antimicrobial agents and chemotherapy*, 50(9), 2976-82.

MM218 is another potent anti-inflammatory inhibitor of CypA, which selectively inhibits the extracellular fraction of CypA. In comparison to CsA, it has a better mode of action and anti-inflammatory effectivity in animal models with acute lung diseases and myocardial repercussion injuries [134].

Due to its involvement in many cellular processes, CypA displays an important target in drug discovery. It is a particular challenge to discover specific therapeutic agents to inhibit particular functions without affecting others and causing cross-reactivity with other cyclophilin isoforms [97] [133] [134].

Apart from drug discovery itself, it is crucial to identify other mechanisms by which CypA achieves its role in inflammatory diseases. It is currently known that CD147 plays a key role with CypA to lead to a number of inflammatory-related pathways. However, it is unlikely that it is the sole receptor to unequivocally mediate all the cellular responses associated with CypA. Therefore, determining other receptors for CypA is essential to understand the CypA pathway

better, and to optimally design therapeutic drugs that prevent the binding of CypA to those receptor targets in the treatment of inflammatory diseases [95] [133] [134]. Further structural characterisation by X-ray crystallography would serve to improve the current understanding of the mechanisms underlying specific interactions of cyclophilins with different targets.

## 3.2 Recombinant expression and purification of CypA from *E.coli* and crystallisation

### 3.2.1 Recombinant expression and purification

Recombinant human CypA (165 residues) was expressed and purified to homogeneity. CypA was expressed from a pET21a ampicillin resistant plasmid using *E.coli* T7 express cells grown in 2 x YT media. A bacterial pellet of 8 g was obtained from 1 L culture. The cells were lysed by sonication. The first step of purification involved ion-exchange chromatography using a 5 mL Q SP Hi-Trap column [83]. The column was equilibrated with buffer A (50 mM HEPES pH 6.8, 1 mM DTT, 2.5 mM EDTA and 100 μM phenylmethanesulfonyl fluoride) and the lysate applied. Protein was eluted using buffer B (50 mM HEPES pH 6.8, 1 mM DTT, 2.5 mM EDTA, 400 mM NaCl and 100 μM phenylmethanesulfonyl fluoride) for elution, as described in 2.5.1.3 (Figure 3.4).

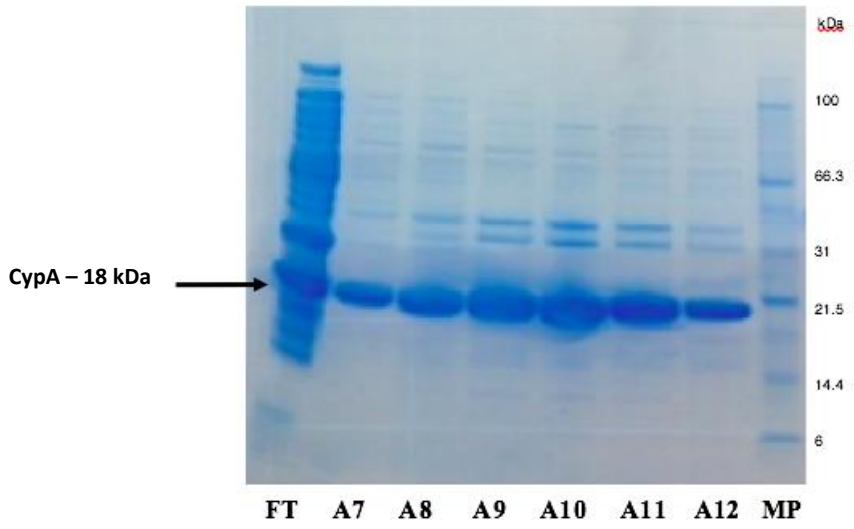



**Figure 3. 4. Purification of CypA by ion-exchange chromatography** Fractions from the ion exchange column applied to a 4-12 % (w/v) Bis-tris SDS-PAGE. The gel shows ~ 70 % purity of CypA at the expected molecular weight of 18 kDa after ion-exchange chromatography. FT: flow-through, MP: marker protein, A7-A12: fractions from ion-exchange of CypA protein.

The pooled fractions were then concentrated down to 2.5 mL. The final step of the purification of CypA was gel filtration chromatography using GF Buffer on a Superdex S-75 Column using 25mM tris-HCl (pH 7.5), 100mM NaCl, 0.5 mM DTT and 0.5mM EDTA for both equilibration and elution [85], as described in 2.5.1.5 (Figure 3.5). The protein was well expressed and the final purified yield was 25 mg per litre.

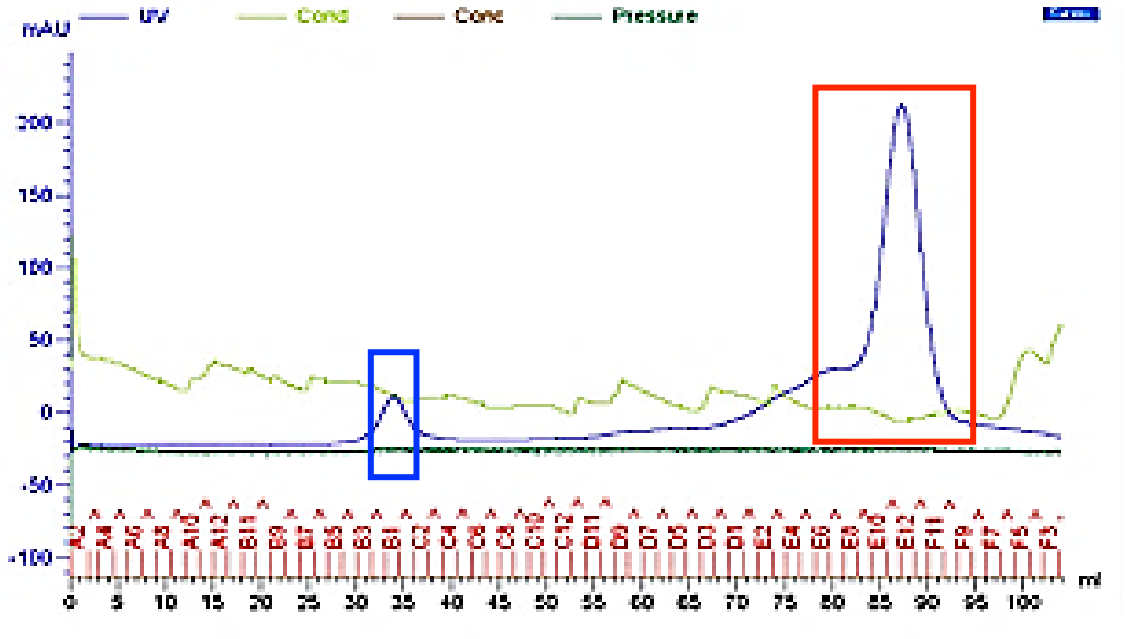



**Figure 3. 5. Elution Profile of CypA using a Superdex S-75 Column Gel Filtration Chromatography** The blue trace shows absorption at 280nm in mAU. The peak at fractions B3 - C2 (blue rectangle) is from the elution of aggregated protein, and peak at fractions E8 - F8 (red rectangle) is from the elution of the pure protein CypA.

## 3.2.2 Protein Crystallisation

The crystallisation of CypA was performed using crystallisation conditions described by Pflugl G. et al. (1993) [135]. CypA was concentrated to 16 mg/mL, equivalent to ~ 0.9 mM for crystallisation (2.5.3.1). The best crystals for diffraction were obtained from conditions with 35% - 44% ammonium sulfate at pH 8.5 after 24 hours at 20 °C (Figure 3.6). The CypA crystals have the morphology of thin needles. The crystals diffracted to 2.5 Å on beamline I04 at a wavelength of 0.97949 Å, and a temperature of 100 Kelvin with a 100 % exposure for 0.1 seconds / 0.1º oscillation.

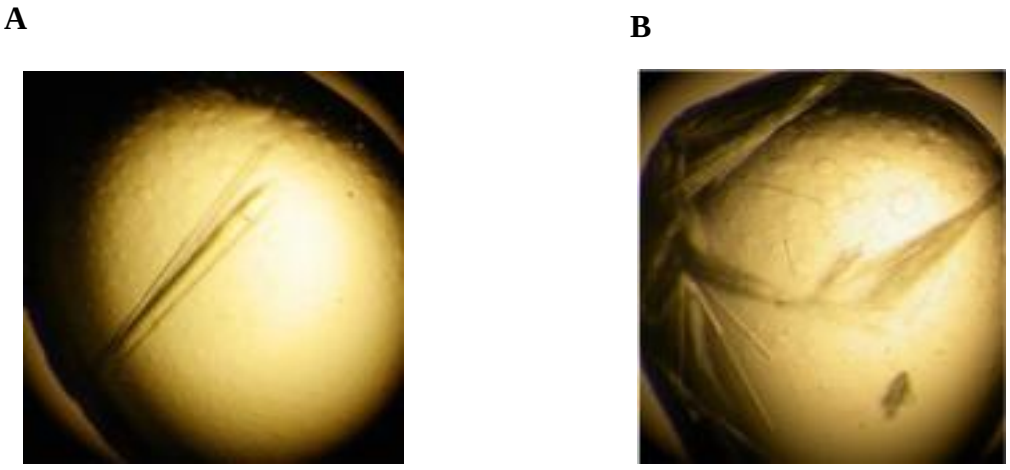



**Figure 3. 6. CypA crystals were grown under conditions of 40 to 41 % ammonium sulphate** A) 40 % and B) 41 % ammonium sulfate pH 8.5.

## 3.3 Crystallographic Solutions of CypA

### 3.3.1 Addition of Ligands

Compounds for soaking were identified by Dr Samantha Lodge (NMR Facility, The University of Kent) from a range of compounds known to bind to the mitochondrial form of CypA (CypD). The three best hits were selected for crystallographic studies and had $K_D$s within the concentrations range of micro-molar to nano-molar (Table 3.1). Previous unpublished work has proven that 4′ sulfamoylbenzoic Acid is able to bind to CypA (Dr Samantha Lodge).

| Ligand | SPR ($K_D$) | Types of crystallisation | Soak | Co-Crystal |
|---|---|---|---|---|
| **3′ aminoacetophenone** | 0.06 mM | Soak / Co-Crystal | × | ✓ |
| **4′ sulfamoylbenzoic Acid** | > 1 mM | Soak | × | × |
| **6′ aminoindazole** | > 1 mM | Soak / Co-Crystal | ✓ | ✓ |

**Table 3. 1. Summary of the soak and co-crystallisation experiments with CypA** ✓ Data obtained, × No data obtained

The three compounds were dissolved to give 30 mM stocks in DMSO and crystal soaks were prepared at a ratio of 3 : 1 ligand to protein excess (0.9 mM) for each of the compounds. Crystals were soaked for 2 days at 20 °C before cryo-cooling in liquid nitrogen with 40 % ammonium

sulfate cryosolvent prior to data collection. Immediate archiving of the CypA-4′ sulfamoylbenzoic acid soak was necessary to prevent the crystals from dissolution. This could be due to the compound binding to the protein, and causing conformational changes leading to alterations in the crystal morphology. The co-crystals were prepared at a ratio of 2 : 1 ligand to protein excess under the same conditions as the apo crystal crystallisation conditions.

The crystals soaked with 3′aminoacetophenone did not diffract, possibly due to DMSO intolerance. No data could be collected for 4′ sulfamoylbenzoic Acid due to bad crystal quality from the soaks and no crystals were grown from co-crystallisation.

### 3.3.2 Data collection and processing

All data was collected at DIAMOND on the IO4.1 beamline at a wavelength of 0.97949 Å, at a temperature of 100 Kelvin with a 100 % exposure for 0.1 seconds / 0.1º oscillation. The space group was orthorhombic $P3_1$ 2 1 (Figure 3.7).

Data was auto-processed through Xia2. A simple molecular replacement with Phaser into $P3_1$ 2 1 was performed using the apo structure of CypA as a search model. Ten rounds of structural refinement using Refmac5 [136] were performed and the difference density maps were examined for presence of the ligand. Once the presence of the ligand was identified, the structure was finalised with Buster [137] [138]. The crystallographic data is summarised in Table 3.2 and 3.3.

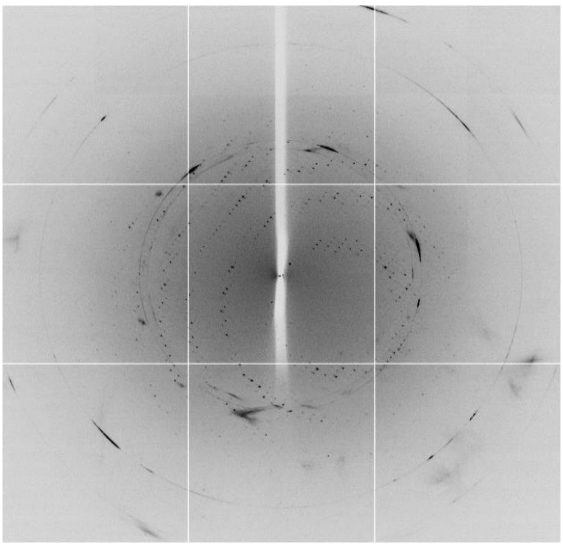


**Figure 3. 7. X-ray diffraction pattern of CypA.** Data reduction software including *xia2* and *fast_dp* were used to detect the space group as P $3_1$ 2 1 [138] [139]. The diffraction intensities observed result from diffraction of the X-ray beam passing through the crystal which is a repeating array of CypA proteins.

## 3.3.3 Crystallographic Data, Refinement and Ramachandran Plot Statistics with ligands

Tables 3.2 and 3.3 outline a summary of the crystallisation data collection and structural refinement details of CypA. The data was solved and refined through Phaser-Molecular Replacement, Refmac5 and Wincoot.

| **Compound** | **APO** |
|---|---|
| Space Group | P $3_1$ 2 1 |
| Presence of Ligand | No |
| Unit cell dimensions (Å) (a,b,c) | 59.98, 59.98, 94.18 |
| Unit Cell Angle (°) | 90, 90, 120 |

| | |
|---|---|
| Number of observed reflections | 13,948 |
| Number of unique reflections | 9,758 |
| Oscillation range per frame (°) | 0.1 |
| Overall rotation (°) | 180 |
| Resolution Range (Å) | 47.09 – 2.25 |
| Multiplicity | 7.2 (6.3) |
| Completeness (overall) (%) | 99.4 |
| Mean I/sigma (I) (overall and last shell) | 14.9 (2.4) |
| Rmerge (%) (overall and last shell) | 16.93 |

**Table 3. 2. Summary of the X-ray crystallography data collection of CypA.**

| **Compound** | **APO** |
|---|---|
| Maximum Resolution ( Å) | 2.5 |
| R-factor (%) (last shell) | 17.65 |
| R-free (%) | 23.07 |
| RMSD Bond Lengths (Å) | 0.0154 |
| RMSD Bond Angles (°) | 1.6791 |
| Average B-factor Protein ($Å^2$) | 38.04 |
| Number of Protein Residues | 164 |

| | |
|---|---|
| Number of water atoms modelled | 22 |
| $SO_4^{2+}$ (number of atoms) | 1 |
| Ramachandran Plot preferred (%) | 96.30 |
| Ramachandran allowed (%) | 3.70 |
| Ramachandran generously allowed (%) | 0.00 |

**Table 3. 3. Summary of the X-ray crystallography structure refinement and Ramachandran Plot statistics of CypA.**

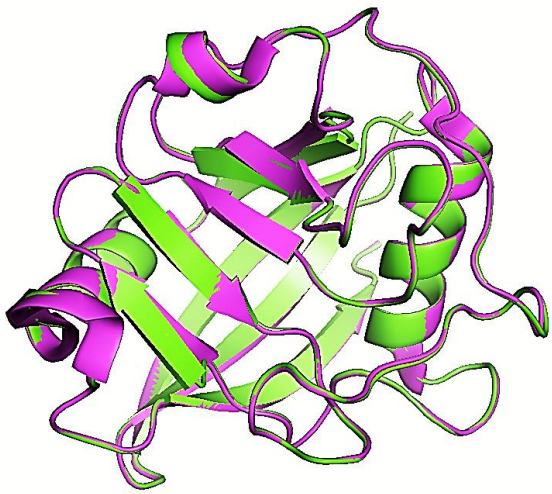


**Figure 3. 8. Overlay of CypA refined coordinates crystal structure as deposited 3K0O.** The crystal structure of CypA consists of two α-helices surrounded by eight-stranded anti-parallel β-barrel structure, as expected (Figure 3.1). The overlay of the ribbon crystal structure of CypA refined coordinates (green) with published crystal structure of ribbon CypA 3K0O (magenta) shows a near perfect match. (all heavy atom RMSD = 0.421) [140] [141].

Figure 3.8 confirms that the crystal structure of CypA matches with the published crystal structure of CypA (3K0O), with the presence of two α-helices surrounded by an eight-stranded anti-parallel β-barrel structure [140].

Neither of the ligands 3′ aminoacetophenone or 6′ aminoindazole bind to CypA. The results suggest that the soak concentration was not high enough for the binding to take place. A closer look into the active site was undertaken to investigate the difference between the ligand binding sites of CypA and CypD (Figures 3.9).

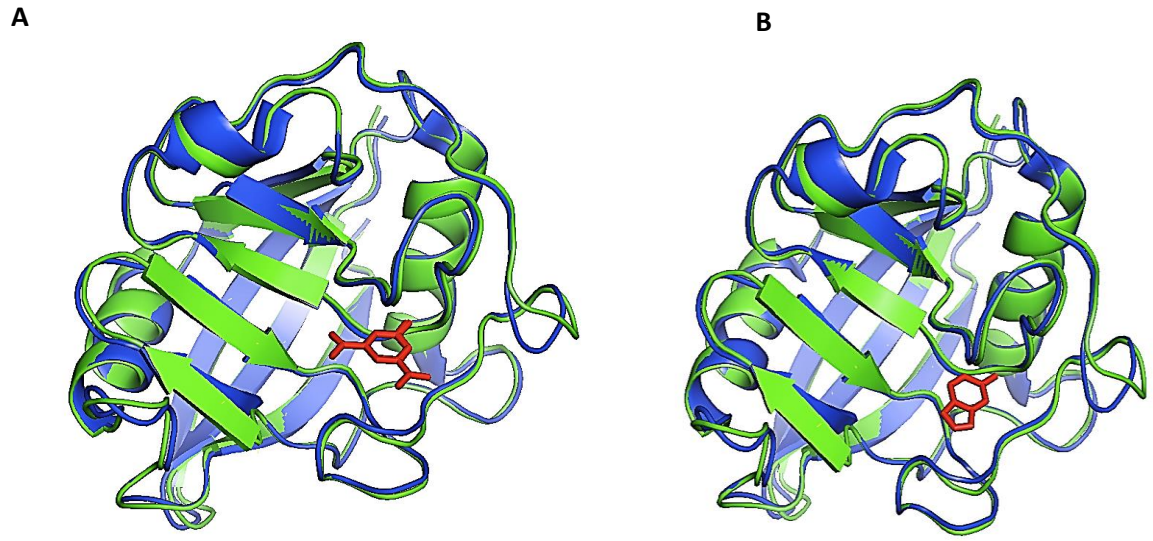



**Figure 3. 9. Comparison of CypA refined coordinates crystal structure with CypD complexed with the ligands**

A) Overlay of CypA refined coordinates (green) with ribbon CypD (blue) complexed with deposited 3′ aminoacetophenone (red stick ligand) (PDB Code: 3R59). B) Overlay of the refined coordinates for CypA (green) and CypD (blue) complexed with deposited 6′ aminoindazole (3R56). The overlays show the absence of 3′ aminoacetophenone and 6′ aminoindazole in the green CypA structure, confirming that the ligands did not bind the protein.

A B

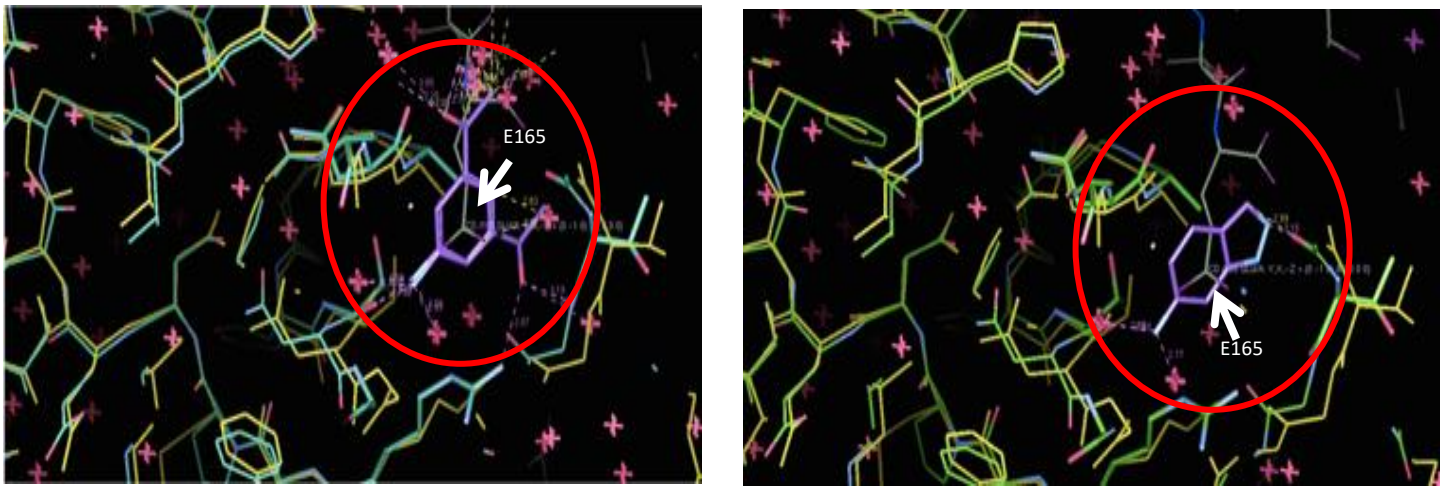



**Figure 3. 10. Comparison of the active site of CypA with CypD complexed with ligands** A) Overlay of CypA (yellow) with CypD complexed with 3′ aminoacetophenone (3R59) (cyan). B) Overlay of CypA (yellow) with CypD complexed with 6′ aminoindazole (3R56) (green).

The overlays displayed in Figure 3.10 show the position of the ligands within active site of CypD. However, the same ligands did not bind to CypA. When comparing CypA to CypD, it is observed that glutamic acid 165 (E165) at the C-terminus of CypA, occludes the active site. The crystal packing of CypA causes E165 to obscure the active site of the protein and prevents the binding of any ligands by blocking the binding site. As a result, an alternative crystal form, such as small block crystals, was sought, which could also be suitable for in-situ crystallisation at room temperature.

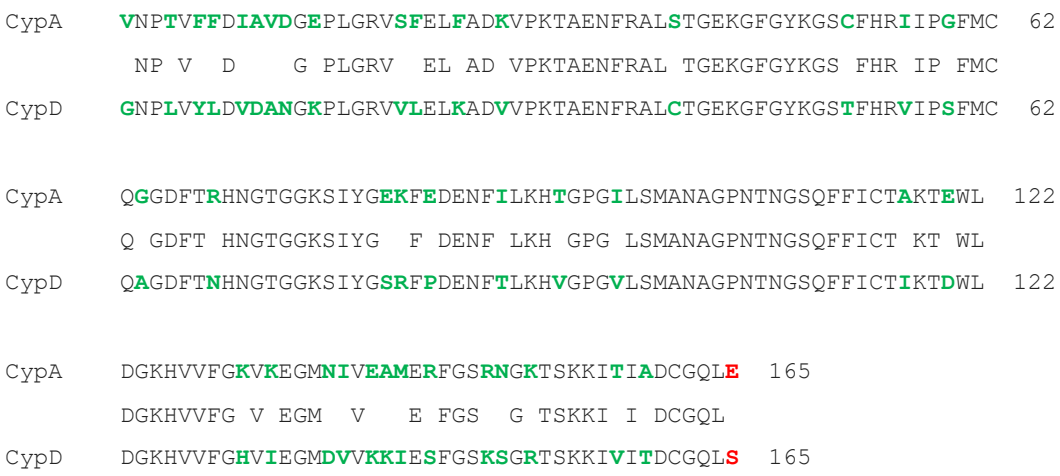



**Figure 3. 11. Comparison of CypA and CypD protein sequences.** The protein sequence of the CypA sample used in the experiments was compared with CypD (Uniprot ID: P30405). The differences between the two sequences are highlighted in green. The most interesting change is shown in red, where E165 is substituted to S165 in CypD, which may explain why the ligands do not bind to CypA.

The protein sequences of CypA and CypD were compared (Figure 3.11) and it was clear that the E165 does not exist in CypD, and therefore it may suggest that this prevents the binding of the ligands to CypD in comparison to CypA.

## 3.4 Protein Crystallisation Optimisation

### 3.4.1 Crystallisation Conditions

Several different conditions under which CypA crystallises have been published. Screening trials were set up around these conditions to improve crystals quality. The different conditions attempted were as follows:

- 10 – 41 % ammonium sulfate Conditions, pH 8.5 [137], [CypA] = 16 mg/mL, at 20 °C [142].
- 14 – 24 % polyethylene glycol (PEG) 8000, pH 7.2 – 8.2 [138], [CypA] = 10 mg/mL and 16mg/mL, at 20 °C [143].
- 1.2 M – 2.2 M malic Acid, pH 7.2 – 8.2 [79], [CypA] = 10 mg/mL at 20 °C (PDB Code: 3K0O) [80].

Unfortunately, the changes in the conditions did not improve the morphology of the CypA crystals.

Sitting-drop experiments were set up with conditions, as described in 3.2.2 and 3.2.3 (2.5.4.2), using 320 μL of reservoir solution [144]. However, CypA crystals grown from sitting-drop experiments did not improve in quality.

### 3.4.2 Micro-seeding experiments

Micro-seeding is an experiment to improve crystal quality. Small splinters of the original crystal are used to introduce a nucleation point. For this purpose, CypA needle crystals were harvested and re-suspended in mother reservoir solution. Seeds can be introduced in two different ways; either by a process called streak micro-seeding or dilution micro-seeding. In the streak micro-seeding method, seeding is achieved by picking a small proportion of the reservoir solution, and touching it to the smashed crystals before returning to the reservoir which is then sealed. Dilution micro-seeding involves the creation of a dilution series from the initial mother liquor containing the micro-seeds [145]. The number of dilutions in the series can be varied. Unfortunately, both of these techniques did not result in an improvement of crystal morphology or quality.

### 3.4.3 Additive Screen Experiments

The Hampton Research I (HR2-138) and II (HR2-428) additive screens were used on an automated Mosquito system TTP LabTech, with 96 different conditions, designed to stabilise the protein and improving the solubility of the sample resulting in an improved crystal packing. The additive screens consisted of 1 mL of 96 different additives, containing multivalent cations, salts, amino acids, dissociating agent linkers, polyamines, chaotropes, co-factors, reducing agents, polymers, chelating agents, carbohydrates, polyols, detergents, non-detergents, amphiphiles, osmolytes, organics (non-volatile) and organics (volatile) [146]. These agents did not help in the improvement of CypA crystal quality.

### 3.4.4 Concentration of the protein sample

In an attempt to improve crystal quality, a series of crystallisation were performed with CypA concentrations [147] ranging from 5 mg/mL (0.3 mM) to 25 mg/mL (1.4 mM) (Figure 3.12). Drops from 5 mg/mL to 8 mg/mL CypA concentration were clear and did not show any evidence of crystallisation. Similarly drops from 18 mg/mL to 25 mg/mL CypA concentration had only evidence of precipitation.

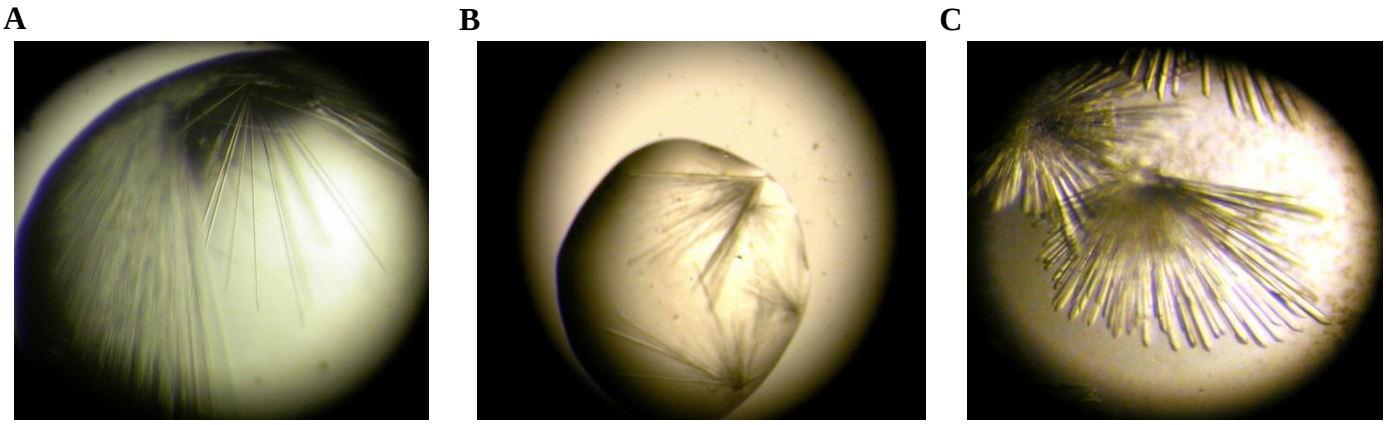



**Figure 3. 12. Variation of the concentration of CypA.** A) 16 mg/mL (1 mM) CypA at 43% ammonium sulfate pH 7.2 20 °C. B) 12 mg/mL (0.7 mM) CypA at 40% ammonium sulfate pH 7.2 20 °C. C) 8 mg/mL (0.5 mM) CypA at 40% ammonium sulfate pH 7.2 20 °C

Although block-like CypA crystals could not be obtained using this approach, sharper and medium-quality crystals were possible with a CypA concentration of 8 mg/mL. However, the crystal packing problem was still not resolved with the new crystal form (3.12C). It was therefore decided to perform in-situ mutagenesis experiments to remove C-terminal E165 residue of the CypA sequence.

## 3.5 Mutagenesis of Cyclophilin A

### 3.5.1 Sequence of wild-type CypA

The uniprot data of wild-type (identifier: P62937-1) did not show any evidence of post-translational modifications occurring at the residue of C-terminal E165 (Figure 3.13). E165 does not contribute to the formation of any secondary structural motifs. Therefore, truncation of E165 is not expected to affect the structure of the protein [148].

```
        10         20         30         40         50
MVNPTVFFDI AVDGEPLGRV SFELFADKVP KTAENFRALS TGEKGFGYKG
        60         70         80         90        100
SCFHRIIPGF MCQGGDFTRH NGTGGKSIYG EKFEDENFIL KHTGPGILSM
       110        120        130        140        150
ANAGPNTNGS QFFICTAKTE WLDGKHVVFG KVKEGMNIVE AMERFGSRNG
       160
KTSKKITIAD CGQLE
```

**Figure 3. 13. Protein sequence of wild-type CypA.** The protein consists of 165 residues, and has a mass of 18,012 Daltons. This sequence belongs to the isoform 1 (identifier: P62937-1) [148].

Interestingly, a number of mutagenesis experiments have already been performed on CypA before. The experiments are all substitution experiments, at positions 89, 106 and 165 [149] – [150]. The residue E165 had been substituted previously for D165. The previous substitution performed at E165 is enough evidence to suggest that the residue can be subjected to directed mutagenesis without causing effects to the function of the protein [150].

## 3.5.2 Directed mutagenesis of wild-type CypA

The CypA gene was subjected to in-situ mutagenesis to remove the C-terminal residue E165. Successful mutagenesis was verified by an analysis carried out by Beckman Coulter Genomics [151]. The gene sequence was translated into a protein sequence and aligned with the wild-type CypA protein sequence in order to identify the differences using the vector NTI software tool (Figure 3.14) [152].

```
Wild-type CypA (1) MVNPTVFFDIAVDGEPLGRVSFELFADKVPKTAENFRALSTGEKGFGYKG (50)
Mutant CypA    (1) MVNPTVFFDIAVDGEPLGRVSFELFADKVPKTAENFRALSTGEKGFGYKG (50)

Wild-type CypA (51) SCFHRIIPGFMCQGGDFTRHNGTGGKSIYGEKFEDENFILKHTGPGILSM (100)
Mutant CypA    (51) SCFHRIIPGFMCQGGDFTRHNGTGGKSIYGEKFEDENFILKHTGPGILSM (100)

Wild-type CypA (101) ANAGPNTNGSQFFICTAKTEWLDGKHVVFGKVKEGMNIVEAMERFGSRNG (150)
Mutant CypA    (101) ANAGPNTNGSQFFICTAKTEWLDGKHVVFGKVKEGMNIVEAMERFGSRNG (150)

Wild-type CypA (151) KTSKKITIADCGQLE----- (170)
Mutant CypA    (151) KTSKKITIADCGQLHHHHHH (170)
```

**Figure 3. 14. Comparison of Wild-type CypA and Mutant CypA protein sequences**

## 3.5.3 Recombinant Expression and Purification of the mutant CypA

Recombinant human mutant CypA (170 residues) was expressed and purified to homogeneity. CypA was expressed from a pET21a ampicillin resistant plasmid using *E.coli* T7 express cells

grown in 2 xYT media. A bacterial pellet of 11 g was obtained from 1 L culture. The cells were lysed by sonication. The first step of purification involved nickel-affinity chromatography using a 5 mL His-Trap HP column [81]. The column was equilibrated with Buffer A (50 mM tris-HCl pH 7.8, 300 mM NaCl, 30 mM imidazole and 1 mM 2-beta mercaptoethanol) and the lysate applied. Protein was eluted using Buffer B (50 mM tris-HCl pH 7.8, 300 mM NaCl, 30 mM imidazole, 500 mM imidazole and 1 mM 2-beta mercaptoethanol) and straight through a 26/10 HiPrep Desalting Column using Buffer C (25 mM $NaH_2PO_4$, 0.5 mM DTT pH 6.5), as described in 2.5.1.2 (Figure 3.15 and Figure 3.16).

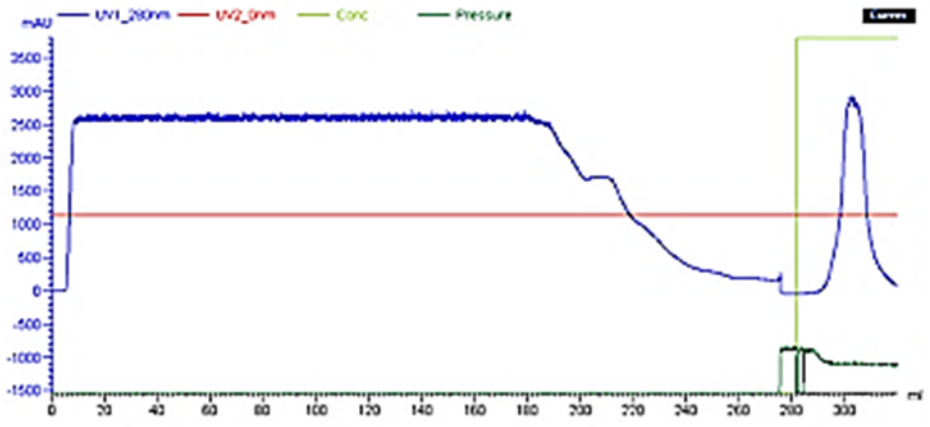



**Figure 3. 15. Elution Profile of mutant CypA on His-Trap 5 mL Immobilised Nickel Affinity Chromatography and 26/10 HiPrep Desalting Column in serial.**

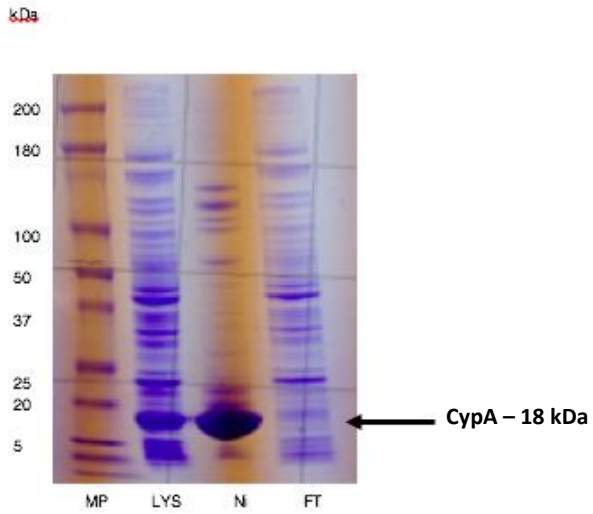



**Figure 3. 16. Purification of mutant His-CypA using Nickel-Affinity Chromatography.** The 4-12 % (w/v) Bis-tris gel shows ~ 70 % purity of CypA at the expected molecular weight of 18 kDa after immobilised nickel affinity chromatography. MP: marker protein, LYS: Lysate, Ni: Nickel Elution, FT: Flow-through.

The final step of the purification of CypA was gel filtration chromatography [85] using GF Buffer (25 mM tris-HCl (pH 7.5), 100 mM NaCl, 0.5 mM DTT and 0.5 mM EDTA) on a Superdex S-75 Column for both equilibration and elution, as described in 2.5.1.5 (Figure 3.17 and Figure 3.18). The protein was well expressed and the final purified yield was 50 mg per litre.

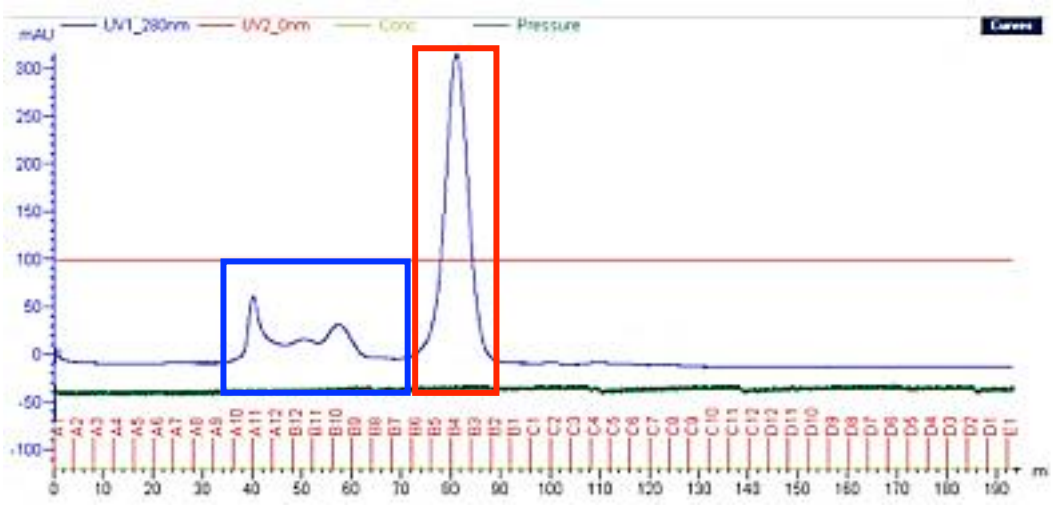



**Figure 3. 17. Elution Profile of mutant CypA on Superdex S-75 Column Gel Filtration Chromatography.** Fractions A10 - B9 (blue rectangle) show the elution of aggregated protein, and fractions B6 - B2 (red rectangle) show the elution of the pure protein, clearly demonstrating a good separation.

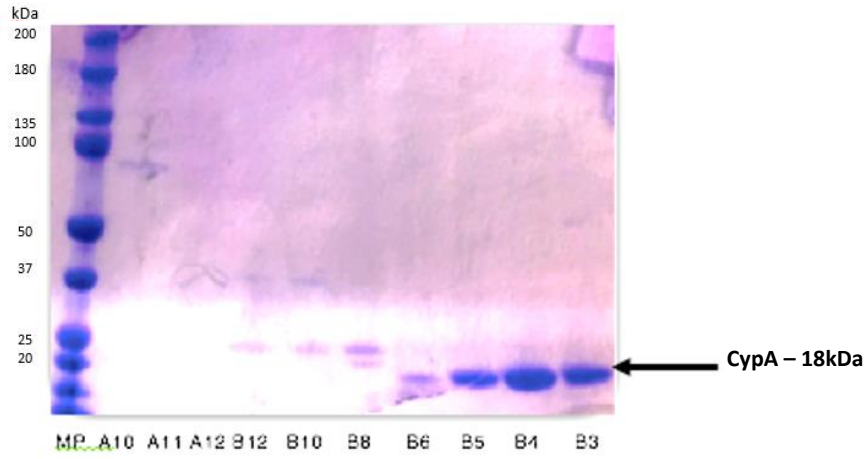



**Figure 3. 18. Purification of mutant His-CypA by Gel Filtration Chromatography.** The purified fractions on the 4-12 % Bis-tris gel shows ~ 90 % purity of protein at the expected molecular weight of 18 kDa after gel filtration chromatography. MP: marker protein. Fractions A10-B6 represent aggregates and fractions B5-B3 represent the mutant protein.

## 3.5.4 Mass spectrometry of mutant CypA

Electrospray mass spectrometry (MS) was performed to measure the molecular mass of the mutant CypA [73] [90] [91]. Figure 3.19 shows the mass spectrum of CypA with the presence of two peaks.

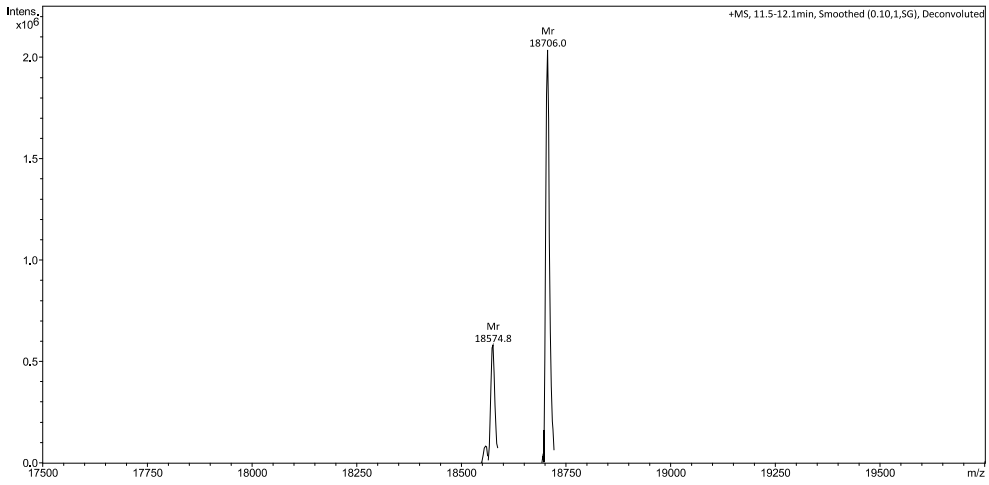



**Figure 3. 19. Mass spectrometry spectrum plot of mutant CypA.** The results show the presence of two observable peaks. The base peak, which is the most intense ion, is Mr = 18,706.0 Da, the second peak has a molecular mass of Mr = 18,574.8 Da.

Two peaks were observed at 18,706.0 Da and 18,574.8 Da. The major peak corresponds to the mutant CypA (18,706.23 Da). The minor peak at 18.5 Da has a difference of 131.2 from the major peak. The relative ratio mass corresponds to the loss of methionine, which has a molecular weight of 131.19 Da. Loss of methionine from the N-terminal due to methionine amino-peptidase is a common post-translational modification in *E.coli* [153].

## 3.5.5 Crystallisation of the mutant CypA

After verification of purity and the exact mass of mutant CypA, the protein was used for crystallisation trials using the same conditions as for the wild-type CypA. Since the residue E165 was removed, a change in crystal packing can be expected. The protein was therefore subjected

to further optimisation, including trials of previously used conditions, protein concentration variation and additive screens. None of these techniques led to crystal formation. It appears that the single amino acid truncation resulted in a complete change in the crystallisation pattern of the protein. Alternatively, the presence of the his-tag could have also prevented the crystallisation of the protein, as the protein sequence is modified. Therefore, although site-directed mutagenesis was successful, the binding of the CypD ligands into the mutant CypA could not be validated using X-ray crystallography.

This result shows the utility of NMR, which does not rely on the additional step of protein crystallisation, and could give an answer of binding modes while the protein is in solution. The lack of a crystal structure should not always be a barrier for understanding the structure of a protein-ligand complex and the interactions of residues with ligands, which in this case could not be achieved using X-ray crystallography with CypA.

## 3.6 NMR screening methods on protein-ligand binding with CypA

Various NMR methodologies were performed to determine ligands "hits" to the protein, including ligand-observed $^{1}$H STD, $^{1}$H WaterLOGSY and $^{1}$H CPMG binding studies, and protein-observed $^{1}$H/$^{15}$N HSQC spectra.

### 3.6.1 NMR Sample Preparation of CypA/3′aminoacetophenone and CypA/ 4′ sulfamoylbenzoic Acid complexes

The binding of 3′aminoacetophenone and 4′sulfamoylbenzoic Acid with CypA was evaluated using STD [56] [59], WaterLOGSY [154] and CPMG experiments [155] from a sample expressed in *E.coli* T7 express cells grown in rich LB media and purified into GF Buffer (25 mM tris-HCl (pH 7.5), 100 mM NaCl, 0.5 mM DTT and 0.5 mM EDTA) as described in Chapter 2.5.1.5. Binding experiments were carried out with 0.2 mM CypA with the addition of either 8 mM 3′aminoacetophenone and 4′ sulfamoylbenzoic Acid.

