



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Rational design of isoform specific ligands

A thesis submitted for the degree of

Doctor of Philosophy

By

Charis Georgiou



THE UNIVERSITY
of EDINBURGH

Abstract

Cyclophilins (Cyp) are proteins that catalyze the interconversion of trans/cis isomers of proline belonging to the peptidyl-prolyl isomerases family (PPIase). In addition to their PPIase activity, Cyps have diverse biological roles and have been implicated in a number of different diseases such as HIV-1 and HCV. Although several Cyp inhibitors have been reported in the literature, none are able to inhibit with high specificity various Cyp isoforms. To facilitate the development of isoform-specific Cyp ligands, we have pursued detailed studies of Cyp dynamics and ligand binding thermodynamics using molecular simulations, biophysical assays and protein X-ray crystallography.

Research efforts were focussed on the identification of novel Cyp inhibitors using X-ray crystallographic studies and Surface Plasmon Resonance (SPR) experiments on fragments from an in-house bespoke library of small compounds. These biophysical studies revealed a number of fragments that are able to bind to diverse Cyp isoforms with high micromolar – low millimolar activity. To further examine the binding of these fragments to cyclophilins, identify interactions with the proteins and explain specificity trends from SPR and X-ray results, molecular dynamics (MD) simulations and free energy calculations were pursued. Models of apo and holo Cyps in complex with fragments that we had experimentally tested were set up using the Amber, AmberTools and FESetup software. Free energy calculations were performed using the thermodynamic integration (TI) technique with the Sire/OpenMM software. The results were analysed with custom scripts. Correlations between computed and measured binding energies, and calculated and observed binding modes were analysed to help develop guidelines for the development of isoform specific cyclophilin ligands.

A detailed comparison of the merits and drawbacks of the experimental and computational techniques used in this work has also been made, and strategies for effective combination of the methodologies in structure-based projects are outlined.

Declaration

The work in this thesis is an original work of the author unless otherwise specified and acknowledged in the thesis. This thesis was submitted from the author for the degree of doctor of Philosophy in the University of Edinburgh, 2016, and had not been submitted in part or in full in any other universities or research body before.

Charis Georgiou

Acknowledgements

First of all I would like to express my gratitude to my supervisors, Prof Malcolm Walkinshaw and Dr Julien Michel for believing in me and giving me the chance to work with them for 4 years. I am grateful for all their guidance, help, support and knowledge that gave to me all these years.

I would like to thank current and past members of 234 office in The school of Chemistry, including and not limited to: George, Remi, Gaetano, Juan, Kevin, Thomas, Doig, Julien, Josh, Stefano, Harris, Jordi, Toni, Alessio, Joan, Lisa, Kanhaya, Cesar, Arun, Georgia and many others for the interesting discussions, support in researching, noisy mornings and afternoons in the office, trips, parties and outings, and for making this experience as pleasant and funny as possible.

I would like to thank each and every one on floor 3 of Swan Building in the ISMB, especially all members of Walkinshaw lab. I would like to thank them for the countless questions that they have answered me, and for teaching a chemist, protein expression, purification and characterisation, SPR and many other biophysical techniques. Special thanks should be given to Dr Iain McNae and Dr Jacqueline Dornan for their valuable help on X-ray crystallography.

Very special thanks should be given to my family. My mother Christia, my father Stelios and my siblings Natasa and Marios, who believe in me, support me and encourage me from the first day of my live. I am grateful for their love and their support throughout the years.

Last but not least, I have to express my love and my gratitude to the light of my live, my wife Pattama Wapeesittipan, for her fathomless love and support every day.

... and glory to God

Abbreviations

3D	Three dimensional
6His-CypA	Cyclophilin A tagged with 6 Histidine residues
6His-CypB	Cyclophilin B tagged with 6 Histidine residues
6His-CypD	Cyclophilin D tagged with 6 Histidine residues
Abu	Abu pocket of cyclophilin
Atm	Atmosphere
C	Concentration
CADD	Computed aided drug discovery
CsA	Cyclosporin A
CV	Coefficient of variation
Cyp	Cyclophilin
Cyps	Cyclophilins
DLS	Dynamic light scattering
DMSO	Dimethyl sulfoxide
EDC	1-ethyl-3-(3-diaminopropyl) carbodiimide hydrochloride
FBDD	Fragment based drug discovery
FDA	Food and drug administration
FEP	Free energy perturbation
FF	Force field
HBA	Hydrogen bond acceptor

HBD	Hydrogen bond donor
H-bond	Hydrogen bond
HCV	Hepatitis C virys
HID	Histidine protonated at the δ nitrogen position
HIE	Histidine protonated at the ϵ nitrogen position
HIS	Histidine
HIV	Human immunodeficiency virus
HMR	Hydrogen mass repartitioning
HTS	High throughput screening
IC ₅₀	Half maximal inhibitory concentration
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K133I CypD	Mutant CypD with Isoleucine instead of Lysine in the 133 amino acid
k_a	Association rate constant
K_a	Equilibrium association constant
k_{cat}	Turnover number
KCl	Potassium Chloride
k_d	Dissociation rate constant
K_d	Equilibrium dissociation constant
K_m	Michaelis constant

LB	L.broth growth medium
LBDD	Ligand based drug discovery
LE	Ligand efficiency
MD	Molecular dynamics
mM	millimolar
MM	Molecular mechanics
MS	Mass spectrometry
MW	Molecular weight
NaCl	Sodium Chloride
NHS	N-hydroxysuccinimide
NMR	Nuclear magnetic resonance
NPT	Isothermal Isobaric ensemble
NTA	Ni ²⁺ -nitrilotriacetic acid
NVT	Canonical ensemble
PBS	Phosphate Buffer Saline
PDB ID	Protein data bank identity
PEG	Polyethylene Glycol
PPIase	Peptidyl prolyl isomerase
Pro	Proline pocket of cyclophilin
QM	Quantum mechanic
R	Response

R/Rfree	Refinement statistic values used in X-ray crystallography
Req	Equilibrium response units
RH	Hydrodynamic radius
Rmax	Maximum response units
rpm	Rounds per minute
RU	Resonance units
SBDD	Structure base drug discovery
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPR	Surface Plasmon Resonance
Std	Standard deviation
TDA	Thermal denaturation assay
TI	Thermodynamic integration
X-ray	X-ray crystallography
ΔG	Binding free energy
ΔG°	Standard state binding free energy
$\Delta\Delta G$	Relative binding free energy

Table of Contents

Abstract	i
Declaration	ii
Acknowledgements	iii
Abbreviations	iv
Table of Contents	viii
Figures	xiv
Tables	xviii
1 Introduction	19
1.1 Aims of this Chapter	19
1.2 Historic overview of drug discovery	19
1.3 Modern drug discovery	20
1.3.1 Structure based drug discovery	22
1.3.1.1 Fragment based drug discovery	25
1.4 Introduction to Cyclophilins	29
1.4.1 Cyclophilin family of proteins	29
1.4.2 Biological role of Cyps	33
1.4.3 Existing Cyp inhibitors and the need of new inhibitors	33
1.5 Project aims and project plan	37
1.6 References	39
2 Selection and preparation of proteins and ligands	46
2.1 Aims of this Chapter	46
2.2 Selection and preparation of proteins	46
2.2.1 Selection of Cyp isoforms	46

2.2.2	Protein expression and purification.....	46
2.2.2.1	Cyp plasmids.....	46
2.2.2.2	Transformation and overexpression of Cyp.....	47
2.2.2.3	Purification of Cyps	48
2.2.3	Protein characterization.....	54
2.2.3.1	SDS-PAGE and protein concentration	54
2.2.3.2	Dynamic light scattering.....	54
2.2.3.3	Thermal denaturation assay	57
2.2.3.4	Enzymatic activity of Cyp	59
2.3	Selection of ligands – the creation of a small compound library.....	62
2.3.1	The hit compound	62
2.3.2	Selection of 98 small compound analogues	63
2.4	Conclusion	65
2.5	References	66
3	Surface plasmon resonance studies on Cyps.....	68
3.1	Introduction and Aims.....	68
3.1.1	Aims of this Chapter	68
3.1.2	Introduction to SPR.....	68
3.1.2.1	Surface Plasmon Resonance	68
3.1.2.2	Kinetics and affinity measurements using SPR.....	70
3.2	Materials and Methods	74
3.2.1	Instrumentation and materials used.....	74
3.2.2	Dilution and Preparation of fragments	76

3.2.3	Single point measurements	76
3.2.4	Dilution series measurements	77
3.2.5	Data processing and fitting.....	77
3.3	Results and Discussion.....	78
3.3.1	Single point measurements	78
3.3.2	Dilution series measurements	82
3.3.2.1	Active compounds and calculated dissociation constants	82
3.3.2.2	Binding stoichiometry of compounds on Cyps	84
3.3.2.3	Selectivity profiles	87
3.4	Conclusions	89
3.5	References	90
4	X-ray crystallography using Cyp complexes	92
4.1	Introduction and Aims.....	92
4.1.1	Aims of this Chapter	92
4.1.2	Introduction to X-ray crystallography.....	92
4.1.2.1	Crystallisation	93
4.1.2.2	Data collection	95
4.1.2.3	Structural Refinement	98
4.2	Materials and Methods	100
4.2.1	Materials.....	100
4.2.2	Methods.....	100
4.2.2.1	Crystallization of CypA	100
4.2.2.2	Crystallization of CypB	101

4.2.2.3	Crystallization of K133I CypD.....	101
4.2.2.4	Co-crystallization of K133I CypD.....	101
4.2.2.5	Soaking of crystals and data collection.....	102
4.2.2.6	Structure determination.....	102
4.3	Results and Discussion.....	103
4.3.1	Crystallization of CypA	103
4.3.2	Binding of compounds in the Abu pocket of CypA.....	104
4.3.2.1	Pyridine, benzene or pyrimidine.....	109
4.3.3	Binding of compounds in the Pro pocket of CypA	113
4.3.4	Crystallization of CypB	115
4.3.5	Crystallization of CypD	117
4.4	Conclusion	119
4.5	References	121
5	Free energy calculations of Cyp complexes.....	123
5.1	Introduction and Aims.....	123
5.1.1	Aims of this Chapter	123
5.1.2	Introduction to Molecular Dynamics (MD).....	123
5.1.2.1	Molecular Dynamics.....	123
5.1.2.2	Molecular Mechanics.....	124
5.1.3	Introduction to free energy calculations.....	125
5.1.3.1	Coupling parameters	128
5.1.3.2	Calculation of relative binding free energy	129
5.2	Materials and Methods	130

5.2.1	Preparation of proteins and ligands.....	130
5.2.2	Docking calculations for complex set up.....	131
5.2.3	Perturbation map and free energy files set up.....	132
5.2.4	Alchemical free energy simulations.....	133
5.2.5	Data processing and fitting.....	135
5.3	Results and Discussion.....	136
5.3.1	Absolute binding free energies of ligand 3	136
5.3.2	Relative binding free energy calculations on CypA	138
5.3.3	Insights in the selectivity profiles of 5 compounds from free energy calculations.....	140
5.3.3.1	Ligand 61	141
5.3.3.2	Ligand 60 and Ligand 98.....	142
5.3.3.3	Ligand 90	144
5.3.3.4	Ligand 4	145
5.4	Interaction of fragments in the Abu and Pro pockets of CypA.....	146
5.5	Conclusion	154
5.6	References	155
6	Discussion, conclusions and future work.....	158
6.1	Aims of this Chapter	158
6.2	Identification of a new class of Cyp inhibitors	158
6.2.1	Aims	158
6.2.2	Results and discussion	159
6.2.3	Future Work	162
6.3	Analysis of the binding of these fragments in the active site of Cyps	162

6.3.1	Aims	162
6.3.2	Results and discussion	163
6.3.3	Future Work	165
6.4	Selectivity profile of fragments binding on Cyps	165
6.4.1	Aims	165
6.4.2	Results and discussion	166
6.4.3	Future Work	169
6.5	Conclusions	170
6.6	References	170
7	Appendix	171
7.1	Sequences of recombinant human Cyps	171
7.2	Purchased compounds	172
7.3	SPR dissociation constants and stoichiometry results	182
7.4	Steady state response curves	185
7.5	Binding of fragments in the active site of CypA.....	206
7.6	Binding of compounds out of the 80's loop of CypA.....	208
7.7	Data collection, refinement and Ramachandran plot statistics for Cyp crystals.....	209
7.8	Cyp X-ray structures used in free energy calculations.....	214
7.9	Perturbation maps for free energy calculations.....	215

Figures

Figure 1-1 Drug discovery at the interface between multiple sciences	21
Figure 1-2 Traditional modern drug discovery process	22
Figure 1-3 The iterative structure based drug discovery process.....	23
Figure 1-4 Comparison between HTS and FBDD	26
Figure 1-5 Comparison between the catalysed and uncatalysed isomerisation of proline	30
Figure 1-6 Amino acid sequence, secondary and tertiary structure of CypA	31
Figure 1-7 3D surface structure of CypA colour coded based on amino acid conservation	32
Figure 2-1 Overexpression of Cyp isoforms	48
Figure 2-2 Sample IMAC elution profiles and SDS-PAGE gels for 6His-Cyps	50
Figure 2-3 Sample SEC elution profiles and SDS-PAGE gels of 6His-Cyps.....	52
Figure 2-4 His-tag cleavage SDS-PAGE gels for CypA, CypB and CypD.....	53
Figure 2-5 Dynamic light scattering analysis of CypA, CypB and CypD	56
Figure 2-6 Sample TDA profile of CypA + CsA and CypA – CsA complexes	58
Figure 2-7 Kinetic mechanism of the PPIase - chymotrypsin coupled assay	59
Figure 2-8 Characteristic PPIase profile of samples with and without CypA	61
Figure 2-9 X-ray structure of 2,3-diaminopyridine binding in the Abu pocket of CypA	63
Figure 2-10 Chemical structures of the 2,3-diaminopyridine hit compound and analogues.....	64
Figure 3-1 Surface plasmon resonance phenomenon.....	69
Figure 3-2 Kinetic parameters of the interaction between two or more components	71
Figure 3-3 Use of steady state fitting for the calculation of binding affinity.....	74
Figure 3-4 Assay and sensor chip set up	75

Figure 3-5 Single point measurement SPR results.....	80
Figure 3-6 Histogram of number of compounds based on their activity on the 3 surfaces of the chip.....	81
Figure 3-7 Correlation between the activity ranking of compounds upon interaction with CypA, CypB and CypD isoforms	82
Figure 3-8 Dilution series SPR results	83
Figure 3-9 Stoichiometry of interaction of active ligands with Cyps	86
Figure 3-10 Histogram of number of active compounds with 1:1 stoichiometry per selectivity profile.....	88
Figure 3-11 Final selected compounds with activity less than 10mM and 1:1 stoichiometry.....	88
Figure 4-1 Hanging and sitting vapour diffusion crystallisation techniques	95
Figure 4-2 Data diffraction pattern of a protein crystal	96
Figure 4-3 Bragg's law with just one atom at each lattice point.....	97
Figure 4-4 Constructive combination of waves incident to different atoms.....	97
Figure 4-5 Fragments that bind CypA based on X-ray crystallography	103
Figure 4-6 Fragments that do not bind CypA based on X-ray crystallography	104
Figure 4-7 Binding of 98 in the Abu pocket of CypA	105
Figure 4-8 Overlay between 1 and 9 in the Abu pocket of CypA.....	106
Figure 4-9 Binding of fragments out of the 80's loop of CypA.....	108
Figure 4-10 Possible binding poses of 3 in the Abu pocket of CypA.....	111
Figure 4-11 Two possible poses of binding of 89 and 91 in the Abu pocket of CypA	113
Figure 4-12 Binding of 16, 60 and 61 in the Pro pocket of CypA.....	114
Figure 4-13 CypB crystals	115

Figure 4-14 Crystal structure of <i>apo</i> CypB solved from an <i>apo</i> crystal of CypB soaked with 97	116
Figure 4-15 K133I CypD co-crystallised with a ligand from Daum et al., 2009.....	118
Figure 4-16 Chemical structure of compound 1 from Daum et al., 2009	119
Figure 5-1 Thermodynamic cycle of free energy.....	127
Figure 5-2 Absolute binding free energy of lig3 to Abu and Pro pockets of CypA, CypB and CypD	137
Figure 5-3 Histogram of calculated and experimental absolute binding free energy of fragments on CypA	139
Figure 5-4 Absolute binding free energies of 61 to the Abu and Pro pockets of Cyp	142
Figure 5-5 Absolute binding free energies of 60 to the Abu and Pro pockets of Cyp	143
Figure 5-6 Absolute binding free energies of 98 to the Abu and Pro pockets of Cyp	144
Figure 5-7 Absolute binding free energies of 90 to the Abu and Pro pockets of Cyp	145
Figure 5-8 Absolute binding free energies of 4 to the Abu and Pro pockets of Cyp	146
Figure 5-9 Histogram of RMSD of 3 from the X-ray crystallographic pose	147
Figure 5-10 RMSD versus time of 3 from the X-ray crystallographic pose.....	149
Figure 5-11 Representative structures of pose A and H-bond distances with key protein and water molecules.....	150
Figure 5-12 Representative structures of pose B and H-bond distances with key protein and water molecules.....	151
Figure 5-13 Representative structure of pose C and H-bond distance with a key protein molecule.....	152

Figure 5-14 Alignment of 3 with compound from (Gelin et al., 2015) in complex with CypA.....	153
Figure 6-1 Venn diagram showing the hit compounds from the three different screening techniques, SPR, X-ray and MD.....	161
Figure 7-1 Steady state response curves	185
Figure 7-2 Binding of compounds 1, 2, 3, 5, 9 and 99 on the Abu pocket of CypA206	
Figure 7-3 Binding of compounds 49, 56, 81, 89, 91, 97 and 98 in the Abu pocket of CypA	207
Figure 7-4 Binding of fragments out of the 80's loop of CypA.....	208

Tables

Table 1-1 Human Cyclophilin family members.....	30
Table 1-2 Percentage sequence identify of all cyclophilins' PPIase domain and comparison of conservation of the residues that define the active side	32
Table 1-3 Compounds reported in the literature as Cyp inhibitors.....	35
Table 2-1 hydrodynamic radius and molecular weight of Cyps as predicted from DLS	55
Table 4-1 Distances between ligands' side chain (R ₂) and protein residues in the 70's and 80's loops	107
Table 4-2 Distances between fragments' side chains (-NH ₂ and -Cl) and protein/water residues	110
Table 6-1 One compound that was "active" in all three techniques	161
Table 7-1 Table of Purchased compounds.....	172
Table 7-2 Table of dissociation constants and stoichiometry results from the dilution series SPR experiment.....	182
Table 7-3 Data collection, refinement and Ramachandran plot statistics for <i>apo</i> CypA CypB and CypD and for CypA in complex with 17 ligands.....	209

1 Introduction

1.1 Aims of this Chapter

The first aim of this chapter is to introduce the process of drug discovery and how this process evolved through the last decade. The process of structure based drug design (SBDD) and fragment based drug design (FBDD) will then be introduced as part of the drug discovery process. The protein family that was studied for the purposes of this research, their function, inhibitors and pharmaceutical significance are subsequently explained. Finally the aims and the project plan of this research will be highlighted.