### 3.6.2 NMR Sample STD Data Acquisition

$^{1}$H STD experiments [56] [59] were acquired on a Bruker Avance III 600 Hz spectrometer at 25°C, with 16,384 points (spectral width 9615 Hz) in the direct F1 dimension ($^{1}$H), with a total number of 256 scans were acquired with an acquisition time of 0.85 seconds. Saturation was achieved using EBURP2 pulses of length 20 ms or 50 ms for 3′aminoacetophenone and 4′sulfamoylbenzoic acid, respectively. The type of water suppression used was excitation sculpting and the spectra were referenced against water. Saturation was applied at -0.7 / -30 ppm for the on and off resonance experiments respectively. The relaxation delay time (D1) in the experiments was 2.5 seconds. Data were zero filled and Fourier transformed using TopSpin 3.2 and an exponential line broadening of 0.30 Hz was used (Figures 3.20 and 3.21).

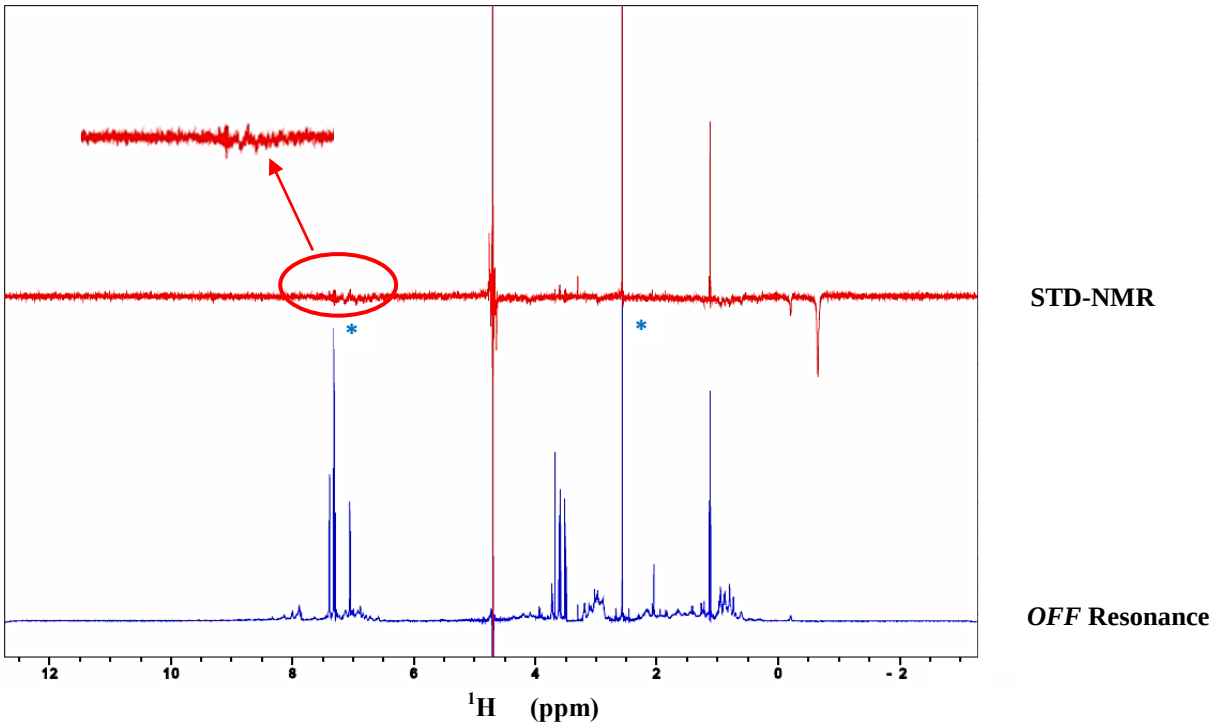



**Figure 3. 20. $^{1}$H STD experiments performed at 25 °C on 0.2 mM CypA protein complexed with 8 mM 3' aminoacetophenone.** Blue spectrum *off*-resonance spectrum; Red spectrum final difference STD spectrum. STD spectrum of 3′aminoacetophenone binding to CypA. The blue asterisk (*) represents the 3′aminoacetophenone peaks.

The principle of STD-NMR is based on saturation transfer from the protein to a bound ligand. Transfer of saturation is achieved by selectively saturating the protein resonances through a soft pulse. Through spin diffusion (which is efficient for large molecules), the saturation is then spread throughout the entire protein molecule. When a ligand binds to the protein, the saturation is transferred to the ligand by inter-molecular $^{1}$H-$^{1}$H cross-relaxation [56] [59]. As a result, protons from the ligand that are closer to the protein receptor surface receive more saturation than protons that are further from the receptor. Similarly, unbound ligand protons will not

receive any saturation at all via this route. Typically, two experiments are carried out, where in one the protein is saturated (*on* resonance) and in the second the saturation pulses are placed a long way off resonance typically -30 ppm (*off* resonance). The *off*-resonance spectrum should show the ligand alone and the *on*-resonance should show the ligand signals with reduced intensities depending on how much saturation is transferred. Thus, the difference between *on*- and *off*-resonance gives the STD difference spectrum, revealing signals from ligand protons, which have bound to the saturated protein receptor [56] – [59].

The 3′ aminoacetophenone peaks (highlighted with *) show binding at 2.5 ppm and no binding at 7.5 ppm (Figure 3.20), where the signals are completely attenuated in the STD difference spectrum. This suggests that 3′ aminoacetophenone is a weak binder to CypA.

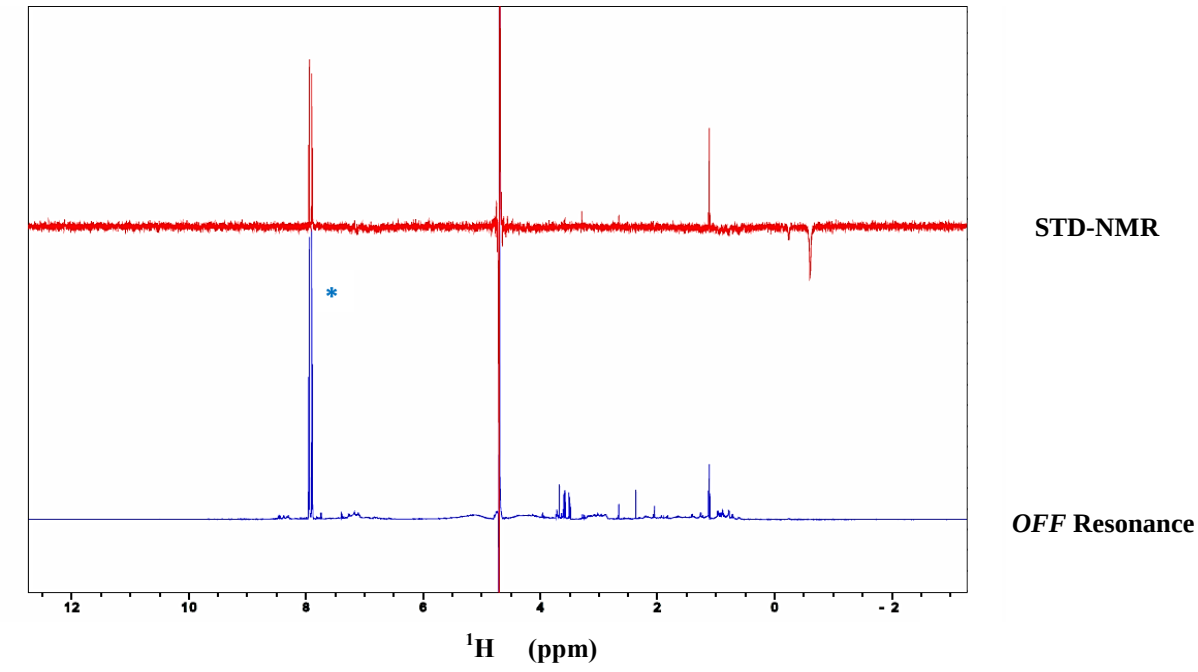



**Figure 3. 21. $^{1}$H STD experiments performed at 25 °C on 0.2 mM CypA protein complexed with 8 mM 4′ sulfamoylbenzoic Acid.** Blue spectrum off-resonance spectrum; Red spectrum STD difference spectrum, the blue asterisk (*) indicates the 4′ sulfamoylbenzoic Acid peak.

The 4′ sulfamoylbenzoic acid peak at 8 ppm shows binding (*, Figure 3.21), this data agrees with the hypothesis that the negative crystallographic result with CypA could be due to damage to the crystals during the binding of the ligand (c.f: 3.3.1).

### 3.6.3 An Introduction to WaterLOGSY

WaterLOGSY is a complementary to STD and is based on the transfer of saturation from bound water molecules instead of the protein. The resonance of the water molecules is selectively excited, and the magnetisation is transferred to protons using the NOE mechanism. The resonance of the water molecules is excited and transferred by one of three mechanisms: 1) direct transfer from water to ligand in protein-bound waters. 2) indirect transfer between bound waters to the protein by exchange followed by spin diffusion and transfer to ligand. 3) direct exchange of protons in the binding site followed by transfer. The signals for unbound and bound ligands are opposite (Figure 3.22 A). This method is useful in understanding protein ligand interactions with ligands at micro-molar affinities in structure-based drug discovery [154] [156].

#### 3.6.3.1 NMR Sample WaterLOGSY Data Acquisition

$^{1}$H WaterLOGSY experiments were acquired on a Bruker Avance III 600 Hz spectrometer at 25°C, with 32,768 points (spectral width 9615 Hz) in the direct F1 dimension ($^{1}$H) with a total number of 256 scans measured within an acquisition time of 1.70 seconds. Water inversion was achieved using a 6.4 ms eSNOB inversion pulse, the nOe mixing period was 1.2 seconds, during which a low power of 10G/cm Z gradient pulse was applied. Further suppression of the protein

signals was achieved using 20 cycles of CPMG with an inter-pulse delay of 1 millisecond. The relaxation delay time (D1) in the experiments was 2.0 seconds. Data was processed using TopSpin 3.2 and line broadening of 1.0 Hz was applied (Figures 3.22B and 3.23). Excitation sculpting was used for water suppression and the spectra were referenced against water [156].

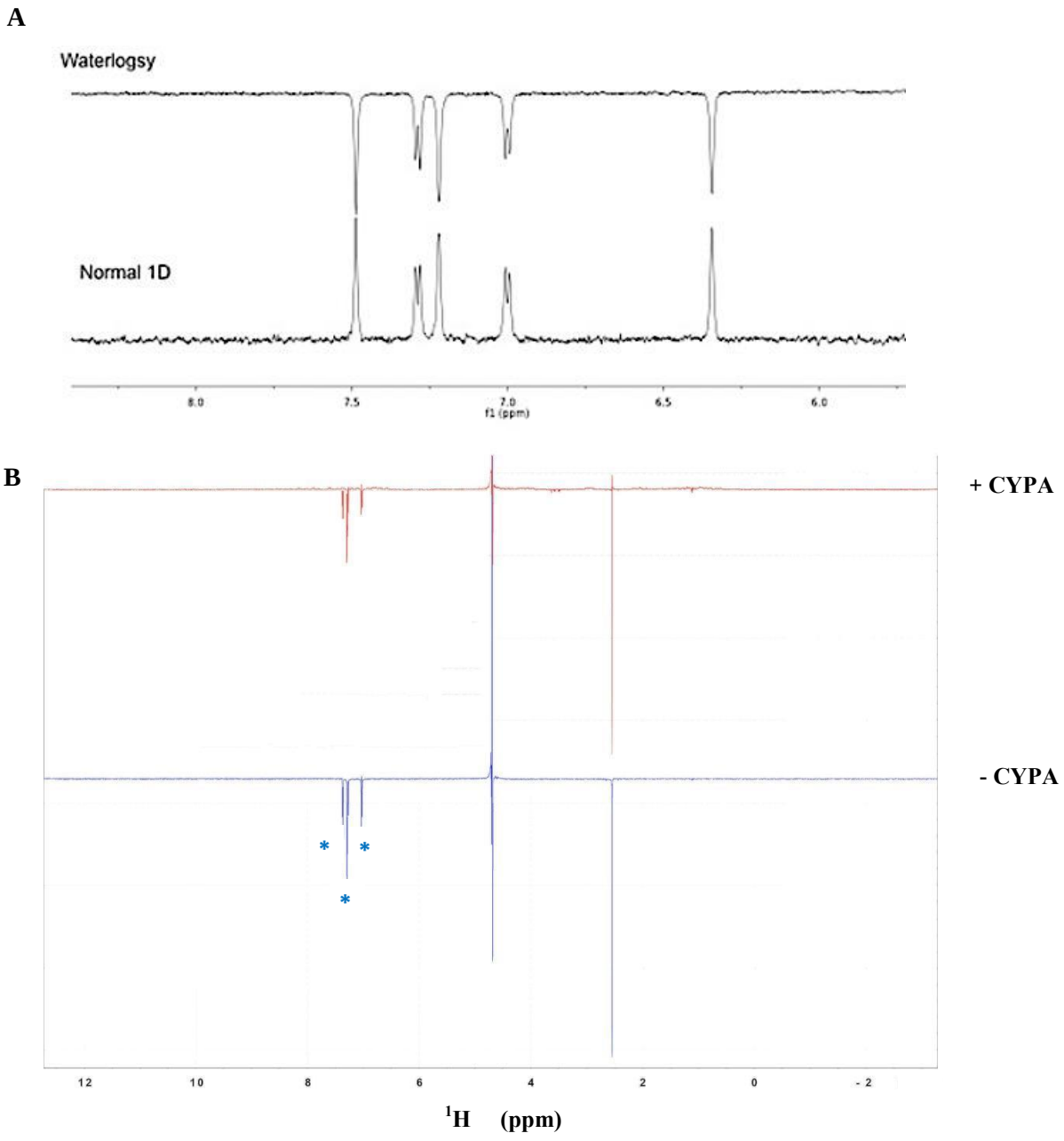



**Figure 3. 22. $^1$H WaterLOGSY experiments performed at 25 °C on 0.2 mM CypA protein complexed with 8 mM 3′ aminoacetophenone.** (A) Typical spectrum of protein-ligand binding observed by WaterLOGSY. Binding

is characterised by inversion of peaks in the WaterLOGSY spectrum (B) Blue spectrum 3′ aminoacetophenone only; Red spectrum 3′ aminoacetophenone complexed with CypA. All the $^{1}$H signal ligand peaks (*) in the red spectrum (+ CypA) are in the same direction as the blue spectrum (- CypA). *Adapted from (A)*: http://fragmentech.univ-lyon1.fr/waterlogsy-experiment/ [154]

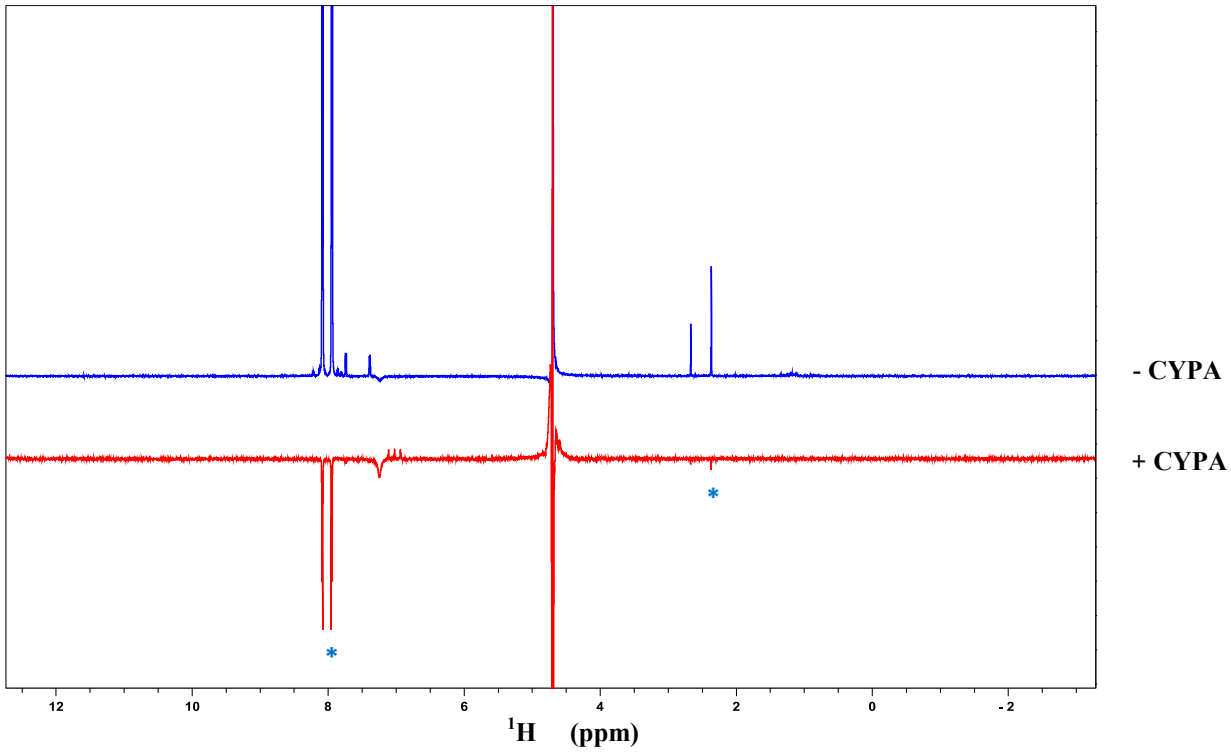



**Figure 3. 23. $^{1}$H WaterLOGSY experiments performed at 25 °C on 0.2 mM CypA protein complexed with 8 mM 4′ sulfamoylbenzoic Acid.** Blue spectrum 4′ sulfamoylbenzoic Acid only; Red spectrum 4′ sulfamoylbenzoic Acid complexed with CypA. All the $^{1}$H signal ligand peaks (*) in the red spectrum (+ CypA) are inverted in the opposite direction compared to the blue spectrum (- CypA).

Attenuation of $^{1}$H peaks shows weak binding of the ligand to the protein in WaterLOGSY experiments [156]. Therefore, Figure 3.22 B) indicates no binding of 3′ aminoacetophenone to CypA. The STD results also showed a weak binding.

Figure 3.23 shows the inversion of peaks (+ CypA), indicating binding of 4′ sulfamoylbenzoic Acid to CypA (both at 2.5 ppm and 8 ppm), validating the STD results.

## 3.6.4 An Introduction to Carr-Purcell- Meiboom Gill (CPMG)

CPMG is another experiment commonly used as a screening tool to identify compounds that bind to a protein. It is based on the measurement of transverse or spin-spin $T_2$ relaxation of any nucleus. Positive binding results in the attenuation of the signal, whereas non-binding does not change the spectrum. Effectively, signals have a higher rate of $T_2$ when bound to a slow tumbling protein [155].

### 3.6.4.1 NMR Sample Carr-Purcell- Meiboom Gill (CPMG) Data Acquisition

$^1$H CPMG experiments [155] were acquired on a Bruker Avance III 600 Hz spectrometer, with 16,384 points (spectral width 9615 Hz) in the direct F1 dimension ($^1$H), with a total number of 128 scans. The CPMG element of the sequence used a total of 600 cycles of refocusing period of 1 millisecond. The relaxation delay time (D1) in the experiments was 0.001 seconds. Data was acquired over an acquisition time of 0.85 seconds. Data was processed in Topspin 3.2 using a line broadening of 1.0 Hz. (Figure 3.24).

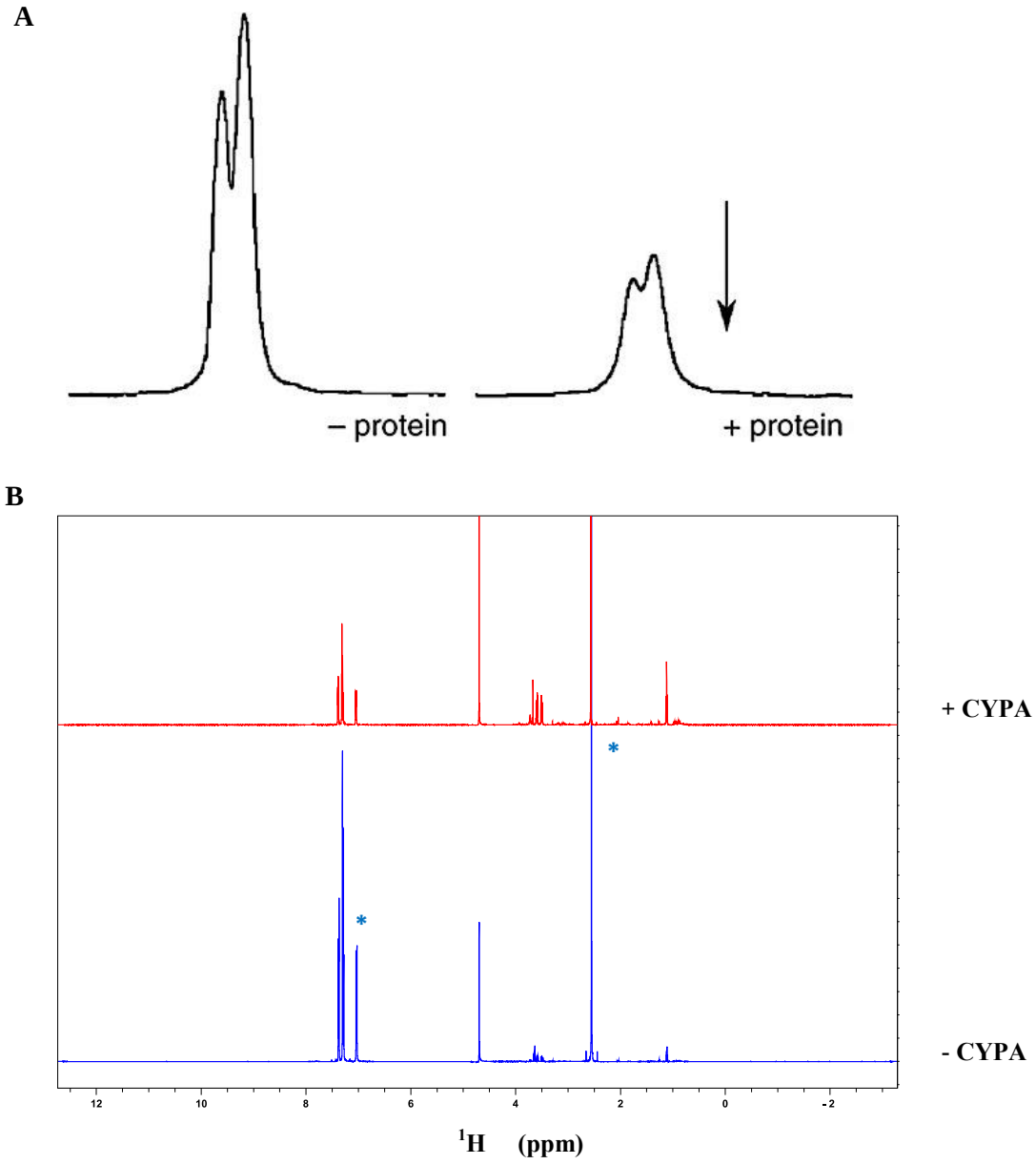



**Figure 3. 24. $^{1}$H CPMG experiments performed at 25 °C on 0.2 mM CypA protein complexed with 8 mM 3′ aminoacetophenone.** (A) Typical spectrum of protein-ligand binding observed by CPMG. Binding is characterised by attenuation of peaks in the CPMG spectrum (B) Blue spectrum 3′ aminoacetophenone only; Red spectrum 3′ aminoacetophenone complexed with CypA. The ligand $^{1}$H signal peaks (*) in the red spectrum (+ CypA) are attenuated in comparison to the blue spectrum (- CypA).

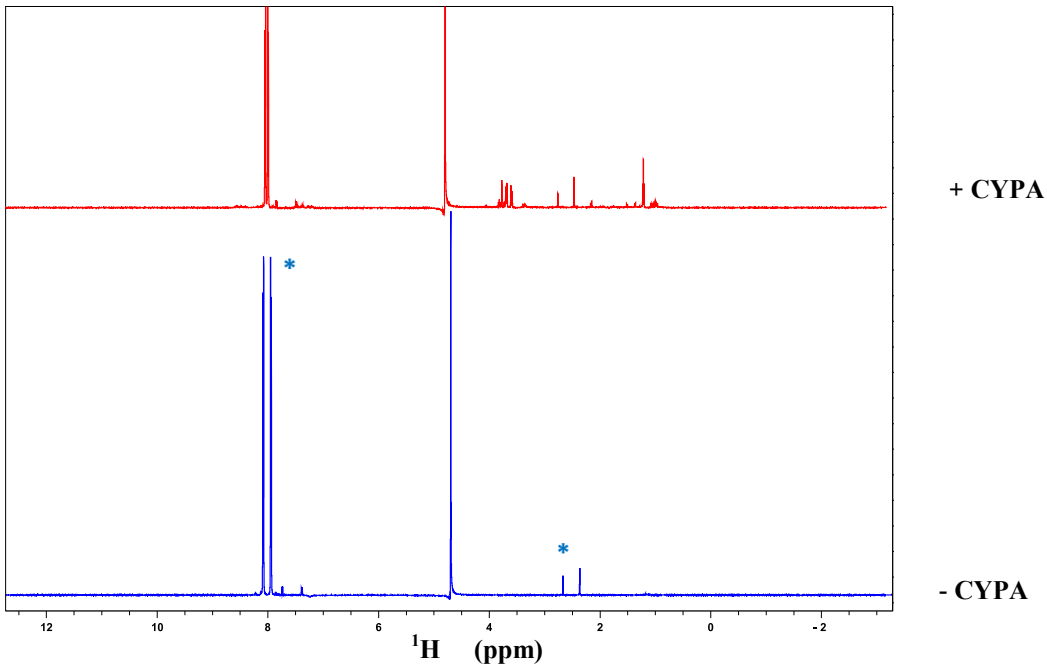



**Figure 3. 25. $^1$H CPMG experiments performed at 25 °C on 0.2 mM CypA protein complexed with 8 mM 4′ sulfamoylbenzoic Acid.** Blue spectrum 4′ sulfamoylbenzoic Acid only; Red spectrum 4′ sulfamoylbenzoic Acid complexed with CypA. The ligand $^1$H signal peaks (*) in the red spectrum (+ CypA) are attenuated in comparison to the blue spectrum (- CypA).

Attenuation of $^1$H peaks shows binding of the ligand to the protein in CPMG experiments [155]. Therefore, Figure 3.24 B) indicates binding of 3′ aminoacetophenone to CypA through the attenuation at 7 ppm and Figure 3.25 indicates binding of 4′ sulfamoylbenzoic Acid of CypA at 7.5 ppm.

These NMR screening methodologies are good tools to identify ligands/fragments that bind to the protein in comparison to X-ray crystallography. However, none of these ligand-based techniques are able to inform us about the key residues in the protein binding site, and therefore are limited to simply identifying ligand “hits”. Consequently, this highlights the need for protein-

observed methods for the prediction of protein ligand interactions in the application in structural-based drug discovery [155].

## 3.7 Characterisation of CypA by NMR Spectroscopy

Analysis of the protein-ligand interactions in NMR is usually undertaken with the full assignment of $^{15}N$ isotopically labelled protein, for which the protein needs to be folded and stable at 25 °C [157] – [159].

### 3.7.1 An Introduction to $^{15}N/^{1}H$ heteronuclear NMR experiments

$^{1}H$-$^{15}N$ HSQC is a heteronuclear experiment performed on proteins, which provides a chemical shift 'fingerprint' which is often used to determine the stability of the protein for further experiments. The sharpness and dispersion of the peaks over the available spectral range correlate with the folding of a protein [157] – [159].

This experiment identifies all the hydrogen-nitrogen correlations. It is based on the transfer of magnetisation from hydrogen to an attached $^{15}N$ nuclei through a large 1 bond J-HN-coupling. A chemical shift is measured from the nitrogen and the magnetisation is then transferred back to the hydrogen for detection again via the J-HN-coupling. Usually, the H-N correlations present in a $^{1}H$-$^{15}N$ HSQC originate mostly from backbone amide groups, but tryptophan side-chain Nε - Hε groups and asparagine/glutamine side chains Nδ - Hδ2/Nε - Hε2 groups can also be identified. Other visible groups in such a spectrum can be Arginine Nε - Hε peaks, but the Nε are usually outside the excited region [157] – [159].

## 3.7.2 NMR Sample Preparation of $^{15}N/^{1}H$ CypA

The stability of CypA was investigated by proton-based experiments at 25°C using a $^{15}N$ isotopically enriched sample expressed in *E.coli* T7 express cells grown in minimal media and purified in GF Buffer (25 mM $NaH_2PO_4$, 100 mM NaCl and 1 mM DTT pH 6.5) as described in Chapter 2.5. The final concentration of the protein was 0.7 mM.

## 3.7.3 $^{15}N/^{1}H$ HSQC NMR Data Acquisition with CypA

A $^{15}N/^{1}H$ Heteronuclear Single Quantum Correlation (HSQC) experiment was acquired on a Bruker Avance III spectrometer at 600 MHz (F2) and 60 MHz (F1) at 25°C, with 1024 points (9000 Hz) in the direct F2 dimension ($^{1}H$) and 256 points (2100 Hz) in F1 ($^{15}N$) indirect dimensions. There was a total number of 8 scans per increment each with an acquisition time of 0.11 seconds in F2 dimension. Watergate was the water suppression technique. The relaxation delay time (D1) in the experiments was 1.2 seconds. The size of the real spectrum was 4096 in the F2 dimension and 512 in the F1 dimension (Figure 3.26).

The samples were in 90% $H_2O$ and 10% $D_2O$.

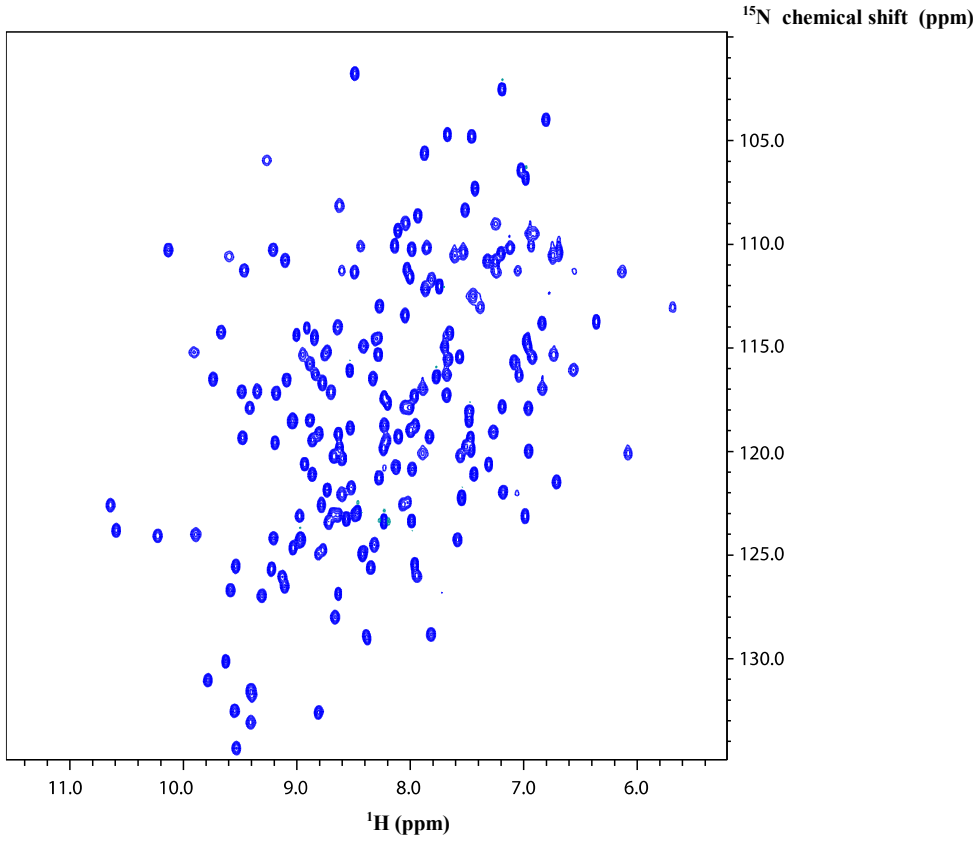



**Figure 3. 26. $^{1}H/^{15}N$ HSQC Spectrum of CypA at 25 °C.** The peaks are all sharp and clear, which suggest that the protein is folded and can be assigned with further triple resonance through-bond experiments.

$^{1}H/^{15}N$ HSQC experiments confirm that the protein is folded and informs the presence of certain types of amino acid groups. However, the spectrum needs to be assigned using through-bond triple resonance experiments to be able to match each peak with a specific amino acid from the CypA protein sequence [157] – [159].

The assignment was not attempted, as the findings could not be compared with a valid X-ray structure, as part of this combined method approach study, because no ligand binding occurred in the crystal structure.

## 3.8 Summary and Discussion

Chapter 3 has attempted to characterise the structure of CypA using X-ray crystallography and both ligand and protein-observed NMR methods. CypA was studied in complex with the ligands 3′ aminoacetophenone, 4′ sulfamoylenzoic acid and 6′ aminoindazole. Previous STD experiments show that the ligands that bind to CypD also bind to CypA. This chapter extends over the NMR characterisation of CypD ligands in CypA.

X-ray crystallography studies revealed that the crystal packing of CypA did not allow the binding of the ligands to CypA. This appears to be due to C-terminal residue E165, obstructing the ligand binding site of CypA. In an attempt to overcome the packing effect hindering the binding site and create a new crystal form, in-situ mutagenesis was performed to remove C-terminal E165, but the crystallisation of mutant CypA did not yield suitable crystal samples of 3′ aminoacetophenone, 4′ sulfamoylbenzoic Acid and 6′ aminoindazole complexed with CypA; this might be explained by the presence of the His-tag, which was not removed due to absence of a thrombin cleavage site.

NMR screening methodologies, including $^{1}$H STD, $^{1}$H WaterLOGSY and $^{1}$H CPMG were performed to show that 3′ aminoacetophenone and 4′sulfamoybenzoic acid bind to CypA. However, none of these ligand-based techniques are able to inform about the key residues in the

protein binding site, and therefore are limited to simply identifying ligand “hits”. CypA was then studied using 2D-based heteronuclear $^{1}H/^{15}N$ experiments.

CypA was an intended as an example to illustrate the need for an advanced NMR method to determine ligand binding sites. This study highlighted the requirement for full protein assignment [159] and how NMR interaction of $^{1}H$ ligand-based observations can be used in the absence of an X-ray structure. As the protein- and ligand-observed NMR studies could not be compared with X-ray structure for validation purposes, this target protein was abandoned.

# Chapter 4

## The Determination of the Structure of TbrPDEB1 and Ligand Binding Sites using X-ray Crystallography and NMR

This chapter's main emphasis is to understand the interactions of TbrPDEB1 with the ligands VUF 14234, PPS 59083, VUF 14450 and PPS 60036 using X-ray crystallography and various NMR methodologies, and to characterise the protein residues involved in ligand binding.

## 4.1 Introduction about Phosphodiesterase B1 in *Trypanosoma brucei* (TbrPDEB1)

TbrPDEB1 is a cyclic nucleotide phosphodiesterase expressed in *Trypanosoma brucei*, and along with TbrPDEB2, it has been genetically and pharmacologically characterised as an important virulence factor and therapeutic target for the neglected disease, human African Trypanosomiasis (HAT) or African Sleeping Sickness [164] – [166]. The demands for the development of new drugs for this target are increasing, and the current treatment options available lead to serious and even sometimes lethal side effects [166] [167]. *Trypanosoma* is a type of protozoa unicellular parasite that belongs to the order of kinetoplastida, due to the presence of a kinetoplast. A kinetoplast is an organelle containing DNA in the mitochondria of the organism. The *Trypanosoma* parasite is unique by the presence of a single flagellum, which is critical for motility, cell division, morphogenesis, attachment to host and vector cells, immune evasion, and the viability and pathogenicity of the parasite [164] – [169].

### 4.1.1 Functional Studies of TbrPDEB1

Both TbrPDEB1 and TbrPDEB2 are catalytic enzymes, selectively mediating hydrolysis of the ubiquitous secondary messenger cyclic adenosine mono-phosphate (cAMP) to AMP.

$CO_2$-regulated adenylate cyclases, embedded in membranes or in the cytoplasm of eukaryotic cells, convert adenosine tri-phosphate (ATP) to cAMP. cAMP signal activates the regulatory subunit of protein kinase A (pkA), cAMP-gated ion channels, or the guanidine-nucleotide exchange proteins EPAC1 and EPAC2 (Figure 4.1). The cAMP signal is usually terminated by PDEs. These PDEs are pivotal for cell signalling processes [164] – [166].

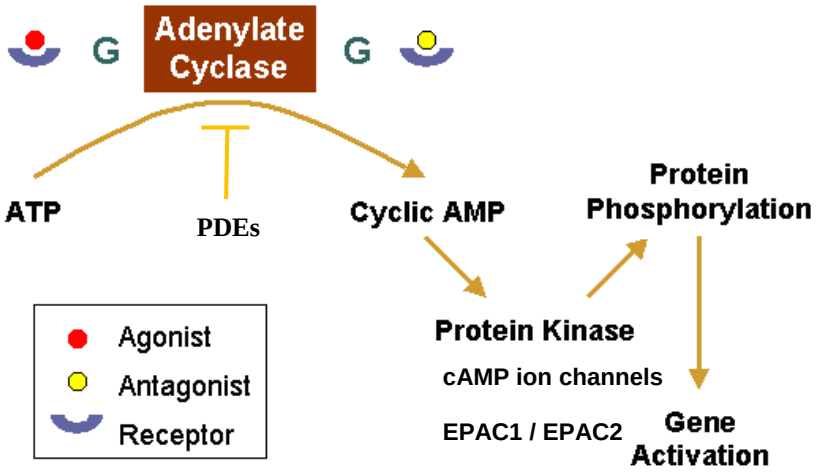



**Figure 4.1. PDEs regulate the hydrolysis of cAMP.** Adenylate cyclase is responsible for the conversion of ATP to cyclic AMP. Cyclic AMP then activates protein kinase enzymes which results in phosphorylation of proteins responsible for gene activation. Phosphodiesterases mediate the hydrolysis of cAMP and can terminate the cAMP signal. *Adapted from*: Janicak PG et al: Principles and Practice of Psychopharmacotherapy, 2nd Edition, Lippincott, Williams & Wilkins, 1997 [170]

Seebeck's group showed that RNA interference (RNAi) against either of TbrPDEB1 or TbrPDEB2 revealed an observable phenotype in the development of the parasite. However, simultaneous RNAi against both of the enzymes resulted in complete cell lysis within 48 hour of RNAi induction. The simultaneous RNAi of both enzymes caused a 1000-fold increase in cAMP concentration within few hours, which was not observed in RNAi of individual

TbrPDEs. Impaired division of trypanosomes and cell lysis occur following the accumulation of intracellular cAMP. Both of the enzymes can functionally complement each other, highlighting the importance of both enzymes *in-vitro* and *in-vivo* [165]. Therefore, drug discovery and treatment should be strongly targeted at both enzymes at the same time to significantly alter cAMP levels and cause the parasitic death [164] – [169].

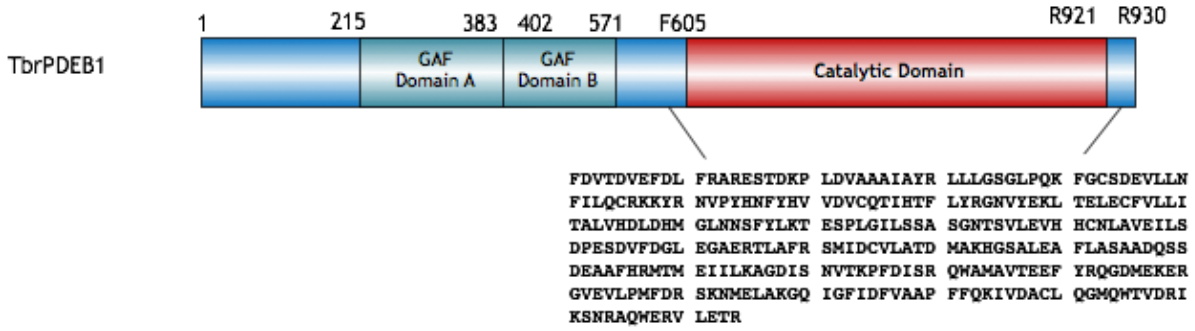



**Figure 4.2. The construct of PDEB1-Tbr in pET28a was provided by the Free University of Amsterdam (VU).** TbrPDEB1 and TbrPDEB2 reveal very close sequence identity. The TbrPDEB1 enzyme consists of two ubiquitous N-terminal GAF domains, where cAMP binds (D234 - E554) and a catalytic domain (V586 - R908), causing the hydrolysis of cAMP into AMP.

Although TbrPDEB1 and TbrPDEB2 reveal very close sequence identity, they are both located at distinct sub-cellular localisation. TbrPDEB1 is located in the flagellum within the flagellar skeleton, and most of TbrPDEB2 is found throughout the cell body distributed between the flagellum and cytoplasm. A small portion of TbrPDEB2 is also found in the flagellum. cAMP signaling occurs in the flagellum in Tbr. Luginbuehl's transfection of pro-cyclic trypanosomes with green-fluorescent protein (GFP) reporters studies show that the N-termini of both enzymes are essential for determining their corresponding sub-cellular localisations. Mutation of residues in the N-terminal regions of those enzymes showed that

single amino acid substitutions can re-locate the enzymes from cell body to flagellar skeleton [168] [169].

### 4.1.2 Structure of TbrPDEB1

The TbrPDEB1 enzyme consists of two ubiquitous N-terminal GAF domains, where cAMP binds (D234 - E554) and a catalytic domain (V586 - R908), causing the hydrolysis of cAMP into AMP (Figure 4.2). Inhibition of the catalytic domain of TbrPDEB1 and inhibition of the full-length occur at the same inhibitor concentrations, indicating that the catalytic domain is the driver of the enzyme activity. The catalytic domain (37 kDa) consists of sixteen α- helices and no β- strands (Figure 4.3). The active site of the protein consists of two divalent metal atoms, zinc and magnesium, each forming an octahedral geometry. TbrPDEB1 has also a parasite-specific P-pocket directly adjacent to the active site and consists of the invariant Q874 on helix H15 and residues from the M-loop and helix H14 on the opposing sides. The parasite-specific P-pocket is also found in *Leishmania major* PDEB1 (Lmj) and *Trypanosoma cruzi* PDEC (Tc) [164] – [169].

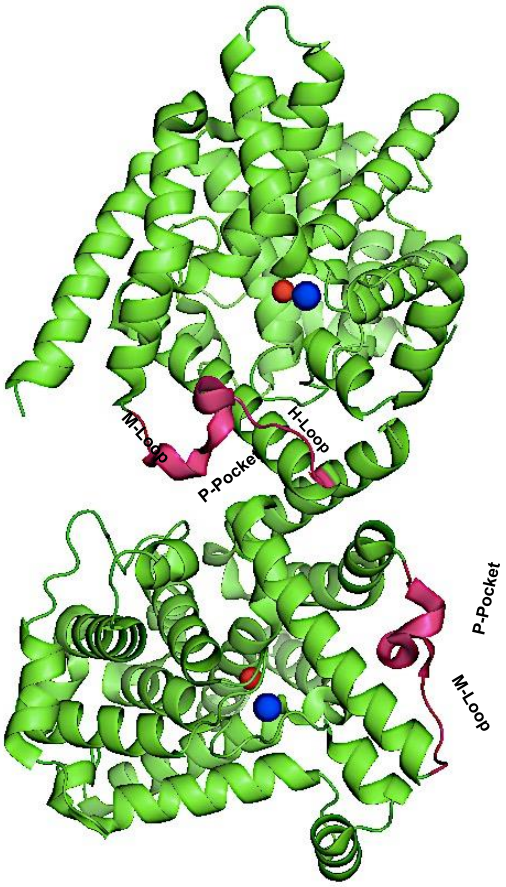



**Figure 4.3. TbrPDEB1 crystal structure (PDB: 4I15)** The catalytic domain (37 kDa) consists of sixteen α-helices and no β- strands (Figure 4.3). The active site of the protein consists two divalent metal atoms, zinc and magnesium, each forming an octahedral geometry. TbrPDEB1 also shows the presence of a parasite-specific P pocket, which is found to be directly adjacent to the active site. Metal ions are shown in blue (magnesium) and red (zinc). The H-loop and M-loop form the P-pocket (pink). The structure is shown as two molecules in the AU and the packing effects near the p pocket as one molecule.

## 4.2 Recombinant expression and purification of TbrPDEB1 from *E.coli* and crystallisation

### 4.2.1 Recombinant expression and purification

The first step of purification involved nickel-affinity chromatography using Buffer A on 5 mL His-Trap Column (20 mM tris-HCl pH 8, 300 mM NaCl, 10 % (v/v) glycerol, 1 mM 2-beta mercaptoethanol, 10 mM imidazole) for equilibration, and Buffer B (20 mM tris-HCl pH 8, 300 mM NaCl, 10 % (v/v) glycerol, 1 mM 2-beta mercaptoethanol, 10 mM imidazole, 250 mM imidazole) for elution, as described in 2.5.1.2 (Figures 4.4 and 4.5).