1.2 Historic overview of drug discovery

The modern drug discovery process started less than 100 years ago. For drug discovery to evolve it was important first for chemistry to mature as a science, pharmacology to become a scientific discipline and analytical chemistry to advance in order to allow the purification of active ingredients. During the 19th century many new theories in these scientific fields evolved and allowed the subsequent evolution of drug discovery (Drews, 2000). Before the 19th century there was not really a drug discovery process. The precursor of drug discovery was mainly based on traditional medicines and natural remedies that were discovered usually by accident.

During the 19th century, Avogadro's atomic theory, the differentiation of compounds in acids, bases and aromatic molecules and the establishment of the periodic table opened the road for new advances in chemistry. Also the benzene theory and the study of coal-tar derivatives, after the industrial revolution, led to the evolution of dye chemistry. The linkage of dyes with medicine came rapidly after Paul Ehrlich discovered that a dye that he was using to stain microscope slides, was able to kill bacteria (Drews, 2000; The Birth of (Synthetic) Dyeing, 2007). Moreover during the same period, at the end of the 19th century, the research of Paul Ehrlich and Langley

Rational design of isoform specific ligands led to drug receptor theory (Drews, 2000; Maehle et al., 2002). This theory revolutionised the field of physiology, the root of pharmacology, and in combination with dye chemistry, opened the way to the modern pharmacology. Furthermore, advances in analytical chemistry came to add new tools in the medicine of the 19th century. The purification and characterisation of the active ingredients from plants and other extracts were not possible until then. The first cases were reported in 1805 with the isolation of morphine from opium extract by F. W. Sertürner (Sertürner, 1805) and in 1848 with the purification of papaverine, again from opium, by Georg Merck (Georg, 1848).

All these advances in chemistry, pharmacology, dye chemistry and analytical chemistry led medicine to a new era in the 20th century, where the active ingredients from natural products could be extracted, tested in whole cells or organisms, and used as drugs. Such an example includes the discovery of the most famous antibiotic to date, penicillin, by Fleming in 1929 (Fleming, 1929). Since then multiple natural products were isolated and purified and used as drugs including ivermectin, lovastatin, Cyclosporin A (CsA) and FK506 (Drews, 2000). During the 20th and 21st centuries the evolution in synthetic chemistry allowed the synthesis of different natural products and libraries of small organic molecules and fragments. At the same time the advances in chemical and molecular biology led to the cloning and synthesis of protein targets and their direct use for compounds screening instead of whole cells or organisms. Finally the successful sequencing of the human genome allowed the development of gene therapy where genes are used as drug targets to treat or prevent diseases.

1.3 Modern drug discovery

Nowadays drug discovery is at the interface of many scientific disciplines (Figure 1-1) including, genomics and proteomics, biology (including structural, molecular and cell biology), synthetic, medicinal and computational chemistry, pharmacology, clinical medicine, biotechnology and many others. All of these disciplines evolved rapidly in the last decades. New techniques and methodologies, products, chemical libraries and

Rational design of isoform specific ligands technologies allow the more efficient sampling of chemical space and synthesis of drug candidates. The easy and fast purification of pharmaceutical targets, the use of computers to facilitate the drug discovery in all steps and the technology to link industry and academia make the drug discovery process more efficient (Drews, 2000; Lombardino and Lowe, 2004).

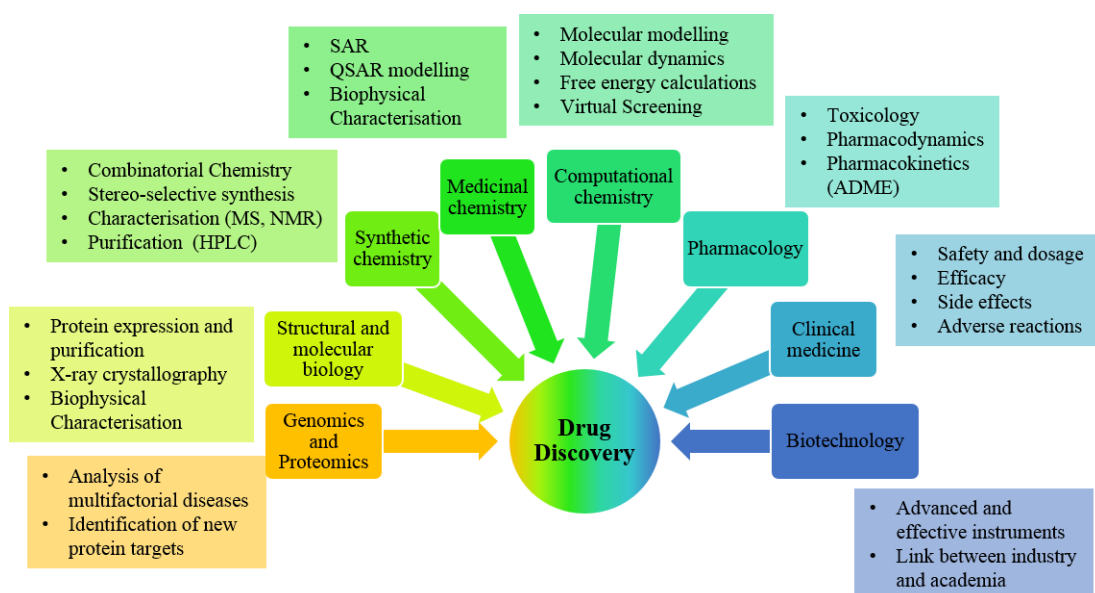


Figure 1-1 Drug discovery at the interface between multiple sciences

The traditional drug discovery process nowadays is very long and very costly. Each drug discovery project in a pharmaceutical industry can last up to 15 years and can cost more than 1 billion euros (DiMasi et al., 2003; Dickson and Gagnon, 2009). Drug discovery is usually divided, broadly, into four steps (Figure 1-2). The first one is the identification and validation of a pharmaceutical target using genomics and proteomics. Once the target is known and validated, the drug design process starts with multiple rounds of compound testing, validation and design. Usually this process starts with thousands of compounds and finishes with only one or two optimised lead compounds.

Rational design of isoform specific ligands

Drug design varies based on what is known for the drug target and the ligands that can be used as possible drug candidates. If the structure of the target is known then the drug design process can be structure based (SBDD) if not, and only the ligands that are inhibiting the target are known, then the process can be ligand based (LBDD).

The third step of drug discovery process is the drug development that involves testing lead compounds in pre-clinical (animals) and three clinical phases (healthy humans and patients). At the end of this process one successful compound that can pass all the tests, may get approval from one or more drug administration agencies and released in the market of the respective agency's jurisdiction.

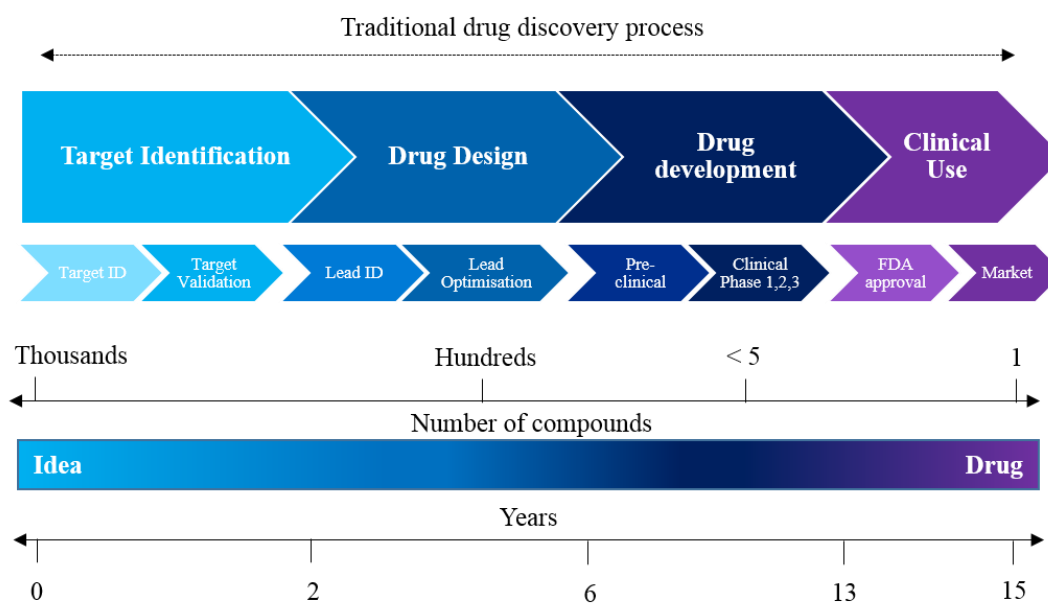


Figure 1-2 Traditional modern drug discovery process

1.3.1 Structure based drug discovery

The SBDD is an iterative process that is repeated multiple times until all or most of the properties of the drug candidates, lead compounds, meet some specific criteria

Rational design of isoform specific ligands based on the guidelines of the drug administration agencies (Anderson, 2003; Lombardino and Lowe, 2004). An example of SBDD process can be seen in the Figure 1-3 below.

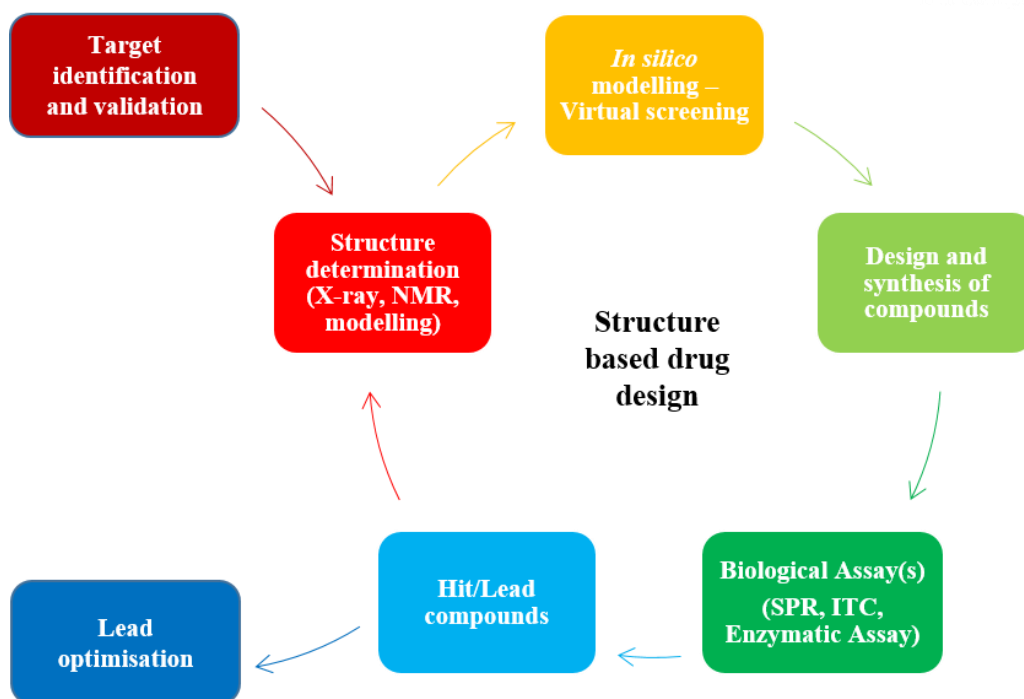


Figure 1-3 The iterative structure based drug discovery process

As was explained above, for the drug design to start, first the biological target should be identified and validated for activity and significance against a specific virus or disease. Structure based drug design starts with the need for a structure of the target protein. This structure can be experimentally observed using NMR or X-ray crystallography. X-ray crystallography is the primary technique used at the moment for the structure determination of biological targets (Blundell et al., 2002; Garman, 2014). Furthermore NMR is gradually becoming more popular as it offers the possibility to study the dynamics of proteins in solution (Pellecchia et al., 2002; Homans, 2004). Homology modelling can be used for a model structure determination, if the experimental observation of the structure is not possible. Homology modeling

Rational design of isoform specific ligands can be effective (Carlsson et al., 2011), especially in the cases where the homolog proteins share similar primary, secondary and tertiary structure with the target protein.

Virtual screening or high throughput screening (HTS) techniques are usually the methods of choice for the first screening of different libraries of compounds. The most active compounds (hit compounds) or the compounds with the higher docking scores (virtual hit compounds) are selected and further tested for activity on the target protein using different biological assays and biophysical measurements, such as Surface Plasmon Resonance (SPR) and Isothermal Titration Calorimetry (ITC) or enzymatic assays. Hit compounds usually have high nM or low μ M activity. X-ray crystallography and/or NMR may be used for the 3D structure determination of the complexes. At the end of each cycle all the results are analyzed to differentiate active from inactive or less active compounds, to identify interactions between the target protein and hit compounds and for the design and synthesis of new compounds. The cycle is then repeated multiple times until the binding activity of the hit compounds on the target proteins is optimized, and the lead compounds with high activity, in the low nM range, and selectivity are identified.

After the lead identification process, or sometimes in parallel with it, a similar process is taking place. This is the lead optimization process and it involves the improvement of the selectivity profile of the lead compounds as well as their ADMETox properties (Absorption Distribution Metabolism Excretion and Toxicology) while maintaining their high activity. This process is cyclic as the lead identification process and involves multiple rounds of design, synthesis and validation of lead compounds.

The choice or the order of different biophysical methods used for the hit/lead identification can change, based on the target protein, the compounds' properties, the project aims and the researcher's expertise. Different techniques give different results. Some of them can have a higher rate of hit compounds (Schiebel et al., 2016) while others give better activity or structure determination (Keseru et al., 2016). As an

Rational design of isoform specific ligands example in Schiebel *et. al.* have shown that the use of X-ray crystallography as the first screening method gives more hit compounds than any of the other six methods that they tested.

Moreover it should be noted that during all phases of the drug discovery process, but especially in the lead identification and optimization phases, computer aided drug discovery (CADD) is a major tool (Ou-Yang *et al.*, 2012). Advances in computing allow the use of different computational software for the identification, secondary structure prediction, folding and modelling of target proteins. Moreover CADD may be used to study the thermodynamics and molecular dynamics of the proteins as well as their binding active sites and allosteric binding sites using molecular dynamic simulations (Durrant and McCammon, 2011). Furthermore computer aided drug discovery can be used for the fast screening of compound libraries for hit identification, the calculation of the free energy binding of different lead compounds on the target protein (Michel *et al.*, 2010) and the prediction of the ADMETox properties of the lead compounds.

1.3.1.1 Fragment based drug discovery

Fragment Based Drug Discovery (FBDD) has now established as an alternative technique to traditional high throughput screening (HTS) for the generation of chemical lead compounds and both are used nowadays from academia and industry in the drug discovery process.

The theory of FBDD was first introduced by Shuker *et al.* and Hajduk *et al.* with their SAR by NMR work (Shuker *et al.*, 1996; Hajduk *et al.*, 1997), while William Jencks in 1981 (Jencks, 1981) described the theory of linking or growing a fragment. According to the theory from Jencks the binding free energy of a ligand AB, with components A and B, can be described from the free energy of binding of each component, ΔG_A and ΔG_B , plus a connection Gibbs energy. So the FBDD process can be described from the Figure 1-4.

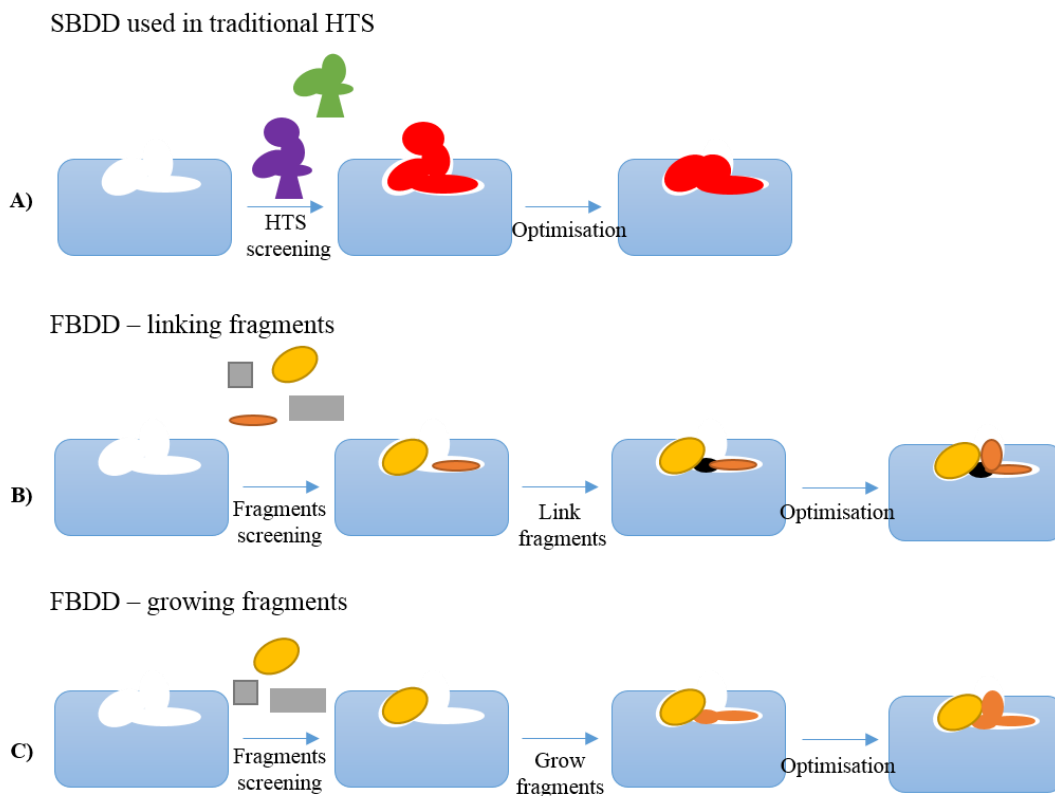


Figure 1-4 Comparison between HTS and FBDD. A) Traditional HTS, B) FBDD – linking fragments and C) FBDD – growing fragments

In this figure, we see that during traditional HTS, researchers usually are using libraries with drug like compounds for hit identification, which are subsequently optimized to improve their free energy of binding. In contrast to HTS, during FBDD researchers are identifying hit compounds from a fragment library and then improve the free energy of binding of these fragments either by linking two or more fragments together (using linker atoms), or by gradually growing one fragment.

The major difference between HTS and FBDD, is the size difference of the ligands used. FBDD uses small molecules (fragments), with low molecular weight, for the identification of drug lead compounds instead of more drug-like molecules that are used in HTS. Fragments can be described from the rule of 3 first described by Congreve *et al.* in 2003. According to this rule, a fragment can be any molecule with

Rational design of isoform specific ligands molecular weight (Mw) \leq 300 Da, number of H-bond acceptors \leq 3, number of H-bond donors \leq 3 and clogP (computed logarithm of compounds partition coefficient between n-octanol and water) \leq 3. On the other hand, usually but not always, researchers are using the Lipinski's rule of 5 to describe a drug like molecule. The rule of 5 was first introduced from Christopher A. Lipinski in 1997 after analyzing the physical and chemical properties of more than 2000 drug and drug candidate compounds. It states that a molecule is more likely to have acceptable absorption or permeation when it has \leq 5 H-bond donors, \leq 10 H-bond acceptors, Mw \leq 500 and cLogP \leq 5 (Lipinski et al., 1997).

Fragments are usually characterized as weak binders, with their binding affinity to be usually around high μ M - low mM, while drug like molecules can have binding affinity down to sub nM range. The problem with "weak" or "strong" binders characterisation is that fragments and drug like molecules don't have the same physical properties, i.e. size or molecular weight. To tackle this issue, researchers came up with a metric, ligand efficiency (LE), which takes in account the binding affinity of the molecule as also as its number of heavy atoms. (Hopkins et al., 2004) LE was first reported from Andrew Hopkins and co-workers and can be calculated from:

$$LE = (\text{ligand's free energy of binding}) / (\text{number of ligand's heavy atoms})$$

Where the free energy of binding is expressed in kilocalories per mole and the number of heavy atoms are all the atoms of the ligand except hydrogens.

The advantages of using FBDD and consequently of small fragments, compared to big drug like molecules, are many and were described before from many researchers (Murray et al., 2012; Hall et al., 2014). Small fragments have low ClogP and MW and can result to lead compounds with better physical properties and LE than lead compounds started from more complex compounds. Also small fragments can cover more efficiently the chemical space instead of more complex compounds and although they have weak potency they can establish high – quality interactions with the protein target (Hall et al., 2014).

Nevertheless dealing with low affinity binders, such as fragments, is not an easy process, since many pitfalls exist. Usually the dissociation constant of small fragments can be up to the mM range, but at this concentration many organic molecules are not soluble. Moreover when an experiment is performed under these conditions the high concentration of fragments can lead to aggregate formation. Aggregates can make the performance of the experiment impossible or even worse they can react with protein molecules giving false positive or negative results. Furthermore the presence of impurities or reactive intermediates in such a high concentrations can lead again to false results.

Because of their low affinity and all the pitfalls related to the use of fragments, it is more challenging to find a technique that can be used for the accurate and precise determination of fragments binding affinity. The selection of the technique used in a FBDD project should be done carefully as it is key for success. In previous years, NMR was one of the most widely used primary techniques, for the identification of fragment binding on proteins and for SAR studies. (Davies, Thomas G.; Hyvönen, 2012; Keseru et al., 2016) This is because NMR has high sensitivity (up to low mM activity) and is a medium throughput technique. “Protein detected NMR” approaches require large quantities of protein and is limited to small proteins (~30 – 40 KDa). On the other hand “ligand detected NMR” requires less protein quantities, but no information about the binding sites of the fragments is obtained and there are significant false positive rates through unspecific binding. (Davies, Thomas G.; Hyvönen, 2012)

SPR is another technique that is widely used as a primary technique for the detection of fragments binding on proteins. SPR is a medium throughput technique as NMR, but with less sensitivity (high μM) and more chances to produce false positives. Moreover a high degree of expertise is needed for the correct set up and run of this experiment.

Rational design of isoform specific ligands

Also other techniques such as ITC, thermal shift assays (TSA), enzymatic or functional assays and mass spectrometry (MS) exist and are increasingly used nowadays in FBDD but usually they are not used as primary techniques but as secondary. (Davies, Thomas G.; Hyvönen, 2012; Keseru et al., 2016) This is because most of them require large quantities of protein (ITC), have very high chance to produce false positives and negatives (TSA and functional assays), and give no structural information of the binding. Furthermore the use of computational methods as described before is increasing continuously in all steps of SBDD but also in the FBDD.

X-ray crystallography technique produce low number of false positives, although some false negatives occur, is highly sensitivity (up to mid mM) and can give structural information. The main downside of X-ray methods, is their lower throughput but nevertheless the throughput of crystallography has greatly increased the last 15 years. (Hartshorn et al., 2005; Davies, Thomas G.; Hyvönen, 2012) Because of these reasons, the X-ray method nowadays gradually becomes more widely used in FBDD. In the last few years, many researchers and journals highlighted the importance of X-ray crystallography in FBDD and multiple successful case studies were published. (Blundell et al., 2002; Hartshorn et al., 2005; Davies, Thomas G.; Hyvönen, 2012; Keseru et al., 2016; Schiebel et al., 2016)

1.4 Introduction to Cyclophilins

1.4.1 Cyclophilin family of proteins

Cyclophilins (Cyps) are a family of proteins that belong to the peptidyl-prolyl isomerases (PPIases) class and are able to catalyze the isomerisation of proline residues (Figure 1-5), promoting and facilitating protein folding (Fischer et al., 1989). Human cyclophilin family counts seventeen members with the archetype and most abundant one being cyclophilin A (CypA) (Davis et al., 2010a). Cyclophilin orthologues can also be found in plants, parasites and animals such as *Plasmodium falciparum* and *C. Elegans* (Bell et al., 2006; Marín-Menéndez et al., 2012).

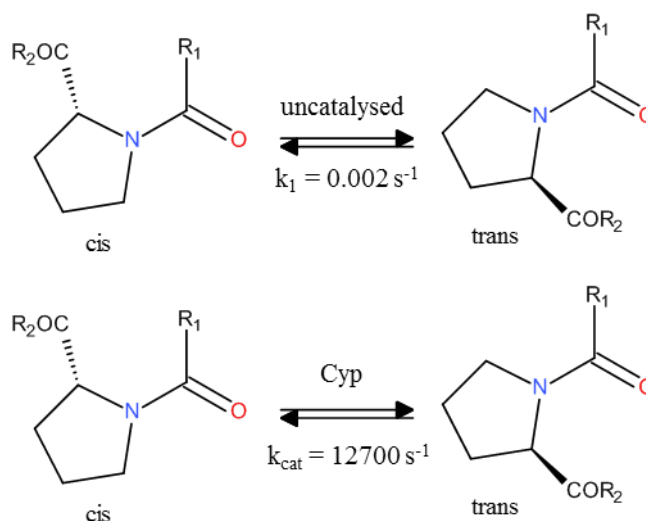


Figure 1-5 Comparison between the catalysed and uncatalysed isomerisation of proline

CypA with some other isoforms have only one domain, the PPIase domain, consisting of a single chain of around 170 amino acids. Others have longer chains and multiple domains, required for their cellular localization or for the protein - protein interactions that they are involved with. The structural characteristics, localization and domains of all human cyclophilins family members can be given in Table 1-1.

Table 1-1 Human Cyclophilin family members

Protein Name (other name)	Size (kDa)	Amino Acids (No)	Localization	Domains (No)
PPIA (CypA)	18.0	165	cytoplasm	PPIase
PPIB (CypB)	23.7	216	endoplasmic reticulum	PPIase / ER
PPIC (CypC)	22.7	212	cytoplasm	PPIase
PPID (Cyp40)	40.7	370	cytoplasm	PPIase / TPR (3)
PPIE (CypE)	33.0	301	nucleus / spliceosome	PPIase / RRM
PPIF (CypD)	22.0	207	mitochondria	PPIase / Mitoch. Dom.
PPIG	88.6	754	nucleus / cytoplasm	PPIase / RS
PPIH	19.2	177	nucleus	PPIase / snRNP
PPIL1	18.2	166	spliceosome	PPIase
PPIL2	58.8	520	nucleus	PPIase / U-Box
PPIL3	18.6	161	spliceosome	PPIase
PPIL4	57.2	492	nucleus / spliceosome	PPIase / RRM
PPIL6	35.2	311	cytoplasm	PPIase
PPIWD1	73.6	646	nucleus	PPIase / WD (4)
NK-TR	165.6	1462	membrane	PPIase / Membr. Dom.
CWC27	53.8	472	spliceosome	PPIase / CC (2)
E3-SUMO	358	3224	nucleus	PPIase/RanBD1 (4)/ RanBD2 zing fingers (8)/ TPR

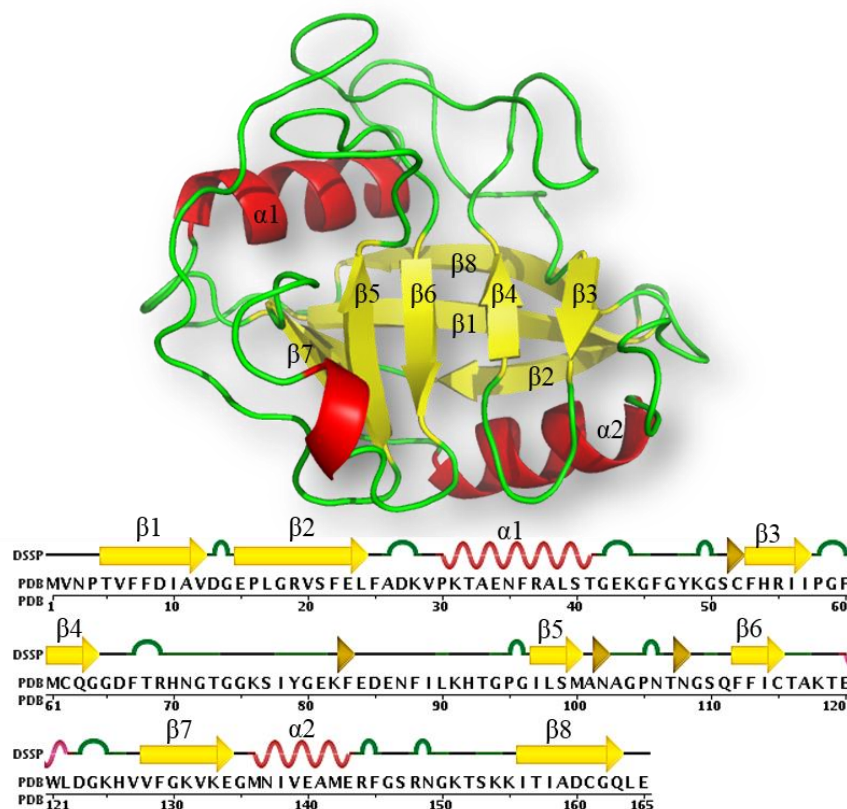


Figure 1-6 Amino acid sequence, secondary and tertiary structure of CypA. β sheets can be seen in yellow arrows, α helices in red spirals, and loops in green. Alignment of sequence and secondary structure was obtained from <http://www.rcsb.org/>.

The crystal structures of thirteen out of seventeen cyclophilins have been determined, with the crystal structures of PPIL4, PPIL6 and E3-SUMO not revealed yet. Nevertheless the amino acid sequences of all cyclophilins are well established as well as their secondary and tertiary structures (Davis et al., 2010a). Cyps have the same secondary and tertiary structure that is a β -barrel consisting of eight anti-parallel β -strands, with two α -helices, one at the top and one at the bottom of the barrel (Figure 1-6). Hydrophobic core residues on four of the strands of the barrel with an α -helical turn form the binding site of cyclophilins. Moreover PPIase domains of all cyclophilins share a very high amino acid sequence identity that is 35 – 76% identical to CypA (Table 1-2 and Figure 1-7).

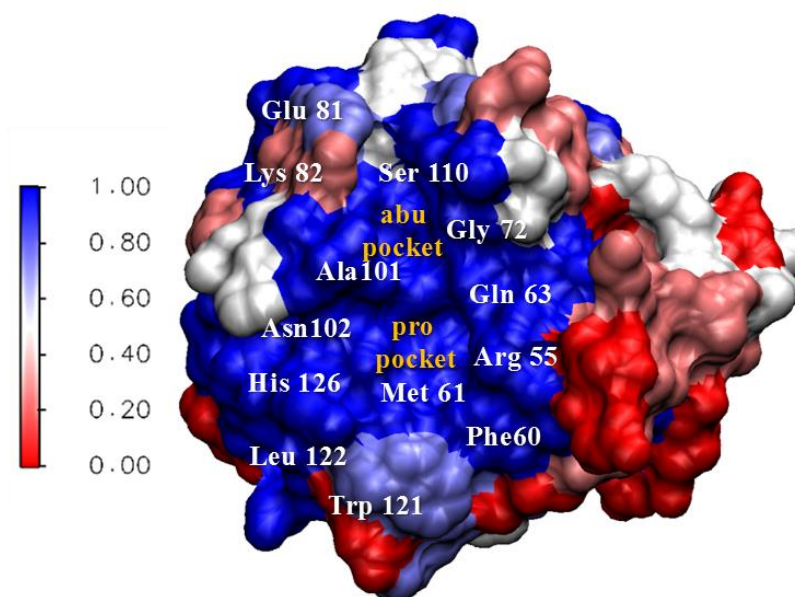


Figure 1-7 3D surface structure of CypA colour coded based on amino acid conservation. Blue and red colours show conserved and non-conserved amino acids respectively.

Table 1-2 Percentage sequence identify of all cyclophilins' PPIase domain and comparison of conservation of the residues that define the active side

Protein name (other name)	Identity to PPIA (%)	Binding site residues of Cyp PPIase domain												
		55	60	61	63	72	81	82	101	102	110	121	122	126
PPIA (CypA)	-	Arg	Phe	Met	Gln	Gly	Glu	Lys	Ala	Asn	Ser	Trp	Leu	His
PPIB (CypB)	64	Arg	Phe	Met	Gln	Gly	Glu	Arg	Ala	Asn	Ser	Trp	Leu	His
PPIC (CypC)	63	Arg	Phe	Met	Gln	Gly	Glu	Thr	Ala	Asn	Ser	Trp	Leu	His
PPID (Cyp40)	60	Arg	Phe	Met	Gln	Gly	Glu	Lys	Ala	Asn	Ser	His	Leu	His
PPIE (CypE)	67	Arg	Phe	Met	Gln	Gly	Lys	Lys	Ala	Asn	Ser	Trp	Leu	His
PPIF (CypD)	76	Arg	Phe	Met	Gln	Gly	Ser	Arg	Ala	Asn	Ser	Trp	Leu	His
PPIG	52	Arg	Phe	Met	Gln	Gly	Gly	Phe	Ala	Asn	Ser	His	Leu	His
PPIH	53	Arg	Phe	Met	Gln	Gly	Gly	Pro	Ala	Asn	Cys	Trp	Leu	His
PPIL1	54	Arg	Phe	Met	Gln	Gly	Lys	Gln	Ala	Asn	Ser	Trp	Leu	His
PPIL2	49	Arg	Phe	Val	Gln	Gly	Lys	Pro	Ala	Asn	Ser	Trp	Leu	His
PPIL3	50	Arg	Phe	Met	Gln	Gly	Lys	Lys	Ala	Asn	Ser	His	Leu	Tyr
PPIL4	36	Asn	Phe	Ile	Gln	Gly	Gly	Leu	Val	Asn	Ser	Tyr	Leu	His
PPIL6	43	Arg	Gly	Met	Gln	Gly	Pro	Thr	Ala	Asn	Ser	Tyr	Leu	Phe
PPWD1	49	Arg	Phe	Met	Gln	Gly	Gly	Glu	Ala	Asn	Ser	Trp	Leu	His
NKTR	50	Arg	Phe	Met	Gln	Gly	Gly	Tyr	Ala	Asn	Ser	Trp	Leu	His
CWC27	43	Arg	Phe	Ile	Gln	Gly	Ala	Pro	Ala	Asn	Ser	Glu	Leu	His
RANBP2	66	Arg	Phe	Val	Gln	Gly	Asp	Lys	Ala	Asn	Ser	Trp	Leu	His

1.4.2 Biological role of Cyps

Both the FKBP family and cyclophilins belong to the immunophilin family of proteins (Takahashi et al., 1989; Dornan et al., 2003; Galat, 2003). This name “immunophilins” derives from their role of blocking the activation of the immune system. Cyps specifically are the *in vivo* receptors of the natural immunosuppressant product cyclosporin A (CsA) (Takahashi et al., 1989). The cyclophilin – cyclosporine complex is commonly used to suppress the activation of T-cells, which is achieved by blocking the action of calcineurin protein through the formation of a cyclophilin – cyclosporine - calcineurin complex. This immunosuppressive activity is used nowadays in organ transplantation to prevent immune response and organ rejection. In addition to their isomerisation and immunosuppressive activity, Cyclophilins are also involved in diverse biological signaling pathways including mitochondrial apoptosis, RNA splicing and different types of cancer but also in the life cycle of different viruses such as Human Immunodeficiency Virus 1 (HIV-1) and Hepatitis C Virus (HCV) (Horowitz et al., 2002; Gaither et al., 2010; Iwasaki, 2012).

More specifically, CypA was reported to be implicated in HCV and HIV infections (Luban et al., 1993; Yang et al., 2008) and different forms of cancer (Yao et al., 2005) as well as neuronal cell death (Dawson et al., 1994; Capano et al., 2002). CypB as CypA was linked to HIV and multiple forms of cancer (Luban et al., 1993; Yao et al., 2005). Moreover mitochondrial CypD (PPIF), which is part of the mitochondrial transition pore, was linked to different neurodegenerative diseases and damage-induced cell death and also to Alzheimer's disease because of its interaction with amyloid beta (A β) (Crompton et al., 1998; Waldmeier et al., 2003; Du et al., 2008; Valasani et al., 2014b).