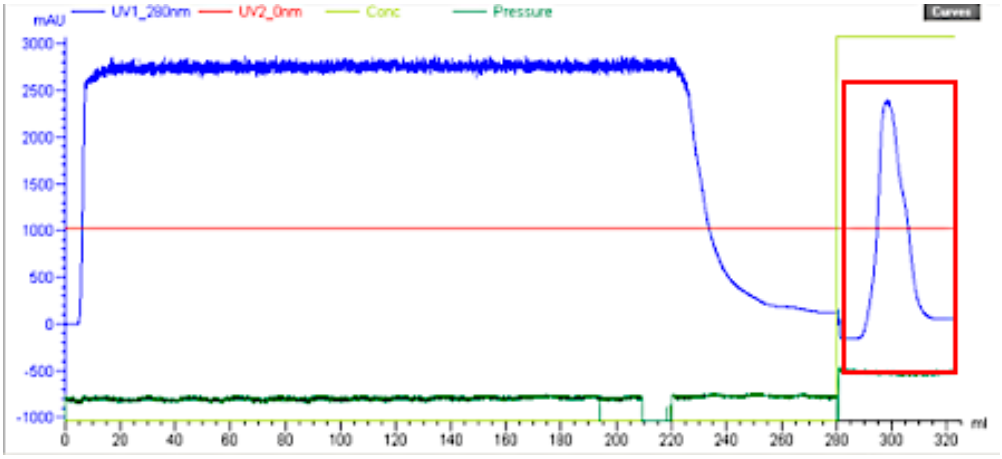



**Figure 4.4. Elution Profile of mutant TbrPDEB1 on His-Trap 5 mL Immobilised Nickel Affinity Chromatography and 26/10 HiPrep Desalting Column** The peak of elution (red rectangle) represent the His-tagged TbrPDEB1 protein. The protein was immediately desalted through 26/10 HiPrep Desalting Column to remove high concentrations of imidazole.

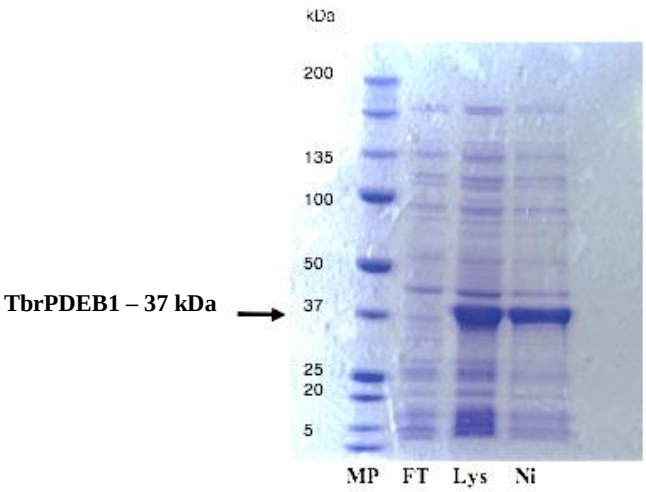



**Figure 4.5. Purification of TbrPDEB1 by Nickel-Affinity Chromatography.** The purification fractions on the 4-12 % (w/v) Bis-tris gel show ~ 70 % purity of TbrPDEB1 at the expected molecular weight of 37 kDa after nickel affinity chromatography (Ni). Lys: lysate, FT: flow-through, Ni: Nickel elution of TbrPDEB1.

The first eluted fraction was subjected to thrombin cleavage (stock: 2 units/µL) cleavage and was added on the basis of 5 units/mL, and the fraction was stored at 4 °C overnight. The next step of purification was ion-exchange chromatography using Buffer A on 5 mL Hi-Trap Column (20 mM tris-HCl pH 7.5, 100 mM NaCl, 5 % (v/v) glycerol, 1 mM 2-beta mercaptoethanol) for equilibration, and Buffer B (20 mM tris-HCl pH 7.5, 1 M NaCl, 5 % (v/v) glycerol, 1 mM 2-beta mercaptoethanol) for elution (Figure 4.6).

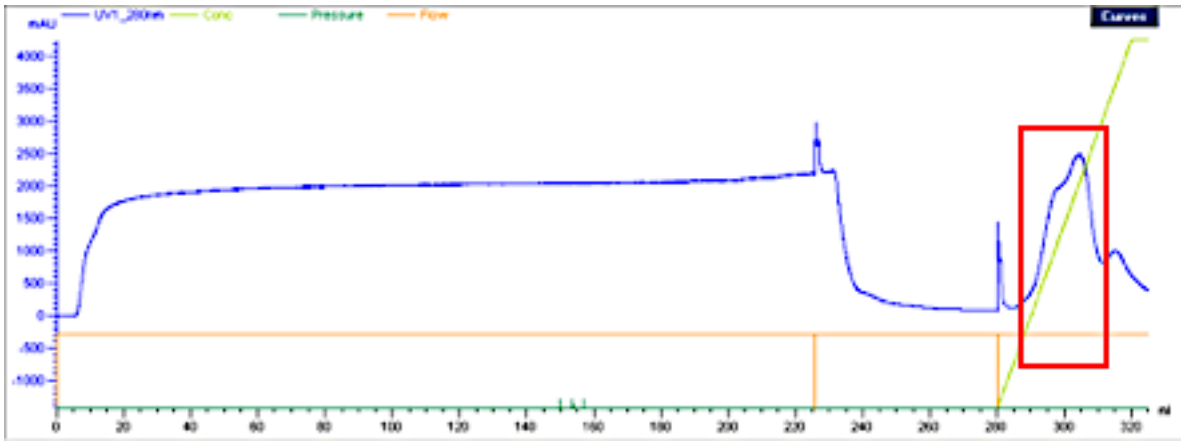



**Figure 4.6. Elution Profile of TbrPDEB1 on Hi-Trap 5 mL Ion Exchange Chromatography**

The final step of the purification of TbrPDEB1 was gel filtration chromatography using GF Buffer on Superdex S-75 Column (20 mM tris-HCl pH 7.5, 50 mM NaCl, 5 % (v/v) glycerol, 1 mM 2-beta mercaptoethanol) for both equilibration and elution, as described in 2.5.1.5 (Figure 4.7). The protein was well expressed and the final purified yield of 10 mg per litre.

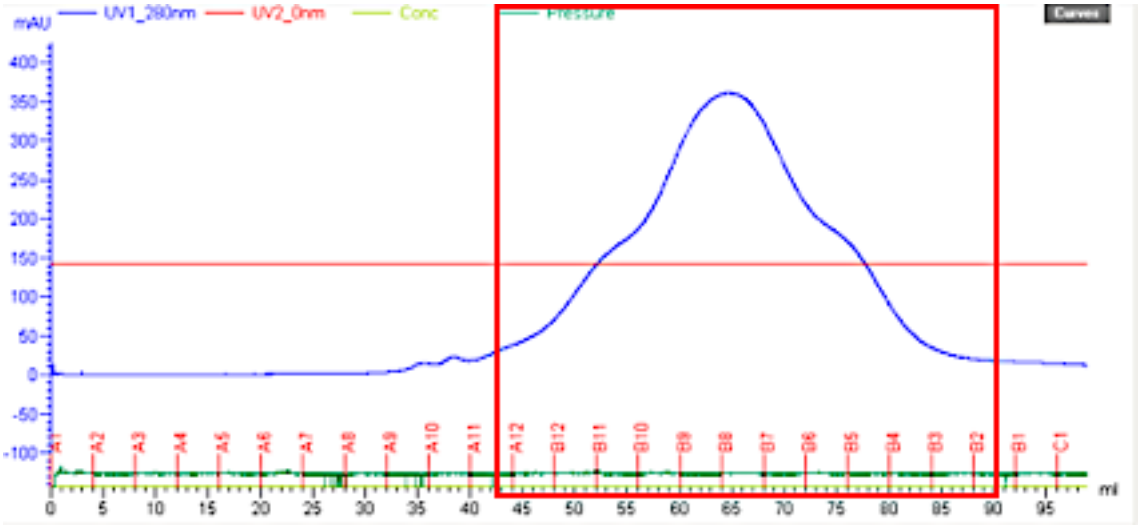



**Figure 4.7. Elution Profile of TbrPDEB1 on Superdex S-75 Column Gel Filtration Chromatography.**

Fractions A12 - B2 (red rectangle) show the elution of the pure protein TbrPDEB1.

As the gel filtration chromatography trace appears to elute three different species, fractions A12-B11, B10-B5 and B4-B2 were run separately on an SDS-PAGE Coomassie gel. All the fractions confirmed to the presence of pure TbrPDEB1 protein and did not show any heterogeneity in the protein sample. Unfortunately, the SDS-PAGE gel was photographed during the time of the experiment. The gel was considered good enough to test the purity of the protein, and therefore mass spectrometry was not performed. In the event that the protein would not crystallise, mass spectrometry would have been done to clarify the level of purity.

## 4.2.2 Protein Crystallisation

The crystallisation of TbrPDEB1 was performed following the crystallisation conditions described by Huanchen Wang et al. (2010). CypA was concentrated at 10 mg/mL, equivalent

to ~ 0.3 mM for crystallisation (2.5.3.1). The best crystals for diffraction were obtained from 14% - 24% PEG3350 pH 6.5 conditions after 24 hours at 20 °C (Figure 4.8).

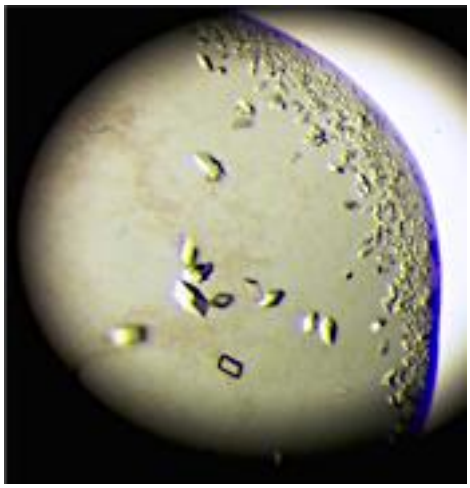


**Figure 4.8. TbrPDEB1 Crystals in 20 % PEG 3350 pH 6.5 Conditions** The TbrPDEB1 crystals always tend to grow as diamonds and can grow up to a diameter of 5 microns. There are micro-crystals at the edge of the drop. The crystals diffracted to 2.0 Å on IO4.1 beamline at a wavelength of 0.97949 Å, temperature of 100 Kelvin with a 100 % exposure for 0.1 seconds / 0.1º oscillation.

Crystals diffracted to 2.0 Å resolution and their space group was C $2_1$. Huanchen Wang studies have shown that this is a soakable protein system [164].

## 4.3 Crystallographic Solutions of TbrPDEB1

### 4.3.1 Addition of Ligands

Compounds for soaking were identified and Surface Plasmon Resonance (SPR) was undertaken by VU Free University (Amsterdam) to detect molecular interactions of ligands with TbrPDEB1, and determine the corresponding binding affinities ($IC_{50}$) using DMSO stock compounds. Thirteen hits were selected for crystallographic studies within the micro-molar to nano-molar ranges. Only four ligands were successfully solved in complex with TbrPDEB1.

| Ligand | SPR ( $IC_{50}$) | Types of crystallisation | Soak | Co-Crystal |
|---|---|---|---|---|
| **PPS 59083** | 100 nM | Soak / Co-Crystal | ✓ | × |
| **PPS 60036** | 0.44 µM | Soak / Co-Crystal | ✓ | × |
| **VUF 14450** | 6.2 µM | Soak / Co-Crystal | ✓ | × |
| **VUF 14234** | 5.6 µM | Soak / Co-Crystal | × | ✓ |

**Table 4. 1. Summary of the soak and co-crystallisation experiments with TbrPDEB1** ✓ Data obtained, × No data could be obtained due to DMSO intolerance in soaks or no crystals were grown in co-crystallisation.

Compound solids were dissolved to 50 mM stock in DMSO. Compounds were used in equimolar ratio with protein concentration (0.9 mM). Crystals were soaked for 2 days at 4 °C before cryo-cooling in liquid nitrogen with 20 % Ethylene Glycol for data collection.
The co-crystals were performed at a ratio of 2 : 1 ligand to protein excess under the same conditions as apo crystal crystallisation conditions.

### 4.3.2 Data collection and processing

All data was collected at DIAMOND on IO4.1 beamline at a wavelength of 0.97949 Å, at a temperature of 100 Kelvin with a 100 % exposure for 0.1 seconds / 0.1º oscillation. The space group was orthorhombic C $2_1$ (Figure 4.9).

Data was auto-processed through Xia2 [138]. A simple molecular replacement with Phaser into C $2_1$ was performed using the apo structure of TbrPDEB1 as search model. Ten rounds of structural refinement using Refmac5 [136] were performed and the difference density maps were examined for presence of the ligand. Once satisfied with the presence of the ligand, the structure was finalised with Buster. The crystallographic data is summarised in Table 4.2 and 4.3.

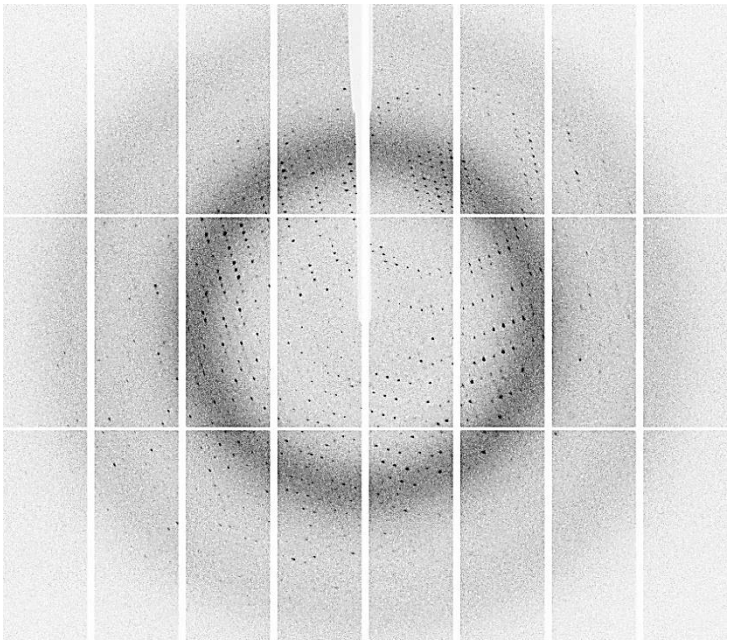


**Figure 4.9. X-ray diffraction pattern of TbrPDEB1** Auto-solving softwares such as *xia2* and *fast_dp* were used to detect the space group of C $2_1$. The diffraction intensities in the diffraction image of TbrPDEB1 represent a reflection of each atom in the TbrPDEB1 protein, generated from the X-ray beam passing through the crystal [138] [139].

## 4.3.3 Crystallographic Data, Refinement and Ramachandran Plot Statistics with ligands

| Compound | VUF14234 | PPS 59083 | VUF 14450 | PPS 60036 |
|---|---|---|---|---|
| Space Group | C $2_1$ | C $2_1$ | C $2_1$ | C $2_1$ |
| Unit cell dimensions (Å) (a,b,c) | 111.74, 115.33, 68.02 | 115.10, 115.34, 68.72 | 115.41, 115.31, 68.43 | 112.820, 115.940, 68.510 |
| Unit Cell Angle (°) | 90, 108.30, 90 | 90, 107.92, 90 | 90, 108.20, 90 | 90, 107.43, 90 |
| Number of observed reflections | 25,442 | 56,129 | 38,635 | 45,186 |
| Oscillation range per frame (°) | 0.1 | 0.1 | 0.1 | 0.1 |
| Overall rotation (°) | 180 | 180 | 180 | 180 |
| Resolution Range (Å) | 64.565 – 2.582 | 79.42 – 2.00 | 57.65 – 2.24 | 57.97 – 2.14 |
| Multiplicity | 3.3 (3.5) | 3.4 (3.5) | 3.0 (1.9) | 3.3 (3.2) |
| Completeness (overall) (%) | 98.9 | 98 | 94.6 | 98.7 |
| Mean I/sigma (I) (overall and last shell) | 8.5 (2.1) | 12.5 (2.1) | 9.8 (2.2) | 13.7 (2.5) |
| Rmerge (overall and last shell) | 0.164 (1.247) | 0.074 (0.651) | 0.066 (0.375) | 0.056 (0.518) |

**Table 4. 2. Summary of the X-ray crystallography data collection of TbrPDEB1 with ligand structures**

| Compound | VUF14234 | PPS 59083 | VUF 14450 | PPS 60036 |
|---|---|---|---|---|
| Maximum Resolution (Å) | 2.6 | 2.0 | 2.4 | 2.2 |
| R-factor (%) (last shell) | 16.2 | 16.4 | 17 | 17.7 |
| R-free (%) | 22.7 | 19.4 | 20.6 | 20.9 |
| RMSD Bond lengths (Å) | 1.15 | 0.97 | 1.01 | 1.01 |
| RMSD Bond Angles (°) | 0.1 | 0.1 | 0.1 | 0.09 |
| Average B-factor Protein ($Å^2$ ) | A=62 B=59.8 | A=38.85 B=38.4 | A=55.41 B=56.6 | A=54.12 B=54.8 |
| Average B-factor Ligand ($Å^2$) | 41.3 | 35.3 | 23.4 | 71 |
| Number of Protein Residues | 665 | 665 | 665 | 665 |
| Number of water atoms modelled | 227 | 649 | 331 | 382 |
| $Zn^{2+}$ (number of atoms) | 2 | 2 | 2 | 2 |
| $Mg^{2+}$ (number of atoms) | 2 | 2 | 2 | 2 |
| Ramachandran Plot preferred (%) | 95.42 | 96.98 | 95.71 | 96.79 |
| Ramachandran allowed (%) | 3.97 | 2.57 | 4.29 | 2.44 |

**Table 4. 3. Summary of the X-ray crystallography structure refinement and Ramachandran Plot statistics of TbrPDEB1 with ligand structures.**

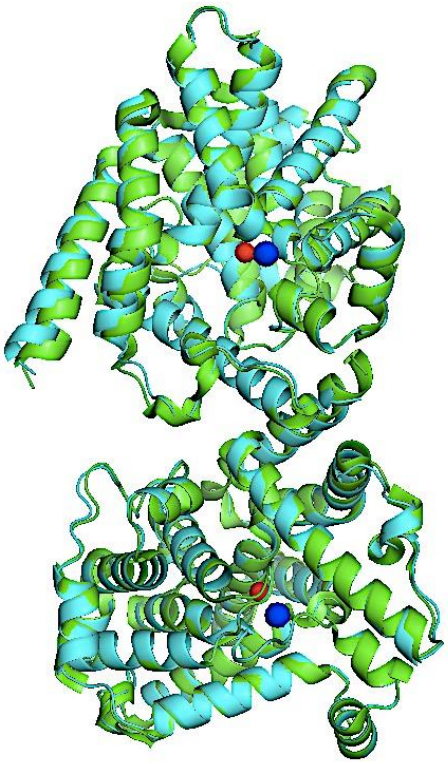


**Figure 4.10. Overlay of TbrPDEB1 refined coordinates crystal structure with deposited PDB: 4I15** The crystal structure of TbrPDEB1 consists of sixteen α-helices and no β-sheets structure, as expected (Figure 4.3). The overlay of the ribbon crystal structure of TbrPDEB1 (green) with ribbon crystal structure of TbrPDEB1 (PDB: 4I15) (cyan) shows a near perfect match [164].

Figure 4.10 confirms that the crystal structure of TbrPDEB1 matches with the published crystal structure of TbrPDEB1 (PDB: 4I15), with the presence of sixteen α-helices and no β-sheets. TbrPDEB1 is formed of two molecules, chain A and chain B, forming a dimer in the crystallographic asymmetric unit in the crystal lattice [164]. Chain A's active site is partly protected from solvent because it is more engaged with the crystal lattice contacts. Since the formation of this dimer is not related to the biological function of the protein, the

dimerisation is considered as a crystallographic artefact. This has not any significant impact on the overall conformation of the protein since both molecules A and B are identical [164].

### 4.3.4 TbrPDEB1 with ligands structures

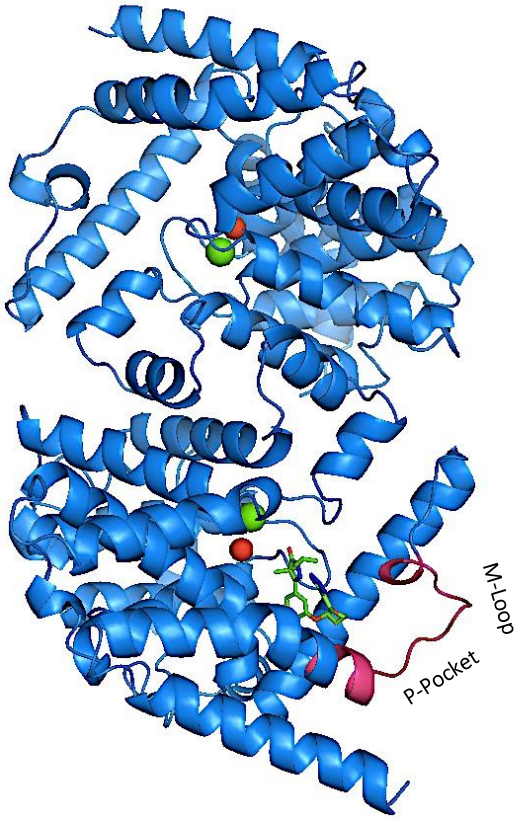



**Figure 4.11. Crystal structure of TbrPDEB1 refined coordinates co-crystallised with the ligand VUF 14234** The ligand VUF 14234 binds to Chain B of TbrPDEB1. A detailed view reveals that VUF 14234 is binding in the active site of TbrPDEB1, consisting of the metal ions $Zn^{2+}$ and $Mg^{2+}$. VUF 14234 is directed towards the P-pocket region [164].

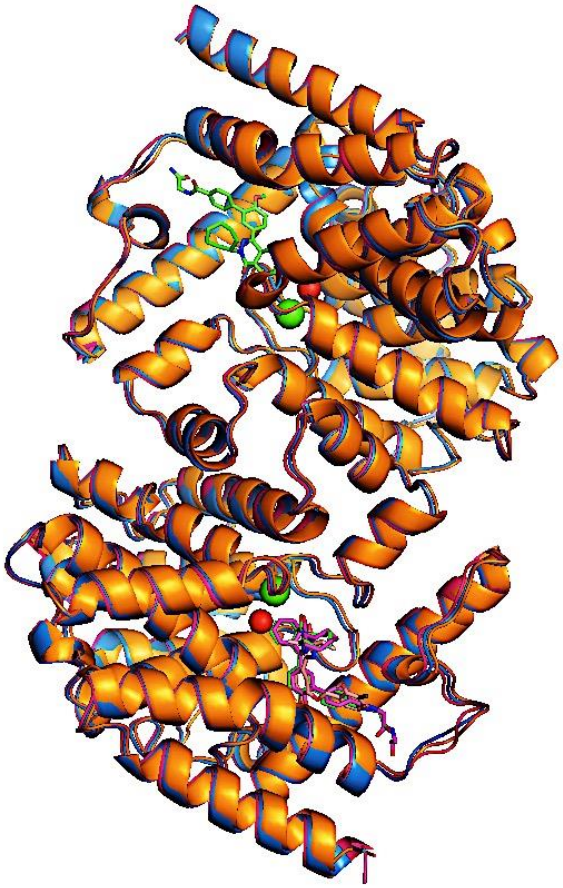


**Figure 4.12. Overlay of crystal structures of TbrPDEB1 in complex with the ligands PPS 60036 (blue), PPS 59083 (red) and VUF 14450 (orange)** The ligands have bound to TbrPDEB1 in the same site but with different binding modes. The ligand PPS 59083 binds to both molecules A and B of TbrPDEB1 but in the same active site. A detailed view into the binding site reveals that ligand binding in the P-pocket of TbrPDEB1 is conserved [164].

All the solved crystal structures of TbrPDEB1 complexed with different ligands either by co-crystallisation or soaking show a common binding site with different modes of binding due to chemotype groups as observed in the known P-pocket binding site with Piclamilast [171]. As such, this suggests that the ligands are suitable for the development of new anti-parasitic inhibitors.

Binding site analysis was required in all complexes to identify key residues and interactions, as well as unique features. Ligplot studies [172] were undertaken to investigate the residues responsible for ligand binding, which would help in the design of potential drugs for the treatment of HAT by inhibiting the activity of TbrPDEB1.

## 4.3.5 Analysis of the TbrPDEB1 active site

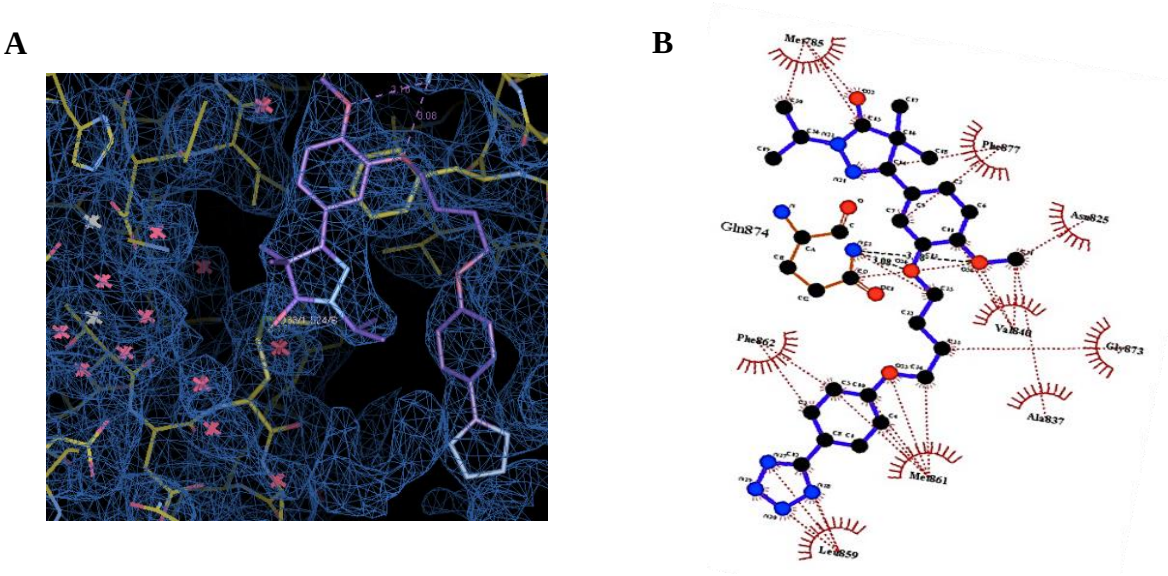



**Figure 4.13. Ligplot representation of the binding site of TbrPDEB1 with VUF 14234** (A) $|2Fo\text{-}Fc|_{\alpha c}$ electron density map of ligand VUF 14234 complexed with TbrPDEB1 contoured at 0.67 σ. (B) Ligplot representation of the interactions of VUF 14234 with key residues in TbrPDEB1 (dashed red semi-spheres lines represent hydrophobic interactions and green dashed lines representing hydrogen bonding)

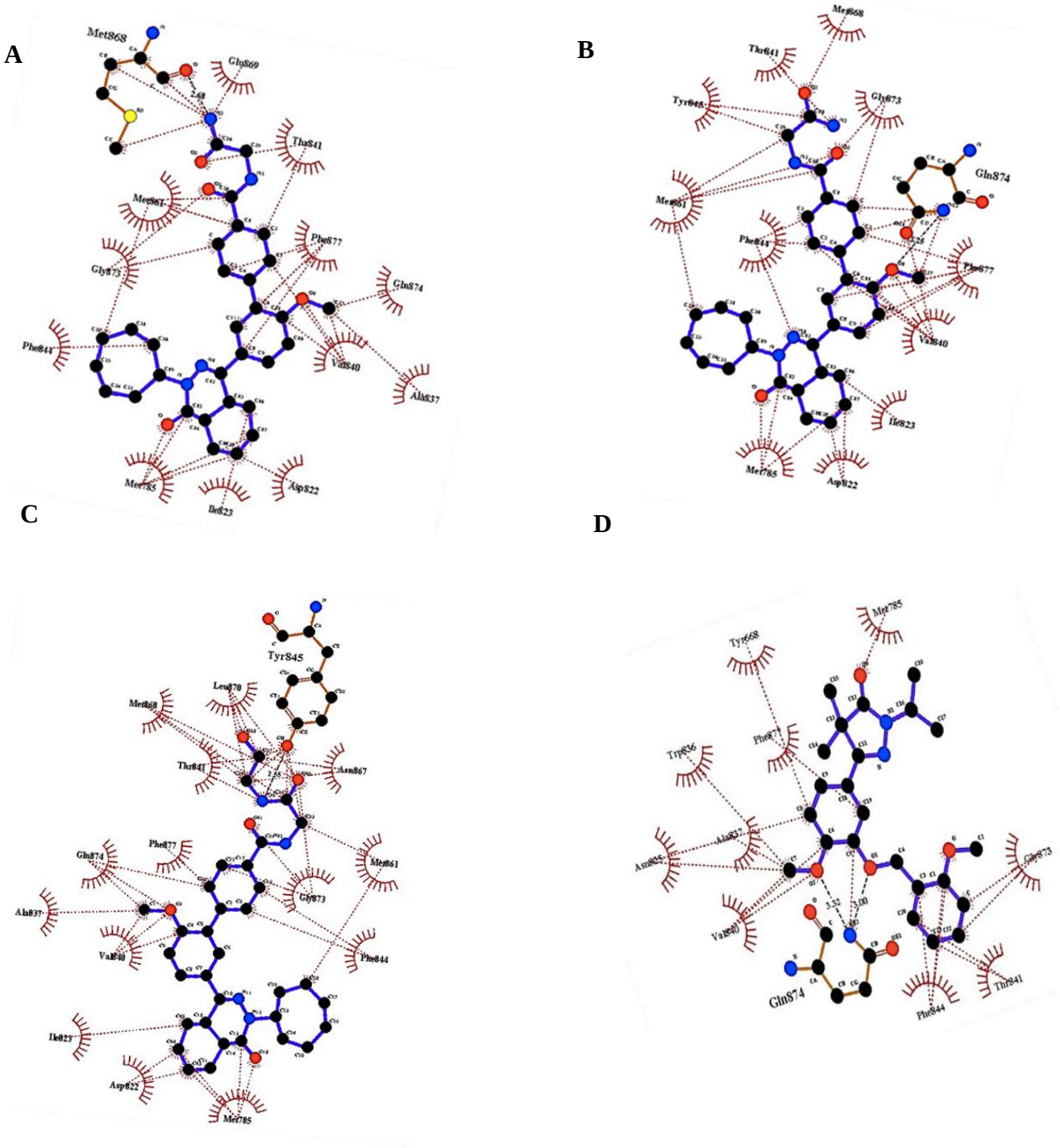



**Figure 4.14. Ligplot representations of the binding site of TbrPDEB1 with different ligands** (A) PPS 59083 in Chain A of TbrPDEB1 (B) PPS 59083 in Chain B of TbrPDEB1 (C) PPS 60036 (D) VUF 14450 Dashed red semi-spheres represent hydrophobic interactions; green interactions represent hydrogen interactions.

All these ligands are binding to TbrPDEB1. Their interactions with the key residues in TbrPDEB1 are being analysed through Ligplot studies. The examples show that the binding is conserved and the key residues

responsible for binding are common in the ligand binding site of TbrPDEB1. All the ligands commonly interact with the residue Q874 from the P-pocket.

Figures 4.13B and 4.14 show that the ligands are predominantly stabilised by hydrophobic interactions. There is also the presence of the conserved hydrophobic interaction between the carbon atoms of the ligands and the residue Q874 from the helix H15 of the P-pocket, as seen in previous studies [164] – [169]. These ligands are good scaffolds for the design of Tbr inhibitors due to their specificity for the P-pocket [173]. Figure 4.13A shows the Wincoot representation of the VUF 14234 ligand bound to the TbrPDEB1, where the incorporation of the ligand is well established. Figures 4.13B and 4.14 confirm that the key residues involved in binding in TbrPDEB1 are M785, D822, I823, N825, A837, V840, T841, L859, M861, F862, G873, F877 and Q874.

Apart from the conserved residues in the ligand binding site of TbrPDEB1, the ligand PPS 59083 also makes hydrogen interactions with the residues Y845, M868 and E869. The ligand PPS 60036 makes hydrogen interactions with the residues L870 and N867. The ligand VUF 14450 makes hydrogen interactions with the residues W836, A837 and Y668.

All these Ligplot interactions have characterised the ligand binding site of TbrPDEB1, which are highly conserved throughout all the studied structures. The residue Q874 plays a major role by forming hydrophobic interaction with each of the ligands studied.

Normal fragment-based approaches for drug screening require high concentration bio-assay or biophysical screening by SPR, NMR, ITC or Thermal Shift Assay to identify bindings followed by X-ray crystallography to determine three-dimensional positions for optimal chemistry follow-up.  X-ray fragment screening alone does not provide binding affinities but

instead can be used as primary screen method. NMR can be used to infer binding affinity information through STD-NMR and titration experiments. NMR can only be used as a three-dimensional technique if the isotopically labelled protein is fully assigned, but the assignment can help in understanding protein-ligand interactions therefore obviating the need for X-ray crystallography.

## 4.4 Characterisation of TbrPDEB1 by NMR Spectroscopy

Analysis of the protein-ligand interactions in NMR is usually undertaken with the full assignment of $^{15}N$ isotopically labelled protein, for which the protein needs to be folded and stable at 25 °C.

### 4.4.1 NMR Sample Preparation of $^{15}N/^{1}H$ TbrPDEB1

The stability of TbrPDEB1 was investigated by proton-based experiments at 25 °C from a $^{15}N$ isotopically enriched sample expressed in *E.coli* T7 express cells grown in minimal media and purified in GF Buffer (25 mM $NaH_2PO_4$, 100 mM NaCl and 1 mM DTT pH 6.5) as described in Chapter 2.5. The final concentration of the protein was 0.3 mM.

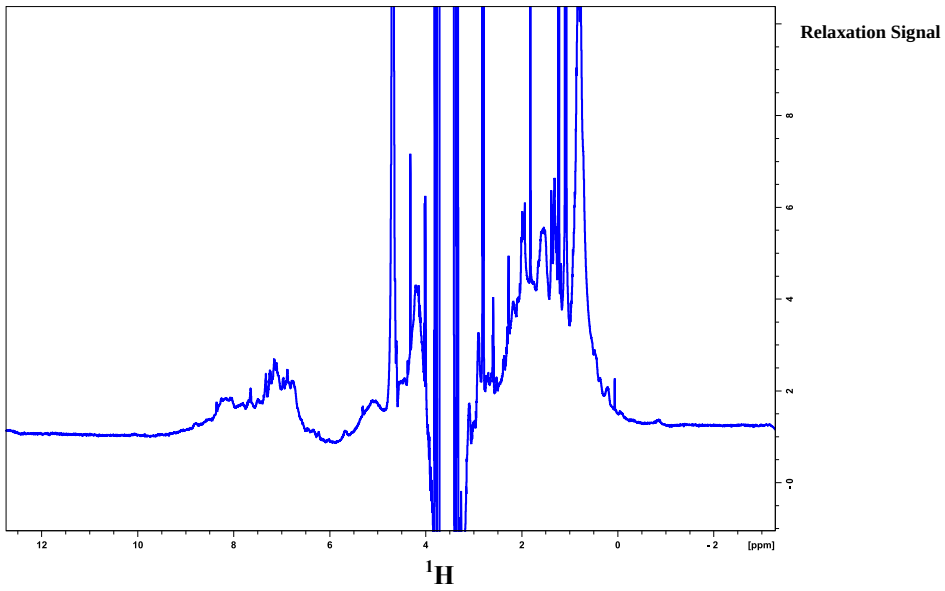



**Figure 4.15. Spectrum of $^{1}$H TbrPDEB1 at 25 °C: proton experiment with water suppression** Spectrum between 6 and 10 ppm represent the amide/aromatic region and side chain NH2; spectrum between 0 and 6 ppm represent the alpha region, water suppression and side chains CH, CH2 and CH3; spectrum between 0 and -2 ppm represent the methyl/aliphatic region. The peaks are broad and not dispersed, revealing that the protein is unfolded [174].

The $^{1}$H TbrPDEB1 experiment (Figure 4.15) shows that the protein peaks are broad and not dispersed, indicating that the protein is unfolded and the GF buffer conditions are not suitable for the protein to remain stable. The broad peaks arise from the overlapping of the residual peaks due to dimerisation of the protein.

The $^{1}$H TbrPDEB1 spectra however contains very sharp and intense resonances that are not protein in nature. These could be derived from non-deuterated DTT present in the final gel filtration buffer. They could also be from glycerol, which was present in Nickel Affinity buffers, desalting buffer and ion-exchange buffers, and not completely filtered out during the final buffer exchange in the gel filtration chromatography. This could also arise from residual water [174].

## 4.4.2 $^{15}N/^{1}H$ HSQC NMR Data Acquisition with TbrPDEB1

$^{15}N/^{1}H$ Heteronuclear Single Quantum Correlation (HSQC) experiment was acquired on the spectrometer Avance III 600 MHz (F2) and 60 MHz (F1) at 25°C, with 1024 points (9000 Hz) in the direct F2 dimension ($^{1}H$) and 256 points (2100 Hz) in F1 ($^{15}N$) indirect dimensions using 50 ms EBurp Pulse 2. There were a total number of scans of 8 within an acquisition time of 0.11 seconds in F2 dimension and 0.07 seconds in the F1 dimension. The relaxation delay time (D1) in the experiment was 1.5 seconds. Watergate was the water suppression used with pulsecal. The size of the real spectrum was 4096 in the F2 dimension and 512 in the F1 dimension.

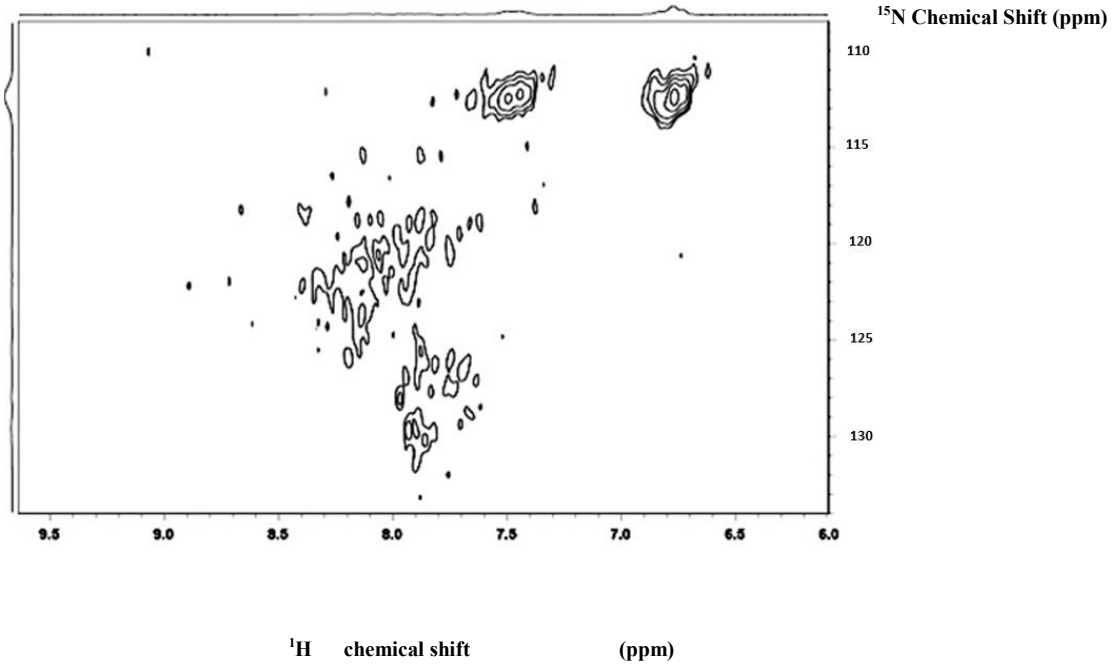



**Figure 4.16. $^{1}H/^{15}N$ HSQC Spectrum of TbrPDEB1 at 25 °C** None of the TbrPDEB1 protein peaks in the spectrum can be identified. The peaks are all broad and unclear, suggesting that the protein is unfolded and cannot be assigned with further triple resonance through-bond experiments.

$^{1}H/^{15}N$ HSQC experiments (Figure 4.16) confirm that TbrPDEB1 is unfolded and does not inform about any of amino acid groups. The spectrum cannot be assigned through-bond triple resonance experiments, as none of peaks can be matched with a specific amino acid from the TbrPDEB1 protein sequence. A perfect spectrum would not be expected for this sample, as it is a 37kDa alpha helical protein and this is not a TROSY spectrum. It could be a result of aggregation and there could be some internal dynamics (a dimer in solution).

### 4.4.3 NMR experiments with temperature variation

The stability of TbrPDEB1 was investigated by proton-based experiments at a temperature range from 25 °C to 50 °C from a $^{15}N$ isotopically enriched sample expressed in *E.coli* T7 express cells grown in minimal media and purified in GF Buffer (25 mM $NaH_2PO_4$, 100 mM NaCl and 1 mM DTT pH 6.5) as described in Chapter 2.5. The final concentration of the protein was 0.3 mM.

Increasing the temperature at which the spectrum of an NMR sample is measured at is a good technique to improve the resolution and decrease the broadness of overlapping NMR peaks by decreasing the rotational correlation time of the protein. However, improvements in the spectrum's quality due to a decrease in the rotational correlation time have to be balanced against a reduction in the stability of the proteins fold as it approaches its thermal unfolding limit leading to aggregation or exchange with unfolded or partially folded species. (Figure 4.17) [175].

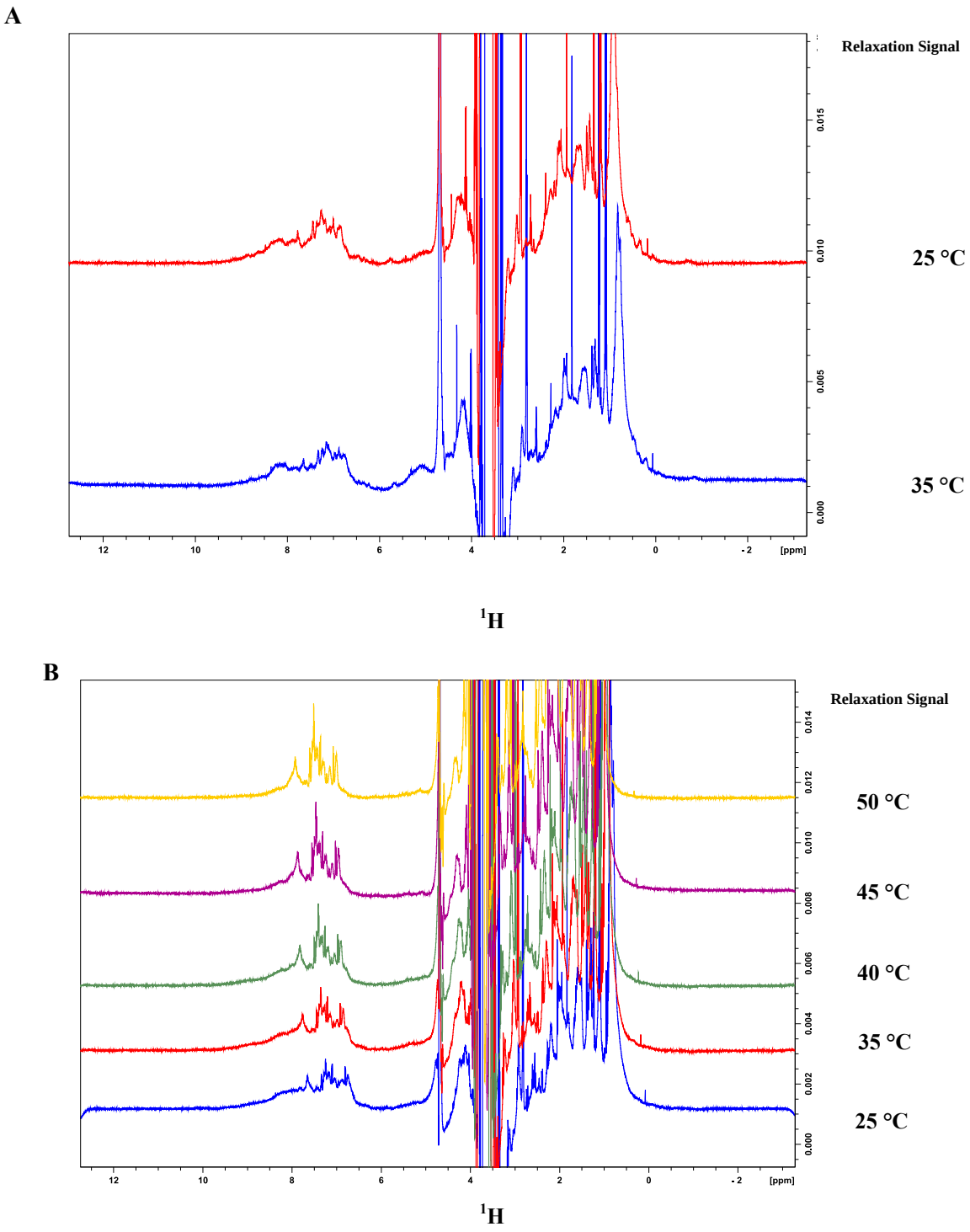



**Figure 4.17. Spectrum of $^1$H TbrPDEB1 at 25 °C to 50 °C** A) The red spectrum represents the data at 25 °C and the blue spectrum at 35 °C. B) Yellow 25 °C, purple 35 °C, green 40 °C, red 45 °C and blue 50 °C. The peaks are broad and not dispersed, revealing that the protein remains unfolded at all temperatures.