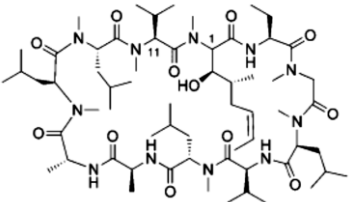
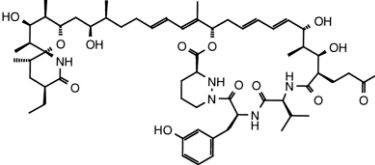
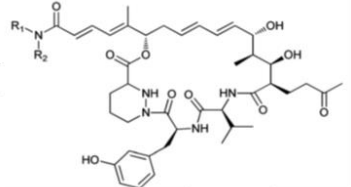
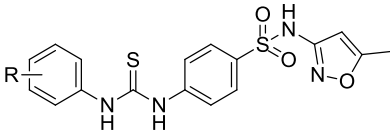
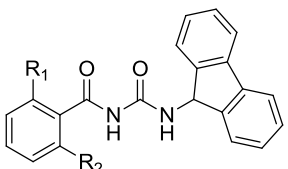
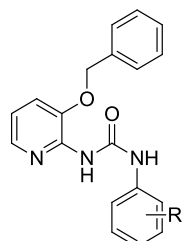
1.4.3 Existing Cyp inhibitors and the need of new inhibitors

Because of their diverse biological role, as described above, cyclophilins are recognized as potential biological targets for the treatment of various diseases,

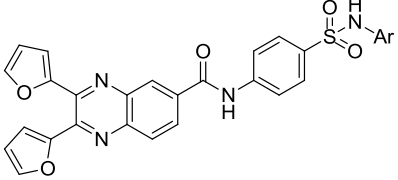
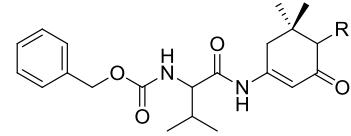
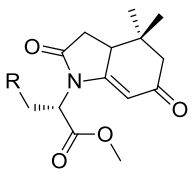
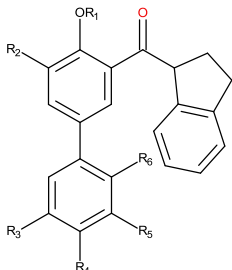
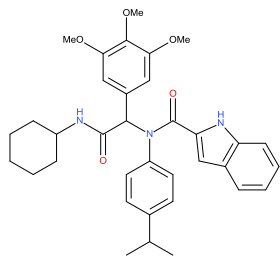
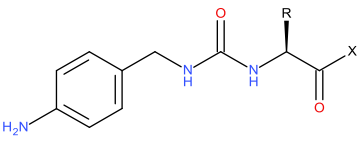
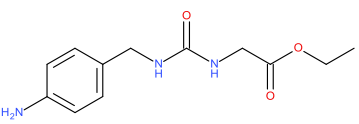
Rational design of isoform specific ligands especially for HCV. Originally the efforts on finding Cyp inhibitors were focused on the natural product inhibitor of Cyp that is CsA (Table 1-3). CsA was firstly discovered back in 1969 in Norway and it is an eleven amino acid cyclic peptide found to be produced naturally from the soil fungus *Tolypocladium inflatum*. It was approved and marketed as drug to be used in organ transplantation by Novartis in 1983 (Svarstad et al., 2000). Since then many researchers have used CsA analogues in order to find new Cyp inhibitors with improved potency, specificity and pharmacokinetics. Some examples include work from A.Scribner *et al.* and M.Walkinshaw *et al.* (Kallen et al., 1998; Scribner et al., 2010). Moreover many pharmaceutical companies, including Scynexis and Novartis, tried to synthesise CsA analogues for the treatment of HCV virus. Compounds such as Alisporivir, NIM811 and SCY-635 were proposed and tested in clinical trials but most of them failed either because of lack of antiviral activity or because of side effects on patients (Gallay, 2012; Hopkins and Gallay, 2012).

Some other macrocycles that were tested as potential anti-HCV drugs are sangamide and different sanglifehrin analogues, most of them showing better antiviral activity and CypA inhibition than CsA, with the best one be Sanglifehrin A (Sanglier et al., 1999; Kallen et al., 2005; Hopkins and Gallay, 2012; Moss et al., 2012). Furthermore, a number of different small molecule Cyp inhibitors have been reported in the literature. Such molecules include compounds with a urea moiety as the central core of the molecule and also several urea analogues such as thiourea or acetyl urea (Guichou et al., 2006; Li et al., 2006; Ni et al., 2009; Chen et al., 2010). Moreover M.Walkinshaw and co-workers, in two articles published on 2007 and 2011, highlighted the effect of dimedone and its derivatives on Cyps and their potential use as Cyp inhibitors (Yang et al., 2007; Dunsmore et al., 2011). Chemical structures of reported small molecule Cyp inhibitors, with their reported K_d and IC_{50} values can be seen in Table 1-3.

Table 1-3 Compounds reported in the literature as Cyp inhibitors

Name	Structure	IC ₅₀ ^{PPIase}	K _d	Selectivity	Primary technique used	Ref.
CsA		K _d = 11.9 ± 2.3 nM	13.4 ± 2.2 nM	No selectivity	Natural product from soil fungi	(Yang et al., 2007)
sanglif ehrin		12.8 nM	IC ₅₀ = 6.9 nm	No selectivity	Natural product from Streptomyces sp.	(Sanglier et al., 1999; Zhang and Liu, 2001)
sanga mide		K _d = 1.1 - 9.6 nM	-	No selectivity	amide derivatives of sanglifehrin A	(Moss et al., 2012)
thioure a derivat ives		0.35 – 0.78 μM	8.24 – 109.5 μM	Not tested	design and synthesis (based on previously published data)	(Chen et al., 2010)
Acylur ea derivat ives		620 – 1.52 nM	-	Not tested	de novo design	(Ni et al., 2009)
Diaryl urea derivat ives		14 – 1400 nM	-	Not tested	virtual screening	(Guichou et al., 2006)

Rational design of isoform specific ligands

Amide linker derivatives		0.25 – 6.43 μM	0.076 – 41 μM	Not tested	combinatorial library design and virtual screening,	(Li et al., 2006)
Dimedone derivatives		$K_d = 22 - 28 \mu\text{M}$	42 – 50 μM	Not tested	In silico screening	(Yang et al., 2007)
Dimedone derivatives		$K_i = 6.8 - 200 \mu\text{M}$	15.9 – 50 μM	Not tested	design and synthesis (based on previously published data)	(Dunsmore et al., 2011)
aryl 1-indanylnketones		$K_i = 0.52 - 21 \mu\text{M}$	-	Selective CypA or CypB inhibition	design and synthesis (based on Pin1 inhibitors)	(Daum et al., 2009a)
bis-amide derivatives		$5.5 \pm 1.6 \text{ nM}$	$570 \pm 20 \text{ nM}$	CypA selectivity over CypB	virtual screening	(Yang et al., 2015)
urea-based derivatives		$K_i = 0.1 - 63 \mu\text{M}$	0.41 – 127 μM	Not tested (designed to inhibit CypD)	design and synthesis (based on previously published data)	(Shore et al., 2016)
urea-linker derivatives		0.01 – 11 μM	-	No selectivity Cyp A,B and D inhibition	virtual screening	(Ahmed-Belkacem et al., 2016)

Despite the large number of compounds reported as successful inhibitors, as reported above, specific inhibition of Cyp isoforms is still a challenging target. Only two research studies until today have claimed some isoform selective inhibition of Cyps using aryl 1-indanylketones and bis-amide derivatives (Daum et al., 2009a; Yang et al., 2015). Specificity in drug design is very important.

For this family of proteins especially, specificity is important not only to avoid side effects from the use of a nonspecific drug molecule, but also for the understanding of the biological role and the signaling pathway that every Cyp isoform is involved in. Achieving high binding specificity is currently considered the most challenging issue for the design of Cyp inhibitors. This is due to the very high degree of similarity of the sequence identity of the active site Cyp residues but also of the tertiary structure, as described before.

1.5 Project aims and project plan

In this FBDD study we are pursuing detailed studies of Cyp dynamics and binding thermodynamics using molecular simulations, biophysical assays and protein x-ray crystallography. Our aims are to:

- I. Identify new small fragments that bind CypA
- II. Carry out detailed studies of CypA - ligand complexes using molecular dynamics (MD) simulations and biophysical experiments.
- III. Analyze intermolecular interactions including hydrogen bonding between the ligand and protein and examine the role played by water. Rationalize the driving forces of ligand binding to CypA
- IV. Examine any differences between ligand binding to CypA and other Cyp isoforms

Experimental research efforts will be focussed on the identification of novel Cyp inhibitors from an in-house bespoke library of small compounds. Initial experiments

Rational design of isoform specific ligands will be focused on the expression, purification and characterisation of Cyp isoforms, as described in Chapter 2. Chapter 2 will also explain the selection and generation of the in-house bespoke library of small compounds. Initially, Surface Plasmon Resonance (SPR) (Chapter 3) and X-ray crystallography (Chapter 4) experiments will be used for the determination of binding affinities and structural binding data respectively. To further examine the binding of any fragments to cyclophilins, identify interactions with the proteins and explain specificity trends from SPR and X-ray results, molecular dynamics (MD) simulations and free energy calculations will be pursued (Chapter 5). Comparison between the outcomes of these experimental and computational techniques will be reported (Chapter 6), and emphasis will be given on opportunities for synergies to facilitate the identification of new Cyp inhibitors.

1.6 References

- Ahmed-Belkacem, A.; Colliandre, L.; Ahnou, N.; Nevers, Q.; Gelin, M.; Bessin, Y.; Brillet, R.; Cala, O.; Douguet, D.; Bourguet, W.; et al. Fragment-Based Discovery of a New Family of Non-Peptidic Small-Molecule Cyclophilin Inhibitors with Potent Antiviral Activities. *Nat. Commun.* **2016**, *7*, 12777.
- Anderson, A. C. The Process of Structure-Based Drug Design. *Chem. Biol.* **2003**, *10* (9), 787–797.
- Bell, A.; Monaghan, P.; Page, A. P. Peptidyl-Prolyl Cis-Trans Isomerases (Immunophilins) and Their Roles in Parasite Biochemistry, Host-Parasite Interaction and Antiparasitic Drug Action. *International Journal for Parasitology*. 2006, pp 261–276.
- Blundell, T. L.; Jhoti, H.; Abell, C. High-Throughput Crystallography for Lead Discovery in Drug Design. *Nat. Rev. Drug Discov.* **2002**, *1* (January), 45–54.
- Capano, M.; Virji, S.; Crompton, M. Cyclophilin-A Is Involved in Excitotoxin-Induced Caspase Activation in Rat Neuronal B50 Cells. *Biochem. J.* **2002**, *363* (Pt 1), 29–36.
- Carlsson, J.; Coleman, R. G.; Setola, V.; Irwin, J. J.; Fan, H.; Schlessinger, A.; Sali, A.; Roth, B. L.; Shoichet, B. K. Ligand Discovery from a Dopamine D3 Receptor Homology Model and Crystal Structure. *Nat. Chem. Biol.* **2011**, *7* (11), 769–778.
- Chen, K.; Tan, Z.; He, M.; Li, J.; Tang, S.; Hewlett, I.; Yu, F.; Jin, Y.; Yang, M. Structure-Activity Relationships (SAR) Research of Thiourea Derivatives as Dual Inhibitors Targeting Both HIV-1 Capsid and Human Cyclophilin A. *Chem. Biol. Drug Des.* **2010**, *76* (1), 25–33.
- Crompton, M.; Virji, S.; Ward, J. M. Cyclophilin-D Binds Strongly to Complexes of the Voltage-Dependent Anion Channel and the Adenine Nucleotide Translocase to Form the Permeability Transition Pore. *Eur. J. Biochem.* **1998**, *258* (2), 729–735.
- Daum, S.; Schumann, M.; Mathea, S.; Aumüller, T.; Balsley, M. A.; Constant, S. L.; De Lacroix, B. F.; Kruska, F.; Braun, M.; Schiene-Fischer, C. Isoform-Specific Inhibition of Cyclophilins. *Biochemistry* **2009**, *48* (26), 6268–6277.

Davies, Thomas G.; Hyvönen, M. *Fragment-Based Drug Discovery and X-Ray Crystallography*; 2012; Vol. 317.

Davis, T. L.; Walker, J. R.; Campagna-Slater, V.; Finerty, P. J.; Finerty, P. J.; Paramanathan, R.; Bernstein, G.; Mackenzie, F.; Tempel, W.; Ouyang, H.; et al. Structural and Biochemical Characterization of the Human Cyclophilin Family of Peptidyl-Prolyl Isomerases. *PLoS Biol.* **2010**, *8* (7).

Dawson, T. M.; Steiner, J. P.; Lyons, W. E.; Fotuhi, M.; Blue, M.; Snyder, S. H. The Immunophilins, FK506 Binding Protein and Cyclophilin, Are Discretely Localized in the Brain: Relationship to Calcineurin. *Neuroscience* **1994**, *62* (2), 569–580.

Dickson, M.; Gagnon, J. P. The Cost of New Drug Discovery and Development. *Discov. Med.* **2009**, *4* (22), 172–179.

DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The Price of Innovation: New Estimates of Drug Development Costs. *J. Health Econ.* **2003**, *22* (2), 151–185.

Dornan, J.; Taylor, P.; Walkinshaw, M. D. Structures of Immunophilins and Their Ligand Complexes. *Curr. Top. Med. Chem.* **2003**, *3* (12), 1392–1409.

Drews, J. Drug Discovery: A Historical Perspective. *Science* (80-.). **2000**, 287 (5460), 1960–1964.

Du, H.; Guo, L.; Fang, F.; Chen, D.; Sosunov, A. A.; McKhann, G. M.; Yan, Y.; Wang, C.; Zhang, H.; Molkenin, J. D.; et al. Cyclophilin D Deficiency Attenuates Mitochondrial and Neuronal Perturbation and Ameliorates Learning and Memory in Alzheimer's Disease. *Nat. Med.* **2008**, *14* (10), 1097–1105.

Dunsmore, C. J.; Malone, K. J.; Bailey, K. R.; Wear, M. A.; Florance, H.; Shirran, S.; Barran, P. E.; Page, A. P.; Walkinshaw, M. D.; Turner, N. J. Design and Synthesis of Conformationally Constrained Cyclophilin Inhibitors Showing a Cyclosporin-A Phenotype in *C. Elegans*. *ChemBioChem* **2011**, *12* (5), 802–810.

Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9*, 71.

Fischer, G.; Wittmann-Liebold, B.; Lang, K.; Kiefhaber, T.; Schmid, F. X. Cyclophilin and Peptidyl-Prolyl Cis-Trans Isomerase Are Probably Identical Proteins. *Nature*.

1989, pp 476–478.

Fleming, A. On the Antibacterial Action of Cultures of a *Penicillium*, with Special Reference to Their Use in the Isolation of *B. Influenzae*. *J. Exp. Pathol.* **1929**, *10* (3), 226–236.

Gaither, L. A.; Borawski, J.; Anderson, L. J.; Balabanis, K. A.; Devay, P.; Joberty, G.; Rau, C.; Schirle, M.; Bouwmeester, T.; Mickanin, C.; et al. Multiple Cyclophilins Involved in Different Cellular Pathways Mediate HCV Replication. *Virology* **2010**, *397* (1), 43–55.

Galat, A. *Peptidylprolyl Cis/trans Isomerases (Immunophilins): Biological Diversity-Targets-Functions.*; 2003; Vol. 3.

Gallay, P. A. Cyclophilin Inhibitors: A Novel Class of Promising Host-Targeting Anti-HCV Agents. *Immunol. Res.* **2012**, *52* (3), 200–210.

Garman, E. F. Developments in X-Ray Crystallographic Structure Determination of Biological Macromolecules. *Science (80-.)*. **2014**, *343* (6175), 1102–1108.

Georg, M. Vorläufige Notiz Über Eine Neue Organische Base Im Opium. *Justus Liebigs Ann. Chem.* **1848**, *66* (1), 125–128.

Guichou, J.-F.; Viaud, J.; Mettling, C.; Subra, G.; Lin, Y.-L.; Chavanieu, A. Structure-Based Design, Synthesis, and Biological Evaluation of Novel Inhibitors of Human Cyclophilin A. *J. Med. Chem.* **2006**, *49* (3), 900–910.

Hajduk, P. J.; Sheppard, G.; Nettlesheim, D. G.; Olejniczak, E. T.; Shuker, S. B.; Meadows, R. P.; Steinman, D. H.; Carrera, G. M.; Marcotte, P. A.; Severin, J.; et al. Discovery of Potent Nonpeptide Inhibitors of Stromelysin Using SAR by NMR. *J. Am. Chem. Soc.* **1997**, *119* (25), 5818–5827.

Hall, R. J.; Mortenson, P. N.; Murray, C. W. Efficient Exploration of Chemical Space by Fragment-Based Screening. *Prog. Biophys. Mol. Biol.* **2014**, *116* (2-3), 82–91.

Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-Based Lead Discovery Using X-Ray Crystallography. *J. Med. Chem.* **2005**, *48* (2), 403–413.

Homans, S. W. NMR Spectroscopy Tools for Structure-Aided Drug Design.

Angewandte Chemie - International Edition. 2004, pp 290–300.

Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: A Useful Metric for Lead Selection. *Drug Discov. Today* **2004**, 9 (10), 430–431.

Hopkins, S.; Gallay, P. Cyclophilin Inhibitors: An Emerging Class of Therapeutics for the Treatment of Chronic Hepatitis C Infection. *Viruses*. 2012, pp 2558–2577.

Horowitz, D. S.; Lee, E. J.; Mabon, S. A.; Misteli, T. A Cyclophilin Functions in Pre-mRNA Splicing. *EMBO J*. **2002**, 21 (3), 470–480.

Iwasaki, A. Innate Immune Recognition of HIV-1. *Immunity*. 2012, pp 389–398.

Jencks, W. P. On the Attribution and Additivity of Binding Energies. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, 78 (7), 4046–4050.

Kallen, J.; Mikol, V.; Taylor, P.; Walkinshaw, M. D. X-Ray Structures and Analysis of 11 Cyclosporin Derivatives Complexed with Cyclophilin A. *J. Mol. Biol.* **1998**, 283, 435–449.