Increasing the temperature did not improve the resolution or the resonance of the $^{1}H$ TbrPDEB1 spectrum. The broadness and low-dispersion of the lines were still observed at all temperatures, indicating that TbrPDEB1 was not suitable for NMR characterisation studies. Again non-protein peaks are present in the spectra, which could be explained by DTT, glycerol and/or residual water.

## 4.5 Summary and Discussion

Chapter 4 has attempted to characterise the structure of TbrPDEB1 using X-ray crystallography and NMR methods. TbrPDEB1 was studied in complex with the ligands VUF 14234, PPS 59083, PPS 60036 and VUF 14450. X-ray crystallography studies revealed a conserved binding mode, where the active site consisted of magnesium and zinc ions. The residue Gln874 was conserved in all the crystal structures, which is the main residue from the parasite-specific P-pocket. In fact, all the ligands investigated in this study were directing themselves towards the P-pocket residue, which suggested that this P-pocket can be used as an important target in the design of therapeutic drugs for the treatment of HAT.

Unfortunately, 1D and 2D-based heteronuclear $^{15}N/^{1}H$ experiments resulted in broad and poorly dispersed peaks. Optimisation by increasing temperatures from 25 °C to 50 °C did not improve the spectra of TbrPDEB1. Therefore, the NMR screening methods could not be performed due to the instability of TbrPDEB1. Running $^{15}N/^{1}H$ experiments with a ligand could have perhaps improved the quality.

TbrPDEB1 was a good crystal example, as the interactions of the ligand with the protein could be characterised with the four different ligands studied. The protein's size, stability in NMR buffer and NMR tumbling rate were considered as limiting factors for such NMR studies. Therefore, this study highlighted that not all proteins can be suitable for the purpose of drug design using the NMR technique. However, this work is part of a publication in press with Nature Chemical Biology for its contribution in the treatment of HAT.

# Chapter 5

## Crystal Structures of BRD4-BD1 in complex with 7 ligands

## 5.1 Introduction

BRD4 (PDB code: 3MUK) (EC 3.1.1.31) is a nuclear protein belonging to the bromodomain and extra-terminal domain (BET) family, and is the most characterised protein within the BET family, which also consists of BRD2, BRD3 and BRDT (Figure 5.1). The BET bromodomain family members' main role is to recognise and selectively bind to discrete patterns of acetylated protein residues. Acetylation of histones is an important post-translational modification, affecting lysine residues in all histones and regulating DNA transcription, replication, repair and recombination [176] - [180]. Among those different post-translational modifications that exist in histones, including methylation, phosphorylation, ubiquitination and sumoylation, acetylation remains the most abundant histone modifications in humans, generally associated with active genes and ruling gene expression [180]. Histones are commonly acetylated by histone acetyltransferases, and deacetylated by histone deacetylases [179]. Histone and non-histone protein acetylations mediate their functions by recruiting bromodomains on acetylated lysine and specifically interacting with them, which enable the modification of the chromatin to render it more or less accessible [176] - [182].

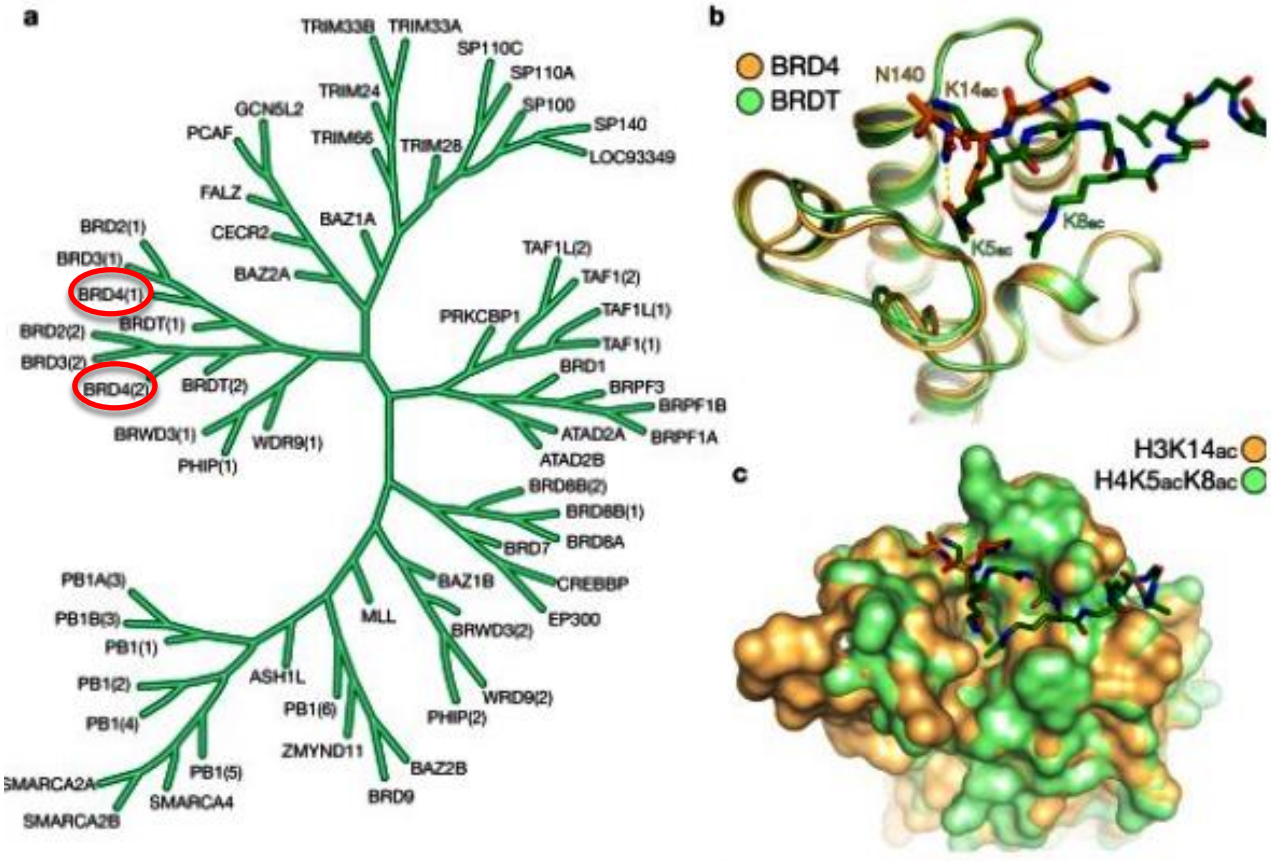



**Figure 5. 1.Phylogenetic tree of the human bromodomain family and substrate recognition of bromodomains.** (A) A total number of 61 bromodomains have been identified in 46 different proteins in the human proteome, which are actively involved in a wide range of cellular functions [176] - [181]. The phylogenetic tree is based on sequence alignments of predicted BRDs. Some BRDs recognise several targets, for which the domains are mentioned in brackets. BRD4 protein is the best characterised within the BET family. (B) Overlay of mouse BRD4 (orange cartoon model) with mouse BRDT (green cartoon model) with mono-acetylated Lys14 in histone H3 (orange sticks) and a di-acetylated H4 peptide mono-acetylated on both Lys5 and Lys8 (green sticks). (C) Surface representation in similar orientation [176] – [181].
*Adapted from*: Susanne Muller, Panagis Filippakopoulos and Stefan Knapp (2011) Bromodomains as therapeutic targets. *Expert Reviews in Molecular Medicine*. 13(29) [183]

All BET bromodomain family members possess two related and tandem bromodomains, BRD4-BD1 and BRD4-BD2 in their N-terminal domain and extra-terminal domain, each

recognising and binding to acetylated lysine residues on histone and non-histone proteins. Unlike the other BET family members, BRD4 contains a unique extended C-terminal domain. The bromodomain proteins are highly conserved and they both contain similar hydrophobic binding pockets (Figure 5.2). [176] – [181].

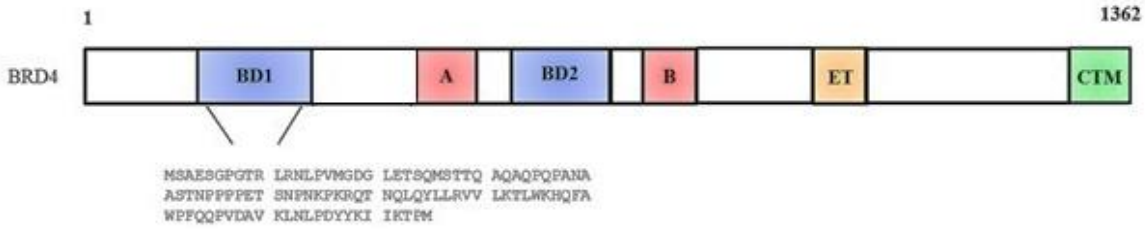



**Figure 5. 2. Domain organisation of the bromodomain BRD4 and extra-terminal domain (BET).** The evolutionarily conserved domains found in BRD4 are bromodomain 1 (BD1), bromodomain 2 (BD2), extra-terminal (ET), motif B and SEED (Ser/Glu/Asp-rich region – not shown in the diagram). Motif A and the carboxyl-terminal motif (CTM) are present only in some family members. Numbers indicate the amino acid limits of the human protein [176] – [181]. *Adapted from*: Chiang CM (2009) Brd4 engagement from chromatin targeting to transcriptional revolution: selective contact with acetylated histone H3 and H4. *F1000 Biol. RFep*. 1(98). [184]

Bromodomains are formed of approximately 110 residues that are arranged into a package of four α-helices, which are known as αZ, αA, αB and αC. The four helices are interconnected through loops, which are named ZA and BC loop regions, and the loops form a deep hydrophobic pocket, which is responsible for binding to acetylated lysine residues (Figure 5.3) [176] – [181] [183].

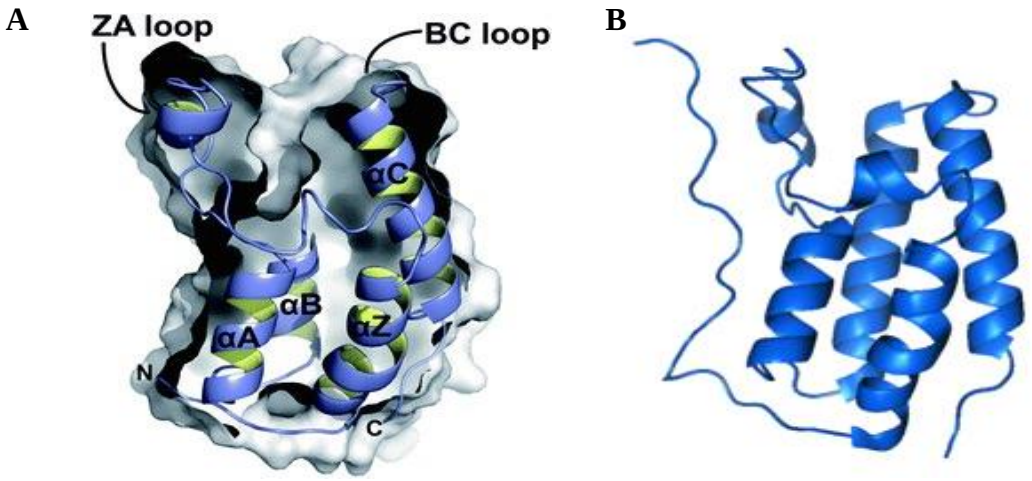



**Figure 5. 3. Comparison of (a) BRD4 crystal structure (PDB: 3MUK) with (b) experimental BRD4 crystal structure and the protein sequence of BRD4.** (A) Crystal structure of BRD4-BD1 (3MUK). The protein surface is shown in gray. The purple ribbons represent the αZ, αA, αB and αC helices, which are interconnected by ZA and BC loops. The loops form the hydrophobic acetyl-lysine binding pocket. (B) Apo crystal structure of BRD4-BD1 is compared with 3MUK, and it is clear that both the structures are similar. *Adapted from*: Silviya D. Furdas, Luca Carlino, Wolfgang Sippl and Manfred Jung (2012) Inhibition of bromodomain-mediated protein-protein interactions as a novel therapeutic strategy. *Med. Chem. Commun*. (3) 123-134 [185]

## 5.1.1 Functional Studies of BET BRD4

BRD4 binds many partners, including acetylated histone 3 (H3) and histone 4 (H4) tails as part of the maintenance of the architecture of chromatin, and acetylated v-rel avian reticuloendotheliosis viral oncogene homolog A (RelA) by binding to acetylated lysine 310 and thereof stabilising nuclear NF-κB in the control of DNA transcription. The unique extra terminal domain of BRD4 can also interact with many chromatin modifiers, such as histone methyltransferase NSD3 [176] – [181] [183]. BD1 and BD2 bind to the acetylated histone tails (H3 and H4), which then forms a complex with the acetylated cyclin T1 subunit of the

positive transcription elongation factor b. As the transcription elongation factor B is responsible for the phosphorylation of RNA Polymerase II by CDK9, bromodomain binding promotes the transcription elongation (Figure 5.4) [176] – [181] [183] [184].

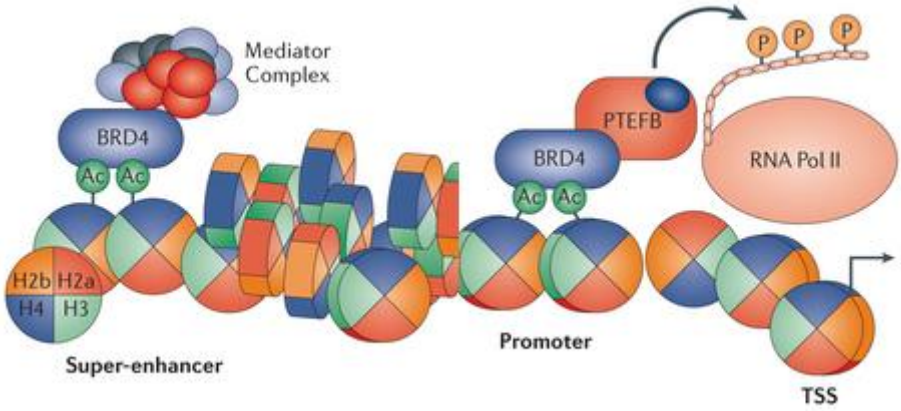



**Figure 5. 4. Mechanism of transcriptional regulation by BRD4 [150]** Bromodomains are transcriptional co-activators. BRD4-BD1 and BD2 bind to acetylated histone tails (H3/H4), and form complex with the acetylated cyclin T1 subunit of the transcription elongation factor B (PTEFB; the complex formed by cyclin-dependent kinase 9 (CDK9) and its activator cyclin T), leading to phosphorylation of the carboxy-terminal heptat repeat region of RNA polymerase II (RNA Pol II). *Adapted from*: Panagis Filippakopoulos and Stefan Knapp (2014) Targeting bromodomains: epigenetic readers of lysine acetylation. *Nature Reviews Drug Discovery*. 13, 337–356 [186].

*Jung et al*. studies proved that the strongest interactions of BRD4 BD1 and BD2 are with di- and tetra- acetylated peptides derived from the histone 4 N-terminal tail. They have also demonstrated that the neighbouring residues around the acetylated lysine significantly influence the binding mode of the bromodomains. Bromodomain binding is highly selective, as they specifically bind to designated patterns of acetylation with appropriate neighbouring amino acids. Moreover, not only promoter regions are recognised by bromodomains, but also

inter- and intra-genic regions. In fact, the BRD4 ET domain also binds to chromatin modifiers, such as the histone methyltransferase NSD3 [181].

### 5.1.2 BET BRD4 as drug target

Discovery of BRD4 inhibitors, such as JQ1 and I-BET151 have helped to have better insights in the understanding of the function of BRD4 (Figure 5.5). It has shown that BRD4 binds to target genes that are part of the cell cycle control and tumour diseases, such as c-Myc, c-Fos and cyclin D1. As such, BRD4 is an attractive target in certain tumours and cellular proliferative disorders *Shi et al*. have shown that a rare cancer NUT midline carcinoma (NMC) is believed to be caused by the genotype of a fusion between BRD4 (or BRD3) and the nuclear protein (NUT) initiated by a mutation of chromosomal translocation. Clinical trials using xenograft models have demonstrated in "*in-vivo*" that addition of the BET inhibitor JQ1 prevents the development of the NMC tumours [176] – [180].

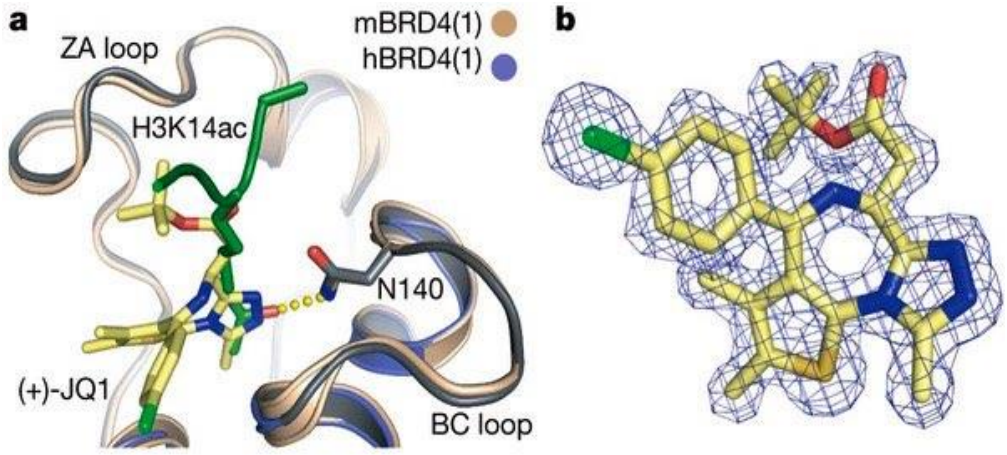



**Figure 5. 5. Interactions of BRD4-BD1 with JQ1** (A) Overlay of the mouse BRD4–H3K14ac peptide complex28 (brown cartoon protein model with green stick peptide model) with the human BRD4-JQ1 (purple cartoon protein model with yellow stick ligand model) complex structure. The hydrogen bond formed to the conserved asparagine (N140) in the peptide complex is shown as yellow dashed lines. (B) $|2Fo\text{-}Fc|_{\alpha c}$ electron density map of inhibitor JQ1 extracted from the BRD4-BD1/JQ1 complex contoured at 2 σ. *Adapted from*: Panagis Filippakopoulos et al. (2010) Selective inhibition of BET bromodomains. *Nature* 468: 1067-1073 [186].

Similarly, several other clinical studies with BRD4 and JQ1 have started in the treatment of haematological malignancies, such as acute myeloid lymphoma, acute lymphoblastic leukaemia, lymphoma, paediatric B precursor acute lymphoblastic leukaemia and multiple myeloma. Additionally, BET inhibitors have played inhibitory roles in solid tumours, including glioblastoma, neuroblastoma, lung cancer and melanoma, confirming that BET inhibitors have anti-proliferative effects [176] – [180].

The other common BET inhibitor, I-BET762, has been shown to protect again endotoxic shock and sepsis, therefore BRD4 may also be implicated in inflammation and inflammatory disorders. Other studies show that herpes and papilloma viruses use BRD4 in the host biotic chromosomes for their propagation during cell division. These multiple examples highlight

that the binding of BET bromodomains to acetylated lysines is a key determinant for the activation of many downstream cellular targets [176] – [180].

### 5.1.3 Structural studies highlighting key residues and interactions in BET BRD4 protein

Structural studies of BET bromodomains co-crystallised with histone-derived peptides demonstrated that there is a conserved asparagine residue in the BC loop of BRD4-BD1 making a hydrogen bond contact with the substrate acetyl-lysine side chain (Figure 5.6). This conserved residue and bonding is also observed in other bromodomains. Similarly, NMR studies of the complex BRD4-BD2 with NF-κB-K310 also revealed important interactions, such as Asn433 from the protein forming a direct hydrogen bonding with the acetylated lysine. Moreover, several other residues from the ZA loop, as well as the αB and αC regions, are also involved in the interactions with acetylated lysine. These "apo" structures were compared with crystal structures complexed with BET inhibitors, including JQ1 or I-BET762, and the comparison highlighted that the inhibitors bind to the loop regions in exactly the same way as the acetylated lysine substrates (Figure 5.6). Therefore, they are both good mimicking inhibitors of the substrate [176] – [180].

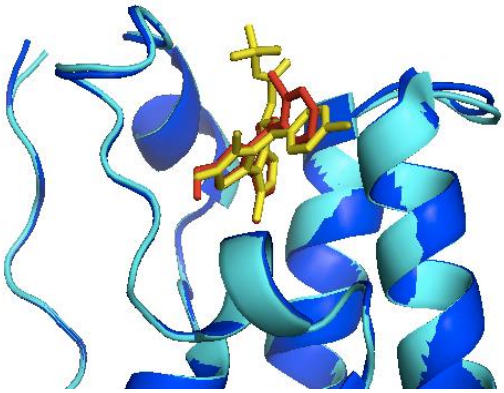


**Figure 5. 6. Comparison of BRD4-BD1/I-BET762 complex with BRD4-BD1/JQ1 complex.** Overlay of BRD4-BD1 (blue cartoon model) complexed with I-BET762 (red stick model) (PDB Code: 4C66) and BRD4-BD1 (cyan) complexed with JQ1 (yellow stick model) (PDB Code: 3MXF). The overlay suggests that the binding mode is conserved regardless of the ligand type.

Mutational analyses of the key residues involved in substrate binding were undertaken to understand the effects on ligand binding, including Tyr139 and Tyr432 in BRD4-BD1, or Tyr139 and Val439 residues in BRD4-BD2. Following the mutations, the mutant bromodomains experienced a change in binding affinity and impaired interaction with acetylated chromatin in comparison with wild-type bromodomains. Other studies investigated the consequences of mutating Asn140 in BRD4-BD1 and Asn433 in BRD4-BD2. Both of the mutations led to the inhibition of binding of di-acetylated H4 peptides, revealing the importance of the hydrogen bond formed with the well-conserved asparagine residue. It has been also revealed that the neighbouring residues to Asn140 are also important for the interaction of BRD4-BD1 with acetylated RelA, such as Tyr139. Interestingly, these studies confirm that certain residues from the active site of BRD4-BD1 are important for the binding of the bromodomain proteins to acetylated substrates, and therefore interference with those residues could easily result in a failure in substrate binding [176] – [180].

Resistance mutations in many cancer drug targets is well described and therefore the development of next-generation inhibitors for bromodomains is essential. Further structural characterisation by X-ray crystallography would serve to improve the current understanding

of the mechanisms underlying specific interactions of bromodomains with chromatin. This would help in the design of future compounds with high potency targeting specifically the BET BRD4 family [176] – [180].

## 5.2 Recombinant expression and purification of BRD4-BD1 from *E.coli* and crystallisation

### 5.2.1 Recombinant expression and purification

Recombinant human pET24a BRD4-BD1 (147 residues) with kanamycin resistance was expressed and purified to homogeneity in *E.coli* T7 express cells grown in 2 xYT media. A yield of 9 g pellet was obtained.

The first step of purification involved nickel-affinity chromatography using Buffer A on 5 mL His-Trap Column (50 mM HEPES pH 7.5, 500 mM NaCl, 5 % (v/v) glycerol, 10 mM imidazole) for equilibration, and Buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 5 % (v/v) glycerol, 250 mM imidazole) for elution (Figure 5.7).

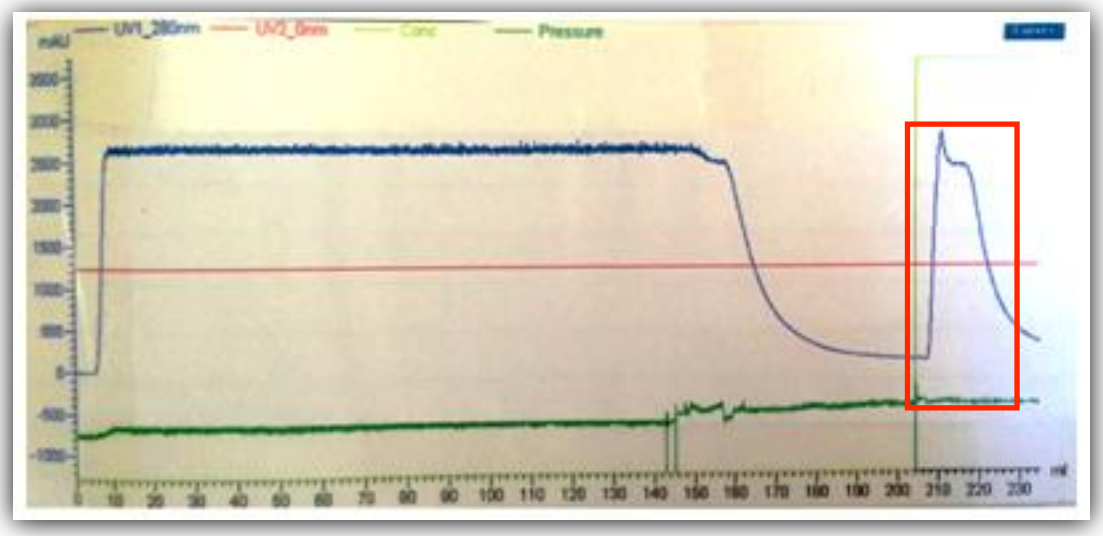



**Figure 5. 7. Elution Profile of BRD4-BD1 on His-Trap 5 mL Immobilised Nickel Affinity Chromatography** The peak of elution (red rectangle) represent the His-tagged BRD4-BD1 protein.

The first eluted fraction was subjected to TEV (stock: 2 units/μL) cleavage and was added on the basis of 2 units/mL, and the fraction was stored at 4 °C overnight. The final step of the purification of BRD4-BD1 was gel filtration chromatography using GF Buffer on Superdex S-75 Column (10 mM HEPES pH 7.5, 50 mM NaCl, 5 % (v/v) glycerol) for both equilibration and elution (Figure 5.8).

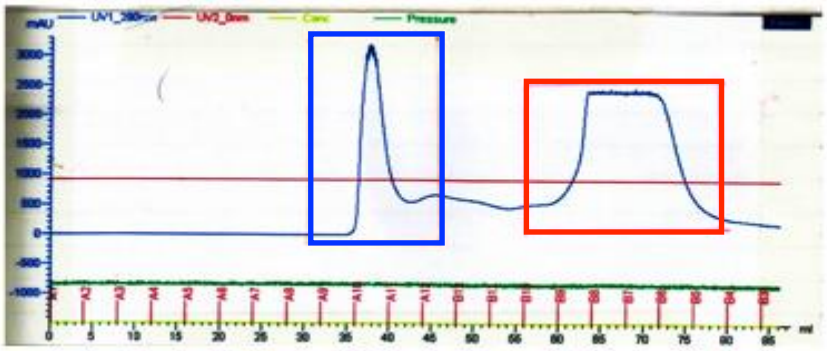


**Figure 5. 8. Elution Profile of BRD4-BD1 on Superdex S-75 Column Gel Filtration Chromatography**

Fractions A9 - A12 (blue rectangle) show the elution of aggregates and fractions B10 - B5 (red rectangle) show the elution of the pure protein BRD4-BD1.

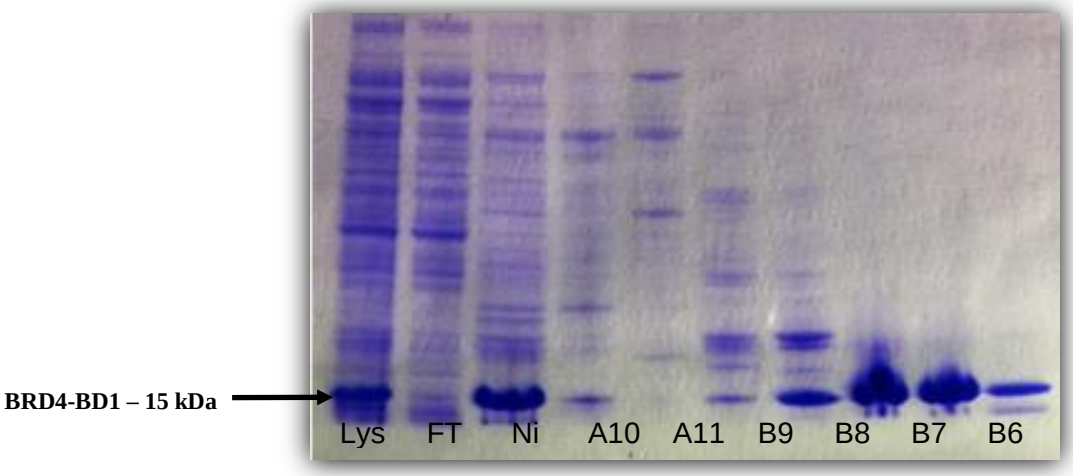



**Figure 5. 9. Purification of BRD4-BD1 by Nickel-Affinity Chromatography and Gel Filtration Chromatography** The purification fractions on the 4-12 % (w/v) Bis-tris gel show ~ 70 % purity of BRD4-BD1 at the expected molecular weight of 15 kDa after nickel affinity chromatography (Ni). The gel shows ~ 90 % purity of protein after gel-filtration chromatography (B8-B5). Lys: lysate, FT: flow-through, Ni: Nickel elution of BRD4-BD1, A10 - B9: Aggregate peaks, B8 - B5: BRD4-BD1 purified protein.

The final purified yield of BRD4-BD1 was 25 mg per litre.

## 5.2.2 Protein Crystallisation

The crystallisation of BRD4-BD1 was performed following the crystallisation conditions described by Vollmuth, F. and Geyer, M [187]. The protein was concentrated at 12 mg/mL, equivalent to ~ 0.8 mM for crystallisation. Best crystals for diffraction were obtained from 3.5 - 5 M sodium formate conditions after 24 hours at 4 °C (Figure 5.10).



**Figure 5. 10. BRD4-BD1 crystals in 3.5 M sodium formate** The BRD4-BD1 crystals always tend to grow with an amount of precipitation in the background. The crystals diffracted up to 1.2 Å on IO4.1 beamline at a wavelength of 0.97949 Å, temperature of 100 Kelvin with a 100 % exposure for 0.1 seconds / 0.1º oscillation.

The space group was $P2_12_12_1$. Previous studies have shown that this is a soakable protein system [187].

## 5.2.3 Addition of Ligands

Compounds for soaking as part of the development of the $^{19}F$ fragment library were identified by Maybridge Ltd (Cambridge, UK) and Charles River Early Discovery. Surface Plasmon

Resonance (SPR) undertaken with Dr. Steve Irving (Charles River Early Discovery) and Kimberley Young was used to detect molecular interactions of ligands with BRD4-BD1, and determine the corresponding binding affinities ($K_D$) using DMSO stock compounds. Seven best hits were selected for crystallographic studies within the micro-molar to nano-molar ranges (Table 5.1).

| Ligand | SPR ($K_D$) (μM) | Types of crystallisation | Soak | Co-Crystal |
|---|---|---|---|---|
| **BTB 01148** | 433 | Soak / Co-Crystal | × | ✓ |
| **BTB 06033** | 590 | Soak / Co-Crystal | ✓ | × |
| **BTB 07004** | 198 | Soak / Co-Crystal | × | ✓ |
| **RJF 00210** | 320 | Soak / Co-Crystal | × | ✓ |
| **TG 00013** | 0.9 | Soak / Co-Crystal | ✓ | × |
| **HTS 05027** | 580 | Soak / Co-Crystal | × | ✓ |
| **RJF 00002** | 750 | Soak / Co-Crystal | ✓ | × |

**Table 5. 1. Summary of the soak and co-crystallisation experiments** ✓ Data obtained, × No data could be obtained due to DMSO intolerance in soaks or no crystals were grown in co-crystallisation.

DMSO is known to bind to BRD4-BD1, therefore compound solids were dissolved to 100 mM stock in ethanol. Compounds were used in equimolar ratio with protein concentration (0.8 mM). Crystals were soaked for 2 days at 4 °C before cryo-cooling in liquid nitrogen with 20 % ethylene glycol cryosolvent for data collection.

In addition, the co-crystals were performed at a ratio of 2 : 1 ligand to protein excess under the same conditions as apo crystal crystallisation conditions.

## 5.3 Crystallographic Solutions of BRD4-BD1

All data was collected at DIAMOND on IO4.1 beamline at a wavelength of 0.97949 Å, at a temperature of 100 Kelvin with a 100 % exposure for 0.1 seconds / 0.1º oscillation. The space group was orthorhombic $P2_12_12_1$.

Data was auto-processed through Xia2 [138]. As the auto-processed data sets missed a $2_1$ screw axis, a simple molecular replacement with Phaser [188] into $P2_12_12_1$ was performed using the apo structure of BRD4-BD1 as search model. Ten rounds of structural refinement using Refmac5 [136] were performed and the difference density maps were examined for presence of the ligand. Once satisfied with the presence of the ligand, the structure was finalised with Buster.

### 5.3.1 BRD4-BD1 with ligand structures

The results of the crystallographic data are summarised in Table 5.2 and 5.3:

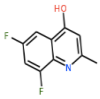
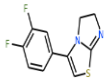
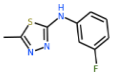
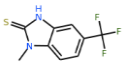
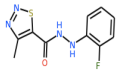
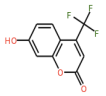
| Compound | BTB01148 | BTB06033 | BTB07004 | RJF00210 | TG00013 | HTS05027 | RJF00002 |
|---|---|---|---|---|---|---|---|
| |  |  |  |  |  |  |  |
| Space Group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit cell dimensions (Å) (a,b,c) | 32.8, 47.24, 78.56 | 46.85, 78.14, 34.62 | 34.11, 47.24, 77.79 | 32.02, 47.09, 79.09 | 32.20, 47.05, 78.64 | 31.96, 47.36, 79.03 | 34.05, 46.92, 78.40 |
| Unit Cell Angle (°) | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 |
| Number of observed reflections | 227,873 (7,018) | 75,905 (5,271) | 84,091 (6,549) | 200,051 (12,625) | 126,225 (9,074) | 108,228 (7,857) | 91,431 (6.639) |
| Number of unique reflections | 39,832 (2356) | 12,490 (893) | 13,641 (976) | 32,336 (2,318) | 21,307 (1,522) | 18,085 (101) | 14,522 (1,030) |
| Oscillation range per frame (°) | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| Overall rotation (°) | 180 | 180 | 180 | 180 | 180 | 180 | 180 |
| Resolution Range (Å) | 32.8 - 1.18 | 40.19 - 1.79 | 31.24 - 1.72 | 30.29 - 1.27 | 39.32 - 1.46 | 40.63 - 1.55 | 26 - 1.69 |
| Multiplicity | 5.7 (3.0) | 6.1 (5.9) | 6.2 (6.7) | 6.2 (5.4) | 5.9 (6.0) | 5.4 (4.8) | 6.3 (6.4) |
| Completeness (overall) (%) | 97.3 | 99.3 | 98.2 | 99.8 | 99.4 | 99.4 | 98.9 |
| Mean I/sigma (I) (overall and last shell) | 15.4 (2.2) | 11.5 (2.0) | 14.9 (2.4) | 17.2 (2.8) | 9.3 (2.6) | 10.30 (1.70) | 9.3 (2.6) |
| Rmerge (%) (overall and last shell) | 5.5 (5.3) | 0.074 (0.763) | 0.058 (0.835) | 0.064 (0.688) | 0.121 (0.693) | 0.136 (0.715) | 0.061 (0.773) |

**Table 5. 2. Summary of the X-ray crystallography data collection of BRD4-BD1 with ligand structures**

The data was solved and refined through Phaser-Molecular Replacement, Refmac5 and Wincoot.

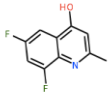
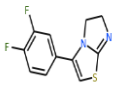
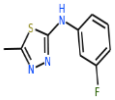
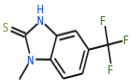
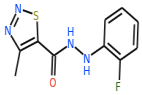
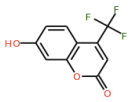
| Compound | BTB01148 | BTB06033 | BTB07004 | RJF00210 | TG00013 | HTS05027 | RJF00002 |
|---|---|---|---|---|---|---|---|
| |  |  |  |  |  |  |  |
| Maximum Resolution | 1.18 | 1.79 | 1.72 | 1.27 | 1.46 | 1.55 | 1.69 |
| R-factor (%) (last shell) | 20 | 25.9 | 28.6 | 14.7 | 19 | 25.1 | 38.0 |
| R-free (%) | 25 | 25.78 | 34.6 | 18.9 | 22 | 27.5 | 40.0 |
| RMSD Bond lengths (Å) | 0.024 | 0.020 | 0.016 | 0.037 | 0.022 | 0.019 | 0.0142 |
| RMSD Bond Angles (°) | 2.437 | 2.099 | 1.942 | 2.878 | 2.097 | 2.077 | 1.714 |
| Average B-factor Protein ($Å^2$ ) | 17.128 | 45.988 | 36.056 | 12.267 | 15.237 | 16.360 | 39.229 |
| Average B-factor Ligand ($Å^2$) | 17.6 | 56.959 | 42.670 | 15.002 | 17.168 | 19.238 | 60.233 |
| Number of Protein Residues | 124 | 122 | 122 | 123 | 126 | 124 | 123 |
| Number of water atoms modelled | 188 | 75 | 94 | 182 | 157 | 58 | 52 |
| Ramachandran Plot preferred (%) | 99.10 | 96.30 | 95.61 | 97.5 | 98.29 | 96.72 | 94.92 |
| Ramachandran allowed (%) | 0.90 | 3.70 | 4.39 | 1.67 | 0.85 | 3.28 | 3.39 |
| Ramachandran generously allowed (%) | 0.00 | 0.00 | 0.00 | 0.83 | 0.85 | 0.00 | 1.69 |

**Table 5. 3. Summary of the X-ray crystallography structure refinement and Ramachandran Plot statistics of BRD4-BD1 with ligand structures.**

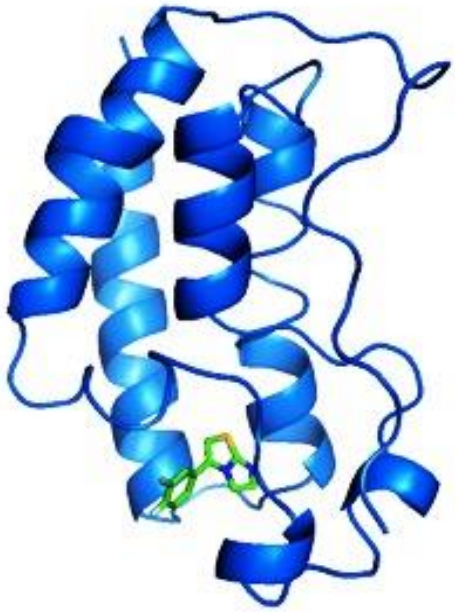


**Figure 5. 11. Crystal structure of BRD4-BD1 co-crystallised with the ligand BTB 07004**

The ligand BTB 07004 binds to BRD4-BD1. A detailed view reveals that BTB 07004 is binding in the loop region of BRD4-BD1.

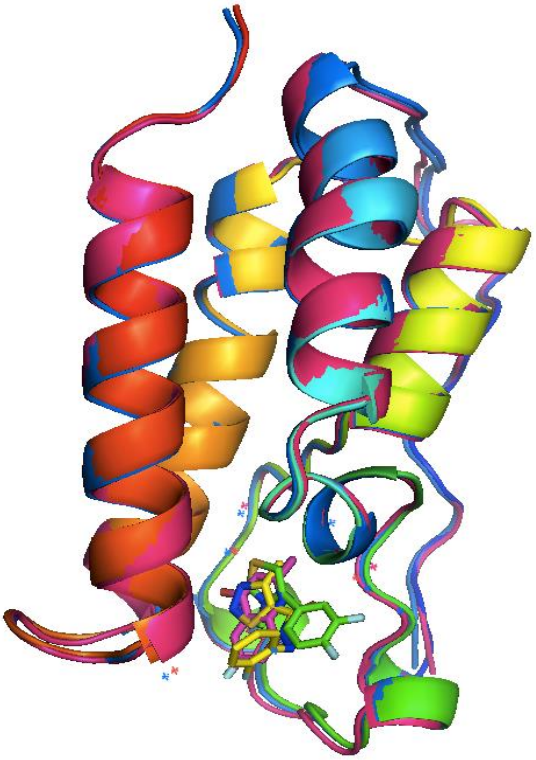


**Figure 5. 12. Overlay of crystal structures of BRD4-BD1 in complex with the ligands BTB 01148 (blue), BTB 06033 (pink), RJF 00210 (yellow), TG 00013 (green), HTS 05027 (orange) and RJF 00002 (cyan)**

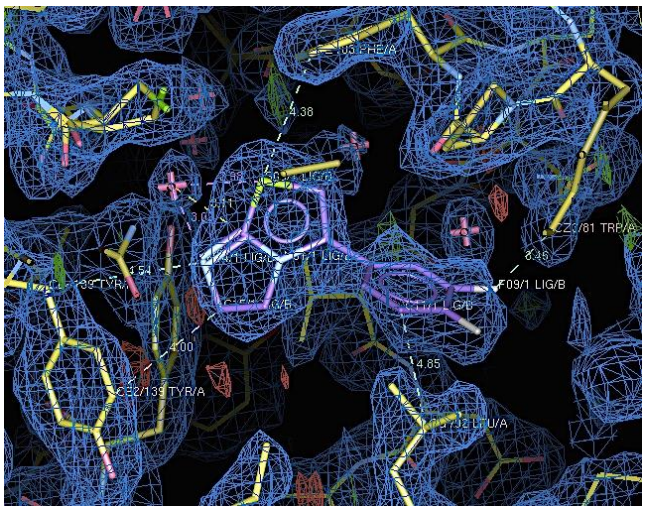
The ligands have bound to BRD4-BD1 in the same site but with different binding modes. The ligand binding in the loop region of the protein is conserved.

All the solved crystal structures of BRD4-BD1 complexed with different ligands either by co-crystallisation or soak show a common binding site with different modes of binding due to chemotype groups as observed in the known acetyl binding site with JQ1 [181]. As such, this suggests that the fragments are suitable for the development of new inhibitors.

Binding site analysis was required in all complexes to identify key residues and interactions, as well as unique features.

## 5.3.2 Analysis of the BRD4-BD1 binding site

**A**




**B**

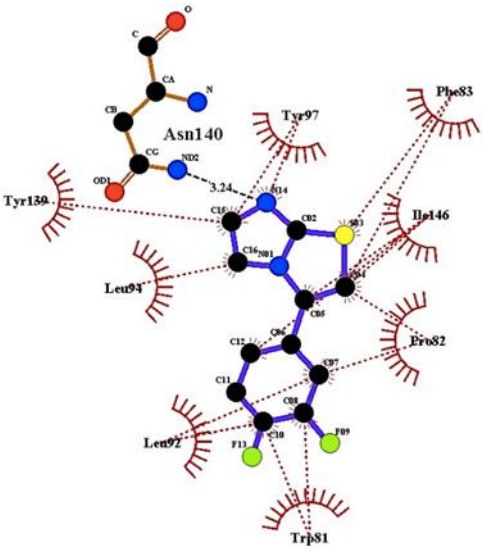
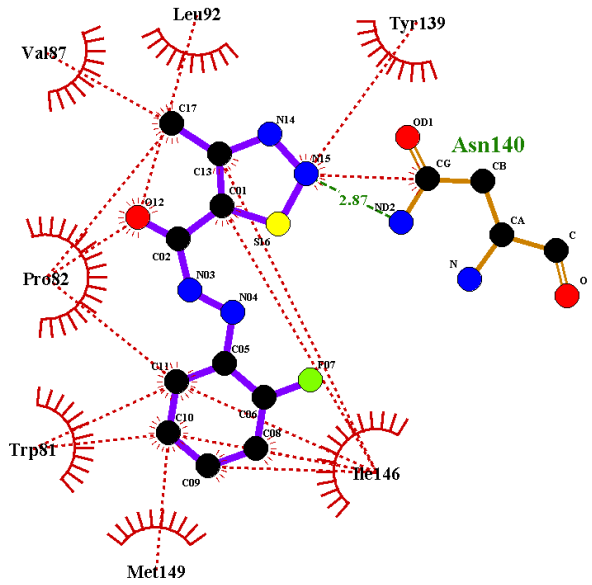
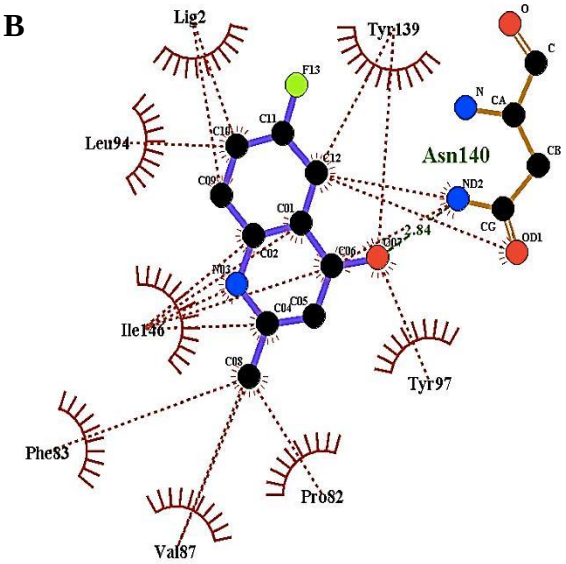



**Figure 5. 13. Ligplot representation of the binding site of BRD4-BD1 with BTB 07004** (A) $|2Fo\text{-}Fc|_{\alpha c}$ electron density map of ligand BTB 07004 complexed with BRD4 contoured at 1.09 σ. (B) Ligplot representation [172] of the interactions of BTB 07004 with key residues in BRD4-BD1 (dashed red semi-spheres lines represent hydrophobic interactions and green dashed lines representing hydrogen bonding).