Kallen, J.; Sedrani, R.; Zenke, G.; Wagner, J. Structure of Human Cyclophilin A in Complex with the Novel Immunosuppressant Sanglifohrin A at 1.6 ?? Resolution. *J. Biol. Chem.* **2005**, 280 (23), 21965–21971.

Keseru, G. M.; Erlanson, D. A.; Ferenczy, G. G.; Hann, M. M.; Murray, C. W.; Pickett, S. D. Design Principles for Fragment Libraries – Maximizing the Value of Learnings from Pharma Fragment Based Drug Discovery (FBDD) Programs for Use in Academia. *J. Med. Chem.* **2016**, acs.jmedchem.6b00197.

Li, J.; Zhang, J.; Chen, J.; Luo, X.; Zhu, W.; Shen, J.; Liu, H.; Shen, X.; Jiang, H. Strategy for Discovering Chemical Inhibitors of Human Cyclophilin A: Focused Library Design, Virtual Screening, Chemical Synthesis and Bioassay. *J. Comb. Chem.* **2006**, 8 (3), 326–337.

Lipinski, C. a; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Developmental Settings. *Adv. Drug Deliv. Rev.* **1997**, 23, 3–25.

Lombardino, J. G.; Lowe, J. A. The Role of the Medicinal Chemist in Drug Discovery-Then and Now. *Nat Rev Drug Discov* **2004**, 3 (10), 853–862.

Luban, J.; Bossolt, K. L.; Franke, E. K.; Kalpana, G. V.; Goff, S. P. Human Immunodeficiency Virus Type 1 Gag Protein Binds to Cyclophilins A and B. *Cell* **1993**, *73* (6), 1067–1078.

Maehle, A.-H.; Prüll, C.-R.; Halliwell, R. F. The Emergence of the Drug Receptor Theory. *Nat. Rev. Drug Discov.* **2002**, *1* (8), 637–641.

Marín-Menéndez, A.; Monaghan, P.; Bell, A. A Family of Cyclophilin-like Molecular Chaperones in Plasmodium Falciparum. *Mol. Biochem. Parasitol.* **2012**, *184* (1), 44–47.

Michel, J.; Foloppe, N.; Essex, J. W. Rigorous Free Energy Calculations in Structure-Based Drug Design. *Mol. Inform.* **2010**, *29* (8-9), 570–578.

Moss, S. J.; Bobardt, M.; Leyssen, P.; Coates, N.; Chatterji, U.; Dejian, X.; Foster, T.; Liu, J.; Nur-e-Alam, M.; Suthar, D.; et al. Sangamides, a New Class of Cyclophilin-Inhibiting Host-Targeted Antivirals for Treatment of HCV Infection. *Medchemcomm* **2012**, *3* (8), 938–943.

Murray, C. W.; Verdonk, M. L.; Rees, D. C. Experiences in Fragment-Based Drug Discovery. *Trends Pharmacol. Sci.* **2012**, *33* (5), 224–232.

Ni, S.; Yuan, Y.; Huang, J.; Mao, X.; Lv, M.; Zhu, J.; Shen, X.; Pei, J.; Lai, L.; Jiang, H.; et al. Discovering Potent Small Molecule Inhibitors of Cyclophilin A Using de Novo Drug Design Approach. *J. Med. Chem.* **2009**, *52* (17), 5295–5298.

Ou-Yang, S.-S.; Lu, J.-Y.; Kong, X.-Q.; Liang, Z.-J.; Luo, C.; Jiang, H. Computational Drug Discovery. *Acta Pharmacol. Sin.* **2012**, *33* (9), 1131–1140.

Pellecchia, M.; Sem, D. S.; Wüthrich, K. NMR in Drug Discovery. *Nat. Rev. Drug Discov.* **2002**, *1*, 211–219.

Sanglier, J. J.; Quesniaux, V.; Fehr, T.; Hofmann, H.; Mahnke, M.; Memmert, K.; Schuler, W.; Zenke, G.; Gschwind, L.; Maurer, C.; et al. Sanglifehrins A, B, C and D, Novel Cyclophilin-Binding Compounds Isolated from Streptomyces Sp. A92-308110. I. Taxonomy, Fermentation, Isolation and Biological Activity. *J Antibiot* **1999**, *52* (5), 466–473.

Schiebel, J.; Radeva, N.; Krimmer, S. G.; Wang, X.; Stieler, M.; Ehrmann, F. R.; Fu,

K.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; et al. Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chem. Biol.* **2016**, acschembio.5b01034.

Scribner, A.; Houck, D.; Huang, Z.; Mosier, S.; Peel, M.; Scorneaux, B. Synthesis and Biological Evaluation of [D-lysine]8cyclosporin A Analogs as Potential Anti-HCV Agents. *Bioorganic Med. Chem. Lett.* **2010**, 20 (22), 6542–6546.

Sertürner, F. No Title. *J. der Pharm. für Aerzte, Apotheker und Chem.* **1805**, 13, 229–243.

Shore, E.; Awais, M.; Kershaw, N.; Gibson, R.; Pandalaneni, S.; Latawiec, D.; Wen, L.; Javed, M. A.; Criddle, D.; Berry, N.; et al. Small Molecule Inhibitors of Cyclophilin D to Protect Mitochondrial Function as a Potential Treatment for Acute Pancreatitis. *J. Med. Chem.* **2016**, acs.jmedchem.5b01801.

Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science (80-.)*. **1996**, 274 (5292), 1531–1534.

Svarstad, H.; Bugge, H. C.; Dhillon, S. S. From Norway to Novartis: Cyclosporin from *Tolypocladium Inflatum* in an Open Access Bioprospecting Regime. *Biodivers. Conserv.* **2000**, 9 (11), 1521–1541.

Takahashi, N.; Hayano, T.; Suzuki, M. Peptidyl-Prolyl Cis-Trans Isomerase Is the Cyclosporin A-Binding Protein Cyclophilin. *Nature* **1989**, 337 (6206), 473–475.

Valasani, K. R.; Vangavaragu, J. R.; Day, V. W.; Yan, S. S. Structure Based Design, Synthesis, Pharmacophore Modeling, Virtual Screening, and Molecular Docking Studies for Identification of Novel Cyclophilin D Inhibitors. *J. Chem. Inf. Model.* **2014**, 54 (3), 902–912.

Waldmeier, P. C.; Zimmermann, K.; Qian, T.; Tintelnot-Blomley, M.; Lemasters, J. J. Cyclophilin D as a Drug Target. *Curr. Med. Chem.* **2003**, 10 (16), 1485–1506.

Yang, F.; Robotham, J. M.; Nelson, H. B.; Irsigler, A.; Kenworthy, R.; Tang, H. Cyclophilin A Is an Essential Cofactor for Hepatitis C Virus Infection and the Principal Mediator of Cyclosporine Resistance in Vitro. *J. Virol.* **2008**, 82 (11), 5269–5278.

Yang, S.; K R, J.; Lim, S.; Choi, T. G.; Kim, J.-H.; Akter, S.; Jang, M.; Ahn, H.-J.;

Kim, H.-Y.; Windisch, M. P.; et al. Structure-Based Discovery of Novel Cyclophilin A Inhibitors for the Treatment of Hepatitis C Virus Infections. *J. Med. Chem.* **2015**, *58* (24), 9546–9561.

Yang, Y.; Moir, E.; Kontopidis, G.; Taylor, P.; Wear, M. A.; Malone, K.; Dunsmore, C. J.; Page, A. P.; Turner, N. J.; Walkinshaw, M. D. Structure-Based Discovery of a Family of Synthetic Cyclophilin Inhibitors Showing a Cyclosporin-A Phenotype in *Caenorhabditis Elegans*. *Biochem. Biophys. Res. Commun.* **2007**, *363* (4), 1013–1019.

Yao, Q.; Li, M.; Yang, H.; Chai, H.; Fisher, W.; Chen, C. Roles of Cyclophilins in Cancers and Other Organ Systems. In *World Journal of Surgery*; 2005; Vol. 29, pp 276–280.

Zhang, L. H.; Liu, J. O. Sanglifehrin A, a Novel Cyclophilin-Binding Immunosuppressant, Inhibits IL-2-Dependent T Cell Proliferation at the G1 Phase of the Cell Cycle. *J. Immunol.* **2001**, *166* (9), 5611–5618.

The Birth of (Synthetic) Dyeing <http://www.open.edu/openlearn/history-the-arts/history/history-science-technology-and-medicine/history-science/the-birth-synthetic-dyeing> (accessed Jun 20, 2006).

2 Selection and preparation of proteins and ligands

2.1 Aims of this Chapter

This chapter aims to explain the selection of Cyp isoforms and to describe the techniques used for the protein expression, purification and characterisation as well as to show some representative results from these experiments. The second aim of this chapter is to describe how the library of small ligands, which was used in this research, was generated.

2.2 Selection and preparation of proteins

2.2.1 Selection of Cyp isoforms

As was explained in the introduction, CypA is the prototype of Cyp family of proteins and the most abundant one. Moreover, CypA alongside CypB and mitochondrial CypD appear to be to date the proteins with the highest biological significance. This is because all of them were recognized to play a major role in the life cycle of many viruses and also to have a key role in many other diseases as described in the introduction. Furthermore CypA, CypB and CypD appeared to have slight differences in their surface electrostatic potential, as described from Davis et al. (Davis et al., 2010b) and they also have very well established production and purification protocols (Wear et al., 2005). For these reasons CypA, CypB and CypD isoforms were selected, out of the 17 Cyp family members, for this study, as our primary biological targets.

2.2.2 Protein expression and purification

2.2.2.1 Cyp plasmids

Plasmids for recombinant human N-terminal hexa-histidine tagged 6His-CypA, 6His-CypB and 6His-CypD were a kind donation from the Edinburgh Protein Production Facility (EPPF) and were prepared as previously described by Wear et al (Wear et al., 2005). 6His-K133I-CypD was a kind donation from Dr Jacqueline Dornan from Prof

Rational design of isoform specific ligands
Malcolm Walkinshaw's group. The sequence for all of proteins used in the experimental section of this study can be seen in Appendix 7.1.

2.2.2.2 Transformation and overexpression of Cyp

Recombinant human Cyps were expressed in *Escherichia coli* C43 (DE3) and C41 (DE3) competent cells from Lucigen. Transformation of *E.Coli* cells was achieved using a similar protocol to the standard transformation protocol from the manufacturer (Lucigen Corporation, 2014). Briefly, 1 – 2 μ L of stock Cyp isoforms plasmids were added to C41 (DE3) or C43 (DE3) competent cells and left in ice for 30 minutes. Cells were heat shocked at 42 °C for 45 seconds and then placed back in ice for another 2 minutes. 500 μ L of L.broth or SOC medium was added to the cells and the cells were incubated at 37 °C for 45 minutes shaking at 250 rounds per minute (rpm). 200 μ L aliquots of cells were plated on agar plates, containing L.Broth (LB) medium and carbenicillin antibiotic 100 μ g ml⁻¹, and were incubated overnight at 37 °C.

Expression of the recombinant Cyp isoforms using the previously transformed *E.Coli* cells was achieved using the standard transformation protocol from BioSilta (BioSilta Ltd, 2014). A single colony from the freshly transformed cells was grown in 2ml of LB media containing 100 μ g ml⁻¹ of carbenicillin at 37 °C for 6 hours with shaking at 250 rpm. 160 μ L of pre-cultured cells were then added to a 500ml round bottom shaking flask containing 50ml sterile water, 2 tablets from one white bag EnPresso media, 25 μ L Reagent A (final concentration 1.5 U/L) and 100 μ g ml⁻¹ of carbenicillin and incubated overnight at 30 °C shaking at 250 rpm. 1mM of IPTG (isopropyl-thiogalactoside) induction agent, 25 μ L Reagent A and a tablet from one black bag EnPresso media was added to the flask and cells were further incubated at 30 °C shaking at 250 rpm for another 24 hours. Cultures were centrifuged at 4500 rpm for 20 min at 6 °C. Supernatant was removed and cultures were flash frozen with liquid nitrogen and saved at -80 °C until further use.

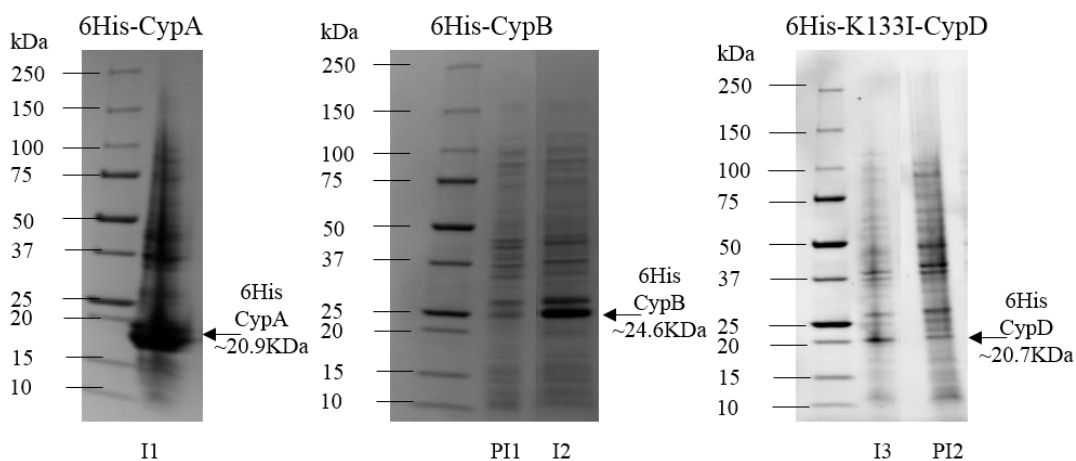


Figure 2-1 Overexpression of Cyp isoforms. SDS-PAGE denaturing gel showing the overexpression of 6His-CypA, 6His-CypB and 6His-K133I-CypD. Columns **I1**, **I2** and **I3** show soluble cell fractions from the induced CypA, CypB and CypD respectively. **PI1** and **PI2** show soluble cell fractions from CypB and CypD respectively. The molecular weight marker from BIO-RAD was used and can be seen in the first column of each gel.

2.2.2.3 Purification of Cyps

Purification of Cyp isoforms was achieved following a similar protocol to the previously published protocol for CypA purification from Martin et al. (Wear et al., 2005). The protocol involves two steps. The first is affinity column chromatography and the second is size exclusion chromatography as described below. Before the first step of Cyp purification, cell pellets were defrosted and re-suspended in Buffer A consisting of 20 mM NaH₂PO₄, 500 mM NaCl, 20 mM imidazole at pH 7.4, plus protease inhibitors cocktail tablets (cOmplete) from Roche. Cells were then lysed with a single passage from a constant cell disruption system (from Constant Systems Ltd) at 22K psi. Cell extract was centrifuged at 50000 g at 4 °C for 1 hour and the supernatant was collected and filtered through a 0.22µm filter before purification. All purification steps follow were performed on an AKTA FPLC (fast protein liquid chromatography) equipment from GE Healthcare at 4 °C.

2.2.2.3.1 Affinity chromatography

Affinity chromatography was performed on a HiTrap IMAC HP 1ml column from GE Healthcare as described in manufacturer's manual (GE Healthcare Bio-Sciences). IMAC (Immobilized Metal Ion Affinity Chromatography) columns are prepacked with Sepharose and are ideal for separation of proteins with histidine tag because of the high affinity of the sepharose and the his tag for the metal ions.

Before purification IMAC column was charged with Ni²⁺ ions, washed and pre equilibrated with buffer A as described in the manual. After loading the cell extract on the column, the column was washed with 20 column volumes (CV) of buffer A, followed by 20 CV of buffer A plus 40mM of imidazole. Protein was eluted using a gradient step of 20 CV, with the buffer changing from Buffer A + 40mM imidazole to Buffer B (Buffer A + 100mM imidazole). Lastly any remaining protein was washed off the column using a further 10 CV wash with Buffer B. Imidazole has a high affinity for the metal ions, and in high concentrations can replace and release the histidine tagged proteins from the Ni²⁺ ions on the column. Protein was eluted as a single peak, collected in 1ml fractions and analyzed with SDS-PAGE. Selected fractions were mixed and spin concentrated, using a Vivaspin with 5KDa exclusion limit, before size exclusion chromatography.

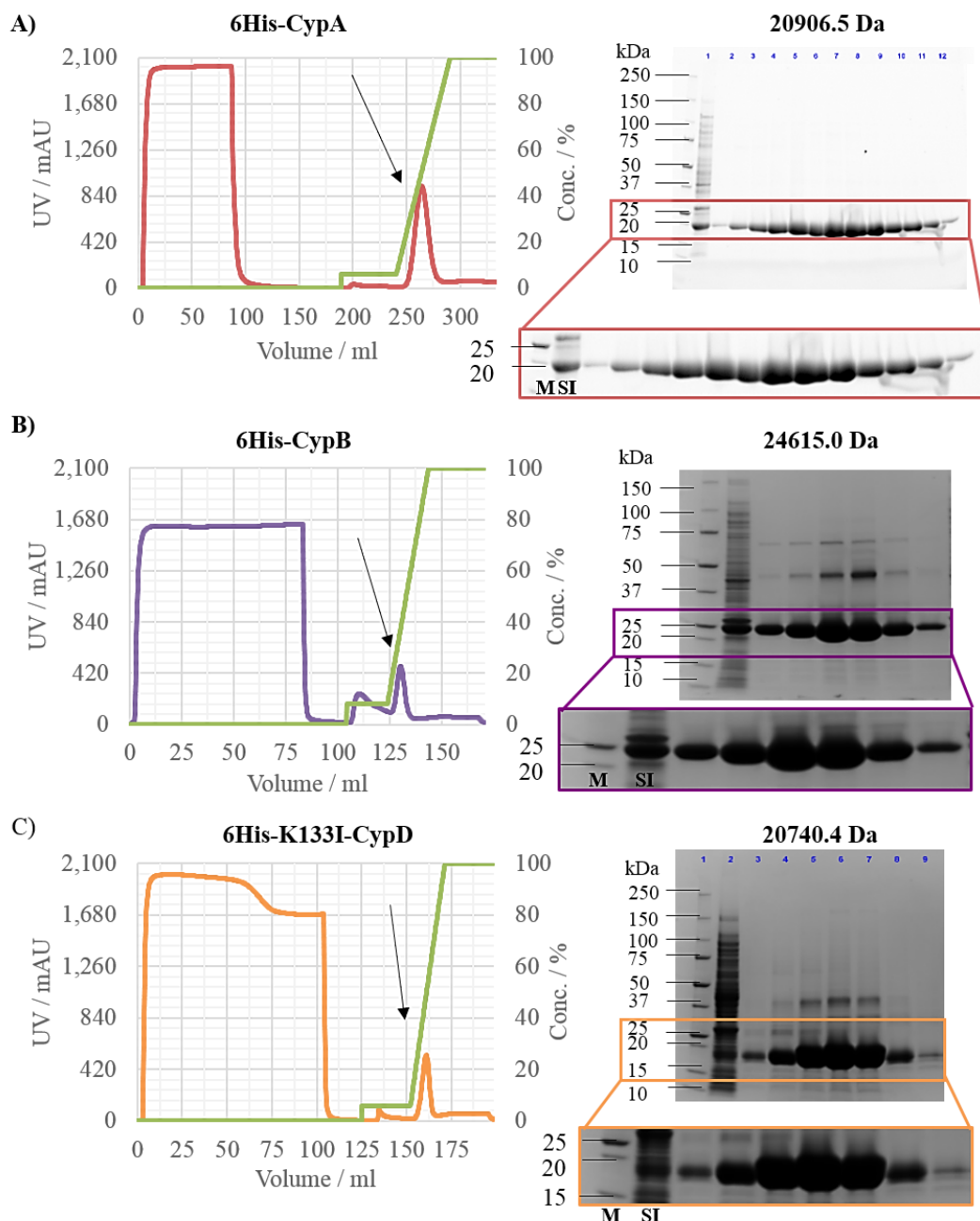


Figure 2-2 Sample IMAC elution profiles and SDS-PAGE gels for 6His-Cyps. A) Elution profile of CypA shown in red, and SDS-PAGE of selected fractions with CypA band at 20.9kDa, B) Elution profile of CypB shown in violet, and SDS-PAGE of selected fractions with CypB band at 24.6kDa. C) Elution profile of CypD shown in orange, and SDS-PAGE of selected fractions with CypD band at 20.7kDa. In all cases Cyps usually elute at ~30% buffer B concentration (shown in green line). In all SDS-PAGE gels above, **M** lane shows the molecular weight marker, **SI** lane shows the sample loaded on the IMAC column and the rest show selected fractions from the single Cyp peak of the chromatogram.

2.2.2.3.2 Size exclusion chromatography

Selected fractions from IMAC were loaded on a HiLoad 26/60 Superdex 75 prep grade column (GE Healthcare Bio-Sciences, 2011). The chromatographic beads of this Superdex column allow the separation of proteins based on their size. This is because smaller proteins and molecules are able to interact and enter in the pores of the beads, making them diffuse slowly, and their retention time increases. The bigger the size of the protein is, the smaller the interaction with the beads and the smaller the retention time is.

The column was pre-equilibrated in PBS, 1 mM DTT, pH 7.4 and pure monomeric Cyp was eluted as a single peak usually after 0.55 – 0.65 CV as can be seen from the Figure 2-3. For the elution of the protein the same buffer, PBS, 1 mM DTT, pH 7.4, was used as for the equilibration of the column. Eluted protein was collected in 1ml fractions and analyzed with SDS-PAGE. Selected fractions were mixed and spin concentrated, using a Vivaspin with 5KDa exclusion limit before any further experiments.

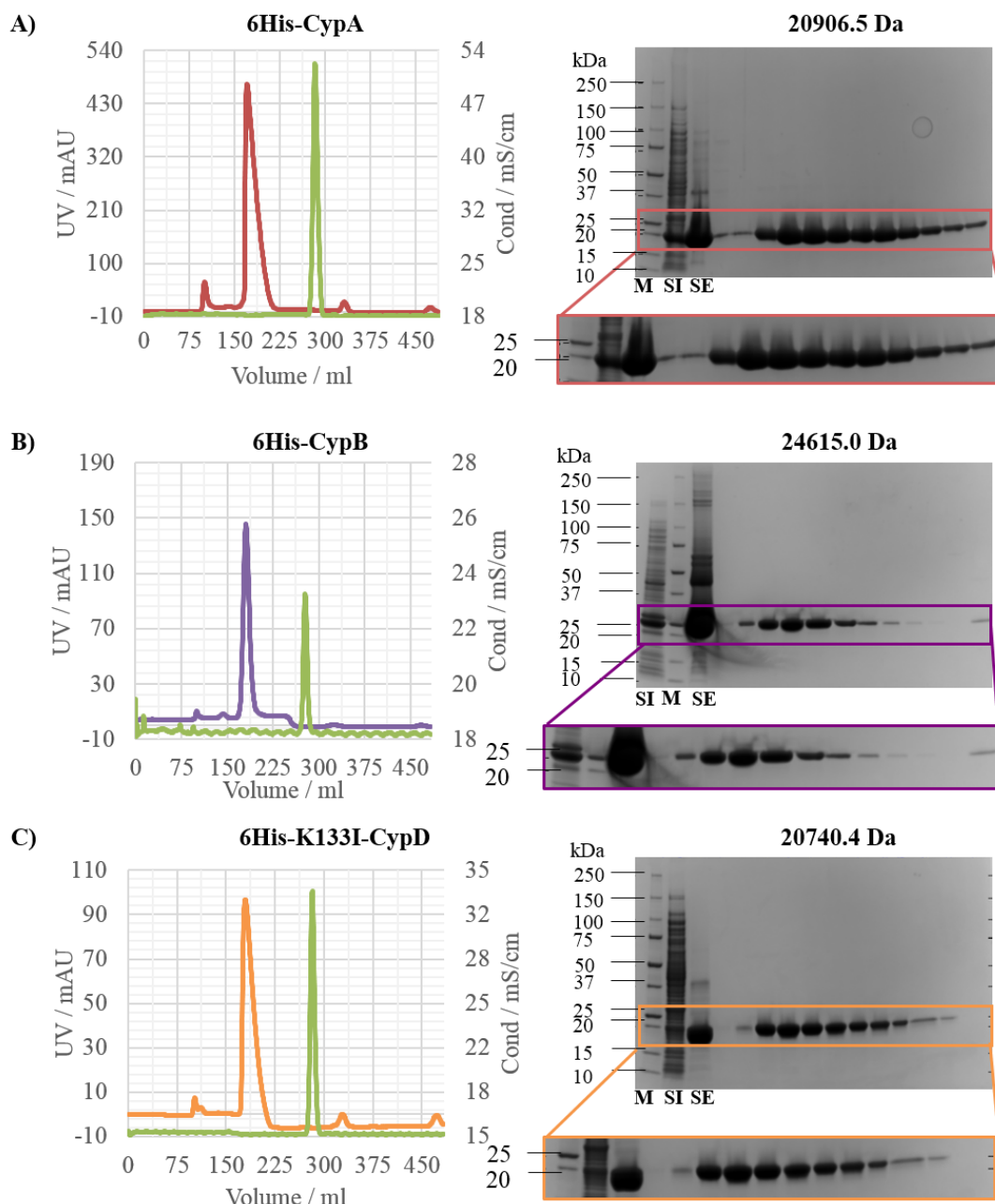


Figure 2-3 Sample SEC elution profiles and SDS-PAGE gels of 6His-Cyps. **A)** Elution profile of CypA shown in red, and SDS-PAGE of selected fractions with CypA band at 20.9kDa, **B)** Elution profile of CypB shown in violet, and SDS-PAGE of selected fractions with CypB band at 24.6kDa. **C)** Elution profile of CypD shown in orange, and SDS-PAGE of selected fractions with CypD band at 20.7kDa. Eluent conductivity in all chromatograms is shown with a green line, and in all cases Cyps elute as a single peak usually after 0.55 – 0.65 CV. In all SDS-PAGE gels above, **M** lane shows the molecular weight marker, **SI** lane shows the sample loaded on the IMAC and the **SE** lane the sample loaded on the size exclusion column while the rest show selected fractions from the single Cyp peak of the chromatogram.

2.2.2.3.3 His tag cleavage of Cyps

The hexa-histidine tag from the purified 6His-Cyp isoforms was cleaved before some biophysical experiments. 6His-Cyps were buffer exchanged, using a Vivaspin with 5KDa exclusion limit, using a cleavage buffer consist of 100 mM Tris, 100 mM NaCl, 5 mM DTT and 1mM EDTA, pH 7.5 and were incubated with TEV (Tobacco Etch Virus) protease (100 ng TEV per 50 μ g of protein) at 30 °C, shaking at 50 rpm, for 4 hours. Proteins were purified using a further step of IMAC to obtain the purified his tag cleaved Cyps, Figure 2-4. Before any further biophysical experiment, Cyps were buffer exchanged to the appropriate buffer using either a Vivaspin with 5KDa exclusion limit or a manual PD10 rapid desalt column from GE Healthcare.

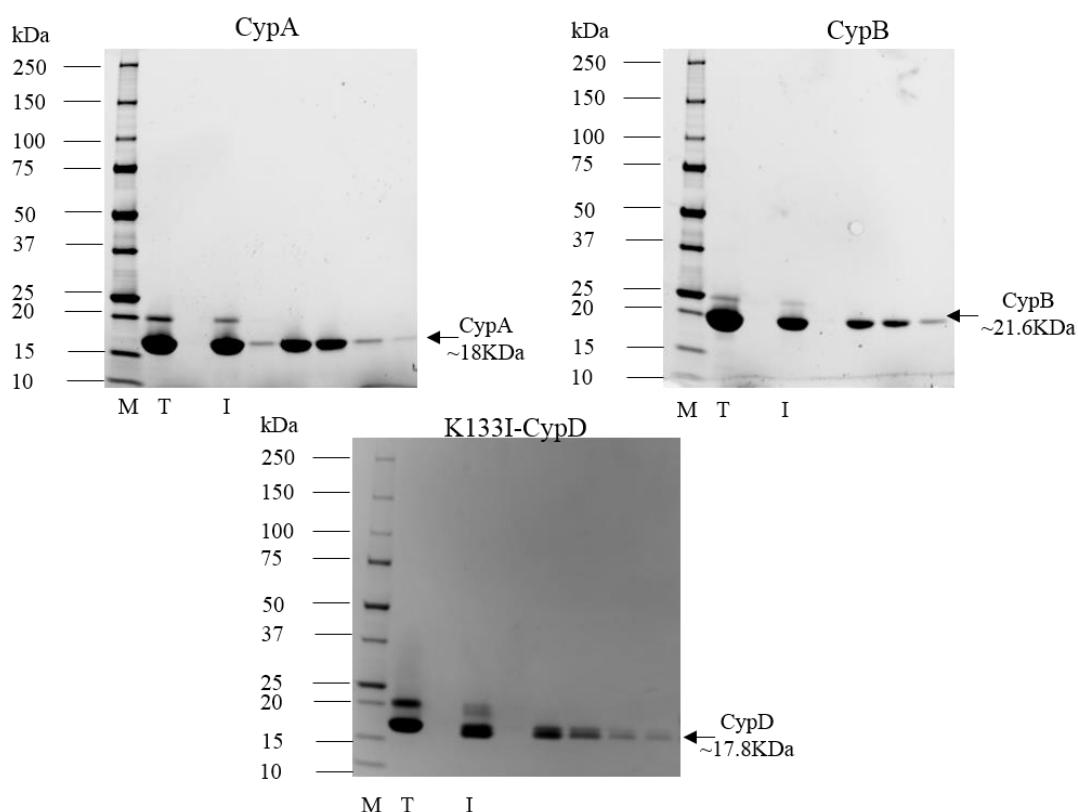


Figure 2-4 His-tag cleavage SDS-PAGE gels for CypA, CypB and CypD. M lane shows the molecular weight marker, lane T shows the protein after 4 hours incubation with TEV and lane I shows the protein sample loaded on IMAC column after a further hour incubation with TEV. The rest lanes show selected fractions after IMAC purification.

2.2.3 Protein characterization

2.2.3.1 SDS-PAGE and protein concentration

Protein samples in all stages of protein expression, production, purification and His tag cleavage, were analyzed for purity using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). SDS-PAGE analysis was performed using precast gels and the mini protean tetra cell system from BIO-RAD, as the manufacturer's protocol describes (BIO-RAD, 2014). Gels were stained using the InstantBlue stain from Expedeon (Expedeon, 2014).

The quantity and concentration of protein at each stage was determined using the NanoDrop Lite Spectrophotometer from Fisher Scientific as explained in the manufacturer's manual (Scientific, 2012). CypA, CypB and CypD are usually expressed very well and their purification is very easy. On average from our experiments, at the end of both purification steps, one pellet, i.e. 50mL in EnPresso media culture, of CypA, CypB and CypD yields 15, 7 and 3 mg of pure his tagged protein respectively. Usually His tag cleavage is also straight forward and on average more than 80% of the His-tagged protein used is collected at the end of this step as a pure his tag cleaved protein.

2.2.3.2 Dynamic light scattering

To verify the quaternary structure of the purified proteins and to examine any possible aggregation between Cyp monomers the dynamic light scattering (DLS) technique was used. The hydrodynamic radius of a protein is related to the intensity fluctuations of laser light that is scattered from protein molecules in solution (Wyatt, 1993). DLS was performed on a Zetasizer APS instrument from Malvern using a 384 well plate with the total final volume of each well used to be 60 μ L. Prior DLS protein samples were buffer exchanged to PBS, concentrated to 1mg ml⁻¹ and spun gently at 13,500 g for 20 minutes at 4 °C. DLS was performed in triplicates for each protein sample at 20 °C and as can be seen from the Figure 2-5, measurements are repeatable with the

Rational design of isoform specific ligands correlation between them high. All protein samples are monodisperse and there are no aggregates in the samples. The hydrodynamic radius (R_H) for each protein was also determined as can be seen in the Figure 2-5 and Table 2-1 below. R_H has the right magnitude for all Cyps and as expected is around $\sim 4.3\text{nm}$ for CypA and CypD and for CypB is slightly higher since CypB has a slightly longer chain. Because of the high quality of the protein samples, the DLS experiment were also able to predict the MW of the proteins. All these data from the DLS experiment show that our protein samples are monodisperse, properly folded and they have the proper R_H and MW.

Table 2-1 hydrodynamic radius and molecular weight of Cyps as predicted from DLS

Protein	Experimental R_H (d.nm)	Experimental MW (KDa)	Theoretical MW (KDa)
CypA	4.3 ± 1.7	20.9 ± 8.3	20.9
CypB	4.6 ± 1.8	24.1 ± 9.8	24.6
CypD	4.3 ± 1.1	20.8 ± 5.3	20.7

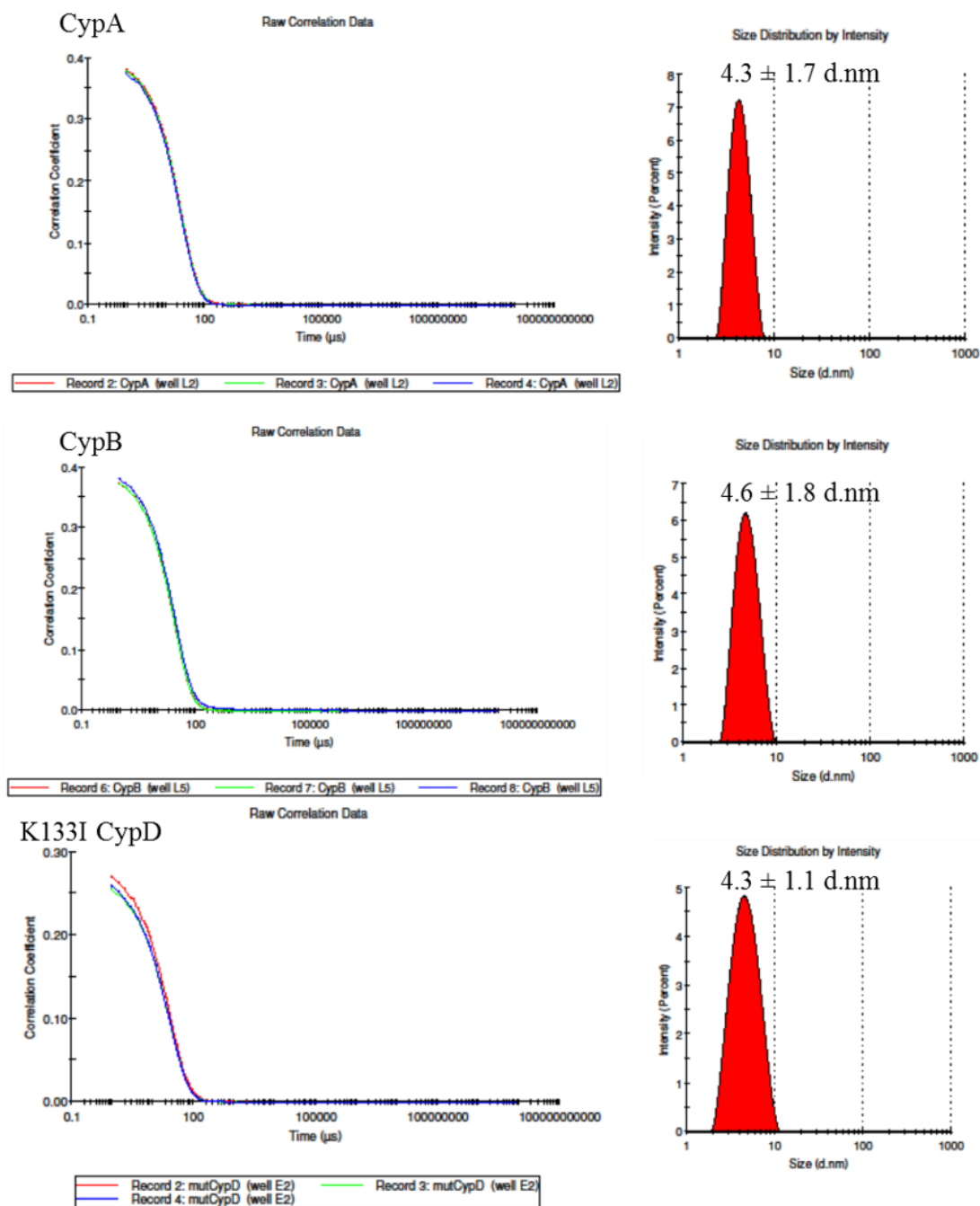


Figure 2-5 Dynamic light scattering analysis of CypA, CypB and CypD. Left figures show the correlation between three repeat experiments, blue, green and red lines, for Cyp isoforms. On the right column we can see the size distribution versus intensity for each isoform. As can be seen correlation between repeats is perfect and protein samples are monodisperse.