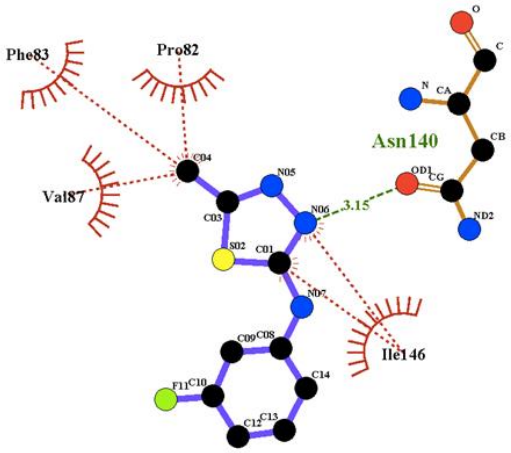
A


Leu92
Val87
Tyr139
Asn140
Pro82
Trp81
Ile146
Met149
2.87

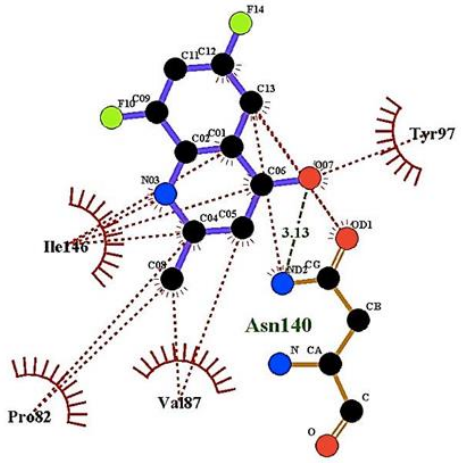
B


Lig2
Tyr139
Leu94
Asn140
Ile146
Tyr97
Phe83
Pro82
Val87
2.84

C


Phe83
Pro82
Asn140
Val87
Ile146
3.15

D


Tyr97
Ile146
Asn140
Pro82
Val87
3.13

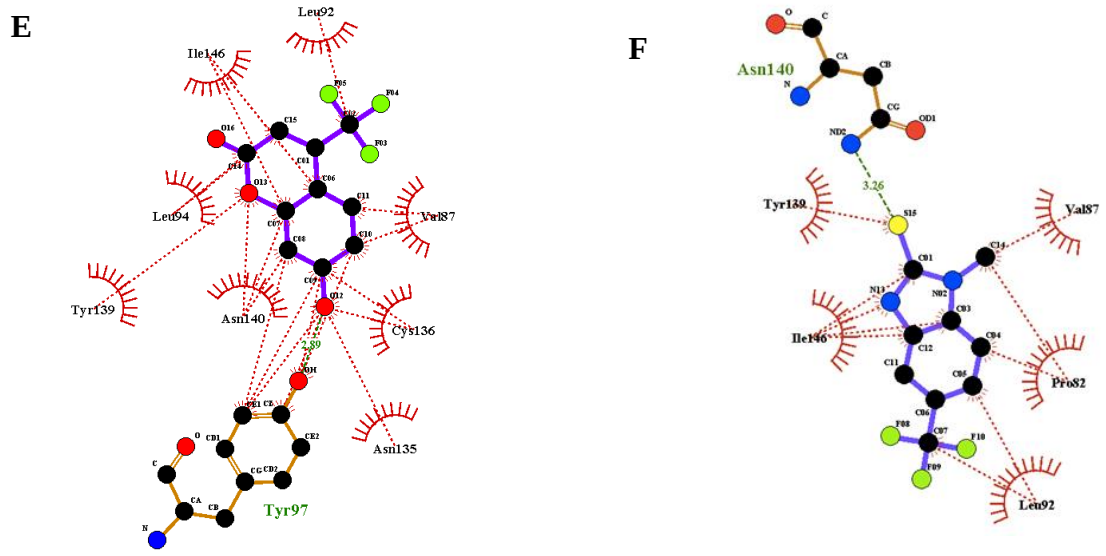



**Figure 5. 14. Ligplot representations of the binding site of BRD4-BD1 with different ligands**

(A) BTB 01148 (B) HTS 05027 (C) RJF 00210 (D) BTB 06033 (E) TG 00013 and (F) RJF 00002. Dashed red semi-spheres represent hydrophobic interactions; green interactions represent hydrogen bond interactions.

All these ligands are binding to BRD4-BD1. Their interactions with the key residues in the protein are being analysed through Ligplot studies. The examples show that the binding is conserved and the key residues responsible for binding are common in the ligand binding site of BRD4-BD1.

Figures 5.13B and 5.14 show that the ligands are predominantly stabilised by hydrophobic interactions. There is also the presence of the conserved hydrogen bond interaction between the nitrogen groups of the ligands and the residue N140, as seen in previous studies [176] – [180], except for the ligand RJF 00002. Figure 5.13a shows the Wincoot representation of the BTB 07004 ligand bound to the protein, where the incorporation of the ligand is well established. Figures 5.13b and 5.14 confirm that the key residues involved in binding in BRD4-BD1 are Y97, F83, I146, P82, W81, L92, L94, Y139 and N140 [189].

The ligands BTB 01148 (Figure 5.14A) and HTS 05027 (Figure 5.14B) connect hydrophobically to other residues not listed above, which are V87 and M149, respectively. As opposed to the other structures, the RJF 0002 ligand (Figure 5.14F) is making an interaction with Y97, which was not observed in the other structures. Although Y97 has already been seen to be involved in the binding process of the other ligands, it seems to be more involved by the formation of hydrogen bonding, in comparison to previous hydrophobic interactions. Moreover, C136 and N135 are also taking part in the binding of the ligand to the protein.

All these Ligplot interactions have characterised the ligand binding site of BRD4-BD1, which are highly conversed throughout all the studied structures. The residue N140 plays a major role by forming hydrogen bond interaction with each of the ligands studied, apart from the ligand RJF 00002, where only Y97 forms a hydrogen bonding. It is important to notice that the crucial interaction of hydrogen bonding with the residue N140 is conserved and observed in all the structures described above.

Normal fragment-based approaches for drug screening require high concentration bio-assay or biophysical screening by SPR, NMR, ITC or Thermal Shift Assay to identify bindings followed by X-ray crystallography to determine three-dimensional positions for optimal chemistry follow-up. X-ray fragment screening alone does not provide binding affinities but instead can be used as primary screen method. NMR can be used rather to infer binding affinity information through STD-NMR and titration experiments. NMR can only be used as a three-dimensional technique if the isotopically labelled protein is fully assigned, but the assignment can help in understanding protein-ligand interactions therefore obviating the need for X-ray crystallography.

## 5.4 Summary and Discussion

BRD4-BD1 was studied as part of the development of the fluorinated fragment-based library in collaboration with Maybridge (Cambridge, UK). The collaborators also provided with several batches of fluorinated compounds, which were then tested on the SPR with Dr Steve Irving and Kimberley Young. The fragments were thought to be well behaved in SPR, with most binders revealing "square wave" fast on/off sensorgrams. The non-specific binds were suppressed from the final list of hits, and a hit rate of 2.9 % was attained for BRD4. Among the fragment hits, seven compounds were chosen on the basis of their $K_D$ and X-ray crystallography was performed. Seven compounds were successfully crystallised in complex with the protein, either by co-crystallisation or soaking techniques.

X-ray crystallographic studies revealed a conserved binding site in the loop region of BRD4-BD1, validating previous studies. The loops of BRD4-BD1 (ZA and BC) are important elements for the ligand binding in the acetyl lysine pocket. The types of amino acid residues forming the loops determine the specificity of binding of the protein to the particular acetylated lysines. The residues forming the bromodomain loops differ in every bromodomain type, which renders them unique, and act as interesting drug targets. Effectively, the inhibitors targeting the loop regions can prevent the interaction with acetylated proteins and reverse the post-translational modifications causing the human diseases [176] – [180] [182]. The fragment ligands used in this study can act as interesting scaffolds that can be developed further as anti-BRD4 drugs for targeting the loop region, and therefore act as a potent inhibitor for the actions of BRD4-BD1. The residue N140 is making an important hydrogen bonding, which was conserved in all the BRD4 structures. Therefore,

N140 plays an important role in the binding pocket, as suggested in previous studies [176] – [180] [182].

Chapter 6 will expand on the NMR assignment of $^{15}N/^{13}C/^{1}H$ BRD4 and the chemical shift mapping of $^{15}N/^{13}C/^{1}H$ BRD4 complexed with BTB 07004. In addition, the $^{19}F$-CPMG was found out to be a very powerful and rapid screening technique.

# Chapter 6

## Ligand binding studies of BRD4-BD1 using NMR

This chapter's main emphasis is to understand the interactions of BRD4 with the ligand BTB 07004 using various NMR methodologies, and to characterise the protein residues involved in ligand binding. The ligand BTB 07004 was chosen because of its affinity of 198 μM by SPR, which is a suitable range for testing binding. STD-NMR is known to be a good technique to understand protein-ligand interactions in the μM range [56] [59].

The different NMR methods used were ligand-observed $^{1}H$ STD and $^{1}H$ CPMG, and protein-observed $^{1}H^{15}N$ chemical shift perturbation (which requires protein backbone assignment). The study also includes $^{19}F$-based NMR methods (CPMG) for detecting ligand binding. The crystal structure of the complex was already described (Chapter 5) and was used to highlight the advantages and disadvantages of each biophysical approach.

## 6.1 Ligand-observed NMR screening methods of BRD4-BD1 with BTB 07004

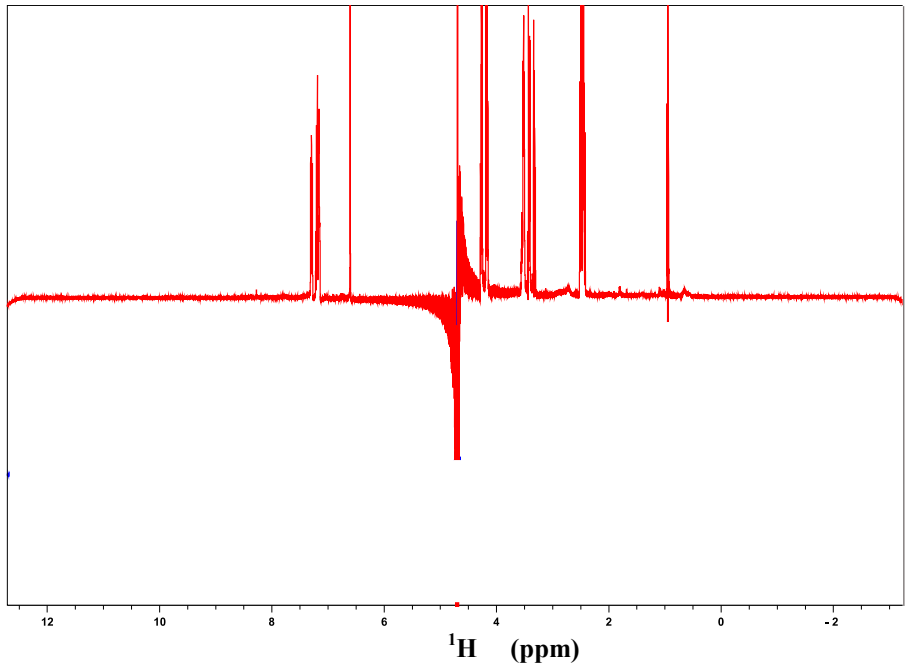
The ligand-observed NMR screening methods used in these studies are $^{1}H$ STD and $^{1}H$ CPMG. The purpose of these experiments is to identify the suitable binders to a protein. They do not require a full protein assignment or any NMR-based information about the receptor [56] [59] [66]. The structure of BTB 07004 bound to BRD4-BD1 is known from crystallography.

## 6.1.1 NMR Sample Preparation of BRD4-BD1/BTB 07004 complex

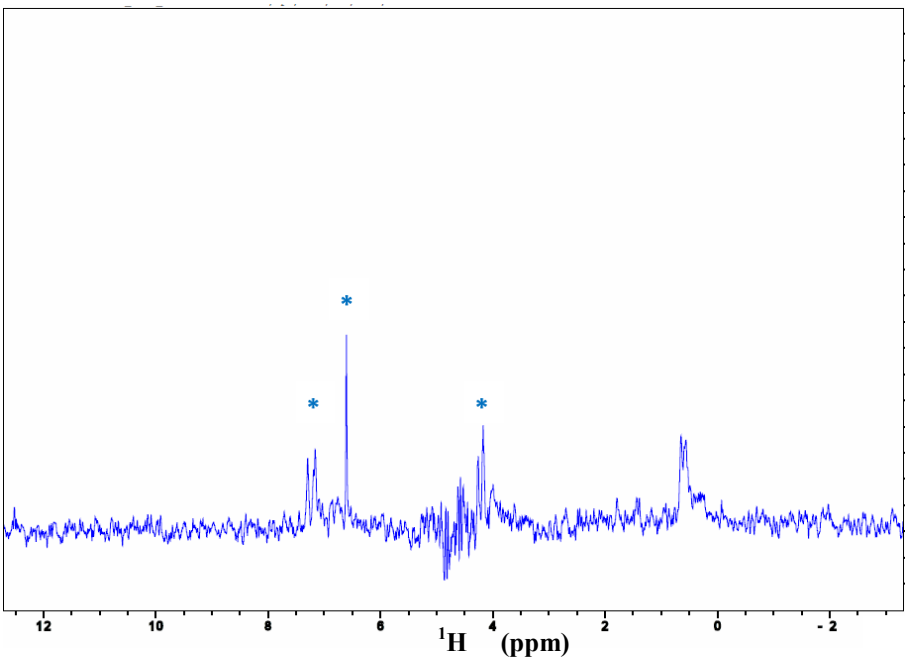
The binding of BTB 07004 with BRD4-BD1 was evaluated with STD and CPMG experiments using an unlabelled sample of BRD4-BD1 expressed in *E.coli* T7 express cells grown in rich LB media and purified in GF Buffer (25 mM $NaH_2PO_4$, 100 mM NaCl and 1 mM DTT pH 6.5) as described in Chapter 2.5. A concentration of 0.1 mM BRD4-BD1 was used with 1 mM BTB 07004.

## 6.1.2 NMR Sample STD Data Acquisition

$^1$H STD experiments were acquired on a Bruker Avance III 600 MHz spectrometer at 25°C, with 16,384 points (spectral width 9615 Hz) in the direct F1 dimension ($^1$H). A total number of 256 scans were acquired with an acquisition time of 0.85 seconds. Saturation was achieved using an EBURP2 pulse of over a 20 ms saturation period, saturation was applied at -0.3 / -30 ppm for the on and off resonance experiments respectively. The type of water suppression used was excitation sculpting and the spectra were referenced against water. The relaxation delay time (D1) in the experiments was 4 seconds. Data were zero filled, Fourier transformed using TopSpin 3.2, and an exponential line broadening of 10.0 Hz was used.




***OFF* Resonance**




***STD Difference***

**Figure 6.1. $^1$H STD-NMR experiments performed at 25 °C on 0.1 mM BRD4-BD1 protein complexed with 1 mM BTB 07004.** Red spectrum off-resonance spectrum; Blue spectrum final difference STD difference spectrum. The blue asterisk (*) indicate the BTB 07004 peaks.

As explained in Chapter 3.6.2, protons from the ligand that are closer to the protein receptor surface receive more saturation than protons that are further from the receptor. The *off*-resonance spectrum should show the ligand alone and the *on*-resonance should show the ligand signals with reduced intensities depending on how much saturation is transferred. Thus, the difference between *on*- and *off*-resonance gives the STD difference spectrum, revealing signals from ligand protons, which have bound to the protein receptor.

Therefore, the BTB 07004 difference peaks (highlighted with *) show binding at 6.5 and 7 ppm. This suggests that BTB 07004 binds to BRD4-BD1, correlating with the X-ray crystallography findings.

### 6.1.3 $^{1}$H Carr-Purcell- Meiboom Gill (CPMG) Data Acquisition and Analysis

$^{1}$H CPMG experiments were acquired on a Bruker Avance III 600 Hz spectrometer at 25°C. Data were collected with 16,384 points and a spectral width of 9615 Hz in the direct F1 dimension ($^{1}$H), with a total number of 128 scans and an acquisition time of 0.85s. The CPMG element used 600 cycles with a delay of 1ms between 180-degree pulses. The relaxation delay time (D1) in the experiments was 0.1 seconds. Water suppression was achieved using a 3:9:19 watergate sequence and 100ms of pre-saturation. Acquired data was processed in TopSpin 3.2 with an exponential line broadening of 1.0 Hz and references against water.

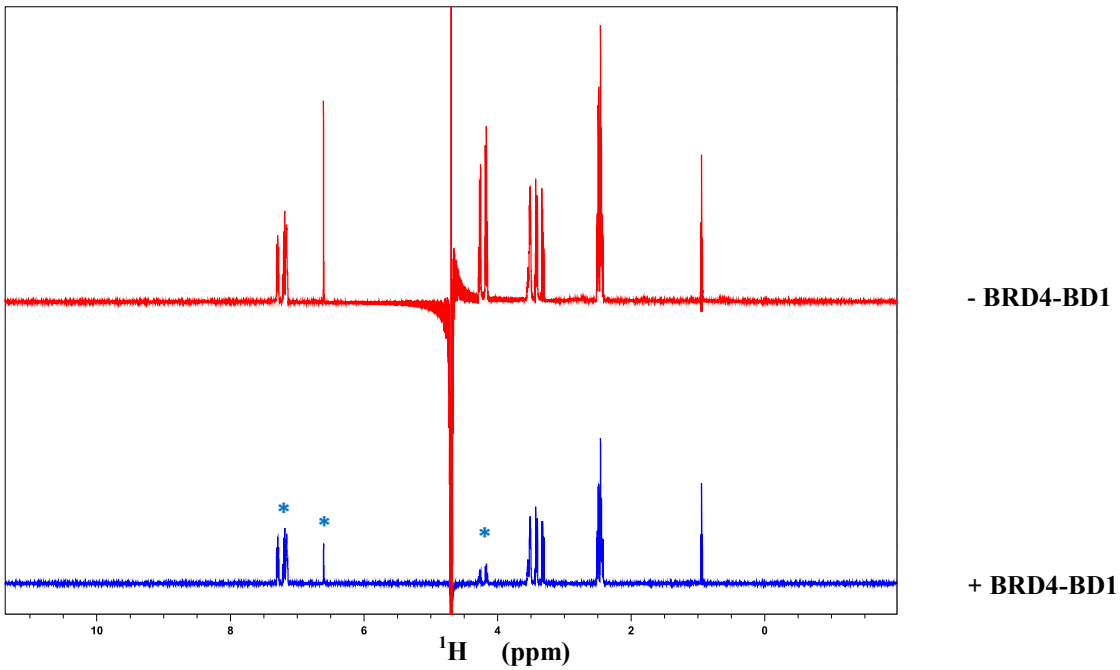



**Figure 6.2. $^{1}$H CPMG experiments performed at 25 °C on 0.1 mM BRD4-BD1 protein complexed with 1 mM BTB 07004.** Red spectrum BTB 07004 only; Blue spectrum BTB 07004 complexed with BRD4-BD1. All the $^{1}$H ligand peaks in the blue spectrum are attenuated in comparison to the red spectrum, indicating that BTB 07004 binds to BRD4-BD1.

As explained in Chapter 3.5.4, attenuation of $^{1}$H peaks shows binding of the ligand to the protein in CPMG experiments. Therefore, Figure 6.2 indicates binding of BTB 07004 to BRD4-BD1. The CPMG experiment results confirm the crystallographic data and $^{1}$H STD experiments of the binding of BTB 07004 to BRD4-BD1.

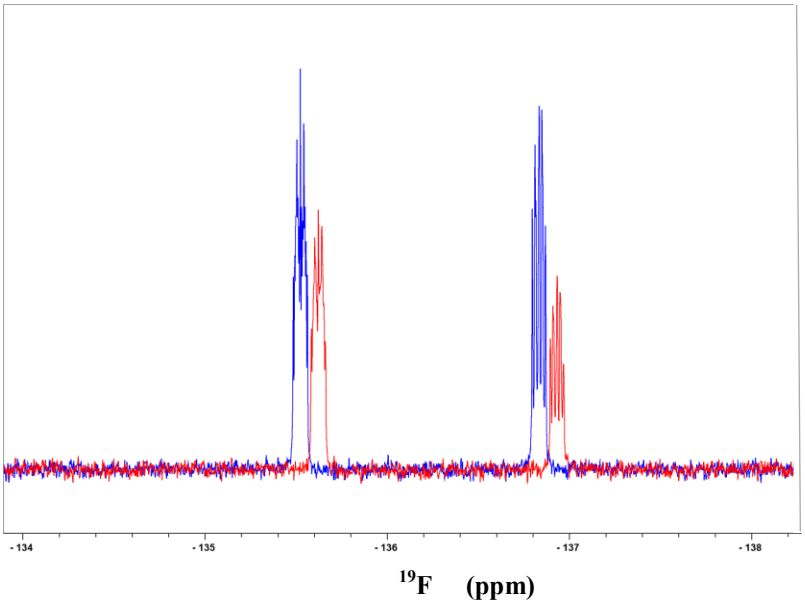
### 6.1.4 An Introduction to $^{19}$F CPMG

As BTB 07004 was part of a fluorinated fragment library, binding of BTB 07004 to BRD4-BD1 was examined by the use of $^{19}$F CPMG, where the measured signal is from a $^{19}$F nucleus rather than $^{1}$H. $^{19}$F CPMG is a similar experiment to $^{1}$H CPMG, where an attenuation of the $^{19}$F signal shows binding of the ligand to the protein. As for $^{1}$H CPMG, differentiation of binding is achieved through the spin-spin $T_2$ relaxation of the nucleus. The $T_2$ relaxation time shows a history of binding and unbinding events and the correlation time.

#### 6.1.4.1 NMR Sample $^{19}$F CPMG Data Acquisition

$^{19}$F CPMG experiments were acquired on a Bruker Avance III 600 Hz spectrometer at 25°C. Data were collected with 32,768 points and a spectral width of 9615 Hz in the direct F1 dimension ($^{1}$H), with a total number of 128 scans. The CPMG sequence used 600 cycles with a delay of 1ms between 180 degree pulses. The relaxation delay time (D1) in the experiment was 1 second. Data was processed with a line broadening of 5 Hz using exponential multiplication in TopSpin 3.2. 0.1 mM BRD4-BD1 was used with 1 mM BTB 07004.

**A**




**B**

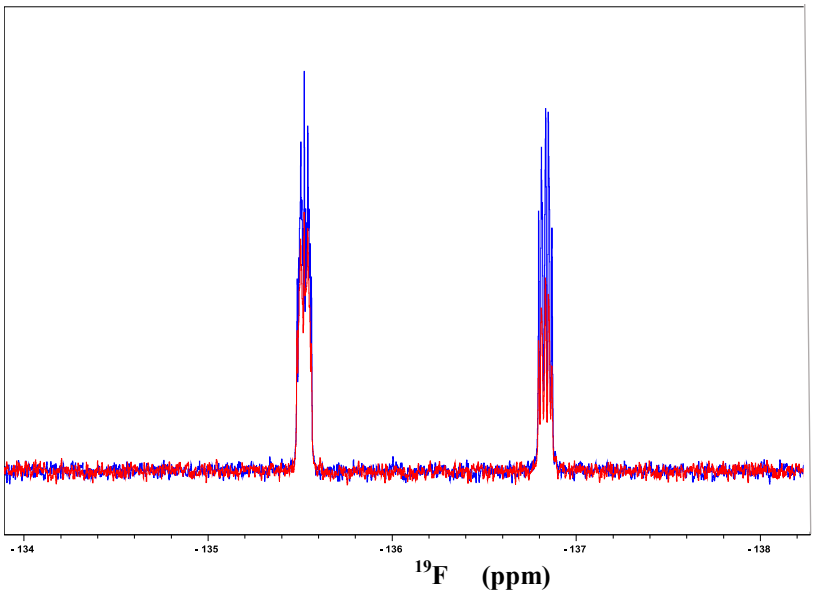



**Figure 6.3. $^{19}$F CPMG experiments performed at 25 °C on 0.1 mM BRD4-BD1 protein complexed with 1 mM BTB 07004.** Blue spectrum BTB 07004 only; Red spectrum BRD4-BD1 and BTB 07004 complex A) with a $^{1}$H shift offset of the two spectra for clarity B) with no offset. The addition of the ligand causes the reduction of signal in the red spectrum, indicating binding of BTB 07004 to BRD4-BD1.

Attenuation of $^{19}F$ peaks shows binding of the ligand to the protein in CPMG experiments. No changes in the $^{19}F$ signals show non-binding. Therefore, the data in Figure 6.3 indicate binding of BTB 07004 to BRD4-BD1. The $^{19}F$ CPMG experiment results confirm the crystallographic data, $^{1}H$ STD and $^{1}H$ CPMG experiments of the binding of BTB 07004 to BRD4-BD1.

All of these NMR screening methodologies are good tools to identify ligands/fragments that bind to the protein. $^{19}F$ CPMG is a powerful and rapid screening technique due to a fast accumulation of spectra with little spectra overlap being present and no interference from buffer components etc. In comparison to $^{1}H$ CPMG, the sensitivity does not differ much but a less crowded spectrum could be observed. $^{19}F$ CPMG offers selective spectra based on the types of fluorines present in the ligand. Therefore, large number of compounds can be looked at in each spectrum of $^{19}F$ CPMG without peaks overlapping. This is a great advantage in reducing the amount of NMR time and increasing throughput [66]. However, none of these ligand-based techniques are able to inform about the key residues in the protein binding site, and therefore are limited to simply identifying ligand “hits”. Consequently, this highlights the need for protein-observed methods for the prediction of protein ligand interactions in the application in structural-based drug discovery.

## 6.2 Protein-observed NMR screening methods of BRD4-BD1 with BTB 07004

Analysis of the protein-ligand interactions in NMR is usually undertaken with assignment of the backbone $^{15}N$ and $^{1}H$ resonances of an isotopically labelled protein, for which the protein needs to be folded, soluble and stable at a concentration of > 300 μM for a period of at least 1-2 days while in use in the spectrometer (typically in the temperature range of 10-25 °C, though lower or higher temperatures maybe used).

### 6.2.1 NMR Sample Preparation of $^{15}N$ BRD4-BD1

The stability of BRD4-BD1 was investigated by proton-based experiments at 25 °C using a $^{15}N$ isotopically enriched sample expressed in *E.coli* T7 express cells grown in minimal media and purified in GF Buffer (25 mM $NaH_2PO_4$, 100 mM NaCl and 1 mM DTT pH 6.5) as described in Chapter 2.5. The final concentration of the protein was 0.7 mM in an NMR sample of 300 μL.

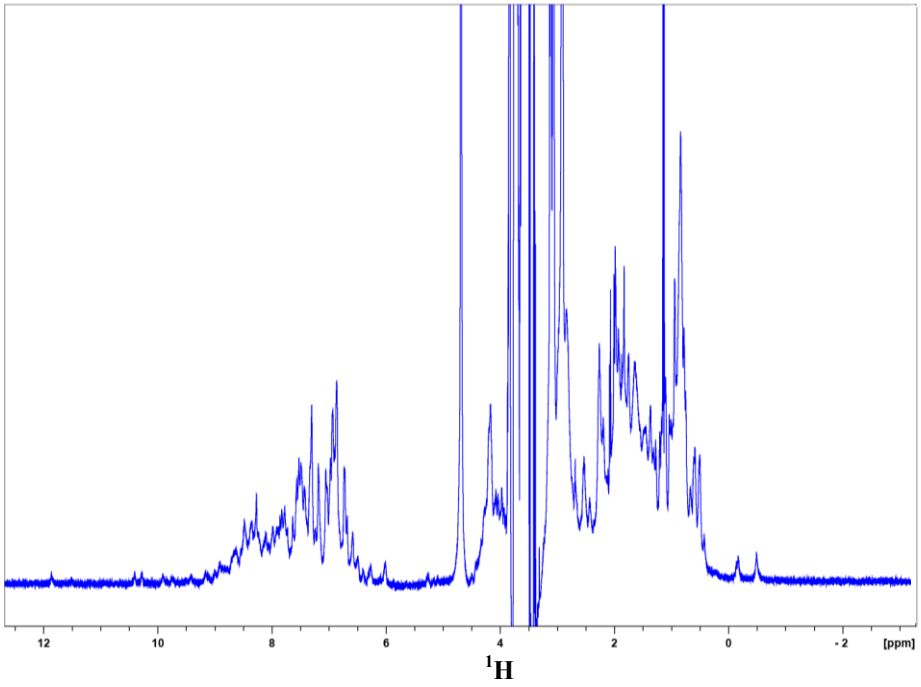



**Figure 6.4. $^{1}$H Spectrum of BRD4-BD1 at 25 °C with water suppression.** The peaks between 6 and 10 ppm originate from the amide/aromatics and side chain NH2s; peaks between 0 and 6 ppm contains the Cαs, water suppression artefacts and side chain CH, CH2 and CH3 resonances; peaks between 0 and -2 ppm originate from methyl groups packed against aromatic residues.

The $^{1}$H BRD4-BD1 experiment (Figure 6.4) shows that the protein peaks are sharp and well dispersed, which suggest that the protein is folded in the buffer conditions used. Therefore, 2D $^{15}$N/$^{1}$H HSQC experiments can be performed at 25 °C in GF buffer.

## 6.2.2 $^{15}$N/$^{1}$H HSQC NMR Data Acquisition with BRD4-BD1

The 2D $^{15}$N/$^{1}$H Heteronuclear Single Quantum Correlation (HSQC) experiment was acquired on a Bruker Avance III spectrometer at 600 MHz (F2) and 60 MHz (F1) at 25°C, with 2048 points (9615 Hz) in the direct F2 dimension ($^{1}$H) and 256 points (1824 Hz) in F1 ($^{15}$N) indirect

dimensions. There was a total of 8 scans per increment. Acquisition times were 0.11 seconds in F2 dimension and 0.07 seconds in the F1 dimension. Watergate was used for water suppression [190]. The relaxation delay time (D1) in the experiments was 1.2 seconds. The size of the processed spectrum was 4096 in the F2 dimension and 512 in the F1 dimension. Data was processing using Bruker TopSpin 3.2 and NMRPipe [191].

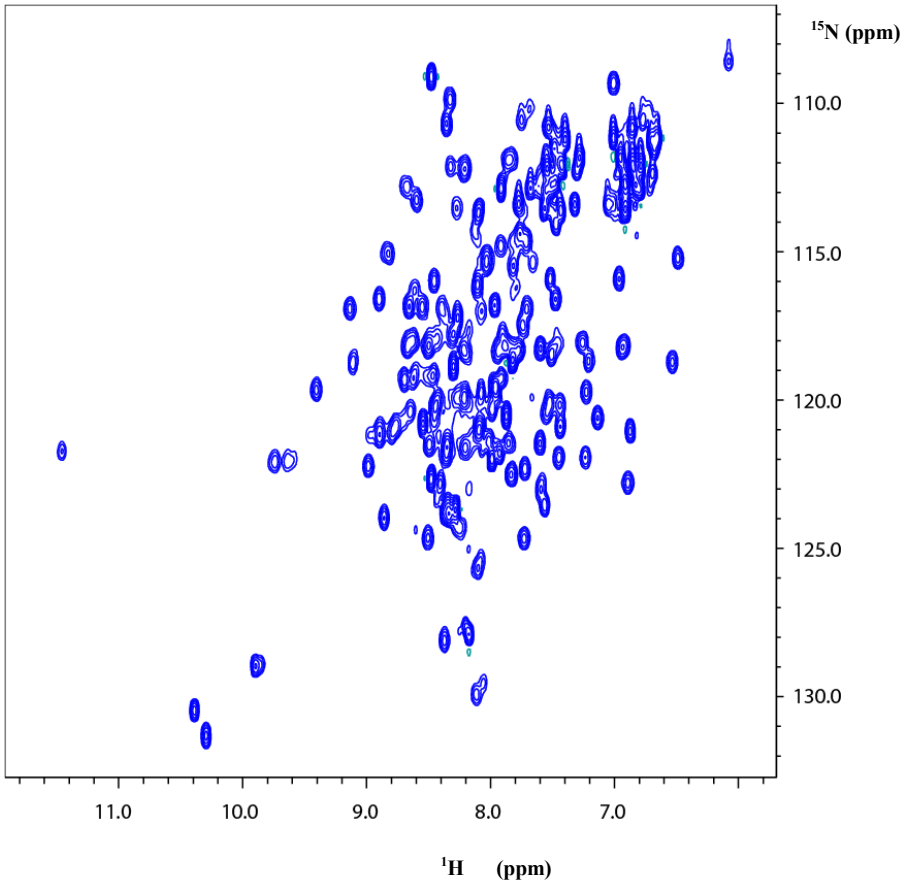



**Figure 6.5. $^{1}H/^{15}N$ HSQC Spectrum of BRD4-BD1 at 25 °C.** There are 120 peaks in the spectrum, which are compatible with 147 amino acids of BRD4-BD1, 6 histidine amino acids from the His-tag and 9 amino acids from the thrombin cleavage site. The peaks are all sharp and clear, and well dispersed which suggest that the protein is folded and can be assigned with further triple resonance through-bond experiments.

$^{1}H/^{15}N$ HSQC experiments confirm that BRD4-BD1 is folded and informs the presence of certain types of amino acid groups. However, the spectrum needs to be assigned using through-bond triple resonance experiments to be able to match each peak with a specific amino acid and amide from the BRD4-BD1 protein sequence. In comparison to Figure 3. 24. $^{1}H/^{15}N$ HSQC Spectrum of CypA at 25 °C, it can be noticed that the BRD4 peaks are broader than CypA. Despite the fact that BRD4 is a smaller protein than CypA, the protein sample of BRD4 was slightly precipitated and cloudy suggesting mild oligomerisation of the protein, which could explain the broadness of the peaks, an alternative source of broadening would be internal dynamics.

## 6.3 Characterisation of BRD4-BD1 by NMR Spectroscopy

### 6.3.1 Introduction

Assignments were performed using triple resonance spectra (HNCA, HN(CO)CA, HNCACB and HN(CO)CACB), which enable the identification of specific resonances of observed NMR active nuclei within a shift. Sequential assignment has always been the main method for amino acid backbone resonance assignment of proteins and has permitted the assignment of many large complex protein molecules [158] [159]. Once assigned, the NMR data for $^{1}H^{15}N$ HSQC and related spectra can be used for a variety of purposes, including measuring chemical shift

perturbations upon ligand binding, titrations to show the effect of changing pH, temperature, pressure and sophisticated experiments for measuring dynamics.

In order to characterise BRD4-BD1 by NMR spectroscopy, it was essential to carry out a sequential backbone assignment by triple resonance. Triple resonance experiments require the protein sample to be $^{15}N/^{13}C$ isotopically enriched, which would then allow measurement of data for $^{15}N$, $^{13}C_O$ $^{13}C_\alpha$ and $^{13}C_\beta$ atoms as well as all connected $^{1}H$ atoms. The NMR spectrum consists of resonances, which are each linked with particular nuclei in the protein being analysed and are assigned to a spin-system and mapped to specific amino acid residues in the protein sequence. Consequently, the resonance should be accurate and assigned to the type of amino acid, and the assignment must also be specific for individual nuclei in each amino acid in the protein sequence [157].

Although BRD4-BD1 has been extensively studied crystallographically, few studies have looked into its characterisation by NMR, hence a complete assignment is not available in the literature or at the BioMagRes Bank (BMRB). The through-bond triple resonance experiments were performed to identify each amino acid by its distinctive spin-system. This was followed by matching the adjacent spin systems to uniquely identify the position within the protein sequence, enabling the complete assignment of BRD4-BD1 [157] – [159].

Typically, the first experiment in the pair associates the amide $^{1}H$ and $^{15}N$ with the $^{13}C$ nuclei in a specific amino acid residue, known as “i”, and the $^{13}C$ nuclei of the preceding residue, known as “i-1” [HNCA, HN(CA)CO, HNCACB]. The second experiment of the pair associates the same amide $^{1}H$ and $^{15}N$ with the $^{13}C$ nuclei of the preceding amino acid residue only (i-1; HN(CO)CA, HN(CO)CACB, HNCO] (Figure 6.6). In this chapter, data from two pairs of

experiment [HNCA, HN(CO)CA, HNCACB, HN(CO)CACB] BRD4-BD1 is assigned. The HNCA correlates the amide $^{1}H$ and $^{15}N$ with the $^{13}C_{\alpha}$ of a residue i, and $^{13}C_{\alpha}$ of the previous residue i-1 (solid blue line - dashed blue line in Figure 6.6). The HN(CO)CA correlates the amide $^{1}H$ and $^{15}N$ with the $^{13}C_{\alpha}$ of a residue and the preceding residue. HNCACB and HN(CO)CACB correlate atoms in a similar way with the addition of information from the CB atoms [157] – [159].

Most of these experiments are carried in an out and back mode where the transfer of information is: $H_{(i)} \rightarrow N_{(i)} \rightarrow C_{\alpha\ (i,\ i-1)}* \rightarrow N_{\ (i)}* - H_{(i)}*$, where chemical shifts are only detected at the * residues. Transfer of magnetisation/information is achieved because of the large 1 bond couplings between the $^{1}H$ and $^{15}N/^{13}C$ labelled atoms [157] – [159].

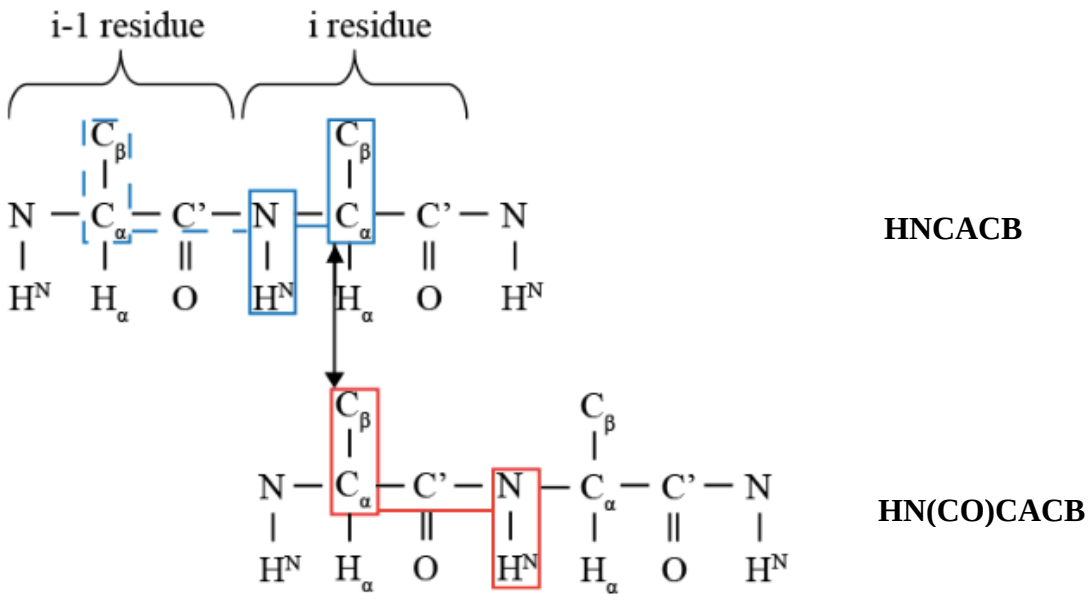



**Figure 6.6. Correlations in HNCACB and HN(CO)CACB experiments in a dipeptide fragment of a protein backbone.** The diagram shows that the HNCACB and HN(CO)CACB experiments enable the correlation of each NH group with $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts of one amino acid (i), and of the preceding residue.

Once the triple resonance data has been used to combine spin systems into sequential fragment, these must be matched to the primary sequence based on distinctive $^{13}C_{\alpha/\beta}$ and $^{15}N$ shifts, e.g.: a glycine has a $^{13}C_{\alpha}$ shift of ~45 ppm and an alanine has a $^{13}C_{\beta}$ shift of ~20 ppm. Fitting of chemical shifts to the proteins primary sequence was achieved using a mixture of Mapper 2 and manual methods [192]. In house python programs for the export of spin systems from CCPN Analysis to Mapper 2 were provided by G.S.Thompson at the University of Kent NMR Facility [157] – [159].

## 6.3.2 The Assignment of BRD4-BD1

### 6.3.2.1 NMR Sample Preparation of $^{15}N/^{13}C/^{1}H$ BRD4-BD1

The backbone resonance assignment of BRD4-BD1 was obtained by through-bond triple resonance experiments from a $^{15}N/^{13}C$ isotopically enriched sample expressed in *E.coli* T7 express cells grown in minimal media and purified in GF Buffer (25 mM $NaH_2PO_4$, 100 mM NaCl and 1 mM DTT pH 6.5) as described in Chapter 2.5. The final concentration of the protein was 0.7 mM.

## 6.3.2.2 NMR Data Acquisition of HNCACB and HN(CO)CACB experiments

HNCA, HN(CO)CA, HNCACB and HN(CO)CACB spectra were acquired at 25°C on a Bruker Avance III Spectrometer. Details of the experiments are summarised below:

| **HNCA / HN(CO)CA** | **F3** | **F2** | **F1** |
|---|---|---|---|
| **Nucleus** | $^{1}H$ | $^{15}N$ | $^{13}C$ |
| **Carrier Frequencies (ppm)** | 4.730 | 119.000 | 54.000 |
| **Frequency (Hz)** | 9615.385 | 1824.293 | 4828.507 |
| **Time Delay (points)** | 2048 (1024) | 64 (32) | 128 (64) |
| **Number of Scans** | HNCA: 16 scans and HN(CO)CA: 16 scans | | |

| **HNCACB / HN(CO)CACB** | **F3** | **F2** | **F1** |
|---|---|---|---|
| **Nucleus** | $^{1}H$ | $^{15}N$ | $^{13}C$ |
| **Carrier Frequencies (ppm)** | 4.699 | 120.000 | 38.997 |
| **Frequency (Hz)** | 9615.385 | 1702.670 | 1136.644 |
| **Time Delay (points)** | 1024 (512) | 60 (30) | 120 (60) |
| **Number of Scans** | HNCACB: 32 scans and HN(CO)CACB: 16 scans | | |

**Table 6.1 NMR Experimental Details for HNCA, HN(CO)CA, HNCACB and HN(CO)CACB experiments**

Data were processed with NMRPipe with a cosine window applied in all dimension, water suppression using a convolution difference filter and linear prediction in the $^{15}N$ dimension to increase the length of data by a factor of 2. All dimensions were zero-filled by a minimum of a factor of 2.

### 6.3.2.3 Amino Acid Sequential Backbone Resonance Assignment

The assignment of BRD4-BD1 using triple resonance relies on the distinctive chemical shifts of $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ for each amino acid. Certain amino acids can easily be identified, due to their characteristic chemical shifts, such as serine/threonine, which have a $^{13}C_{\beta}$ at a higher chemical shift field (60 – 70 ppm range) than the $^{13}C_{\alpha}$. Similarly, alanine has a $^{13}C_{\beta}$ at a lower field than the $^{13}C_{\beta}$ of any other amino acid residues, rendering them easily distinguishable. Glycine is another amino acid that can be easily identified as it does not have a $^{13}C_{\beta}$ and only its $^{13}C_{\alpha}$ can be visualised in the spectrum at approximately 45 ppm. Other residues that can be identified from distinctive chemical shifts include valine, proline and isoleucine, which have higher than normal $^{13}C_{\alpha}$ chemical shifts (Figure 6.7).