2.2.3.3 Thermal denaturation assay

To further characterize CypA protein and to test whether the protein was folded and active, a thermal denaturation assay was used. Thermal denaturation assay is a technique that monitors the unfolding of proteins induced from the increase of temperature. This is achieved with the use of a fluorescent dye, SYPRO (Steinberg et al., 1996). SYPRO, in a hydrophobic environment, has the ability to absorb light at $\lambda_{ex} = 490$ nm and emit light at $\lambda_{em} = 590$ nm. Denaturation of proteins, in response to a temperature increase, expose the highly hydrophobic cores of the proteins, increase the hydrophobicity of the environment and induce an increase in emitted light at $\lambda_{em} = 590$ nm from the fluorescent dye SYPRO (Pantoliano et al., 2001). This technique allows the monitoring of protein unfolding in real time and the effects of buffer changes or any inhibitors, making this technique very useful in drug discovery (Epps et al., 2001; Lo et al., 2004).

TDA was performed on Bio-Rad iCycler IQ instrument with an iQ5 real time detection system using a 96 PCR well plate sealed with iCycler IQ optical tape from Bio-Rad. The total volume of the analyte was 50 μ L and consisted of 5x SYPRO orange (from Invitrogen) and 5 μ M of CypA in PBS buffer. CsA when used was screened at 5 μ M final concentration. Experiments were performed between 20 – 70 °C with a 0.5 °C increment and a 30 seconds hold between increments. Three different measurements were performed: 1) one control without any protein or ligand, just buffer and SYPRO, 2) CypA and SYPRO in buffer and 3) a control with a known ligand, in this case is CsA.

As can be seen from the Figure 2-6 below, measurements were repeated in duplicate and the results were the same. In both cases the melting temperature of CypA was at 51 °C but when CsA was added the melting temperature was shifted by 5 °C to 56 °C. This increase in melting point after addition of CsA shows the stabilisation of the tertiary structure of CypA by CsA. CsA binds in the active site of CypA and delays its

Rational design of isoform specific ligands unfolding. Moreover this binding of CsA on CypA shows the correct folding of the protein.

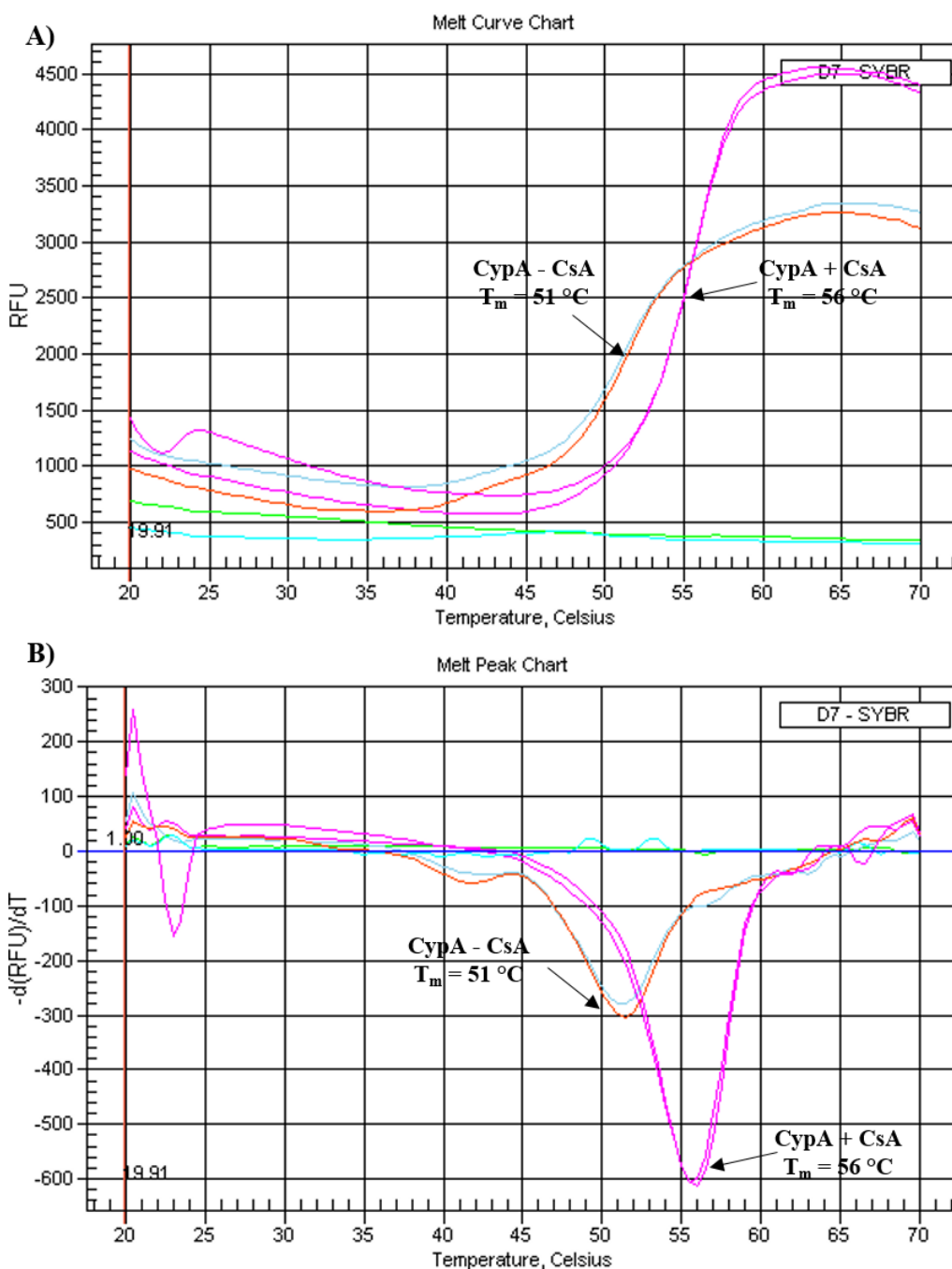


Figure 2-6 Sample TDA profile of CypA + CsA and CypA - CsA complexes. Graph (A) on top shows the raw output data from the TDA assay that is the fluorescent emission plotted against temperature. Graph (B) on the bottom shows the first derivative of the fluorescent emission with respect to temperature.

2.2.3.4 Enzymatic activity of Cyp

Cyps, as described in the introduction, are enzymes and one of their main functions is the isomerisation of proline residues. To further verify the activity of our protein samples, CypA protein sample was tested in a PPIase assay.

A schematic diagram showing the kinetic mechanism of PPIase assay can be seen in Figure 2-7. PPIase enzymatic assay is a standard spectrophotometric chymotrypsin coupled assay, where chymotrypsin enzyme has high conformational selectivity over trans X-Pro-Phe-pNA peptides. Selective hydrolysis of the para-nitroanilide of the trans X-Pro-Phe-pNA peptide from chymotrypsin produce a bright yellow color that absorbs light at 400 nm wavelength, Figure 2-7. Since X-Pro-Phe-pNA peptide in the assay is at equilibrium between cis : trans isomers, usually 70% : 30% in LiCl/TFE solution, selective cleavage of trans isomer will lead to the re-equilibration of the peptide in solution and more cis isomer to be converted to trans.

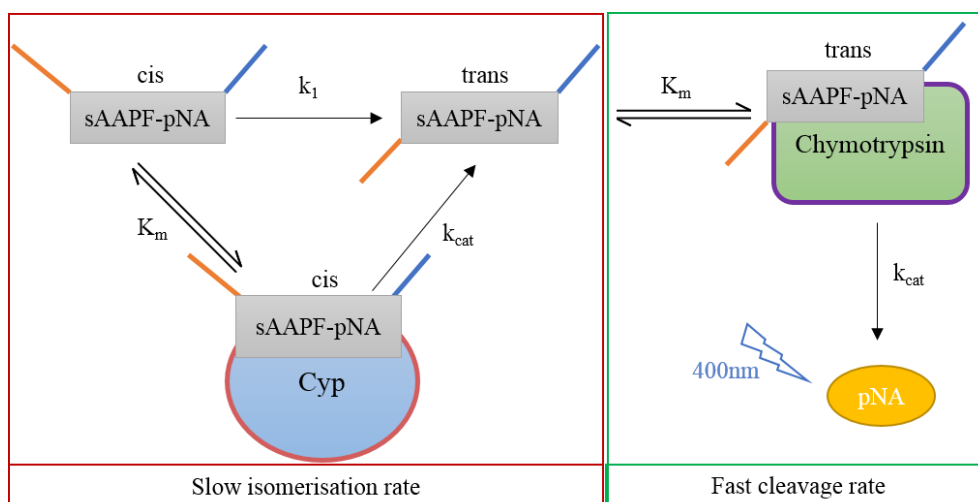


Figure 2-7 Kinetic mechanism of the PPIase - chymotrypsin coupled assay

Rational design of isoform specific ligands
 Peptidyl prolyl isomerase (PPIase) enzymatic assays monitor the isomerization of proline in real time as described from Kofron et al. (Kofron et al., 1991). The progress curves of the cleaved substrates can be described by Equation1.

$$\frac{d[C]}{dt} = -k_1[C] - k_{cat}[E]_p \frac{[C]}{[C] + K_m} \quad \text{Equation1}$$

Where [C] is the concentration of cis substrate at time t, k_1 the first order rate constant for the uncatalysed substrate isomerization process, $[E]_p$ is the Cyp concentration and K_m and k_{cat} are the Michaelis constant and the turnover number for the catalyzed substrate isomerization process. Analytical integration of Equation1, provide Equation2, from which the concentration of pNA product [P], at any time t, can be computed.

$$E' \ln \left(1 - \frac{[P]}{[C]_0 + K_m + E'} \right) + K_m \ln \left(1 - \frac{[P]}{[C]_0} \right) + k_1 t (K_m + E') = 0 \quad \text{Equation2}$$

Where $E' = [E]_p k_{cat} / k_1$, $[P] = [C]_0 - [C]$, and $[C]_0$ is the initial concentration of the cis substrate at time 0.

Measurements were performed as described previously (Kofron et al., 1991). All the reactions took place at 5 °C in a JASCO V-550 UV/Vis spectrometer using the protocol described before by Kofron *et al.* PPIase assay buffer consists of 50 mM HEPES, 100 mM NaCl, pH 8.0 and a total volume of 1 mL was used in each cuvette. N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (sAAPF-pNA) substrate and α -chymotrypsin from bovine pancreas were purchased from Sigma and diluted in LiCl/TFE (0.2g/10ml) and 10mM HCl respectively. The final concentration was 20 nM for the substrate, 6 mg/mL for the chymotrypsin and 10nM for the protein.

All samples of CypA tested were found to be catalytically active, Figure 2-8 below. As can be seen from the curves, the absorbance is increased over time, since more *trans* peptide is cleaved and more *para*-nitroanilide is released in solution. The presence of CypA (green curves) show a dramatic increase in the isomerization reaction rate when compared to the curves when no CypA was used (blue curves). This is because CypA catalyzes the isomerization of the peptide, from *cis* to *trans* and increases the rate of the reaction.

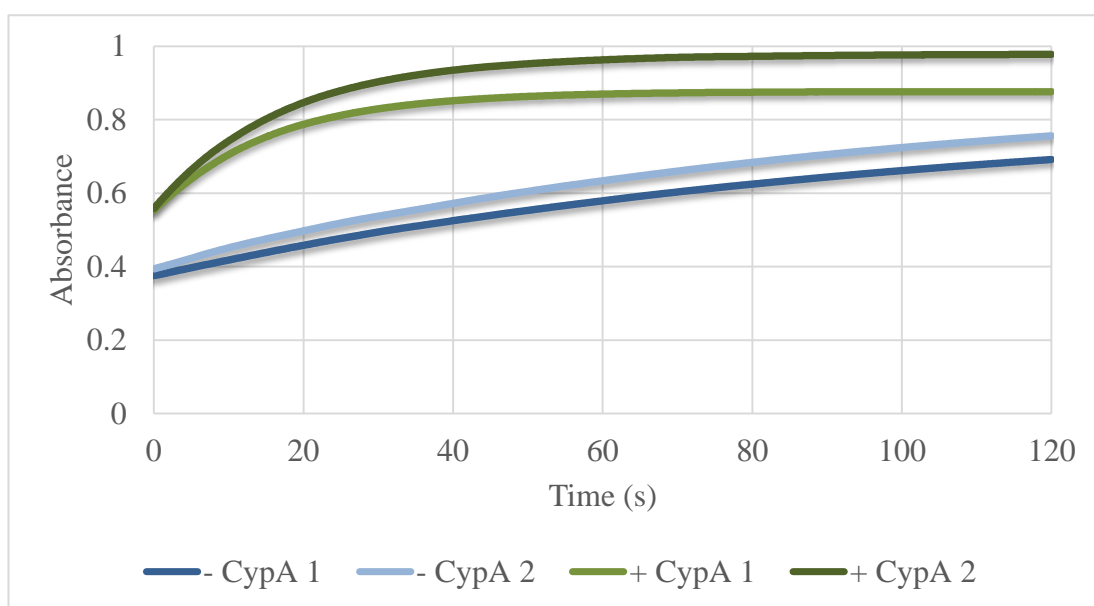


Figure 2-8 Characteristic PPIase profile of samples with and without CypA. Raw results from the PPIase assay of 4 samples, showing the change in absorbance versus time. The two lines in blue show replicates using samples without CypA, prepared only with sAAPF-pNA substrate and Chymotrypsin enzyme. Lines in green show replicates using samples with sAAPF-pNA substrate, Chymotrypsin and CypA enzymes.

As can be seen from the figure, the samples prepared without Cyp, shown in blue, have a shallow slope, and they almost have the form of a straight line. On the other hand the lines in green, i.e. samples prepared with CypA, have a very steep slope. The slope of each line represent the rate of product [P] formation (i.e. the *trans* substrate formation).

From the results it is obvious that samples with CypA have a much steeper slope, indicating the higher rate of *trans* substrate formation and the enzymatic activity of our protein samples.

2.3 Selection of ligands – the creation of a small compound library

2.3.1 The hit compound

Before the start of this project researchers in Prof Malcolm Walkinshaw's group, in the University of Edinburgh, have been studying Cyps for many years. During previous studies on Cyps, multiple experiments have been performed. Experiments included SPR studies on an in house library of 500 compounds. These SPR studies have been performed by Dr Martin Wear in Walkinshaw's group and from those experiments only one compound came out as a hit.

This compound was 2,3-diaminopyridine (Figure 2-10), which was identified to bind CypA with binding affinity on the high μM low mM scale and with 1:1 stoichiometry. Subsequently, and for the identification of the binding site of this compound on CypA, Dr Iain McNae from the same group, has used X-ray crystallographic studies to get structural data about the CypA – 2,3-diaminopyridine complex. From those studies, the structure of 2,3-diaminopyridine was found to bind in the Abu pocket of CypA as Figure 2-9 shows below. As can be seen from the figure below the main interaction that this compound establishes in the active site of the pocket, are H-bond interactions with the two tightly bonded water molecules (W1 and W2) and is also very close (3.5 Å) to the backbone carbonyl of Gly74. Water molecules W1 and W2 act as H-bond bridges between the protein and the ligand.

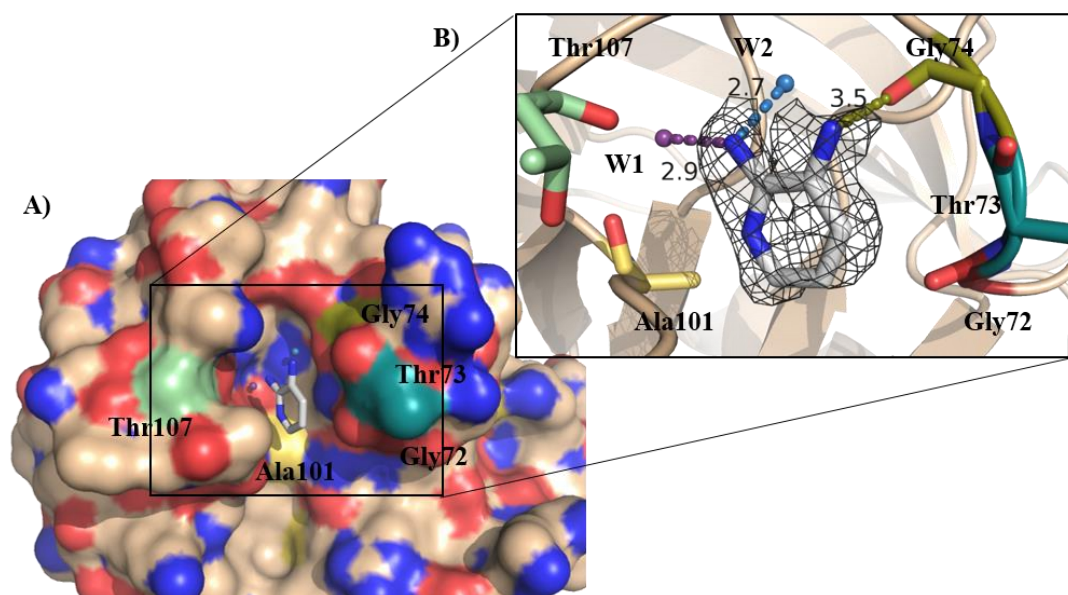


Figure 2-9 X-ray structure of 2,3-diaminopyridine binding in the Abu pocket of CypA. **A)** 3D surface structure of 2,3-diaminopyridine binding in the Abu pocket of CypA. **B)** 2Fo – Fc electron density map (at 1.0 σ shown in black) of 2,3-diaminopyridine bonded to the Abu pocket CypA. Distances between ligand with water and CypA molecules are highlighted and colour coded based residue colour. All figures above were generated using a crystal structure generated in this study, as described in the X-ray chapter of this thesis, which is the same as the original from Dr Iain McNae.

2.3.2 Selection of 98 small compound analogues

As was described in the Introduction chapter, one of the goals of this project was the identification of novel small compounds that bind Cyps. Subsequently the second aim was the use of these complexes for the study of dynamics and binding thermodynamics of Cyps using molecular simulations, biophysical assays and protein x-ray crystallography. Based on what we knew at the time of starting this project about Cyp inhibitors, the binding affinity from SPR and the structural data from X-ray crystallography of 2,3-diaminopyridine, this molecule was selected as the starting point for this project.