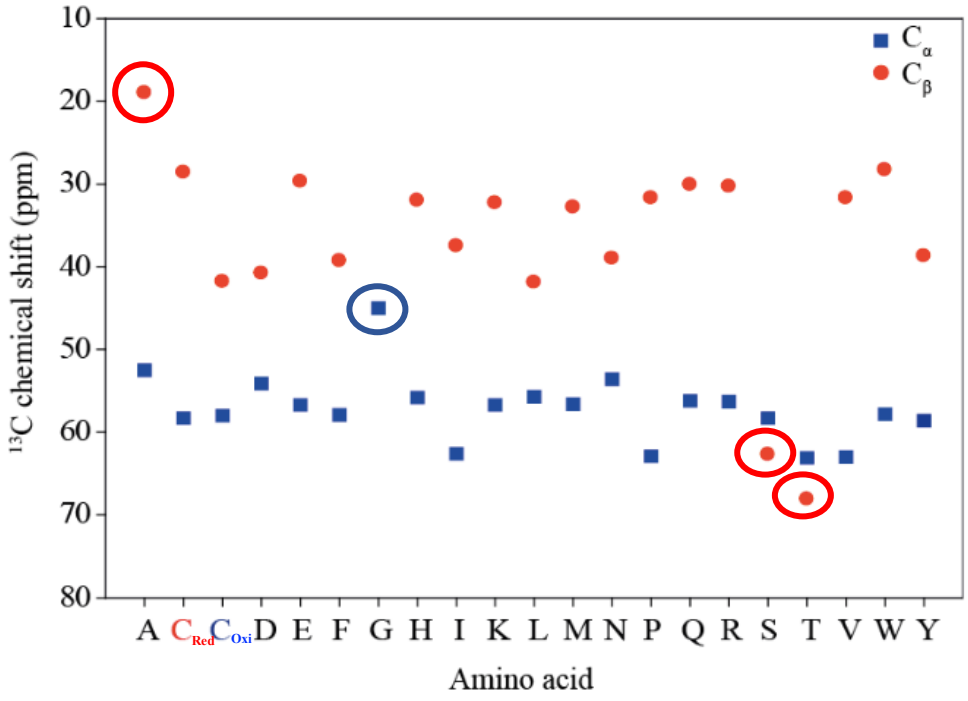



**Figure 6. 7. The NMR chemical shift values of $^{13}C_\alpha$ and $^{13}C_\beta$ resonances for each amino acid in random coil conformation.** The diagram clearly shows the distinctive patterns of serine/threonine, alanine and glycine in comparison to the other residues. Cysteine is shown in both its oxidised (Coxi) and reduced (Cred) forms. The most distinctive chemical shifts are circled. *Adapted from*: Wishart et al. (1995). $^{1}$H, $^{13}$C and $^{15}$N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbour effects. J Biomol NMR, 5, 67-81 [193].

The spectra obtained from the HNCACB and HN(CO)CACB experiments with the BRD4-BD1 protein were assigned using Analysis Version 2.3. The software helps analysing the data sets into spin systems and matching between the suitable spin systems. Linked spin systems (fragments) were then connected to the primary sequence using Mapper 2. The assignment of each of the amino acid residues was achievable with the $^{1}H_N$, $^{15}N$, $^{13}C_\alpha$ and $^{13}C_\beta$ resonances using the distinctive chemical shift patterns of serine/threonine, proline, glycine and alanine [190]. Spin

systems and fragments were manually and semi-automatically used with the software Mapper 2, which was used to place the fragments within the primary sequence.

### 6.3.2.4 Sequential Backbone Assignment of BRD4-BD1

The complete analysis of the triple resonance experiments for BRD4-BD1 led to the sequential backbone assignment of $^{1}H_{N}$, $^{15}N$, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ nuclei. These are listed in Appendix 6.1. Excluding the His-tag and thrombin cleavage site, 80 % of amide $^{1}H_{N}$ and $^{15}N$, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ shifts was assigned, though the absolute placement of one spin system within the primary was not definitive (2 possibilities). An example of the assignment from H77 to W81 shows the sequential links identified through the triple resonance experiments using both $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ (Figure 6.8). Of the remaining 20 % un-assigned residues, most were found to be residues adjacent to prolines (prolines are not visible on standard triple resonance assignment spectra because they lack an NH). As a result, the sequential link could not be established with those residues. One threonine residue could not be located either, possibly as it was situated next to another threonine, which can lead to a degeneracy and near identical chemical shifts.

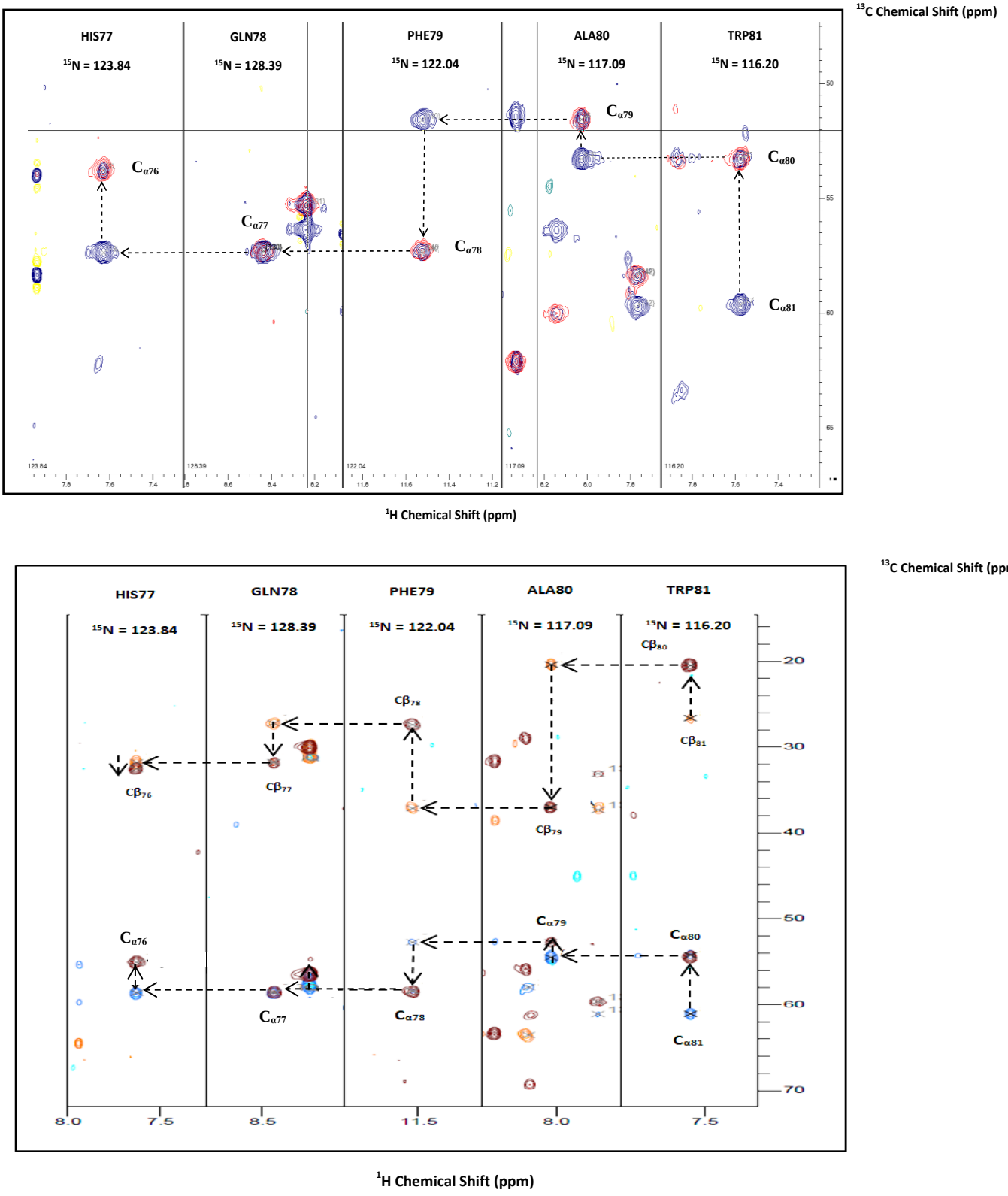



**Figure 6.8. Sequential assignment example from backbone triple resonance assignment of BRD4-BD1 from H77 to W81.** The red peaks HN(CO)CA, blue peaks HNCA, brown peaks HN(CO)CACB, orange peaks HNCACB.

HN(CO)CA/ HNCA and HN(CO)CACB/ HNCACB show the assignment stretch for residues H77 to W81. Sequential matches are shown using the dotted arrows. The HN(CO)CACB/ HNCACB experiment shows the $C_\alpha$ and $C_\beta$ atoms of the preceding amino acid. The HN(CO)CA/ HNCA strips show the $C_\alpha$ as positive peaks in blue and the $C_\beta$ as negative peaks in orange. The horizontal lines show inter-residue connections and the vertical lines show the intra-residue connections. The relevant $^{15}N$ chemical shift position of each plane is shown above the respective strips. The connectivities for both the $C_\alpha$ and $C_\beta$ peaks in the HN(CO)CACB and HNCACB experiments are shown in the bottom panel.

The $^{15}N/^{1}H$ HSQC spectrum with the residues of BRD4-BD1 assigned is shown in Figure 6.9.

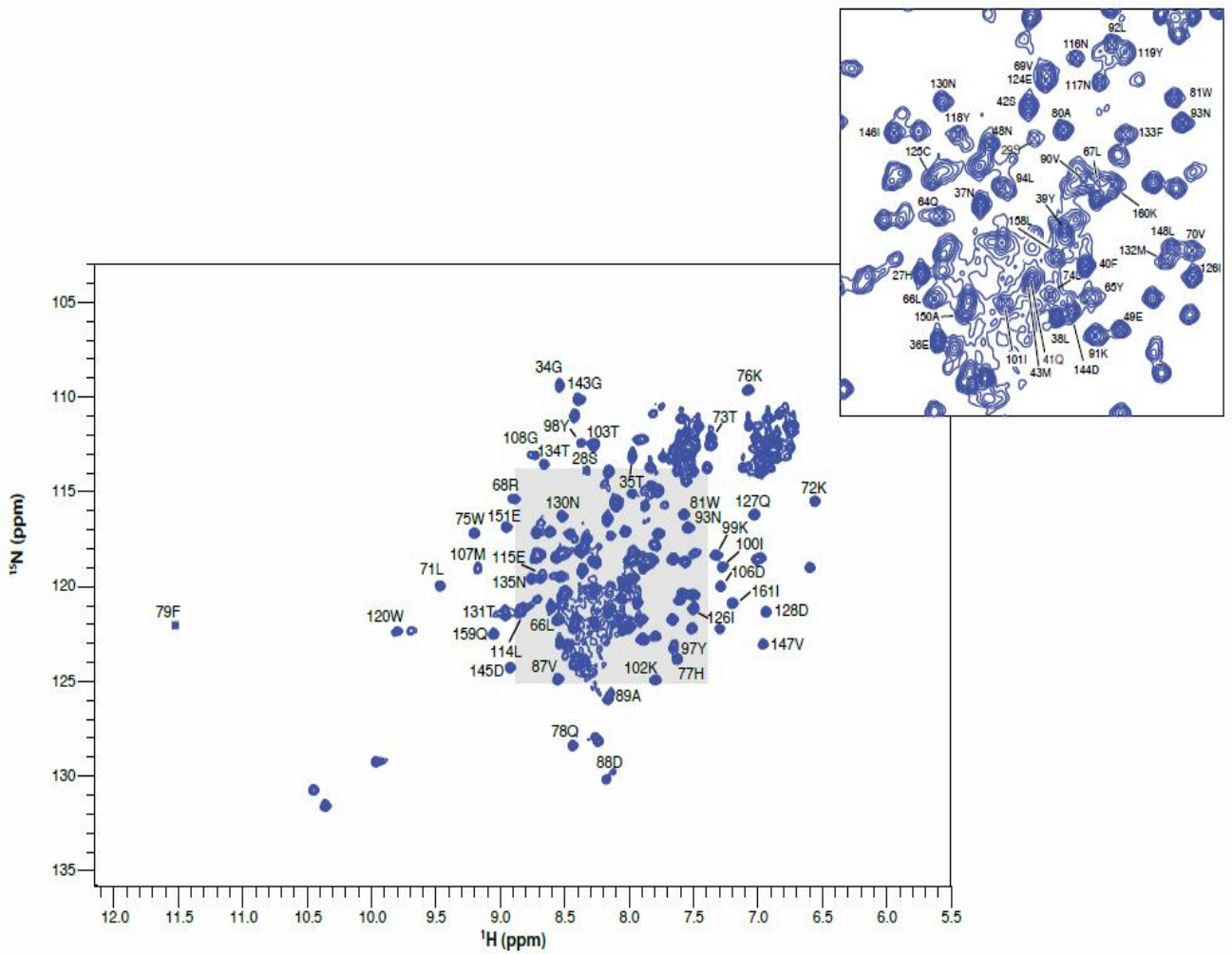



**Figure 6.9. $^{15}$N/$^{1}$H HSQC spectrum for BRD4-BD1.** All the identified amide resonances are annotated on the spectrum. Experiments were run at 25°C with the protein sample at pH 6.5.

### 6.3.2.5 Secondary Structure Prediction from Chemical Shifts using TALOS

The secondary structure of BRD4-BD1 was predicted using TALOS (Torsion Angle Likelihood Obtained from Shift and sequence similarity) [194] [195]. TALOS is a database software used for the prediction of phi and psi backbone torsion angles from HA, CA, CB, CO and N chemical shifts of a protein sequence (Fig 6.10). TALOS arises from an algorithm that employs a Bayesian interference statistical way to predict the secondary structure of protein using the types and chemical shifts of an amino acid. TALOS also calculates the neighbouring amino acid residues leading to the formation of a structural chemical shift database for related sequences [196] [197].

A protein backbone adopts various secondary structures (α-helices and β-sheets) through the rotation of specific dihedral angles at the C-$C_\alpha$ and $C_\alpha$-N bonds, Ψ (Psi) and Φ (Phi), respectively.

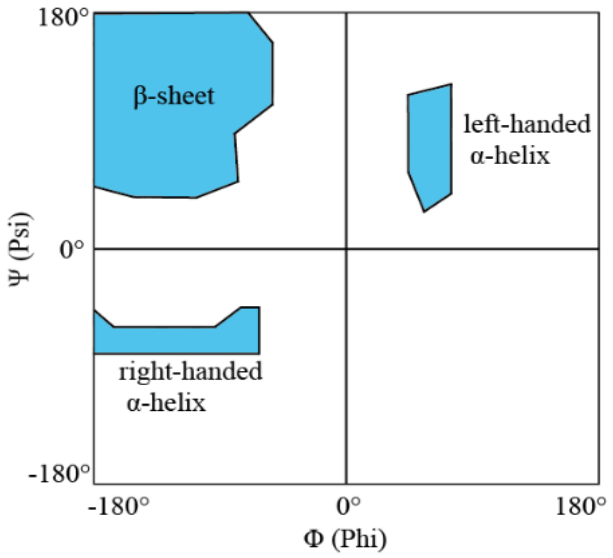



**Figure 6.10. Schematic representation of the Ramachandran Plot outlining the angles the Ψ (Psi) between the C and $C_\alpha$ bond against Φ (Phi) between $C_\alpha$ and N bond to characterise different types of secondary structural elements.** Angles and secondary structural motifs of the protein backbone are shown by the blue regions.

*Adapted from*: http://knowledgesun.com/ramachandran-plot/ [198] [199]

The chemical shifts from the triple resonance sequential backbone assignment of $^{1}H_N$ and $^{15}N$, $^{13}C_\alpha$ and $^{13}C_\beta$ in BRD4-BD1 were used with TALOS software to assign the secondary structure of the protein. Once the predictions using TALOS were successfully completed, the results were summarised in the form a sequential diagram (Fig 6.11 and Fig 6.12). The TALOS predictions are lists in Appendix 6.2.

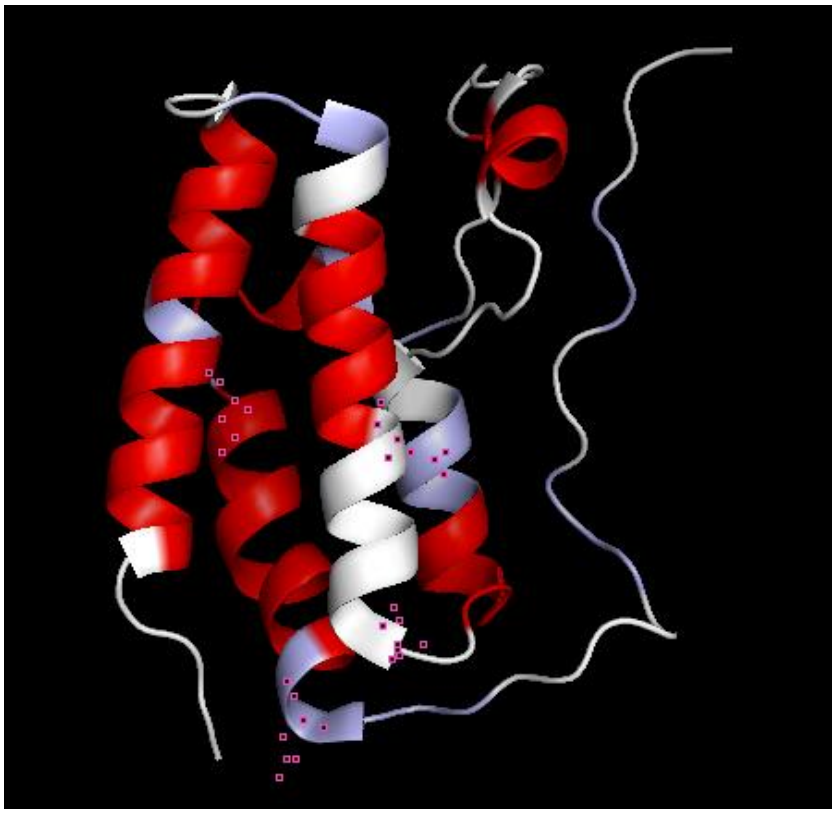
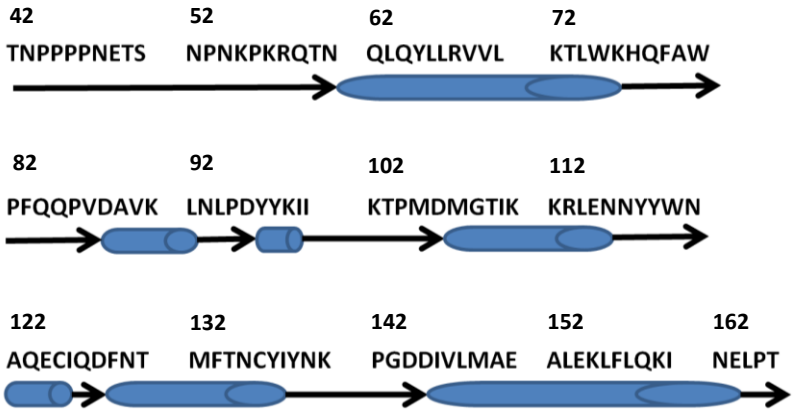


**Figure 6.12. A) Predictions of secondary structure of BRD4-BD1 using TALOS for BRD4-BD1.** Structure coloured by dangle predictions from dangle, red is helix, blue is from unassigned residues and white is assigned residues with no dangle predictions.

**Experimental Crystal Structure**




**NMR Secondary Structure**

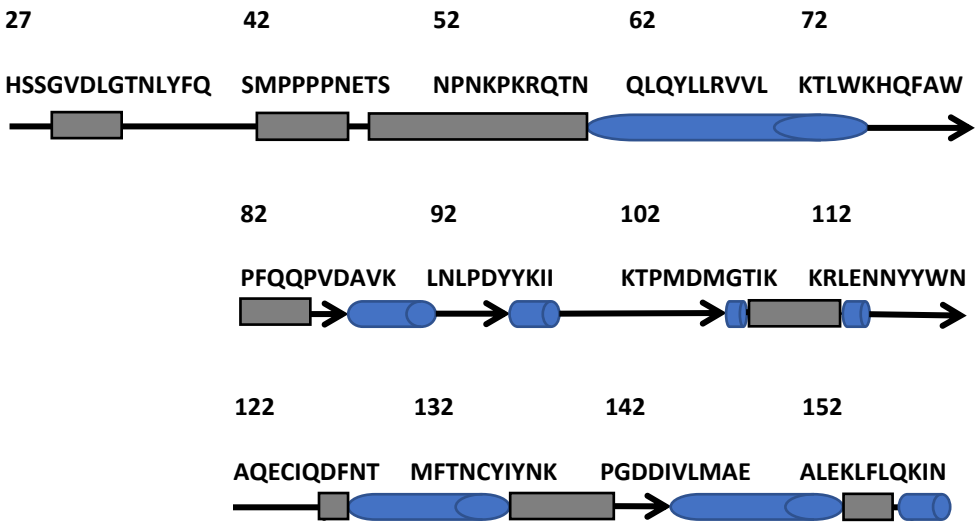



**Figure 6.12. B) Comparison of secondary structure predictions of BRD4-BD1 from X-Ray Crystallography and NMR methods.** The predicted structures are shown as a) BRD4-BD1 Experimental Crystal Structure b) NMR Structure, where α-helices are shown as blue cylinders and meanders shown as black arrows. Unassigned regions are represented as grey boxes.

TALOS was successful in predicting the secondary structures for BRD4-BD1. The X-ray structure was in agreement with the NMR structure for this construct. The structure is predominantly helical and containing loops and meanders. However, small differences were observed in the TALOS prediction between the two structures. Please note that there is a variation between the two constructs (T42/N43 are substituted as S42/M43 in the NMR structure). Both the constructs are the same from P44 to N162.

M107 was shown as helix in the X-ray crystallography structure, but as a meander in the NMR structure. The sections from T109 to R113, A122 to I126, and E154 to F157 were represented as helices in the X-ray crystallography structure, whereas as meanders in the NMR structure. Therefore, only 15 residues were different between the X-ray structure and NMR structure, which suggests that TALOS is an efficient secondary structure prediction tool.

## 6.4 The Chemical Shift Map of BRD4-BD1 complexed with BTB 07004

### 6.4.1 The Theory of Chemical Shift Mapping

As opposed to common units of frequency and wavelength in spectroscopy, NMR spectroscopy uses chemical shifts (ppm) as units, as absolute frequencies of NMR lines are directly proportional to the magnetic field strength. When doubling the magnetic field strength, the frequency is also doubled and therefore the comparison of absolute absorption frequencies

between spectrometers of different field strengths becomes impossible. However, when peak positions are described using chemical shift, they are independent of the field strength, making comparison between different spectrometers possible [200].

It is important that the $^{15}N/^{1}H$ HSQC data of the protein, which have been perturbed by a ligand, are obtained under identical conditions as $^{15}N/^{1}H$ HSQC native protein for comparison purposes. The distances from the signals in the fully assigned spectrum to the closest peak of the unassigned spectrum is measured. The chemical shift mapping results in the generation of a graph representing the minimal chemical shift difference for each residue. This is called "Conservative Chemical Shift Mapping" [201] [202].

The minimal chemical shift difference is measured using the following equation 1:

$$\boldsymbol{shift\ difference} = \sqrt{(\Delta^{\mathbf{1}}\boldsymbol{H_N})^{\mathbf{2}} + (\frac{\mathbf{1}}{\mathbf{5}}\Delta^{\mathbf{15}}\boldsymbol{N})^{\mathbf{2}}}$$

where, $\Delta^{1}H_N$ is the chemical shift change in the $^{1}H_N$ dimension, $\Delta^{15}N$ is the chemical shift change in the $^{15}N$ dimension, in ppm. Since all the values are always squared, the shift difference remains a positive value regardless of $\Delta^{1}H_N$ or $\Delta^{15}N$. The $^{15}N$ difference is multiplied by 1/5 to compensate for the chemical shift range of $^{15}N$ being 25 ppm while that of $^{1}H_N$ is 5 ppm. The measured value is effectively the Pythagorean or Euclidean distance between two peaks.

## 6.4.2 Chemical Shift Mapping of BRD4-BD1 with BTB 07004

### 6.4.2.1 NMR Sample Preparation of $^{15}$N BRD4-BD1 with BTB 07004

Chemical shift mapping of BRD4-BD1 was obtained utilising $^{15}$N/$^{1}$H HSQC in the presence and absence of the ligand BTB 07004. The protein sample was obtained from a $^{15}$N isotopically enriched sample expression in *E.coli* T7 express cells grown in minimal media and purified in GF Buffer (25 mM NaH2PO4, 100 mM NaCl and 1 mM DTT pH 6.5) as described in Chapter 2.5. The final concentration of the protein was 0.7 mM in complex with 3 mM ligand BTB 07004 (KD = 198 μM) added in ethanol to give a final concentration of 5 % ethanol. 5% $D_2O$ as used for the lock.

### 6.4.2.2 NMR Data Acquisition of $^{15}$N BRD4-BD1 in the presence and absence of BTB 07004

HSQC experiments were acquired at 25°C on a Bruker Avance III Spectrometer, with 2048 points (9615 Hz) in the direct F2 dimension ($^{1}$H), 256 points (2007 Hz) in F1 ($^{15}$N) indirect dimensions. Carrier frequencies for the experiments were set to 4.730 ppm, and 119.000 ppm for $^{1}$H, and $^{15}$N respectively. Data were processed with NMRPipe with a cosine window applied in all dimension [191], water suppression using a convolution difference filter and linear prediction in the $^{15}$N dimension to increase the length of data by a factor of 2. All dimensions were zero-filled by a minimum of a factor of 2.

The ligand binding site of BRD4-BD1 has been mapped through NMR chemical shift perturbation in the presence of the chemical ligand, BTB 07004 (Figure 6.13). Some peaks were not used due to the lower quality of the data in comparison to Figure 6.9.

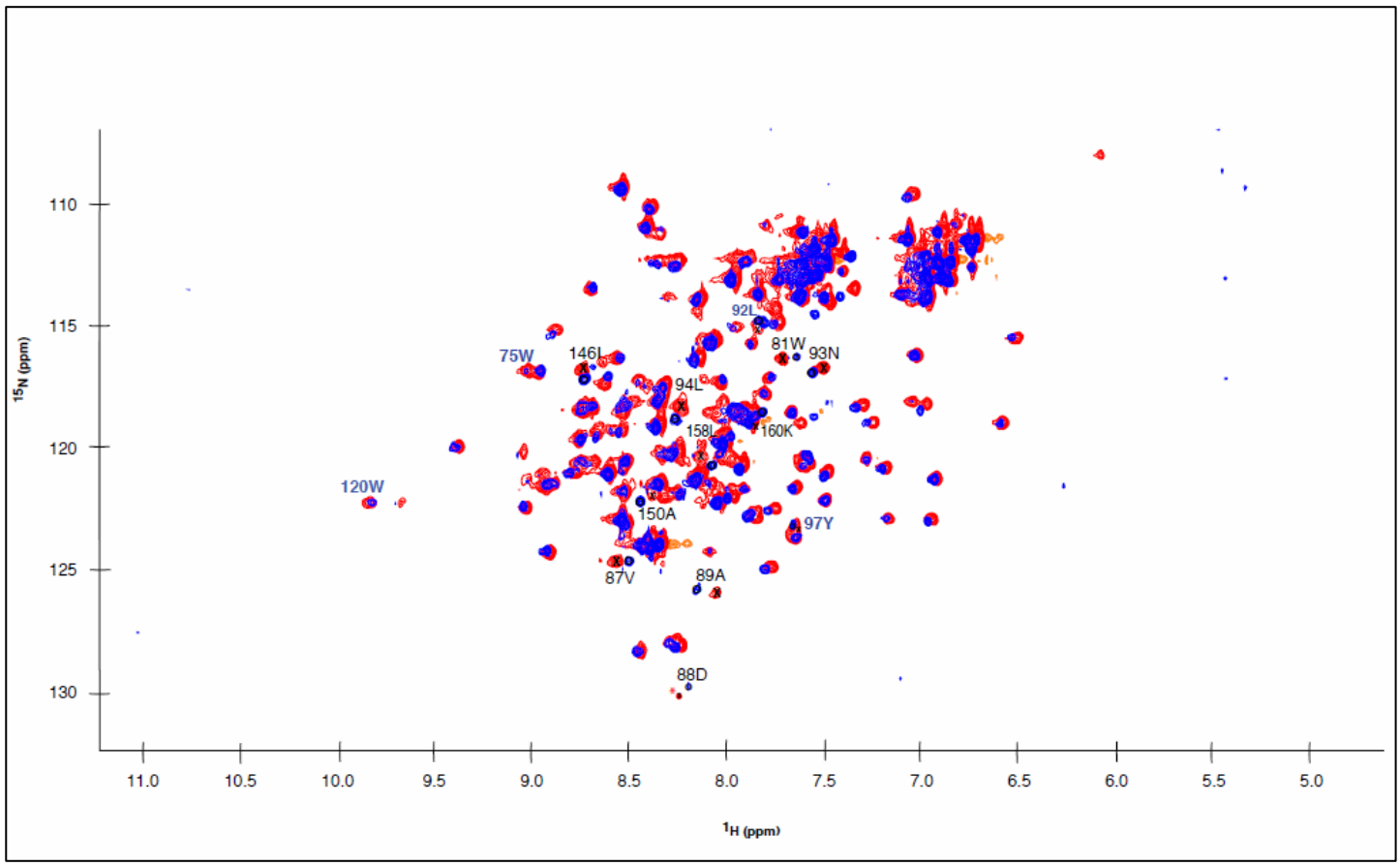



**Figure 6. 13. Overlaid $^{15}$N/$^{1}$H HSQC Spectrum of 0.7 mM BRD4-BD1 in the presence and absence of 3 mM BTB 07004 at 25 °C.** The blue spectrum represents the BRD4-BD1 and the red spectrum shows BRD4-BD1 in complex with BTB 07004. Addition of the unlabelled ligand to BRD4-BD1 caused chemical shift change in the protein. The residues with an observed chemical shift above the cut-off in Figure 6.15 are annotated in black. Residues that were expected to produce a chemical shift, but did not show any effects after the addition of the ligand are annotated in blue (this is in comparison with the crystal structure c.f.: section 5.3.2).

In order to observe the effect of BTB07004 binding on the BRD4-BD1 protein, the assignments were used to follow the chemical shift in the protein. Minimal maps can be constructed by

measuring distances from signals in an assigned spectrum, in this case BTB07004, to the nearest peak in the unassigned spectra of BRD4-BD1. Figure 6.14 shows the shifts that occurred due to the binding of BTB07004 to BRD4-BD1.

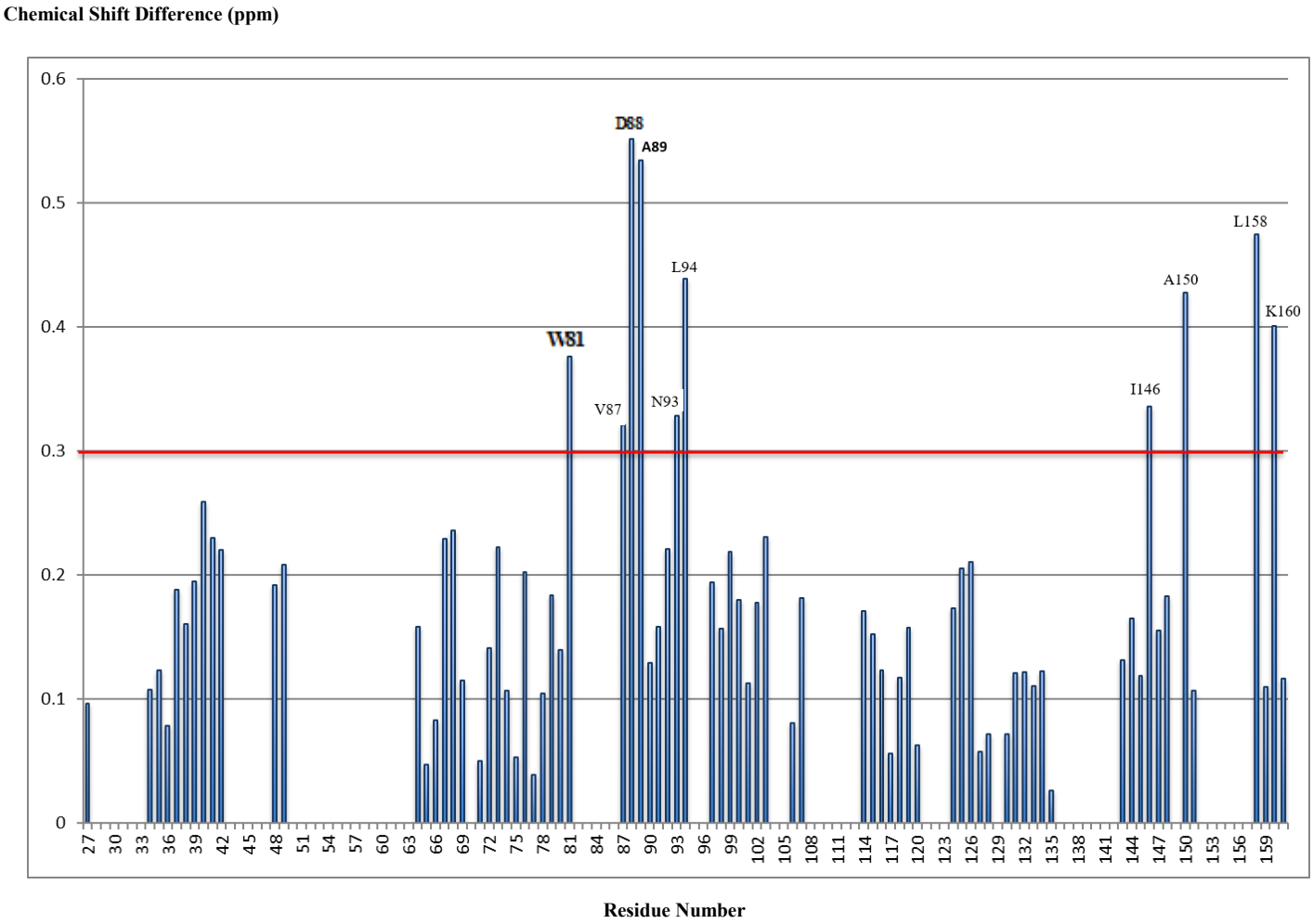



**Figure 6.14. The minimal chemical shift perturbations between BRD4-BD1 and BTB 07004 as calculated by Equation 1.** This plot shows the BRD4-BD1 chemical shift perturbations between BRD4-BD1 with and without an excess of BTB 07004. Chemical shift differences ranging above or equal to 0.3 ppm were considered significant and matched as the residues present in the ligand binding site of BRD4-BD1 (above the red line).

The threshold for significant chemical shift changes was defined by looking for consistent sequence of residues that have a perturbation, which is localised on the protein surface.

The key residues that showed a major chemical shift change are summarised in the table below with a comparison of the residues identified crystallographically (Chapter 5.3.3):

| X-ray Crystallography | NMR |
|---|---|
| **W81** | **W81** |
| P82 | - |
| F83 | - |
| **V87** | **V87** |
| - | D88 |
| - | A89 |
| L92 | - |
| - | N93 |
| **L94** | **L94** |
| Y97 | - |
| C136 | - |
| Y139 | - |
| N140 | - |
| **I146** | **I146** |
| - | A150 |
| - | L158 |
| - | K160 |

**Table 6.1. Comparison of key residues in the active site BRD4-BD1 involved in binding of BTB 07004, characterised by both X-ray crystallography and NMR**. The comparison confirms that the data obtained from NMR chemical shift mapping is in agreement with the X-ray crystallography data. Major chemical shift differences ranging above 0.3 ppm were observed at W81, V87, D88, A89, N93, L94, I146, A150, L158 and K160, from which 4 residues match with the X-ray crystallography structure where there are residues directly in contact with the

ligand. Some of the residues which show shifts in the NMR data but are not part of the ligand binding site such as L158 and K160 are not easily explained by ligand binding but could be due to long range dynamics, changes in conditions/oligomerization state on addition of ligands or the effect of ethanol which is also known to bind to the protein.

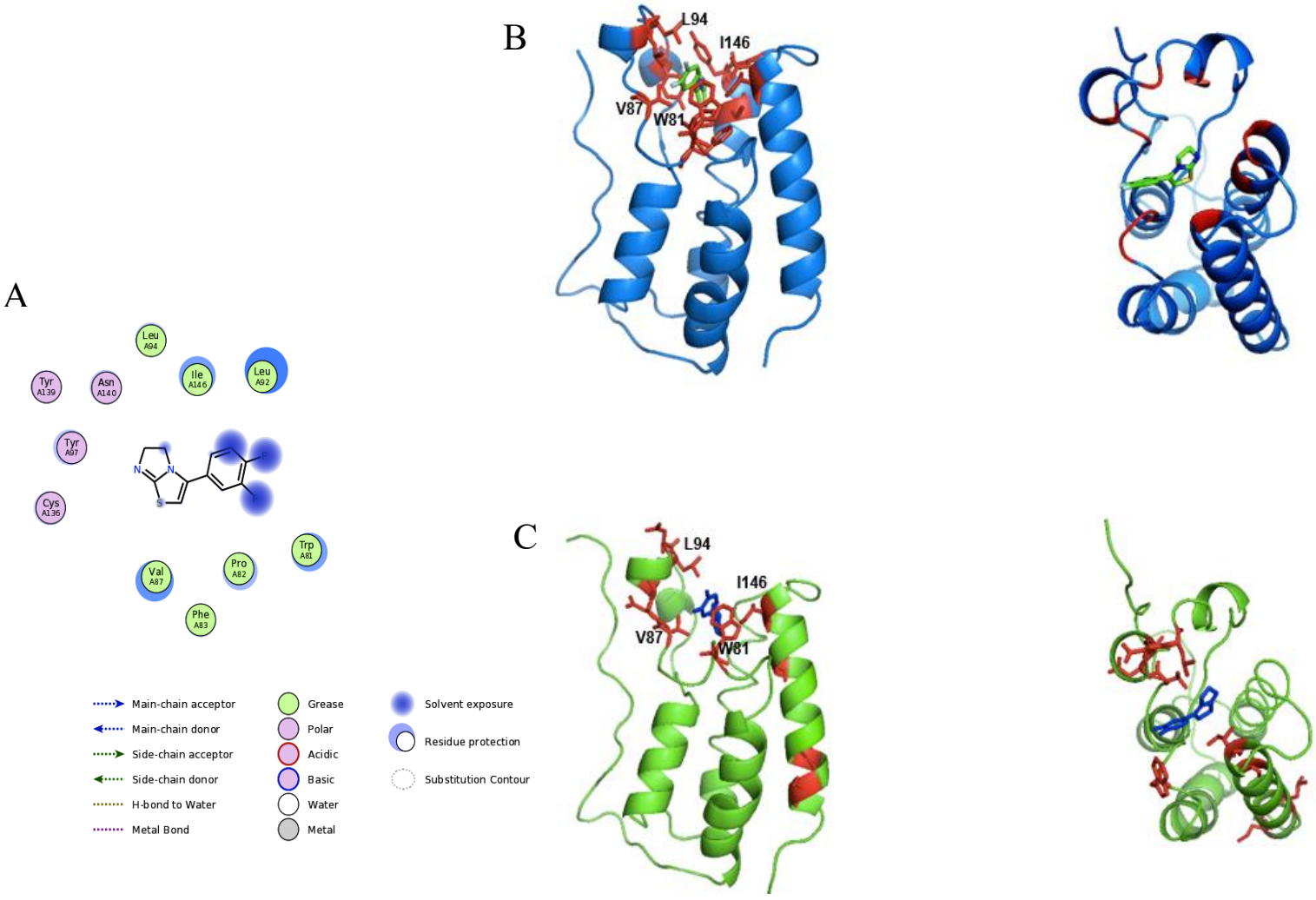



**Figure 6.15. (A) The interactions of BTB 07004 with its environment residues in the binding site of BRD4-BD1 based on the crystal structure from WinCoot Flev. Comparison of the BRD4-BD1 binding-site (B) Crystal structure (C) Minimal chemical shift map to the nearest peak for backbone amide $^{1}H$ and $^{15}N$ resonances of BRD4-BD1 with BTB 07004, mapped on the crystal structure pictures prepared using Pymol.** (A) The amino acids W81, V87, L94, Y97, C136, Y139, N140 and I146 are shown to be the key environment residues of BTB 07004 when bound to BRD4-BD1. (B) (C) Comparison of BRD4-BD1 ligand binding site between crystal and NMR structures. Residues involved in binding are shown in red. As shown in Table 6.1, **W81, V87, L94** and **I146** are the common residues between the crystal and NMR data for the binding site of BRD4-BD1.

The results in Figure 6.16 suggested that W81, V87, L94 and I146 were strongly involved in the binding process of the ligand, as their chemical shift difference ranged above 0.3 ppm, which characterises a major shift. Other residues identified in the crystal structure, such as P82, F83, L92, C136, Y139 and N140 were not identified in the NMR chemical shift mapping. P82 is present in the NMR spectrum as it doesn't have an NH group and its adjacent residue F83 was not assigned. The residues D88, A89, N93 and A150 were also found to have shifted in the NMR spectrum. These residues still reside within the binding pocket region and could also be involved in ligand binding. However, the residues L158 and K160 are very distant from the binding site, and their shifts are not so easily explained. This could be due to changes in dynamics or changes in aggregation state. In addition, this is based on a minimal shift mapping, so measured shifts can be far less than those that actually occur.

### 6.4.2.3 Chemical Shift Mapping Results of BRD4-BD1 in the presence of Ethanol

Because DMSO can bind to the active site of BRD4-BD1, 5% Ethanol (final concentration) was used as a solvent. In order to ensure the fidelity of the data and rule out that the chemical shifts observed in Figure 6.17 were not due to the EtOH, an $^{15}N/^{1}H$ HSQC experiment was performed with 3mM BRD4-BD1 with the presence of 4.5 % EtOH, as the protein is quite promiscuous in its binding.

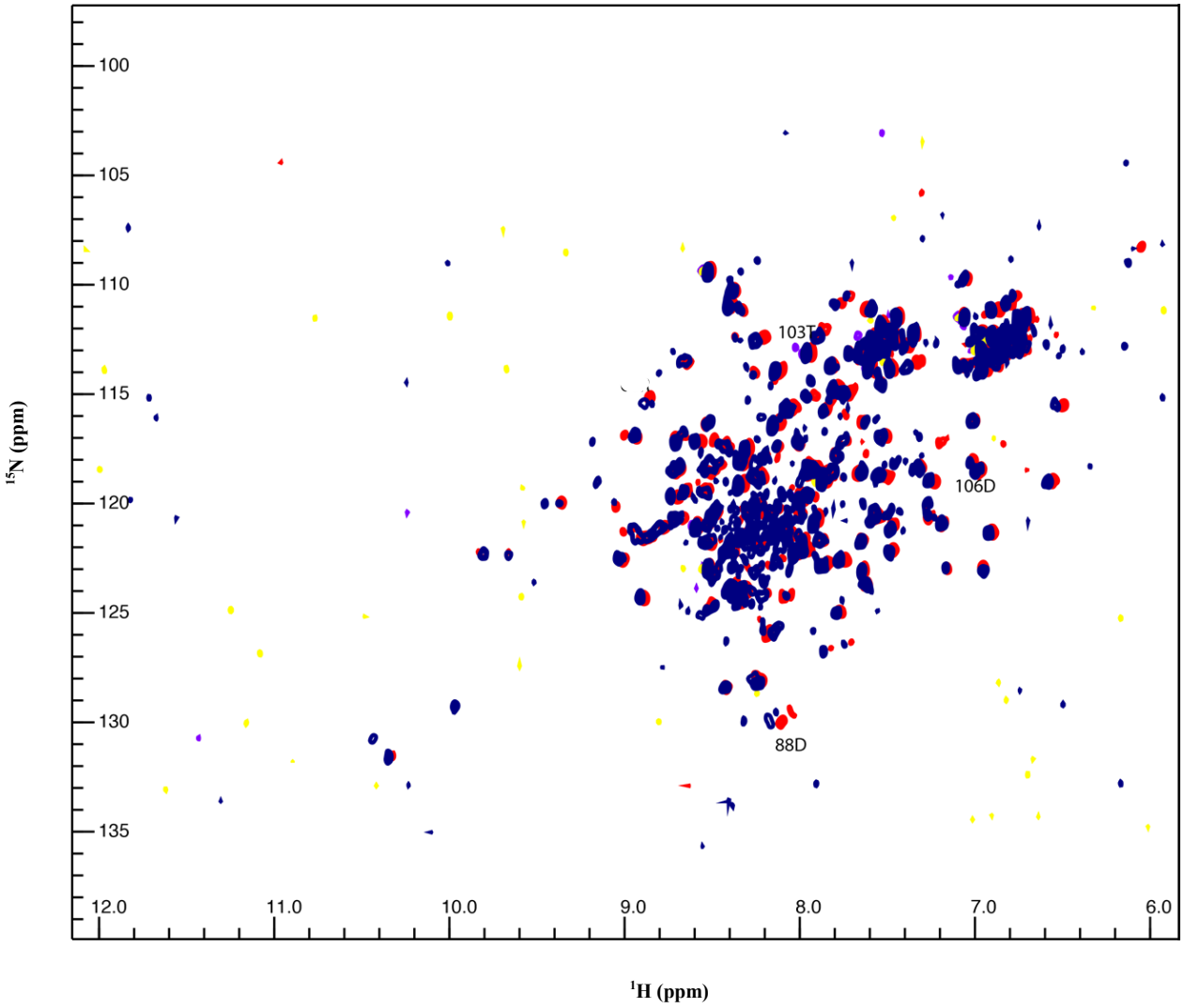



**Figure 6. 16. Overlaid $^{15}$N/$^{1}$H HSQC Spectrum of 0.7 mM BRD4-BD1 in the presence and absence of 4.5 % EtOH at 25 °C.** The blue spectrum represents the BRD4-BD1 and the red spectrum shows BRD4-BD1 in complex with EtOH. Addition of the 4.5 % EtOH to BRD4-BD1 has caused chemical shift change in the protein. The important residues with an observed chemical shift above the cut-off in Figure 6.18 are annotated in black.

The addition of 4.5 % EtOH to BRD4-BD1 has caused a chemical shift perturbation in the protein. Therefore, the shifted residues were compared to Fig. 6.16.

The comparison shows that the residues V70, L71, D88. T103 and L158 have shifted in response to the addition of 4.5 % EtOH. D88 and L158 are the only common residues when

comparing to the chemical shifts of BRD4-BD1 with BTB 07004. Moreover, the shifts with EtOH are not as strong as the ones observed with BTB 07004, which clearly confirms that the chemical shifts seen in Fig 6.13 are in response to BTB 07004 and not an effect of the EtOH. The average chemical shift was 0.08 ppm and the standard deviation was 0.10 ppm.

BRD4-BD4 in 4.5% EtOH was not used as the control experiment to measure the chemical shift changes seen on the addition of BTB07004, as the quality of the spectrum was poorer and not all the peaks were assigned. Moreover, no significant changes were observed within the ligand-binding site.

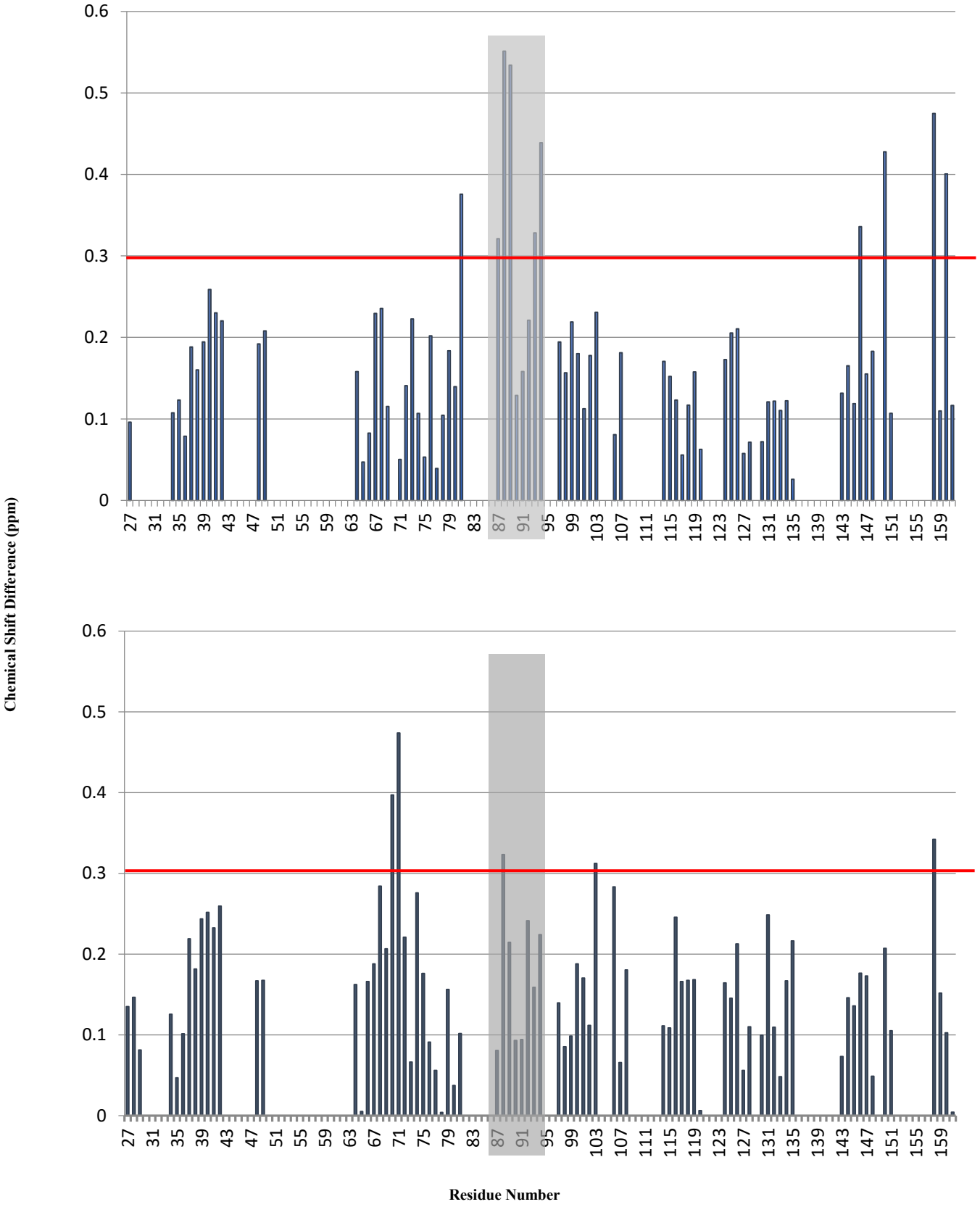



**Figure 6.17. The comparison of minimal chemical shift perturbations between BRD4-BD1, BTB 07004 (top) and EtOH (bottom) as calculated by Equation 1.** The greyed section represents the binding site of BRD4-BD1 with BTB 07004.

Table 6.2 compares the difference between the residues shifted in response to BTB 07004 and EtOH.

| BRD4 + BTB 07004 | BRD4 + EtoH |
|---|---|
| - | V70 |
| - | L71 |
| W81 | - |
| V87 | - |
| **D88** | **D88** |
| A89 | - |
| N93 | - |
| L94 | - |
| - | T103 |
| I146 | - |
| A150 | - |
| **L158** | **L158** |
| K160 | - |

**Table 6.2. Comparison of key residues in the active site BRD4-BD1 involved in binding of BTB 07004 and BRD4-BD1 with EtOH**

Figure 6.17 indicates and compares the chemical shift changes observed between the BRD4-BD1 complexed with BTB07004 and BRD4-BD1 with 4.5 % EtOH. Although D88 and L158 are both largely shifted in response to EtOH, the shift is not as big as with the ligand BTB07004, which confirms that these chemical shift changes are due to the ligand.

# 6.5 $^{19}$F-Fluoroindole-labelled BRD4-BD1 Spectrum

BRD4-BD1 was also studied with $^{19}$F isotopic enrichment. The chemical shift mapping and crystallographic data indicated in the previous sections that W81 was present in the ligand binding site of BRD4-BD1. W81 and the other two tryptophan residues from BRD4-BD1 can then be utilised as an NMR screening tool to monitor ligand binding using $^{19}$F chemical shift correlations and shift changes. In comparison to section 6.1.4, where ligand-observed $^{19}$F CPMG was performed, and ligand binding was examined using the fluorinated fragments of BTB 07004, this section focuses rather on fluorinating the tryptophan residues, particularly W81 of BRD4-BD1 and approaching the ligand-binding experiment on a protein-observed basis.

## 6.5.1 Introduction

$^{1}$H and $^{15}$N/$^{13}$C labelled methods are more commonly used in NMR studies, despite the fact that $^{19}$F is highly sensitive and 100 % abundant because $^{19}$F does not naturally occur in proteins. $^{19}$F NMR methods are less often applied due to the need of making an extra $^{19}$F labelled protein sample [203] [204].

## 6.5.2 The expression and purification of $^{19}$F-Fluoroindole BRD4-BD1

Previous studies have shown that not all isomers of fluoroindole can be accommodated within a protein, and therefore the most integrable fluoroindole needed to be identified. 4′ fluoroindole, 5′ fluoroindole and 6′ fluoroindole were tried (Figure 6.18).

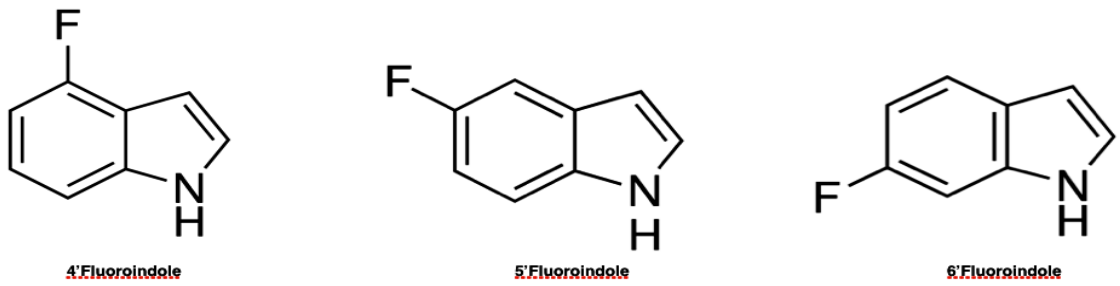



**Figure 6. 18. The different types of fluoroindoles used to be accommodated within BRD4-BD1.**

4′ fluoroindole, 5′ fluoroindole and 6′ fluoroindole were used to determine the most integrable fluoroindole with BRD4-BD1.

Protein samples was obtained from a $^{19}F$ isotopically enriched expression in *E.coli* T7 express cells grown in minimal media and purified in GF Buffer (25 mM $NaH_2PO_4$, 100 mM NaCl and 1 mM DTT pH 6.5) as described in Chapter 2.5. 60 ug/mL The final concentration of fluoroindole was added 15 minutes before the IPTG induction. The expression levels of BRD4-BD1 with the different types of fluoroindoles were assessed on SDS-PAGE gel with the pre - and post - induction samples for each fluorindole (Figure 6.19).

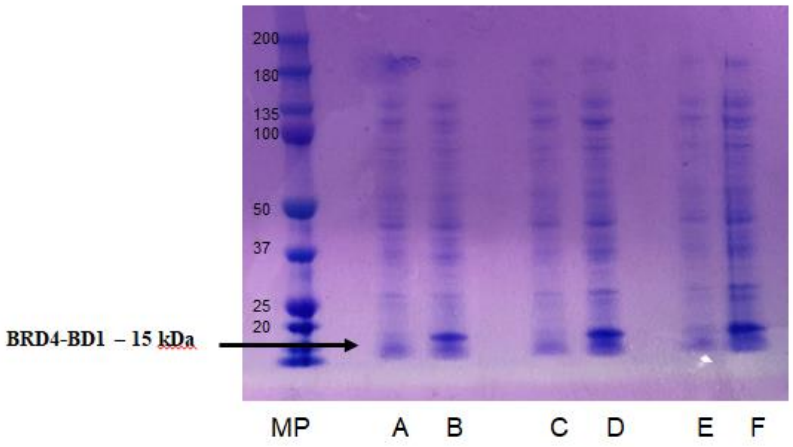



**Figure 6. 19. BRD4-BD1 with the presence of different fluoroindoles.** MP : molecular weight marker. A) Pre-induction sample of BRD4-BD1 with **4′ fluoroindole**. B) Post-induction sample of BRD4-BD1 with **4′ fluoroindole**. C) Pre-induction sample of BRD4-BD1 with **5′ fluoroindole**. D) Post-induction sample of BRD4-BD1 with **5′ fluoroindole**. E) Pre-induction sample of BRD4-BD1 with **6′ fluoroindole**. F) Post-induction sample of BRD4-BD1 with **6′ fluoroindole**.

The SDS-PAGE gel in Figure 6.19 shows that, as anticipated, BRD4-BD1 is well expressed with the presence of all three fluoroindoles. The band intensity was observed to be greater with 5′ fluoroindole, indicating greater expression. This amino acid was used for further experiments.

## 6.5.3 MS experiments of $^{19}$F BRD4-BD1

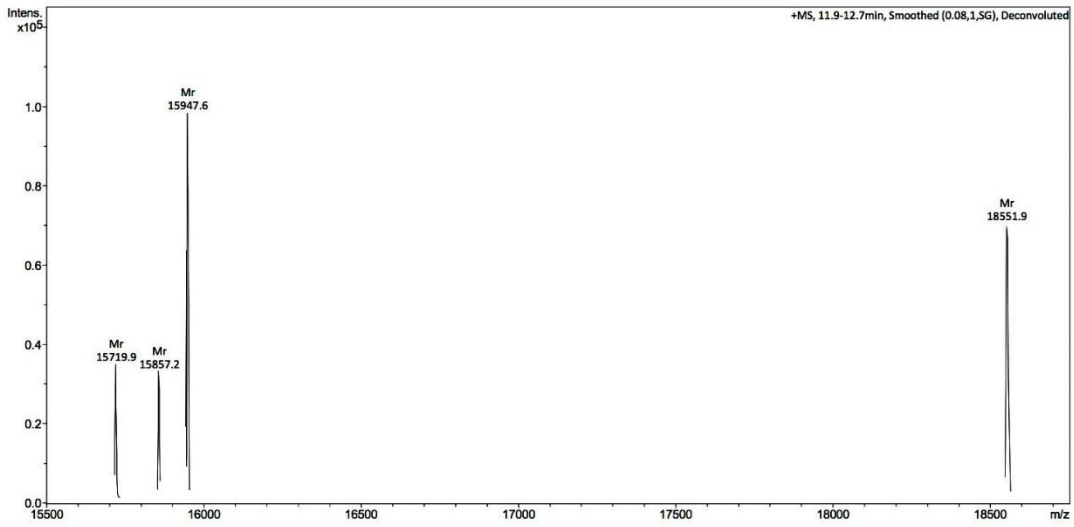



**Figure 6. 20. Mass spectrometry spectrum plot of $^{19}$F BRD4-BD1.** The results show the presence of four different peaks. The expected native of full labelled 13C-15N-19F-BRD4-BD1 is 15993Da allowing for 100% labelling. The most intense ion, at Mr = 15,947.6Da, is consistent with a fully labelled BRD4-BD1 with a labelling efficiency of ~95%. Two lower mass peaks have a molecular mass of Mr = 15,857.2 and 15,719.9, and are lighter than the base peak suggesting they are degradation products of the protein. The fourth peak is very far away from the actual base peak with Mr = 18,551.9 and is assumed to be a contaminant.

The mass spectrometry in Figure 6.20 results confirms the expected mass of the BRD4-BD1. The expected mass of $^{19}$F BRD4-BD1 is 15947.6 Da, which is slightly lighter than the fully labelled expected mass at 15993 Da. However, the presence of multiple peaks in the spectrum suggests that the protein must have degraded. The protein therefore appears to possibly be slightly less stable on inclusion of fluoroindole.

## 6.5.4 NMR Data Acquisition of $^{19}$F BRD4-BD1

$^{19}$F CPMG experiments were acquired on a Bruker Avance III 600 Hz spectrometer at 25°C. Data were collected with 32,768 points and a spectral width of 9615 Hz in the direct F1 dimension ($^{19}$F), with a total number of 128 scans. The CPMG sequence used 60 cycles with a delay of 1ms between 180 degree pulses. Data was processed with a line broadening of 5 Hz using exponential multiplication. 0.1 mM BRD4-BD1 was used with 1 mM BTB 07004.

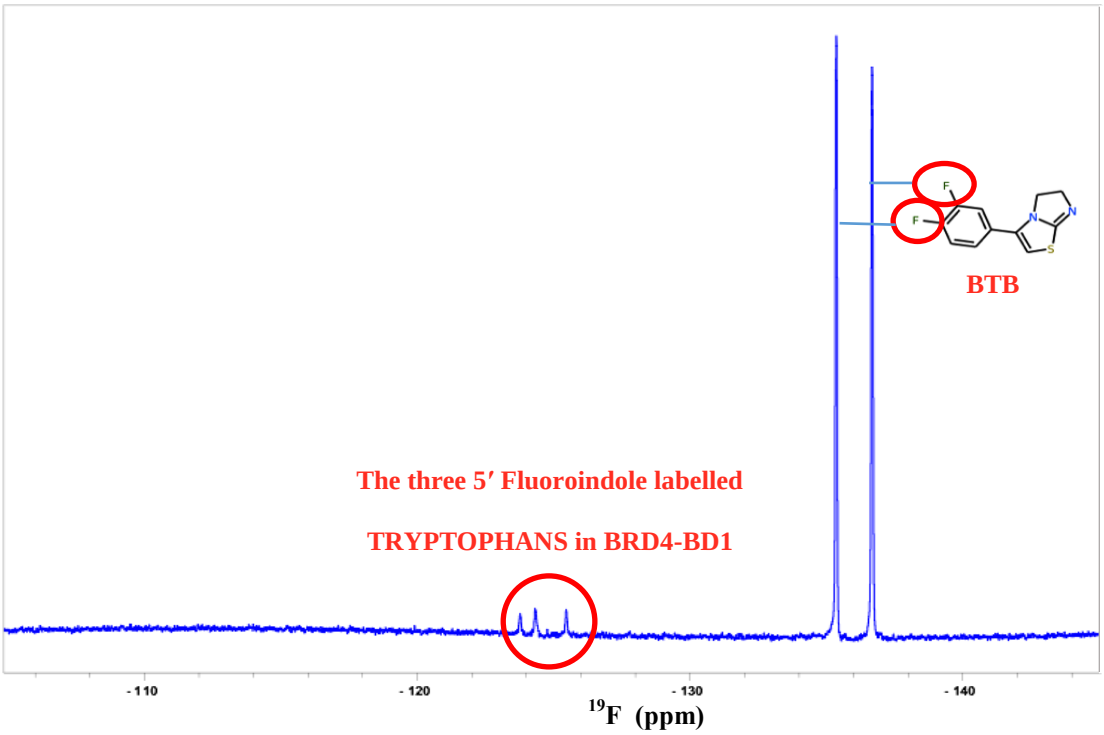



**Figure 6. 21. CPMG spectrum of $^{19}$F BRD4-BD1 in complex with BTB 07004**

The three peaks at approximately - 125 ppm (red region) derive from the three 5′ fluoroindole-labelled tryptophan residues within the protein molecule, and the two large peaks at - 135 ppm arise from the fluorine atoms from BTB 07004. This spectrum confirms that the tryptophan residues from the BRD4-BD1 protein sequence are labelled with $^{19}$F.

Figure 6.21 shows that the tryptophan residues from the BRD4-BD1 protein sequence are labelled with $^{19}F$ (three peaks in the range of 125 ppm). However, it does not inform us which one of the peaks corresponds to W75, W81 or W120. The two peaks in the 135 ppm range, correlate with the two fluorine atoms from BTB 07004. This experiment enables to observe both $^{19}F$ BRD4-BD1 and BTB 07004 in the same spectrum, and therefore ligand binding can be observed from both the protein and ligand perspectives.

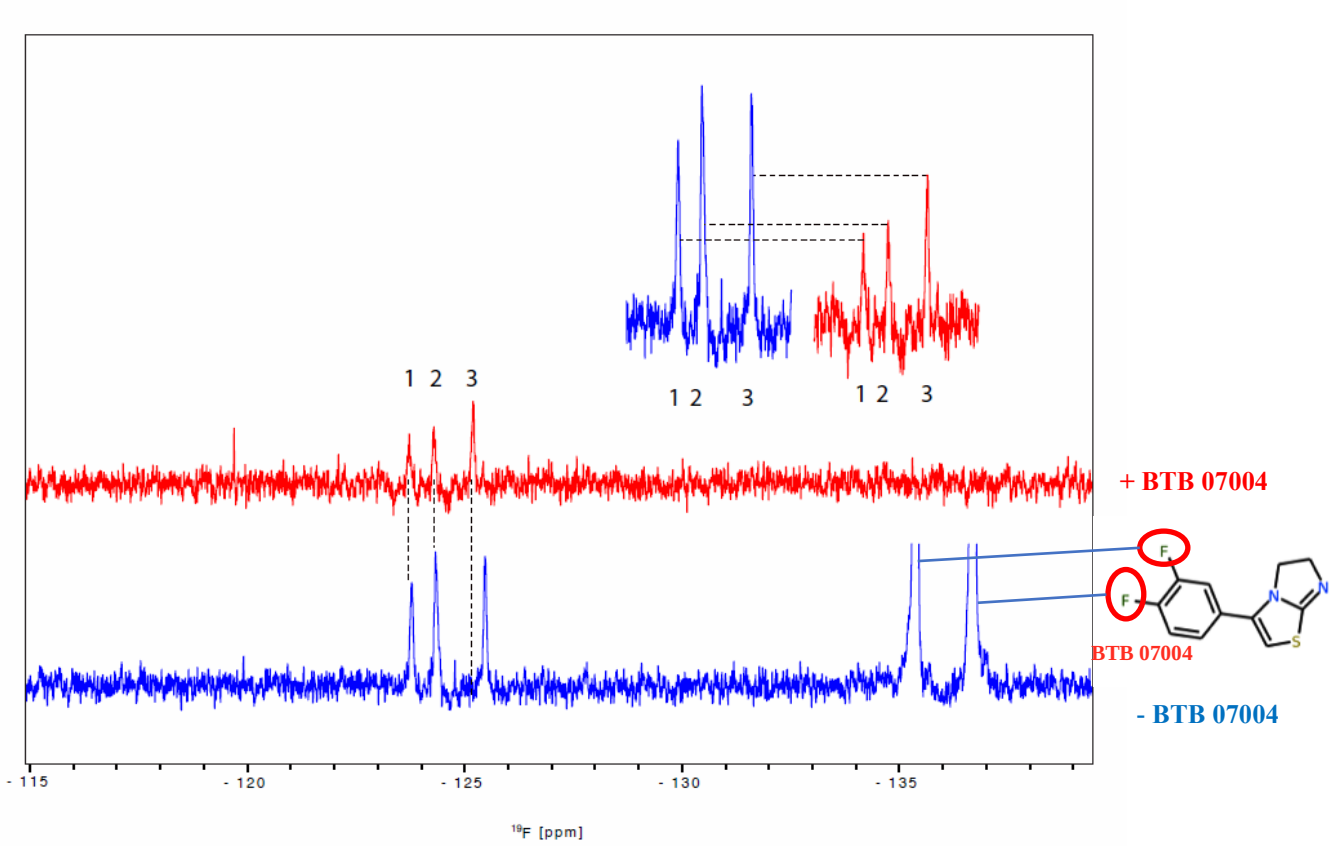



**Figure 6. 22. $^{19}F$ CPMG spectrum of $^{19}F$ BRD4-BD1 with the presence and absence of BTB 07004.** The blue spectrum represents $^{19}F$ BRD4-BD1 and the red spectrum represents $^{19}F$ BRD4-BD1 complexed with BTB 07004. The signals from the three fluoroindole-labelled tryptophan residues (1, 2 and 3) have reduced intensity with the addition of BTB 07004, confirming binding and highlighting the role of the tryptophan residues in the binding of BTB 07004 to BRD4-BD1. Peak 1 = ?, Peak 2 = ? and Peak 3 = W81, based on crystallographic and chemical shift mapping data.

The binding of BTB 07004 with BRD4-BD1 was confirmed with $^{19}F$ CPMG experiment (Figure 6.22). The intensity of the signals from the three tryptophan residues (1, 2 and 3) was reduced intensity with the addition of BTB 07004, confirming binding and highlighting the role of the tryptophan residues in the binding of BTB 07004 to BRD4-BD1. It should be noted that peak 3 has even shifted, indicating strong interaction of the residue with the ligand.

The ligand binding site of BRD4-BD1 includes W81, and this residue associates with the chemical shift following the addition of BTB 07004. The crystallographic data show a hydrophobic interaction of BTB 07004 with W81, which is consistent with this result.

Neither of the residues W75 and W120 have shown any significant $^{1}H$ chemical shifts in the mapping of BRD4-BD1 with BTB 07004. However, W75 is the next closest from the ligand binding site after W81. Based on these, the highly attenuated and shifted peak 3 from Figure 6.22 could be corresponding to W81. The least attenuated peak 1 could be W120, which is the furthest from the ligand binding site. As a result, peak 2 is assumed to be W75. This result is not conclusive.

The control spectrum for this experiment did not contain 4.5% ethanol.

## 6.6 Summary and Discussion

Chapter 6 has characterised the binding of BTB 07004 to BRD4-BD1 using both ligand and protein-observed NMR methods utilising $^{1}H$ and $^{19}F$ nuclei.

NMR studies revealed that the protein was folded, in monomeric state and stable in the GF buffer conditions, rendering it suitable for NMR studies. NMR screening methodologies, including $^{1}H$ STD-NMR, $^{1}H$ CPMG and $^{19}F$ CMPG were performed to show that BTB 07004

binds to BRD4-BD1. In comparison to the other protein systems studied in this project, BRD4 was the most suitable protein system for the comparison of both the X-ray crystallography and NMR methods due to its affinity. Its small size of 15 kDa and stability were advantageous and no gross oligomerisation was present.

BRD4-BD1 was then studied using 2D and 3D-based heteronuclear $^{15}N/^{13}C/^{1}H$ experiments. The triple resonance experiments HNCACB, HNCA, HN(CO)CA, HN(CO)CACB and HNCO were used to make an 80 % sequential backbone assignment of $^{1}H_{N}$, $^{15}N$, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ nuclei in BRD4-BD1. Chemical shift mapping experiments using $^{1}H$-$^{15}N$ HSQC spectra with the addition of the ligand BTB 07004 showed chemical shift differences, indicating ligand binding and highlighting the key residues in the binding site of BRD4-BD1.

Major chemical shift differences ranging above 0.3 ppm were observed for the residues W81, V87, D88, A89, N93, L94, I146, A150, L158 and K160. The chemical shift mapping was in agreement with the binding site in the BRD4-BD1 crystal structure, particularly for the residues W81, V87, L94 and I146.

Further 1D heteronuclear $^{19}F$-based experiments were performed. BRD4-BD1 was isotopically enriched with 5′ fluoroindole to label the tryptophan residues of BRD4-BD1. $^{19}F$ CMPG experiments rapidly confirmed that W81 is part of the binding site, and W75 and W120 are not part of the ligand binding site of BRD4-BD1.

This work has established BRD4-BD1 as a good model system for comparing NMR results with X-ray crystallographic data so as to gain a better understanding of the protein's binding site and the protein ligand interactions.

# Chapter 7

## Discussion

## 7.1 General Summary

Generally structural information for biomolecules and especially proteins has so far been determined by single methods, e.g. X-ray crystallography, Electron Microscopy (EM) or NMR only, although combined NMR X-ray and NMR EM refinements have also been made [79] and these approaches are becoming more popular as methods to look at the internal dynamics of proteins. Analysis of the PDB shows that 90 % of structures have been solved by X-ray crystallography, and these have been used to understand proteins three-dimensional structure, protein function and ligand interactions. Hydrophobic proteins from the lipid bilayer of cellular membranes have also been solved by X-ray crystallography, such as G-protein coupled receptors, ion channels and transporters. However, X-ray crystallography requires the purification of large amount of very pure protein which then critically needs crystallising, these processes can take several months to years. Failure in the crystallographic pipeline can occur at any stages of the process: protein production, crystallisation, X-ray diffraction, X-ray data acquisition, X-ray data phasing and X-ray data refinement.

The main aim of this thesis was developing novel approaches combining X-ray crystallography and isotope-edited ligand and protein-observed NMR methods to understand protein binding sites and interactions in proteins. Significant effort was focused on a number of "model" systems to this end, including CypA, TbrPDEB1 and BRD4-BD1 to compare X-ray crystallography and both ligand and protein-observed NMR methods.

The main results have been summarised in the following table:

| | X-ray Crystallography | NMR |
|---|---|---|
| **APO CypA Protein** | √ | √<br>$^{1}H/^{15}N$ HSQC |
| **CypA and Ligands** | ***X***<br>No liganded complexes observed. | √<br>$^{1}H$ STD-NMR, $^{1}H$ WaterLOGSY and $^{1}H$ CPMG. No further experiments done. |
| **APO TbrPDEB1 Protein** | √ | ***X***<br>No spectra obtained. |
| **TbrPDEB1 and Ligands** | √<br>4 liganded crystal structures. | ***X***<br>Not done |
| **APO BRD4-BD1** | √ | √<br>$^{1}H/^{15}N$ HSQC<br>Majority of backbone HN, Cα ,Cβ & N atoms assigned. |
| **BRD4-BD1 and Ligands** | √<br>7 liganded crystal structures. | √<br>$^{1}H$ STD-NMR, $^{1}H$ WaterLOGSY and $^{1}H$ CPMG with ligand BTB 07004. $^{19}F$ CPMG with ligand BTB 07004.<br>Chemical Shift Mapping with ligand BTB 07004 |

**Table 7.1. Summary of the main results with CypA, TbrPDEB1 and BRD4-BD1 using X-ray crystallography and NMR**√ Experiment done and successful; X Experiment done but unsuccessful.

### 7.1.1 CypA Conclusions

CypA (18 kDa) was studied in complex with the ligands 3′Aminoacetophenone, 4′ Sulfamoylbenzoic Acid and 6′Aminoindazole using X-ray crystals. Results revealed that the crystal packing of CypA did not allow the binding of the ligands to CypA. This appears to be due to C-terminal residue Glu165, obstructing the ligand binding site of CypA. In an attempt to overcome the packing effect hindering the binding site and create a new crystal form, In-situ mutagenesis was performed to remove the C-terminal Glu165 (CypA-Δ165-6HIS), but the crystallisation of mutant CypA did not yield suitable crystal samples of 3'aiminoacetophenone, 4′ Sulfamoylbenzoic Acid and 6′ Aminoindazole complexed with His-tagged CypA.

NMR screening methodologies, including $^{1}$H STD, $^{1}$H Waterlogsy and $^{1}$H CPMG were performed on both CypA-Δ165-6HIS and wild type CypA, which clearly indicate that 3′Aminoacetophenone and 4′Sulfamoylbenzoic Acid bind to CypA. This showed that both the wild type and mutant were binding competent and that that main problem was to achieve crsytallisation of their complex. The mutant CypA-Δ165-6HIS was also studied using 2D/3D-based heteronuclear $^{15}$N/$^{1}$H experiments ($^{1}$H/$^{15}$N HSQC) all of which gave good quality spectra.

CypA is an example that illustrates the need for an advanced NMR method to determine ligand binding sites. This study highlighted the requirement for full protein assignment [133] and interaction studies of ligand-based observations in the absence of an X-ray structure. As the protein- and ligand-observed NMR could not be compared with X-ray structure for validation purposes, this target protein was abandoned.

## 7.1.2 TbrPDEB1 Conclusions

TbrPDEB1 (37 kDa) was studied in complex with the ligands VUF 14234, PPS 59083, PPS 60036 and VUF 14450. X-ray crystallography studies revealed a conserved binding mode, where the active site contained the ligand and magnesium and zinc ions. The residue Gln874 conserved interaction all the crystal structures, and is the main residue which forms the parasite-specific P-pocket. In fact, all the ligands investigated in this study were directing themselves towards the P-pocket residue, which suggested that this P-pocket can be used as an important target in the design of therapeutic drugs for the treatment of HAT.

Unfortunately, 1D and 2D-based heteronuclear $^{15}N/^{1}H$ experiments resulted in broad and poorly dispersed peaks. Although thermal shift assays suggested that the melting temperature of TbrPDEB1 is between 51.5 °C and 55 °C, optimisation by increasing temperatures from 25 °C to 50 °C did not improve the spectra of TbrPDEB1. Therefore, the NMR screening methods could not be performed due to the instability of TbrPDEB1.

TbrPDEB1 becomes stabilised with the presence of a ligand, highlighting the need of a ligand to remain stable and in the conditions of native state. CD experiments also revealed that TbrPDEB1 remains predominantly α-helical at room temperature, as observed crystallographically.

TbrPDEB1 was a good crystal example, as the interactions of the ligand with the protein could be characterised with the four different ligands studied. The protein's size, stability in NMR buffer and NMR tumbling rate were considered as limiting factors for such NMR

studies. Therefore, this study highlighted that not all proteins can be suitable for the purpose of drug design using the NMR technique.

### 7.1.3 BRD4-BD1 Conclusions

BRD4-BD1 (15 kDa) was studied in complex with the ligands BTB 01148, BTB 06033, BTB 07004, RJF 00210, TG 00013, HTS 05027 and RJF 00002. X-ray crystallographic studies revealed a conserved binding site in the loop region of BRD4-BD1 with a conserved interaction with N140 in all studied structures.

BRD4-BD1 was also studied using both ligand (BTB 07004) and protein-observed NMR methods utilising $^{1}H$ and $^{19}F$ nuclei. NMR studies revealed that the protein was folded, in monomeric state and highly stable in the GF buffer conditions, rendering it suitable for NMR studies. NMR screening methodologies, including $^{1}H$ STD-NMR, $^{1}H$ CPMG and $^{19}F$ CMPG were performed to show that BTB 07004 binds to BRD4-BD1.

Ligand binding of BTB 07004 to BRD4-BD1 was observed through the change of intensity of proton signal, while the protein was in folded state. BRD4-BD1 was then studied using 2D-based heteronuclear $^{15}N/^{13}C/^{1}H$ experiments and the pair of triple resonance experiments HNCA, HN(CO)CA, HNCACB, HN(CO)CACB for BRD4-BD1 led to the sequential backbone assignment of $^{1}H_{N}$, $^{15}N$, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ nuclei. Major chemical shift differences ranging above 0.3 ppm were observed for the residues W81, V87, D88, A89, N93, L94, I146, A150, L158 and K160. The chemical shift mapping was in agreement with the binding site in the BRD4-BD1 crystal structure, particularly for the residues W81, V87, L94 and I146.

Additionally, 1D $^{19}F$ experiment were performed to look for site specific chemical shift perturbations. BRD4-BD1 was isotopically enriched with 5'fluoroindole to label the tryptophan residues of BRD4-BD1. Chemical shifts observed in $^{19}F$ CMPG experiments rapidly confirmed that W81 is part of the binding site, and W75 and W120 are not part of the ligand binding site of BRD4-BD1. This showed that fluorine labelling of proteins can be used as a tool to selectively study protein-ligand interaction when a Tryptophan is present in or has close proximity to the ligand binding site.

BRD4-BD1 was a good model system for comparing NMR results with X-ray crystallographic data and gain a better understanding of the protein's binding site and the protein ligand interactions and showed good agreement between the observed NMR X-ray crystallography results.

Overall its be concluded from a small sample of 3 systems, 1 was amenable to NMR studies but not X-ray crystallography due to problems with crystallisation and was expected to given good information on the ligand binding site. One was only amenable to X-ray crystallography but noy NMR to problems with the size of the system and the quality of spectra. In the third case (BRD4) the system was studied but both NMR and X-ray crystallography which gave information on ligand binding hits and the ligand binding site which were in good agreement. Clearly from this analysis a joint NMR and X-ray crystallography approach to analysing ligand binding is an effective approach with each technique having pros and cons and providing complimentary data.

## 7.2 Future Studies: The development of Group Epitope Mapping

The work in this thesis can be continued with the development of an advanced NMR technique for the creation of a Group Epitope Mapping (GEM) methodology. GEM determines the distances between the ligand and the receptor, creating the precise map of binding-sites for the target protein. GEM should inform about the ligand binding sites and interaction sites of any proteins without necessarily the crystal structure of a protein. It will be a good method for proteins with unknown crystal structures to find their ligand-binding sites and become a potential target method for drug discovery.

## 7.3 Conclusions

The characterisation of CypA, TbrPDEB1 and BRD4-BD1 structures and their ligand binding sites has permitted to compare the advantages and disadvantages of X-ray crystallography and NMR techniques.

| | X-ray Crystallography | Nuclear Magnetic Resonance (NMR) |
|---|---|---|
| **Protein Supply** | Requires large amounts (milligrams) of high quality protein that crystallises. Well-ordered crystals must diffract to obtain a three-dimensional structure. | Requires large amounts (milligrams) of soluble and isotopically-labelled protein. |
| **Sample Environment** | Solid-frozen structure. Can be purified in any buffers. Samples exposed to radiation damage. | In buffered solution. Highly suitable in phosphate buffer. Limited due to temperature restrictions. |
| **Protein Size** | Any size proteins and macromolecule complexes can be studied. | Currently only proteins with less than a 50 kDa mass can easily be studied by $^{1}H$-$^{15}N$ methods. Large proteins are difficult to study by NMR due to signal overlap and signal broadening in the spectra due to slower tumbling rates. |
| **Experiment time** | Long time required for screening, crystallisation and optimisation for high quality crystals (weeks/months). Short time required for data acquisition and processing (minutes/hours). | Short time required for sample preparation and data acquisition (minutes/hours). Long time required for data analysis and NMR protein assignment (weeks/months). |
| **Experiment quality** | Highly precise atomic resolution data can be obtained from well-ordered crystals. Good local resolutions and global resolution for | Highly precise data can be obtained from pure protein sample. Solution-state dynamic studies can be obtained. Good local and lower global resolutions. |

| | | |
|---|---|---|
| | disordered/flexible regions. | |
| **Conformational issues** | Crystal packing problems can result in change of protein conformation. | Oligomerisation can be an issue and buffers can influence the folded state of a protein. |
| **Liganded structures** | Very quick (minutes/hours). Once the crystal structure of a protein is fully refined (hours), the binding of a ligand can be quickly determined (minutes). | Quick (days/months). 100 – 1,000 ligands can be identified as ligand "hits" using high-throughput $^{19}F$ and $^{1}H$ STD, WaterLOGSY and CPMG NMR experiments (minutes). Once a $^{1}H/^{15}N$ HSQC spectrum is fully assigned (months), chemical shift mapping can be done quickly for one ligand (1 day). |

**Table 7.2. Comparison of X-ray crystallography and NMR**

X-ray crystallography and NMR are the most widely used techniques in protein structure determination and understanding of ligand binding interactions. Although they both have limitations, they mutually complement each other. For example, NMR might not be suitable for the study of large macromolecules, but can supplement a crystal structure with dynamics' information, binding information and conformational changes in solution. Therefore, combining both the techniques can lead to advanced structural information about a protein and ligand-binding interaction topology. The evolution of advanced NMR techniques such 1D and 2D qSTD allow for the quick determination of distances between the ligand and may have a important role to play the future. It should also be noted that recent developments of high resolution electron cryo-microscopy have highlighted this approach as a key tool in the structural biologists' arsenal in understanding protein-protein and protein-ligand interactions.

# E. Appendix Contents

**Appendix 3.1 List of recombinant products used**

**Appendix 6.1 Resonance Assignment for BRD4-BD1**

**Appendix 6.2 TALOS Secondary Structure Predictions**

**Appendix 6.3 Chemical Shift Mapping of BRD4-BD1 with BTB 07004**

**Appendix 6.4 Chemical Shift Mapping of BRD4-BD1 with EtOH**

# APPENDIX 3.1 List of recombinant products used

| Protein | Amino Acid Sequence | Calculated Average Mass (Da) | Description in the thesis |
|---|---|---|---|
| **Cyclophilin A (CypA)** | $^{2}$VNPTVFFDIAVDGEPLGRVSFELFADKVPKTAE NFRALSTGEKGFGYKGSCFHRIIPGFMCQGGDFT RHNGTGGKSIYGEKFEDENFILKHTGPGILSMAN AGPNTNGSQFFICTAKTEWLDGKHVVFGKVKEGM NIVEAMERFGSRNGKTSKKITIADCGQLE$_{165}$ | **17881.30** | **Chapter 3.2** |
| **Phosphodiesterase B1 expressed in *T brucei* (PDEB1-Tbr)** | $^{605}$FDVTDVEFDLFRARESTDKPLDVAAAIAYRLL LGSGLPQKFGCSDEVLLNFILQCRKKYRNVPYHN FYHVVDVCQTIHTFLYRGNVYEKLTELECFVLLI TALVHDLDHMGLNNSFYLKTESPLGILSSASGNT SVLEVHHCNLAVEILSDPESDVFDGLEGAERTLA FRSMIDCVLATDMAKHGSALEAFLASAADQSSDE AAFHRMTMEIILKAGDISNVTKPFDISRQWAMAV TEEFYRQGDMEKERGVEVLPMFDRSKNMELAKGQ IGFIDFVAAPFFQKIVDACLQCMQWTVDRIKSNR AQWERVLETR$_{918}$ | **35548.61** | **Chapter 4.2** |
| **Bromodomain 4 – BD1 (BRD4-BD1)** *For X-ray Crystallography studies* | $^{42}$TNPPPPNETSNPNKPKRQTNQLQYLLRVVLKTL WKHQFAWPFQQPVDAVKLNLPDYYKIIKTPMDMG TIKKRLENNYYWNAQECIQDFNTMFTNCYIYNKP GDDIVLMAEALEKLFLQKINELPT$_{166}$ | **14822.11** | **Chapter 5.2** |

| | | | |
|---|---|---|---|
| **Bromodomain 4 – BD1 (BRD4-BD1)**<br><br>*For NMR studies* | $^{22}$**HHHHHHSSGVDLGTENLYFQSMPPPPNETSNPN KPKRQTNQLQYLLRVVLKTLWKHQFAWPFQQPVD AVKLNLPDYYKIIKTPMDMGTIKKRLENNYYWNA QECIQDFNTMFTNCYIYNKPGDDIVLMAEALEKL FLQKINELPTEE**$_{168}$ | **17417.88** | **Chapter 6.2** |

# APPENDIX 6.1 Resonance Assignment for BRD4-BD1

The chemical shifts have been referenced in the C dimension by +1.25 ppm.

The final errors in expected shifts vs observed shifts were Cα -0.34 Cβ 0.47 NH 0.40 (N not corrected) according to PANAV (Bowei Wang, Yunjun Wang and David S. Wishart. "A probabilistic approach for validating protein NMR chemical shift assignments". Journal of Biomolecular NMR. Volume 47, Number 2 / June 2010: 85-99).

| Residue | Residue Name | $H_N$ (ppm) | $N_H$ (ppm) | $C_\alpha$ (ppm) | $C_\beta$ (ppm) |
|---|---|---|---|---|---|
| 27 | His | 8.61 | 121.06 | 55.14 | 28.59 |
| 28 | Ser | 8.33 | 113.89 | 59.99 | 68.21 |
| 29 | Ser | 8.15 | 117.30 | 56.36 | 62.59 |
| 34 | Gly | 8.54 | 109.46 | 43.94 | - |
| 35 | Thr | 7.98 | 113.19 | 60.49 | 67.93 |
| 36 | Glu | 8.54 | 123.05 | 55.56 | 28.24 |
| 37 | Asn | 8.37 | 119.22 | 51.76 | 37.04 |
| 38 | Leu | 8.05 | 122.27 | 54.07 | 40.48 |
| 39 | Tyr | 8.02 | 119.82 | 56.59 | 36.79 |
| 40 | Phe | 7.94 | 120.83 | 56.41 | 37.65 |
| 41 | Gln | 8.16 | 121.33 | 54.47 | 27.55 |
| 42 | Ser | 8.17 | 116.37 | 56.89 | 62.04 |
| 43 | Met | 8.17 | 121.34 | 53.75 | 31.37 |
| 48 | Asn | 8.33 | 117.50 | 51.49 | 37.25 |

| | | | | | |
|---|---|---|---|---|---|
| 49 | Glu | 7.80 | 122.53 | 52.39 | 31.17 |
| 64 | Gln | 8.54 | 119.41 | 57.77 | 26.83 |
| 65 | Tyr | 7.92 | 121.75 | 60.19 | 36.88 |
| 66 | Leu | 8.55 | 121.78 | 56.97 | 39.90 |
| 67 | Leu | 7.89 | 118.79 | 56.33 | 41.23 |
| 68 | Arg | 8.89 | 115.36 | 56.95 | 29.31 |
| 69 | Val | 8.10 | 115.49 | 62.61 | 30.47 |
| 70 | Val | 7.50 | 120.42 | 66.89 | 28.90 |
| 71 | Leu | 9.47 | 119.91 | 57.14 | 39.79 |
| 72 | Lys | 6.56 | 115.49 | 58.13 | 30.86 |
| 73 | Thr | 7.37 | 112.19 | 65.25 | 67.18 |
| 74 | Leu | 8.08 | 121.61 | 57.24 | 40.59 |
| 75 | Trp | 9.20 | 117.14 | 59.37 | 27.73 |
| 76 | Lys | 7.07 | 109.60 | 53.81 | 31.13 |
| 77 | His | 7.63 | 123.82 | 57.36 | 30.52 |
| 78 | Gln | 8.44 | 128.34 | 57.33 | 25.97 |
| 79 | Phe | 11.53 | 122.03 | 51.47 | 35.87 |
| 80 | Ala | 8.03 | 116.99 | 53.25 | 19.13 |
| 81 | Trp | 7.58 | 116.11 | 59.83 | 25.38 |
| 87 | Val | 8.55 | 124.89 | 62.76 | 30.61 |
| 88 | Asp | 8.18 | 130.13 | 50.60 | 37.94 |
| 89 | Ala | 8.16 | 125.92 | 52.99 | 17.80 |
| 90 | Val | 7.95 | 118.42 | 63.87 | 30.10 |

| | | | | | |
|---|---|---|---|---|---|
| 91 | Lys | 7.90 | 122.79 | 57.25 | 30.92 |
| 92 | Leu | 7.83 | 114.67 | 53.09 | 40.41 |
| 93 | Asn | 7.54 | 116.83 | 52.17 | 35.12 |
| 94 | Leu | 8.28 | 118.67 | 49.68 | 40.51 |
| 97 | Tyr | 7.66 | 123.26 | 62.25 | 37.76 |
| 98 | Tyr | 8.37 | 112.39 | 58.05 | 34.92 |
| 99 | Lys | 7.32 | 118.31 | 56.15 | 31.08 |
| 100 | Ile | 7.27 | 118.96 | 60.66 | 37.93 |
| 101 | Ile | 8.26 | 121.83 | 55.89 | 31.96 |
| 102 | Lys | 7.80 | 124.91 | 55.94 | 31.21 |
| 103 | Thr | 8.27 | 112.42 | 55.57 | 67.87 |
| 106 | Asp | 7.29 | 119.99 | 51.23 | 42.96 |
| 107 | Met | 9.18 | 119.05 | 58.27 | 33.71 |
| 108 | Gly | 8.73 | 113.07 | 46.06 | - |
| 114 | Leu | 8.86 | 121.33 | 56.52 | 42.14 |
| 115 | Glu | 8.69 | 119.48 | 57.20 | 28.29 |
| 116 | Asn | 7.98 | 115.02 | 51.11 | 36.84 |
| 117 | Asn | 7.88 | 115.73 | 53.01 | 35.02 |
| 118 | Tyr | 8.46 | 117.21 | 59.16 | 38.78 |
| 119 | Tyr | 7.77 | 114.91 | 55.64 | 37.76 |
| 120 | Trp | 9.80 | 122.30 | 57.01 | 28.95 |
| 124 | Glu | 8.10 | 115.69 | 55.05 | 30.50 |
| 125 | Cys | 8.56 | 118.46 | 61.26 | 34.91 |

| | | | | | |
|---|---|---|---|---|---|
| 126 | Ile | 7.50 | 121.12 | 53.67 | 36.80 |
| 127 | Gln | 7.03 | 116.11 | 53.14 | 27.83 |
| 128 | Asp | 6.94 | 121.33 | 51.82 | 41.33 |
| 130 | Asn | 8.52 | 116.23 | 55.31 | 36.68 |
| 131 | Thr | 8.96 | 121.51 | 65.26 | 67.20 |
| 132 | Met | 7.61 | 120.67 | 58.46 | 31.86 |
| 133 | Phe | 7.77 | 117.26 | 59.84 | 36.06 |
| 134 | Thr | 8.66 | 113.53 | 65.65 | 67.04 |
| 135 | Asn | 8.76 | 119.59 | 53.40 | 36.41 |
| 143 | Gly | 8.39 | 110.13 | 43.26 | - |
| 144 | Asp | 8.00 | 122.04 | 53.08 | 39.91 |
| 145 | Asp | 8.93 | 124.24 | 56.58 | 38.33 |
| 146 | Ile | 8.72 | 117.16 | 61.45 | 35.97 |
| 147 | Val | 6.96 | 123.02 | 65.03 | 29.60 |
| 148 | Leu | 7.58 | 120.27 | 56.45 | 39.39 |
| 150 | Ala | 8.42 | 122.13 | 54.16 | 17.73 |
| 151 | Glu | 8.96 | 116.81 | 57.73 | 28.25 |
| 158 | Leu | 8.06 | 120.53 | 55.97 | 40.72 |
| 159 | Gln | 9.05 | 122.50 | 57.60 | 27.07 |
| 160 | Lys | 7.83 | 118.60 | 58.34 | 30.70 |
| 161 | Ile | 7.20 | 120.87 | 55.89 | 40.32 |

There were few ambiguities in the assignment:

- Fragment 27 can be assigned to 21-25 or 136-138 or 154-156 (listed in shift table)

- Fragment 52 can be assigned to 47-49 or 53-55 (listed in shift table)

## Appendix 6.2 TALOS Secondary Structure Predictions

DATA SEQUENCE HSSGVDLGTE NLYFQSMPPP PNETSNPNKP KRQTNQLQYL LRVVLKTLWK
DATA PREDICTED LLLXXXXLLL LLLLLLXXXX XLLXXXXXXX XXXXXXXHHH HHHHHHHHL

DATA SEQUENCE HQFAWPFQQP VDAVKLNLPD YYKIIKTPMD MGTIKKRLEN NYYWNAQECI
DATA PREDICTED LLLLLXXXXX LLHHHHHLLXX HHHHLLLXXL LHXXXXXHHL LLLLXXXLLL

DATA SEQUENCE QDFNTMFTNC YIYNKPGDDI VLMAEALEKL FLQKIN
DATA PREDICTED LLHHHHHHHX XXXXXXLLHH HHHHHXXXXX XHHHLX

H-Helix E-Strand L-Coil X-Unidentifiable

## Appendix 6.3 Chemical Shift Mapping of BRD4-BD1 with BTB 07004

| Residue Number | Minimal Shift (ppm) |
|---|---|
| 27 | 0.10 |
| 28 | 0.00 |
| 29 | 0.00 |
| 30 | 0.00 |
| 31 | 0.00 |
| 32 | 0.00 |
| 33 | 0.00 |
| 34 | 0.11 |
| 35 | 0.12 |
| 36 | 0.08 |
| 37 | 0.19 |
| 38 | 0.16 |
| 39 | 0.19 |
| 40 | 0.26 |
| 41 | 0.23 |
| 42 | 0.22 |
| 43 | 0.00 |
| 44 | 0.00 |
| 45 | 0.00 |
| 46 | 0.00 |

| | |
|---|---|
| 47 | 0.00 |
| 48 | 0.19 |
| 49 | 0.21 |
| 50 | 0.00 |
| 51 | 0.00 |
| 52 | 0.00 |
| 53 | 0.00 |
| 54 | 0.00 |
| 55 | 0.00 |
| 56 | 0.00 |
| 57 | 0.00 |
| 58 | 0.00 |
| 59 | 0.00 |
| 60 | 0.00 |
| 61 | 0.00 |
| 62 | 0.00 |
| 63 | 0.00 |
| 64 | 0.16 |
| 65 | 0.05 |
| 66 | 0.08 |
| 67 | 0.23 |
| 68 | 0.24 |
| 69 | 0.12 |

| | |
|---|---|
| 70 | 0.00 |
| 71 | 0.05 |
| 72 | 0.14 |
| 73 | 0.22 |
| 74 | 0.11 |
| 75 | 0.05 |
| 76 | 0.20 |
| 77 | 0.04 |
| 78 | 0.10 |
| 79 | 0.18 |
| 80 | 0.14 |
| 81 | 0.38 |
| 82 | 0.00 |
| 83 | 0.00 |
| 84 | 0.00 |
| 85 | 0.00 |
| 86 | 0.00 |
| 87 | 0.32 |
| 88 | 0.55 |
| 89 | 0.53 |
| 90 | 0.13 |
| 91 | 0.16 |
| 92 | 0.22 |

| | |
|---|---|
| 93 | 0.33 |
| 94 | 0.44 |
| 95 | 0.00 |
| 96 | 0.00 |
| 97 | 0.19 |
| 98 | 0.16 |
| 99 | 0.22 |
| 100 | 0.18 |
| 101 | 0.11 |
| 102 | 0.18 |
| 103 | 0.23 |
| 104 | 0.00 |
| 105 | 0.00 |
| 106 | 0.08 |
| 107 | 0.18 |
| 108 | 0.00 |
| 109 | 0.00 |
| 110 | 0.00 |
| 111 | 0.00 |
| 112 | 0.00 |
| 113 | 0.00 |
| 114 | 0.17 |
| 115 | 0.15 |

| | |
|---|---|
| 116 | 0.12 |
| 117 | 0.06 |
| 118 | 0.12 |
| 119 | 0.16 |
| 120 | 0.06 |
| 121 | 0.00 |
| 122 | 0.00 |
| 123 | 0.00 |
| 124 | 0.17 |
| 125 | 0.21 |
| 126 | 0.21 |
| 127 | 0.06 |
| 128 | 0.07 |
| 129 | 0.00 |
| 130 | 0.07 |
| 131 | 0.12 |
| 132 | 0.12 |
| 133 | 0.11 |
| 134 | 0.12 |
| 135 | 0.03 |
| 136 | 0.00 |
| 137 | 0.00 |
| 138 | 0.00 |

| | |
|---|---|
| 139 | 0.00 |
| 140 | 0.00 |
| 141 | 0.00 |
| 142 | 0.00 |
| 143 | 0.13 |
| 144 | 0.17 |
| 145 | 0.12 |
| 146 | 0.34 |
| 147 | 0.16 |
| 148 | 0.18 |
| 149 | 0.00 |
| 150 | 0.43 |
| 151 | 0.11 |
| 152 | 0.00 |
| 153 | 0.00 |
| 154 | 0.00 |
| 155 | 0.00 |
| 156 | 0.00 |
| 157 | 0.00 |
| 158 | 0.47 |
| 159 | 0.11 |
| 160 | 0.40 |
| 161 | 0.12 |

## Appendix 6.4 Chemical Shift Mapping of BRD4-BD1 with EtOH

| Residue Number | Minimal Shift |
|---|---|
| 27 | 0.14 |
| 28 | 0.15 |
| 29 | 0.08 |
| 30 | 0.00 |
| 31 | 0.00 |
| 32 | 0.00 |
| 33 | 0.00 |
| 34 | 0.13 |
| 35 | 0.05 |
| 36 | 0.10 |
| 37 | 0.22 |
| 38 | 0.18 |
| 39 | 0.24 |
| 40 | 0.25 |
| 41 | 0.23 |
| 42 | 0.26 |
| 43 | 0.00 |
| 44 | 0.00 |
| 45 | 0.00 |
| 46 | 0.00 |

| | |
|---|---|
| 47 | 0.00 |
| 48 | 0.17 |
| 49 | 0.17 |
| 50 | 0.00 |
| 51 | 0.00 |
| 52 | 0.00 |
| 53 | 0.00 |
| 54 | 0.00 |
| 55 | 0.00 |
| 56 | 0.00 |
| 57 | 0.00 |
| 58 | 0.00 |
| 59 | 0.00 |
| 60 | 0.00 |
| 61 | 0.00 |
| 62 | 0.00 |
| 63 | 0.00 |
| 64 | 0.16 |
| 65 | 0.00 |
| 66 | 0.17 |
| 67 | 0.19 |
| 68 | 0.28 |
| 69 | 0.21 |

| | |
|---|---|
| 70 | 0.40 |
| 71 | 0.47 |
| 72 | 0.22 |
| 73 | 0.07 |
| 74 | 0.28 |
| 75 | 0.18 |
| 76 | 0.09 |
| 77 | 0.06 |
| 78 | 0.00 |
| 79 | 0.16 |
| 80 | 0.04 |
| 81 | 0.10 |
| 82 | 0.00 |
| 83 | 0.00 |
| 84 | 0.00 |
| 85 | 0.00 |
| 86 | 0.00 |
| 87 | 0.08 |
| 88 | 0.32 |
| 89 | 0.21 |
| 90 | 0.09 |
| 91 | 0.09 |
| 92 | 0.24 |

| | |
|---|---|
| 93 | 0.16 |
| 94 | 0.22 |
| 95 | 0.00 |
| 96 | 0.00 |
| 97 | 0.14 |
| 98 | 0.09 |
| 99 | 0.10 |
| 100 | 0.19 |
| 101 | 0.17 |
| 102 | 0.11 |
| 103 | 0.31 |
| 104 | 0.00 |
| 105 | 0.00 |
| 106 | 0.28 |
| 107 | 0.07 |
| 108 | 0.18 |
| 109 | 0.00 |
| 110 | 0.00 |
| 111 | 0.00 |
| 112 | 0.00 |
| 113 | 0.00 |
| 114 | 0.11 |
| 115 | 0.11 |

| | |
|---|---|
| 116 | 0.25 |
| 117 | 0.17 |
| 118 | 0.17 |
| 119 | 0.17 |
| 120 | 0.01 |
| 121 | 0.00 |
| 122 | 0.00 |
| 123 | 0.00 |
| 124 | 0.16 |
| 125 | 0.15 |
| 126 | 0.21 |
| 127 | 0.06 |
| 128 | 0.11 |
| 129 | 0.00 |
| 130 | 0.10 |
| 131 | 0.25 |
| 132 | 0.11 |
| 133 | 0.05 |
| 134 | 0.17 |
| 135 | 0.22 |
| 136 | 0.00 |
| 137 | 0.00 |
| 138 | 0.00 |

| | |
|---|---|
| 139 | 0.00 |
| 140 | 0.00 |
| 141 | 0.00 |
| 142 | 0.00 |
| 143 | 0.07 |
| 144 | 0.15 |
| 145 | 0.14 |
| 146 | 0.18 |
| 147 | 0.17 |
| 148 | 0.05 |
| 149 | 0.00 |
| 150 | 0.21 |
| 151 | 0.10 |
| 152 | 0.00 |
| 153 | 0.00 |
| 154 | 0.00 |
| 155 | 0.00 |
| 156 | 0.00 |
| 157 | 0.00 |
| 158 | 0.34 |
| 159 | 0.15 |
| 160 | 0.10 |
| 161 | 0.00 |

# F. References


[1] www.aminoacid-studies.com/amino-acids/what-are-amino-acids.html

[2] www.nlm.nih.gov/medlineplus/ency/article/002222.htm

[3] www.medicalnewstoday.com/articles/196279.php

[4] www.proteinstructures.com/Structure/protein-structure1.html

[5] George A. Ordway, Daniel J. Garry (2004) Myoglobin: an essential hemoprotein in striated muscle. *Journal of Experimental Biology* 207: 3441-3446

[6] www.vivo.colostate.edu/hbooks/genetics/biotec/basics/prostruct.html

[7] Beat Blattmann and Patrick Sticher (2009) Growing crystals from protein. *Science in School Spring* 11: 30-36

[8] Angela Cronenborn and Marius Clore (1990) Protein Structure Determination in solution by two-dimensional and three-dimensional nuclear magnetic resonance spectroscopy. Anal. Chem. 62: 2-15

[9] Yue WW, Froese DS, Brennan PE (2014) The role of protein structural analysis in the next generation sequencing era. *Top Curr Chem*. 336: 67-98.

[10] www.nigms.nih.gov/about/budget/statements/Pages/February25_1999.aspx

[11] http://www.ebi.ac.uk/interpro/entry/IPR011953

[12] Schubert HL, Raux E, Wilson KS, Warren MJ (1999) Anaerobic Cobalt Chelatase in Cobalamin Biosynthesis from Salmonella Typhimurium. *Biochemistry*. 38: 10660-9

[13] Medlock AE, Dailey TA, Ross TA, Dailey HA, Lanzilotta WN (2007) Wild type human ferrochelatase crystallized with MnCl2. *Journal Molecular Biology*. 373: 1006-16 [14] Cornuejols D (2009) Biological crystals: at the interface between physics, chemistry and biology. Science in School 11: 70-76

[14] http://www.beamer.org/courses/biology/textbook/unit2/proteinsandproteomics

[15] Molecular Biology of the Cell by Alberts B, Johnson A, Lewis J et al. 4th Edition (2002)

[16] Dominguez C1, Boelens R, Bonvin AM. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*. 125(7):1731-7

[17] James Fraser et al. (2011) Assessing protein conformational ensembles using room-temperature x-ray crystallography. *Proc. Natl. Acad. Sci* 108(39): 16247-16252

[18] Linus Pauling, Robert Corey et al. (1951) The structure of proteins: two-hydrogen bonded helical configurations of the polypeptide chain. *PNAS* 37(4): 205-211

[19] Andrea Cavalli et al. (2006) Protein Structure determination from NMR chemical shifts. *PNAS* 104(23): 9615-9620

[20] Jeffery Kelly (1988) New tools for probing protein structure. *Nature Biotechnology* 6: 125-129

[21] http://www.ebi.ac.uk/2can/tutorials/structure/struct3.html

[22] Leis S, Schneider S, Zacharias M (2010). In silico prediction of binding sites on proteins. *Curr Med Chem.* 17;(15):1550-62.

[23] Dietz MS, Fricke F, Krüger CL, Niemann HH, Heilemann M (2014). Receptor-ligand interactions: binding affinities studied by single-molecule and super-resolution microscopy on intact cells. *Chemphyschem.* 15(4):671-6.

[24] Dunn, M. F. (2010). Protein–Ligand Interactions: General Description. *Encyclopedia of Life Sciences*.

[25] vlab.amrita.edu,. (2012). Retrieving Motif Information of a Protein Using Prosite. Retrieved 17 November 2015, from *vlab.amrita.edu/?sub=3&brch=273&sim=1426&cnt=1*

[26] Jacob L., Vert JP. (2008) Protein-ligand interaction prediction: an improved chemogenomics approach. Bioinformatics. 24(19):2149-56.

[27] www.richmond.edu/~jbell2/chapter-17.pdf

[28] Paul Labute and Martin Santavy (2015) Locating Binding Sites in Protein Structures. *Chemical Computing Group Inc.*

[29] Shiming Lin, Ji-Liang Chen, Long-Sun Huang and Huan-We Lin (2005) Measurements of the Forces in Protein Interactions with Atomic Force Microscopy. *Current Proteomics*. 2(1):55-81

[30] Deepshikha Datta (2002) Protein-Ligand Interactions: Docking, Design and Protein Conformational Change. *Thesis Submitted in California Institute of Technology Pasadena, California, USA.*

[31] http://sites.fas.harvard.edu/~lsci1a/10-12notes.pdf

[32] Laurie AT, Jackson RM. (2006) Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. *Curr Protein Pept Sci.* 7(5):395-406.

[33] González-Ruiz D., Gohlke H. (2006) Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr Med Chem.* 13(22):2607-25

[34] Roger S. Macomber (1992) An introduction to NMR titration for studying rapid reversible complexation. *J. Chem. Educ.* 69 (5).

[35] http://www.slideshare.net/jayswan/chapter-8-metabolism-and-enzymes

[36] Bush EC, Clark AE, DeBoever CM, Haynes LE, Hussain S, et al. (2013) Modeling the Role of Negative Cooperativity in Metabolic Regulation and Homeostasis. *PLoS ONE* 8(9): 10.1371

[37] Laurie AT, Jackson RM. (2006) Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. *Curr Protein Pept Sci.* 7(5):395-406.

[38] Jhoti H and Leach A (2007) Structure-based Drug Discovery, *Springer-Verlac New York Inc.*

[39] Glen RC, Allen SC. (2003) Ligand-protein docking: cancer research at the interface between biology and chemistry. *Curr Med Chem.* 10(9):763-7

[40] Von der Osten C, Branner S, Hastrup S, Hedegaard L, Rasmussen MD, Bisgård-Frantzen H, Carlsen S, Mikkelsen JM (1993) Protein engineering of subtilisins to improve stability in detergent formulations. *J Biotechnol*. 28(1): 55-68.

[41] http://www.rcsb.org/pdb/101

[42] Niedzialkowska, E., Gasiorowska, O., Handing, K. B., Majorek, K. A., Porebski, P. J., Shabalin, I. G., Minor, W. (2016). Protein purification and crystallization artifacts: The tale usually not told. *Protein Science: A Publication of the Protein Society*, 25(3), 720–733.

[43] Mark R. Sanderson, Jane V. Skelly (2007) Macromolecular Crystallography: Conventional and High-throughput Methods. *Oxford University Press*. Chapter 3.1 pp45.

[44] www.bio.davidson.edu/molecular/MolStudents/spring2003/Kogoy/protein.html

[45] www.proteinstructures.com/Experimental/Experimental/crystallization-tools.html

[46] Niesen FH, Berglund H, Vedadi M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc.* 2(9):2212-21

[47] http://www.chem.uci.edu/~dmitryf/manuals/Fundamentals/CD%20spectroscopy.pdf

[48] https://www.photophysics.com/resources/tutorials/circular-dichroism-cd-spectroscopy

[49] Cole JL1, Lary JW, P Moody T, Laue TM. (2008) Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol*. 84:143-79.

[50] Stout G, Jensen L (1989) X-ray Structure Determination: A practical guide, *Wiley*

[51] Giacovazzo C (2011) Fundamentals of Crystallography (International Union of Crystallography Monographs on Crystallography), *Oxford University Press*

[52] Babine R (2003) Protein Crystallography in Drug Discovery, *Wiley*

[53] http://users.aber.ac.uk/ruw/teach/334/bravais.php

[54] http://www.academia.edu/16183346/MLE2101_Lab_Report

[55] Mayer, M and Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie (International ed Print)* 38(12):1784-1788

[56] Cintia D.F Milagre et al. (2011) STD NMR Spectroscopy: a case study of fosfomycin binding interactions in living bacterial cells. J. Braz. Chem. Soc 22(2): 195-381

[57] http://www.chem.wisc.edu/areas/reich/nmr/notes-8-tech-2-NOE.pdf

[58] http://www.usp.br/massa/2014/qfl2144/pdf/22_Maio_2014.pdf

[59] Aldino Viegas et al. (2011) Saturation-Transfer Difference (STD) NMR: a simple and fast method for ligand screening and characterization of protein binding. *J Chem. Educ* 88(7): 990-994

[60] Wagstaff JL et al. (2010) Two-dimensional heteronuclear saturation transfer difference NMR reveals detailed integrin αvβ6 protein-peptide interactions. Chem. Commun (Camb) 46(40): 7533-7535

[61] Mayer, M and Meyer, B (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J Am Chem Soc 123(25): 6108-6117

[62] Chi-Fon Chang (2009) Saturation transfer difference NMR study on the specific binding of ligand to protein. *Analytical Biochemistry* 385(2): 380-382

[63] Hiraishi N, Tochio N, Kigawa T, Otsuki M, Tagami J (2013) Monomer-Collagen interactions studied by Saturation Transfer Difference NMR. *Journal of Dental Research* 92(3):284-8

[64] Schade M. (2006) NMR fragment screening: Advantages and applications. *IDrugs*. 9(2):110-3.

[65] Vogtherr M., Fiebig K. (2003) NMR-based screening methods for lead discovery. *EXS*. (93):183-202.

[66] Skinner AL., Laurence JS. (2008) High-field solution NMR spectroscopy as a tool for assessing protein interactions with small molecule ligands. *J Pharm Sci*. 97(11):4670-95.

[67] Elisa Barile and Maurizio Pellecchia (2014) NMR-Based Approaches for the Identification and Optimization of Inhibitors of Protein–Protein Interactions. *Chemical Reviews*. 114(9): 4749–4763.

[68] https://chemistry.osu.edu/~foster.281/fosterlab/papers/Kovacs_pnmrs05.pdf

[69] http://www.cellbio.med.howard.edu/research_facilities/molecular_e_mi

[70] Slavica Jonic et al. (2009) Protein structure determination by electron cryo-microscopy. Current opinion in pharmacology 9: 636-642

[71] Aebi U et al. (1984) Electron Microscopy and image processing applied to the study of protein structure and protein-protein interactions. *J Ultractrastruct Res* 88(2): 143-176

[72] Wesley C. Van Voorhis et al. (2009) The role of medical structural genomics in discovering new drugs for infectious disease. *Plos Computational Biology* 5(10): e1000530

[73] Richard Kriwacki et al. (2004) Protein structure characterization with mass spectrometry. *Spectroscopy* 18: 37-47

[74] Biemann K (1998) Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* 16(1-12):99-111

[75] http://www.eurekalert.org/features/doe/2001-06/drnl-pib061902.phb

[76] P. Singh et al. (2010) Chemical cross-linking and mass spectrometry as a low resolution protein structure determination technique. *Anal Chem* 82(7): 2636-2642

[77] Vicki Wysocki et al (2005) Mass spectrometry of peptides and proteins. *Methods* 35: 211-222

[78] www.panalytical.com/index.cfm?pid=143

[79] www.saxier.org/aboutus/saxs.shtml

[80] James Fraser et al. (2009) Hidden alternate structures of proline isomerase essential for catalysis. *Nature* 462(7273): 669-673

[81] Bornhorst JA., Falke JJ. (2000) Purification of proteins using polyhistidine affinity tags. *Methods Enzymol*. 326:245-54.

[82] www.fisher.co.uk/1/1/172991-hiprep-26-10-desalting-column-53ml.html

[83]www.separations.us.tosohbioscience.com/ServiceSupport/TechSupport/ResourceCenter/PrinciplesofChromatography/IonExchange

[84] Xiong S., Zhang L., He QY. (2008) Fractionation of proteins by heparin chromatography. *Methods Mol Biol*. 424:213-21

[85] http://www.sigmaaldrich.com/life-science/proteomics/protein-chromatography/gel-5filtration-chromatography.html

[86] Hagel L. (2001) Gel-filtration chromatography. *Curr Protoc Mol Biol*. Chapter 10:Unit 10.9

[87] http://www.ruf.rice.edu/~bioslabs/studies/sds-page/gellab2.html

[88] F Chevalier (2011) Analytical Methods and Electrophoresis. *Encyclopedia of Dairy Sciences*, Second Edition, vol. 1, pp.185–192. San Diego: Academic Press.

[89] http://sopnmr.ucsd.edu/sample-preparation.htm

[90] http://www.astbury.leeds.ac.uk/facil/MStut/mstutorial.htm

[91] http://hyperphysics.phy-astr.gsu.edu/hbase/magnetic/maspec.html

[92] http://www.chem.uci.edu/~dmitryf/manuals/Fundamentals/CD%20spectroscopy.pdf

[93] https://www.photophysics.com/resources/tutorials/circular-dichroism-cd-spectroscopy

[94] Ping Wang and Joseph Heitman (2005) The cyclophilins. *Genome Biology* 6(7): 226

[95] Nigro P., Pompilio G., Capogrossi MC. (2013) Cyclophilin A: a key player for human disease. *Cell Death Dis*. 4: e888.

[96] Zheng-Gen Jm et al. (2004) Cyclophilin A is a proinflammatory cytokine that activates endothelial cells. *Atherioscler Thromb Vasc Biol* 24: 1186-1191

[97] Artur Kaul et al. (2009) Essential role of cyclophilin A for hepatitis C virus replication and virus production and possible link to polyprotein cleavage kinetics. *PLos Pathogen* 5(8)

[98] Ping Wang and Joseph Heitman (2005) The cyclophilins. *Genome Biology* 6(7): 226

[99] Structural Analysis of Cyclophilin and inhibitory activity of cyclosporine A on germination and growth of Moniliophtora Perniciosa" Paulo S. Menzani et al. (2011) *Pharmaceutica Analytica Acta* 2153:2453

[100] Marks AR (1996) Cellular functions of immunophilins 76(3): 631-649

[101] Fei Song et al. (2011) Cyclophilin A (CypA) induces chemotaxis independent of its peptidylprolyl cis-trans isomerase activity. *J Biol Chem* 286(10): 8197-8203

[102] Fischer G, Wittmann-Liebold B, Lang K, Kiefhaber T, Schmid FX: Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins (1989). *Nature*. 337:476-478.

[103] Paulo S. Menzani et al. (2011) Structural Analysis of Cyclophilin and inhibitory activity of cyclosporine A on germination and growth of Moniliophtora Perniciosa. *Pharmaceutica Analytica Acta* 2153:2453

[104] Stamnes MA, Shieh BH, Chuman L, Harris GL, Zuker CS (1991). The cyclophilin homolog ninaA is a tissue-specific integral membrane protein required for the proper synthesis of a subset of Drosophila rhodopsins. *Cell*. 65:219–227.

[105] Colley NJ, Baker EK, Stamnes MA, Zuker CS (1991). The cyclophilin homolog ninaA is required in the secretory pathway. *Cell*. 67:255–263.

[106] Suzuki J, Jin ZG, Meoli DF, Matoba T, Berk BC (2006). Cyclophilin A is secreted by a vesicular pathway in vascular smooth muscle cells. *Circ Res*. 98:811–817.

[107] Satoh K, Matoba T, Suzuki J, O'Dell MR, Nigro P, Cui Z, et al. Cyclophilin A mediates vascular remodeling by promoting inflammation and vascular smooth muscle cell proliferation (2008). Circulation. 117:3088–3098.

[108] Satoh K, Nigro P, Matoba T, O'Dell MR, Cui Z, Shi X, et al. Cyclophilin A enhances vascular oxidative stress and the development of angiotensin II-induced aortic aneurysms (2009). *Nat Med*. 15:649–656.

[109] Satoh K, Nigro P, Zeidan A, Soe NN, Jaffre F, Oikawa M, et al. (2011) Cyclophilin A promotes cardiac hypertrophy in apolipoprotein E-deficient mice. Arterioscler Thromb Vasc Biol. 31:1116–1123.

[110] Helekar SA., Char D., Neff S., Patrick J. (1994) Prolyl isomerase requirement for the expression of functional homo-oligomeric ligand-gated ion channels. *Neuron*. 12:179–189.

[111] Helekar SA., Patrick J. (1997) Peptidyl prolyl cis-trans isomerase activity of cyclophilin A in functional homo-oligomeric receptor expression. *Proc Natl Acad Sci USA*. 94:5432–5437.

[112] Golledge J, Norman PE. Atherosclerosis and abdominal aortic aneurysm: cause, response, or common risk factors (2010). *Arterioscler Thromb Vasc Biol*. 30:1075–1077.

[113] Ramachandran S, Venugopal A, Sathisha K, Reshmi G, Charles S, Divya G, et al. Proteomic profiling of high glucose primed monocytes identifies cyclophilin A as a potential secretory marker of inflammation in type 2 diabetes (2012). *Proteomics*. 12:2808–2821.

[114] Bosco DA, Eisenmesser EZ, Pochapsky S, Sundquist WI, Kern D. Catalysis of cis/trans isomerization in native HIV-1 capsid by human cyclophilin A (2002). *Proc Natl Acad Sci USA*. 99:5247–5252.

[115] Braaten D, Luban J. Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells (2001). *EMBO J*. 20:1300–1309.

[116] Sokolskaja E, Sayah DM, Luban J. Target cell cyclophilin A modulates human immunodeficiency virus type 1 infectivity (2004). *J Virol*. 78:12800–12808.

[117] Yang H, Chen J, Yang J, Qiao S, Zhao S, Yu L. Cyclophilin A is upregulated in small cell lung cancer and activates ERK1/2 signal (2007). *Biochem Biophys Res Commun*. 361:763–767.

[118] Cecconi D, Astner H, Donadelli M, Palmieri M, Missiaglia E, Hamdan M, et al. Proteomic analysis of pancreatic ductal carcinoma cells treated with 5-aza-2'-deoxycytidine. (2003) *Electrophoresis*. 24:4291–4303.

[119] Wong CS, Wong VW, Chan CM, Ma BB, Hui EP, Wong MC, et al. Identification of 5-fluorouracil response proteins in colorectal carcinoma cell line SW480 by two-dimensional electrophoresis and MALDI-TOF mass spectrometry (2008). *Oncol Rep*. 20:89–98.

[120] Han X, Yoon SH, Ding Y, Choi TG, Choi WJ, Kim YH, et al. Cyclosporin A and sanglifehrin A enhance chemotherapeutic effect of cisplatin in C6 glioma cells (2010). *Oncol Rep*. 23:1053–1062.

[121] Bell RD, Winkler EA, Singh I, Sagare AP, Deane R, Wu Z, et al. Apolipoprotein E controls cerebrovascular integrity via cyclophilin A (2012). *Nature*. 485:512–516.

[122] Billich A, Winkler G, Aschauer H, Rot A, Peichl P. Presence of cyclophilin A in synovial fluids of patients with rheumatoid arthritis (1997). *J Exp Med*. 185:975–980.

[123] Kim H, Kim WJ, Jeon ST, Koh EM, Cha HS, Ahn KS, et al. Cyclophilin A may contribute to the inflammatory processes in rheumatoid arthritis through induction of matrix degrading enzymes and inflammatory cytokines from macrophages (2005). *Clin Immunol*. 116:217–224.

[124] Yang Y, Lu N, Zhou J, Chen ZN, Zhu P. Cyclophilin A up-regulates MMP-9 expression and adhesion of monocytes/macrophages via CD147 signaling pathway in rheumatoid arthritis. *Rheumatology (Oxford)*. 47:1299–1310.

[125] Wang L, Wang CH, Jia JF, Ma XK, Li Y, Zhu HB, et al. Contribution of cyclophilin A to the regulation of inflammatory processes in rheumatoid arthritis (2010). *J Clin Immunol*. 30:24–33.

[126] Tegeder I, Schumacher A, John S, Geiger H, Geisslinger G, Bang H, et al. Elevated serum cyclophilin levels in patients with severe sepsis (1997). *J Clin Immunol*. 17:380–386.

[127] Gwinn WM, Damsker JM, Falahati R, Okwumabua I, Kelly-Welch A, Keegan AD, et al. Novel approach to inhibit asthma-mediated lung inflammation using anti-CD147 intervention (2006). *J Immunol*. 177:4870–4879.

[128] Liu L, Li C, Xiang J, Dong W, Cao Z. Over-expression and potential role of cyclophilin A in human periodontitis (2013). *J Periodontal Res*. 48:615–622.

[129] Gromov P, Skovgaard GL, Palsdottir H, Gromova I, Ostergaard M, Celis JE. Protein profiling of the human epidermis from the elderly reveals up-regulation of a signature of interferon-gamma-induced polypeptides that includes manganese-superoxide dismutase and the p85beta subunit of phosphatidylinositol 3-kinase (2003). *Mol Cell Proteomics*. 2:70–84.

[130] Li J, Xie H, Yi M, Peng L, Lei D, Chen X, et al. Expression of cyclophilin A and CD147 during skin aging (2011). Zhong Nan Da Xue Xue Bao Yi Xue Ban. 36:203–211

[131] Kallen J, Spitzfaden C, Zurini MG, Wider G, Widmer H, Wuthrich K, Walkinshaw MD. (1991) Structure of human cyclophilin and its binding site for cyclosporin A determined by X-ray crystallography and NMR spectroscopy. Nature. 353:276–279

[132] Ke HM, Zydowsky LD, Liu J, Walsh CT. Crystal structure of recombinant human T-cell cyclophilin A at 2.5 A resolution. (1991) Proc Natl Acad Sci USA. 88:9483–9487.

[133] Jin L, Harrison SC. Crystal structure of human calcineurin complexed with cyclosporin A and human cyclophilin (2002). *Proc Natl Acad Sci USA*. 99:13522–13526.

[134] Huai Q, Kim HY, Liu Y, Zhao Y, Mondragon A, Liu JO, Ke H. Crystal structure of calcineurin-cyclophilin-cyclosporin shows common but distinct recognition of immunophilin-drug complexes (2002). *Proc Natl Acad Sci USA*. 99:12037–12042.

[135] Pflugl, G., Kallen, J., Schirmer, T., Jansonius, J. N., Zurini, M. G. & Walkinshaw, M. D. (1993). *Nature* (London), 361, 91–94.

[136] Murshudov GN, Skubák P, Lebedev AA, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica* 67(Pt 4):355-367.

[137] Blanc E., Roversi P., Vonrhein C., Flensburg C., Lea SM., Bricogne G (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* 60:2210-2221.

[138] G.Winter (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* 43: 186-190.

[139] Graeme Winter, Katherine E. McAuley (2011) Automated data collection for macromolecular crystallography. *Methods in Structural Proteomics* 55(1): 81-93

[140] Mikol, V., Kallen, J., Pflugl, G. & Walkinshaw, M. D. (1993). X-ray structure of a monomeric cyclophilin A-cyclosporin A crystal complex at 2.1 A resolution. *J. Mol. Biol*. 234, 1119–1130

[141] Nadine Merkley, Ian Burton, Tobias Karakach and Raymond T. Syvitski (2013). Magnetic Resonance Technologies: Molecules to Medicine, Using Old Solutions to New Problems - Natural Drug Discovery in the 21st Century, Dr. Marianna Kulka (Ed.)

[142] Mikol V, Duc D. (1994) Crystallization of the complex between cyclophilin A and cyclosporin derivatives: the use of cross-seeding. *Acta Crystallogr D Biol Crystallogr*. 50(4):543-9.

[143] Kallen J, Sedrani R, Zenke G, Wagner J. (2005) Structure of human cyclophilin A in complex with the novel immunosuppressant sanglifehrin A at 1.6 A resolution. *J Biol Chem.* 280(23):21965-71.

[144] McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. Acta Crystallographica. Section F, *Structural Biology Communications*, 70(1):2–20.

[145] D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W., & Marsh, M. (2014). Microseed matrix screening for optimization in protein crystallization: what have we learned? Acta Crystallographica. Section F, *Structural Biology Communications*, 70(9):1117–1126.

[146] https://hamptonresearch.com/product_detail.aspx?sid=36&pid=27

[147] McPherson, A., & Cudney, B. (2014). Optimization of crystallization conditions for biological macromolecules. *Acta Crystallographica*. Section F, Structural Biology Communications, 70(11):1445–1467.

[148] https://www.uniprot.org/uniprot/P62937

[149] Gerhard DS et al. (2004) The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Research*, 14(10):2121-7.

[150] Ebert L., Schick M., Neubert P., Schatten R., Henze S., Korn B. (2004) Cloning of human full open reading frames in Gateway(TM) system entry vector (pDONR201). *EMBL/GenBank/DDBJ databases.*

[151] http://www.beckmangenomics.com/

[152] https://www.thermofisher.com/us/en/home/life-science/cloning/vector-nti-software.html

[153] Varland, S., Osberg, C., & Arnesen, T. (2015). N-terminal modifications of cellular proteins: The enzymes involved their substrate specificities and biological effects. *Proteomics*, 15(14), 2385–2401.

[154] http://fragmentech.univ-lyon1.fr/waterlogsy-experiment/

[155] Kleckner, I. R., & Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta*, 1814(8), 942–968. http://doi.org/10.1016/j.bbapap.2010.10.012

[156] Dalvit, C., Fogliatto, G., Stewart, A. et al. J Biomol NMR (2001) WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. Journal of Biomolecular NMR. 21(4): 349–359

[157] http://www.protein-nmr.org.uk/

[158] http://www.protein-nmr.org.uk/solution-nmr/assignment-theory/

[159] http://www.nmr.chem.uu.nl/~abonvin/tutorials/Assignment-Data/assignment.html

[160] Lee, P. Y., Costumbrado, J., Hsu, C.-Y., & Kim, Y. H. (2012). Agarose Gel Electrophoresis for the Separation of DNA Fragments. *Journal of Visualized Experiments : JoVE*, (62), 3923.

[161] https://www.thermofisher.com/uk/en/home/life-science/dna-rna-purification-analysis/nucleic-acid-gel-electrophoresis/dna-stains/sybr-safe.html

[162] Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., Deng, S.-J. J., Shewchuk, L. M. (2007). Crystallization of protein–ligand complexes. *Acta Crystallographica Section D: Biological Crystallography*, 63(1):72–79.

[163] Müller, I. (2017). Guidelines for the successful generation of protein–ligand complex crystals. *Acta Crystallographica. Section D, Structural Biology*, 73(2):79–92.

[164] Jansen C., Wang H, Kooistra AJ., de Graaf C., Orrling KM., Tenor H., Seebeck T., Bailey D., de Esch IJ., Ke H., Leurs R. (2013) Discovery of novel Trypanosoma brucei phosphodiesterase B1 inhibitors by virtual screening against the unliganded TbrPDEB1 crystal structure. *J Med Chem*. 56(5):2087-96.

[165] Oberholzer M, Marti G, Baresic M, Kunz S, Hemphill A, Seebeck T. (2007) The Trypanosoma brucei cAMP phosphodiesterases TbrPDEB1 and TbrPDEB2: flagellar enzymes that are essential for parasite virulence. *FASEB J*. 21:720–731.

[166] Shakur Y., de Koning HP., Ke H., Kambayashi J., Seebeck T. (2011) Therapeutic potential of phosphodiesterase inhibitors in parasitic diseases. *Handb Exp Pharmacol*. (204):487-510.

[167] Sharron H. Francis, Marco Conti, Miles D. Houslay (2011) Phosphodiesterases as Drug Targets. *Springer* pp499.

[168] Harry P. de Koning, Matthew K. Gould, Geert Jan Sterk, Hermann Tenor, Stefan Kunz, Edith Luginbuehl, and Thomas Seebeck (2012) Pharmacological Validation of Trypanosoma brucei Phosphodiesterases as Novel Drug Targets. *J Infect Dis*. 206(2): 229–237.

[169] Fèvre, E. M., Picozzi, K., Welburn, S. C., Maudlin, I. (2006) Human African trypanosomiasis: epidemiology and control. *Adv. Parasitol*. 61,168-221

[170] Janicak PG et al: Principles and Practice of Psychopharmacotherapy, 2nd Edition, Lippincott, Williams & Wilkins, 1997

[171] Bland ND, Wang C, Tallman C, Gustafson AE, Wang Z, Ashton TD, Ochiana SO, McAllister G, Cotter K, Fang AP, Gechijian L, Garceau N, Gangurde R, Ortenberg R, Ondrechen MJ, Campbell RK, Pollastri MP (2011) Pharmacological validation of Trypanosoma brucei phosphodiesterases B1 and B2 as druggable targets for African sleeping sickness. *J Med Chem*. 54(23): 8188-94.

[172] https://www.ebi.ac.uk/thornton-srv/software/LigPlus/

[173] Antoni R. Blaazer,et al. (2018) Targeting a Subpocket in Trypanosoma brucei Phosphodiesterase B1 (TbrPDEB1) Enables the Structure-Based Discovery of Selective Inhibitors with Trypanocidal Activity. *Journal of Medicinal Chemistry* 61(9):3870-3888

[174] http://www.bio.anl.gov/molecular_and_systems_biology/Sensor/sensor2.html

[175] http://u-of-o-nmr-facility.blogspot.co.uk/2014/03/variable-temperature-to-improve-nmr.html

[176] Shi J, Vakoc CR. (2014) The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Molecular Cell* 54 (5): 728–36.

[177] Itzen F, Greifenberg AK, Bösken CA, Geyer M (2014). Brd4 activates P-TEFb for RNA polymerase II CTD phosphorylation. *Nucleic Acids Research* 42 (12).

[178] Mujtaba S., He Y, Zeng L., Yan S, Plotnikova O., Sachchidanand, Sanchez R., Zeleznik-Le NJ, Ronai Z., Zhou MM. (2004) Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol Cell*. 13(2):251-63

[179] Gabrielle A. Josling, Shamista A. Selvarajah, Michaela Petter and Michael F. Duffy (2012) The Role of Bromodomain Proteins in Regulating Gene Expression. *Genes*. 3:320-343

[180] Das C., Roy S., Namjoshi S., Malarkey CS., Jones DN., Kutateladze TG., Churchill ME., Tyler JK. (2014) Binding of the histone chaperone ASF1 to the CBP bromodomain promotes histone acetylation. *Proc Natl Acad Sci U S A*. 111(12):E1072-81

[181] Jung M., Philpott M., Müller S., Schulze J., Badock V., Eberspächer U., Moosmayer D., Bader B., Schmees N., Fernández-Montalván A., Haendler B. (2014) Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J Biol Chem*. 289(13):9304-19.

[182] http://www.origene.com/Human_ORF/RC218845L2.aspx

[183] Susanne Muller, Panagis Filippakopoulos and Stefan Knapp (2011) Bromodomains as therapeutic targets. *Expert Reviews in Molecular Medicine*. 13(29)

[184] Chiang CM (2009) Brd4 engagement from chromatin targeting to transcriptional revolution: selective contact wth acetylated histone H3 and H4. *F1000 Biol. RFep*. 1(98).

[185] Silviya D. Furdas, Luca Carlino, Wolfgang Sippl and Manfred Jung (2012) Inhibition of bromodomain-mediated protein-protein interactions as a novel therapeutic strategy. *Med. Chem. Commun.* (3) 123-134

[186] Panagis Filippakopoulos and Stefan Knapp (2014) Targeting bromodomains: epigenetic readers of lysine acetylation. *Nature Reviews Drug Discovery*. 13, 337–356

[187] Vollmuth F1, Geyer M. (2010) Interaction of propionylated and butyrylated histone H3 lysine marks with Brd4 bromodomains. Angew Chem Int Ed Engl. 49(38):6768-72.

[188] McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser . *Acta Crystallographica Section D: Biological Crystallography*, 63(1):32–41.

[189] http://www.ccp4.ac.uk/download/#os=windows

[190] Maili Liu et al. (1998) Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy. Journal of Magnetic Resonance. 132 (1): 125-129

[191] https://nmr.chem.ucsb.edu/protocols/nmrPipe.html

[192] Marinescu, V. D., Kohane, I. S., & Riva, A. (2005). The MAPPER database: a multi-genome catalog of putative transcription factor binding sites. *Nucleic Acids Research*, 33: 91-97.

[193] Wishart et al. (1995). $^{1}H$, $^{13}C$ and $^{15}N$ random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbour effects. *J Biomol NMR*, 5, 67-81

[194] Shen, Y. & Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR*, 56: 227

[195] https://spin.niddk.nih.gov/bax/software/TALOS/

[196] https://www.ccpn.ac.uk/meetings/past-meetings/presentations/bill-broadhurst-university-of-cambridge

[197] Carlos M. Cerda-García-Rojas (1990) A PC program for calculation of dihedral angles from $^{1}H$ NMR data. *Tetrahedron Computer Methodology* 3 (2): 113-118

[198] Hollingsworth, S. A., & Karplus, P. A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular Concepts*, 1(3-4), 271–283.

[199] http://knowledgesun.com/ramachandran-plot/

[200] http://www.protein-nmr.org.uk/general/chemical-shifts/chemical-shift-mapping/

[201] Mike P. Williamson (2013) Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* 73:1-16.

[202] Amin, N. T., Wallis, A. K., Wells, S. A., Rowe, M. L., Williamson, R. A., Howard, M. J., & Freedman, R. B. (2013). High-resolution NMR studies of structure and dynamics of human ERp27 indicate extensive interdomain flexibility. *Biochemical Journal*, 450(2), 321–332.

[203] Luis Pablo Calle, Juan Félix Espinosa (2016) An improved 19F-CPMG scheme for detecting binding of polyfluorinated molecules to biological receptors. *Magn. Res. Chem.* 55:355-385

[204] Aramini, J. M., Hamilton, K., Ma, L.-C., Swapna, G. V. T., Leonard, P. G., Ladbury, J. E., Montelione, G. T. (2014). $^{19}$F NMR Reveals Multiple Conformations at the Dimer Interface of the Non-Structural Protein 1 Effector Domain from Influenza A Virus. Structure (*London, England : 1993*), 22(4), 515–525.