2,3-diaminopyridine was used as the hit compound from which a small library of almost 100 molecules was generated. Compounds were selected and purchased from publicly available sources based on availability and also on the binding site - binding

Rational design of isoform specific ligands mode and interactions that 2,3-diaminopyridine compound has in the Abu pocket of CypA as described in paragraph 2.3.1 and in the Figure 2 9. These compounds include analogues of 2,3-diaminopyridine with different substituents on the R1 ,R2 ,R3 and R4 position of the ring (Figure 2 10). Moreover analogues include compounds with different aromatic rings, including and not limited to pyrazine, pyrimidine, fused rings and non-aromatic rings. The main questions we are trying to address with these analogues are:

- If the pyridine ring is essential for the binding or if any other aromatic or non-aromatic cycles or heterocycles are tolerated.
- How important is the amine substituent, on the R1 position of the pyridine ring, for the binding of this compound?
- What chemical moieties can be favoured in the Abu pocket and especially at the 3rd position of the ring? Moreover are larger groups such as amines, esters or acids tolerated in this position, or only small single atom side chains are preferred?
- If and what side chains can be placed at the R3 or R4 position of the ring
- Can we identify longer side chain analogues on the R3 and R4 position, to facilitate the bridging between Abu and Pro pockets?.

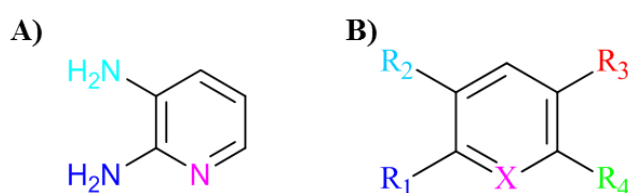


Figure 2-10 Chemical structures of the 2,3-diaminopyridine hit compound and analogues. A) shows the chemical structure of the 2,3-diaminopyridine hit compound while B) shows the chemical structure of the analogues purchased based on the 2,3-diaminopyridine. R₁ = R₂ = R₃ = R₄ = NH₂, I, Br, Cl, F, NO₂, OH, acid or amide linkers and X = C, N

Table 7 1, in Appendix 7.2, shows the chemical structure of all the selected compounds, the ligand number for each one that is used in this study, IUPAC name, MW, manufacturer and code number.

2.4 Conclusion

In this chapter CypA, CypB and CypD were selected as the primary biological targets in this project because of their biological significance and their differences in their surface electrostatic potential. Each protein was expressed, purified and characterized using SDS-PAGE, DLS, TDA and PPIase enzymatic assays. Finally a small library of ~100 compounds was designed and compounds were purchased based on 2,3-diaminopyridine molecule.

2.5 References

BioSilta Ltd. EnPresso B. 2014, pp 1–2.

Davis, T. L.; Walker, J. R.; Campagna-Slater, V.; Finerty, P. J.; Finerty, P. J.; Paramanathan, R.; Bernstein, G.; Mackenzie, F.; Tempel, W.; Ouyang, H.; et al. Structural and Biochemical Characterization of the Human Cyclophilin Family of Peptidyl-Prolyl Isomerases. *PLoS Biol.* **2010**, *8* (7).

Epps, D. E.; Sarver, R. W.; Rogers, J. M.; Herberg, J. T.; Tomich, P. K. The Ligand Affinity of Proteins Measured by Isothermal Denaturation Kinetics. *Anal. Biochem.* **2001**, *292* (1), 40–50.

GE Healthcare Bio-Sciences. HiTrap™ IMAC HP. pp 1–24.

GE Healthcare Bio-Sciences. HiLoad 16 / 600 and 26 / 600 Superdex 75 Prep Grade. 2011, pp 26–31.

Kofron, J. L.; Kuzmic, P.; Kishore, V.; Colon-Bonilla, E.; Rich, D. H. Determination of Kinetic Constants for Peptidyl Prolyl Cis Trans Isomerases by an Improved Spectrophotometric Assay. *Biochemistry* **1991**, *30* (25), 6127–6134.

Laemmli, U. K. Cleavage of Structural Proteins during Assembly of Head of Bacteriophage T4. *Nature* **1970**, 227.

Laskowski, R. A.; Swindells, M. B. LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **2011**, *51* (10), 2778–2786.

Lo, M. C.; Aulabaugh, A.; Jin, G.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G. Evaluation of Fluorescence-Based Thermal Shift Assays for Hit Identification in Drug Discovery. *Anal. Biochem.* **2004**, *332* (1), 153–159.

Lucigen Corporation. OverExpress™ Chemically Competent Cells. 2014, pp 1–9.

Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. a; Lane, P.; et al. High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *J. Biomol. Screen. Off. J. Soc. Biomol. Screen.* **2001**, *6* (6), 429–440.

Scientific, T. F. NanoDrop Lite User Guide. 2012, pp 1–50.

Rational design of isoform specific ligands

Steinberg, T. H.; Haugland, R. P.; Singer, V. L. Applications of SYPRO Orange and SYPRO Red Protein Gel Stains. *Anal. Biochem.* **1996**, *239* (2), 238–245.

Wear, M. A.; Patterson, A.; Malone, K.; Dunsmore, C.; Turner, N. J.; Walkinshaw, M. D. A Surface Plasmon Resonance-Based Assay for Small Molecule Inhibitors of Human Cyclophilin A. *Anal. Biochem.* **2005**, *345* (2), 214–226.

Wyatt, P. J. Light Scattering and the Absolute Characterization of Macromolecules. *Analytica Chimica Acta*. 1993, pp 1–40.

3 Surface plasmon resonance studies on Cyps

3.1 Introduction and Aims

3.1.1 Aims of this Chapter

The aims of the SPR study of Cyps were the investigation and ranking of any potential interaction between different Cyp isoforms and small molecules from our fragment library. Any compounds that have a positive response on Cyp surfaces would then be further investigated to determine their binding affinity, stoichiometry and if possible their kinetic parameters (association and dissociation constants) of binding to Cyps. A secondary aim was the study of any selectivity profiles of binding of these small compounds to different Cyp isoforms. Any potential hit compounds that would show good affinity and stoichiometry on any Cyp isoform would then be used in X-ray crystallography and free energy calculation studies.

3.1.2 Introduction to SPR

3.1.2.1 Surface Plasmon Resonance

For the Surface Plasmon Resonance phenomenon (Markey, 1999; GE Healthcare, 2008) two things are necessary as shown in Figure 3-1. The first is to generate polarised light. The second is to have an electrically contacting gold layer. This gold layer should be in between two media with different refractive indices. On the bottom is a glass surface, of the sensor chip, with high refractive index, and on the top is the buffer (passing over the gold surface) with low refractive index. SPR occurs when the polarised light hits the gold surface at a specific angle.

When the polarised light hits the glass surface at specific angle and wavelength, waves are generated and travel from the glass surface to the free electrons of the gold layer. These electrons absorb the wave and in turn they produce charge density waves, called plasmons. These plasmons cause a reduction in the intensity of the refracted polarised light at a specific angle that is called resonance angle.

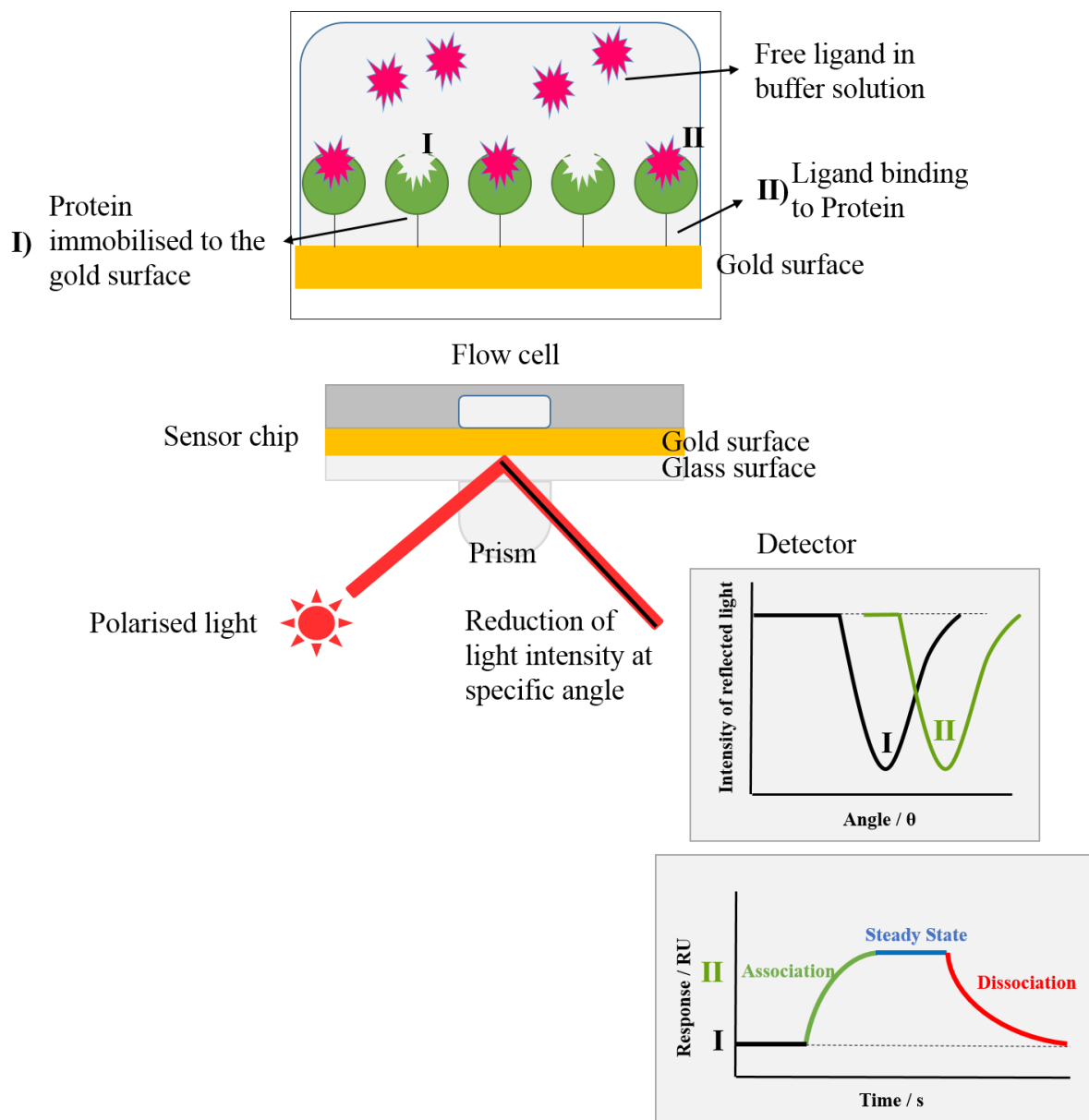


Figure 3-1 Surface plasmon resonance phenomenon. I) Stable and active protein is immobilised on the chip surface, producing a specific intensity and resonance angle. II) Ligand binds to the protein changing the intensity and the resonance angle of the reflected light.

Protein or molecules passing in the flow cell and interacting with the surface cause a change on the refractive index of the solution which in turn changing the intensity and resonance angle of the refracted light reported as response during the SPR experiment (Markey, 1999). As long as the molecules in solution interact with the surface of the

Rational design of isoform specific ligands sensor the detector shows an increasing response on the sensorgram that reaches a constant value when this interaction reaches equilibrium. When the molecules are washed off the surface using a buffer solution, then the sensorgram shows a decrease in the response units.

This change in angle and refractive index is converted to response R (in resonance units RU) versus time as can be seen from Equation3 (Davis and Wilson, 2000).

$$R_{obs} = n * X = X * \left[\left(\frac{\partial n}{\partial C} \right)_{lig} * C \right] \quad \text{Equation3}$$

Where R_{obs} is the observed instrument response, n is the refractive index at the surface and X is a factor to convert n to R_{obs} . Also $\left(\frac{\partial n}{\partial C} \right)_{lig}$ is the refractive index increment (RII) of the bonded ligand and C is the concentration of the bound ligand (in mass per volume) on the surface of the chip. Usually in Biacore instruments 1000 RU corresponds approximately to an angle change of $\sim 0.1^\circ$ and a concentration of 1 ng/mm^2 of surface protein concentration (GE Healthcare, 2008; GE Healthcare Life Sciences, 2012).

SPR is able to show recognition, association and dissociation in real time and is very useful for the study of protein-protein and protein-molecule interactions (Myszka and Rich, 2000; Myszka, 2004). It can be used for the accurate calculation of binding affinity, kinetic profile, and stoichiometry between interacting partners (Myszka, 1997, 2000). Also it can be used to determine specificity and the concentration of the interacting molecules in buffer.

3.1.2.2 Kinetics and affinity measurements using SPR

As described above, the SPR technique can be used just to monitor recognition between molecules or for more complicated experiments such as the calculation of the kinetic parameters of this interaction or the binding affinity of the interacting molecules. Binding affinities can be calculated using two different techniques. One is

Rational design of isoform specific ligands
the kinetic analysis of the interaction, in order to determine the association and dissociation constants and the other one is done by using a steady state approximation. Both techniques will be described briefly below.

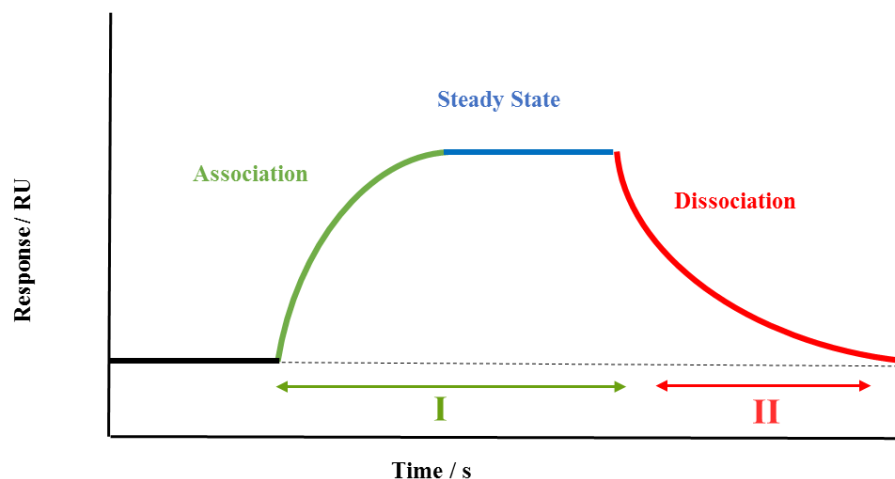


Figure 3-2 Kinetic parameters of the interaction between two or more components. The kinetics of the interaction can be described in three phases as can be seen from the coloured sensorgram. These phases are the Association, Steady state and Dissociation phases. I) Injection of ligand over the chip surface with the immobilised protein, allows the association of the protein with the ligand. II) Buffer flows over the surface allowing the dissociation of the ligand from the protein.

The kinetic evaluation involves the calculation of the dissociation and the association constants of the interaction. Taking as an example a protein – ligand interaction, association is the complex formation PL between a protein P and a ligand L while dissociation is exactly the reverse process. Dissociation is breaking of the PL complex to give the starting materials ligand L and protein P. For the kinetic evaluation first we need to monitor the response change over time, as can be seen in Figure 3-2, while injecting different concentrations of ligand over the sensor surface with the immobilized protein. The last step is to fit the experimental data into a 1:1 interaction model between Protein P and Ligand L, as shown in Equation 4, only for the phase of interest. As an example if we want to calculate the dissociation or the association constant only the curve in red or green respectively will be used.



Where k_a is the association rate constant in $M^{-1} s^{-1}$ and k_d is the dissociation rate constant in s^{-1}

The total rate of complex formation in the experiment during association phase can be expressed from the Equation 5 :

$$\frac{d[PL]}{dt_{ass}} = k_a[P][L] - k_d[PL] \quad \text{Equation 5}$$

While the complex formation during dissociation phase is defined from:

$$\frac{d[PL]}{dt_{diss}} = -k_d[PL] \quad \text{Equation 6}$$

Instead of concentration units, to calculate the rate constants in SPR, response units from the sensorgram can be used. As explained in Chapter 3.1.2.1, response is linearly related to the consecration of ligand binding on the surface of the chip. Moreover as also is explained in (Müller et al., 1998) the concentration of the ligand bound to the protein, in $mol L^{-1}$, can be calculated from the concentration in resonance units and the dextran matrix layer on the chip (approximately ~ 100 nm) using:

$$C_{\left(\frac{mol}{litre}\right)} = \frac{C_{(RU)}}{100 * M_W} \quad \text{Equation 7}$$

So protein ligand complex concentration [PL] can be measured during the experiment as response R (in response units RU). Ligand concentration [L] is the total concentration [C] added during the injection of the analyte. Protein concentration (P) that is free on the chip can be calculated if we subtract the concentration that is in complex (RU) from its maximum feasible number R_{max} . So by substituting the concentration terms with response from the sensorgram, the Equation 5 and Equation 6 can be written again as:

$$\text{During association:} \quad \frac{dR}{dt_{ass}} = k_a \cdot C \cdot (R_{max} - R) - k_d \cdot R \quad \text{Equation 8}$$

During dissociation:
$$\frac{dR}{dt_{diss}} = -k_d \cdot R \quad \text{Equation 9}$$

Linearizing and integrating the kinetic Equation 8 and Equation 9 above, the resulting equations can then be fitted to the experimental data and calculate the k_d and k_a rate constants. From the rate constants then the dissociation constant can be calculated based on Equation 10.

$$K_D = \frac{k_d}{k_a} \quad \text{Equation 10}$$

On the other hand using a steady state approximation we assume that the injection of ligand over the immobilized protein surface is long enough that the rate of complex (PL) formation remains unchanged and the response reach an equilibrium (R_{eq}). That means that $dR/dt = 0$ and Equation 8 can be rewritten as:

$$\frac{dR}{dt_{eq}} = k_a \cdot C \cdot (R_{max} - R_{eq}) - k_d \cdot R_{eq} = 0 \quad \text{Equation 11}$$

Rearranging and solving Equation 11 for R_{eq} and setting $K_a = k_a/k_d$ then:

$$R_{eq} = \frac{K_a \cdot C \cdot R_{max}}{K_a \cdot C + 1} \quad \text{Equation 12}$$

Equation 12 can then be used to fit the steady state approximation curve to the plot of R_{eq} against concentration C , as can be seen in the Figure 3-3. From the plot the K_a and hence K_D of the interaction can be calculated.

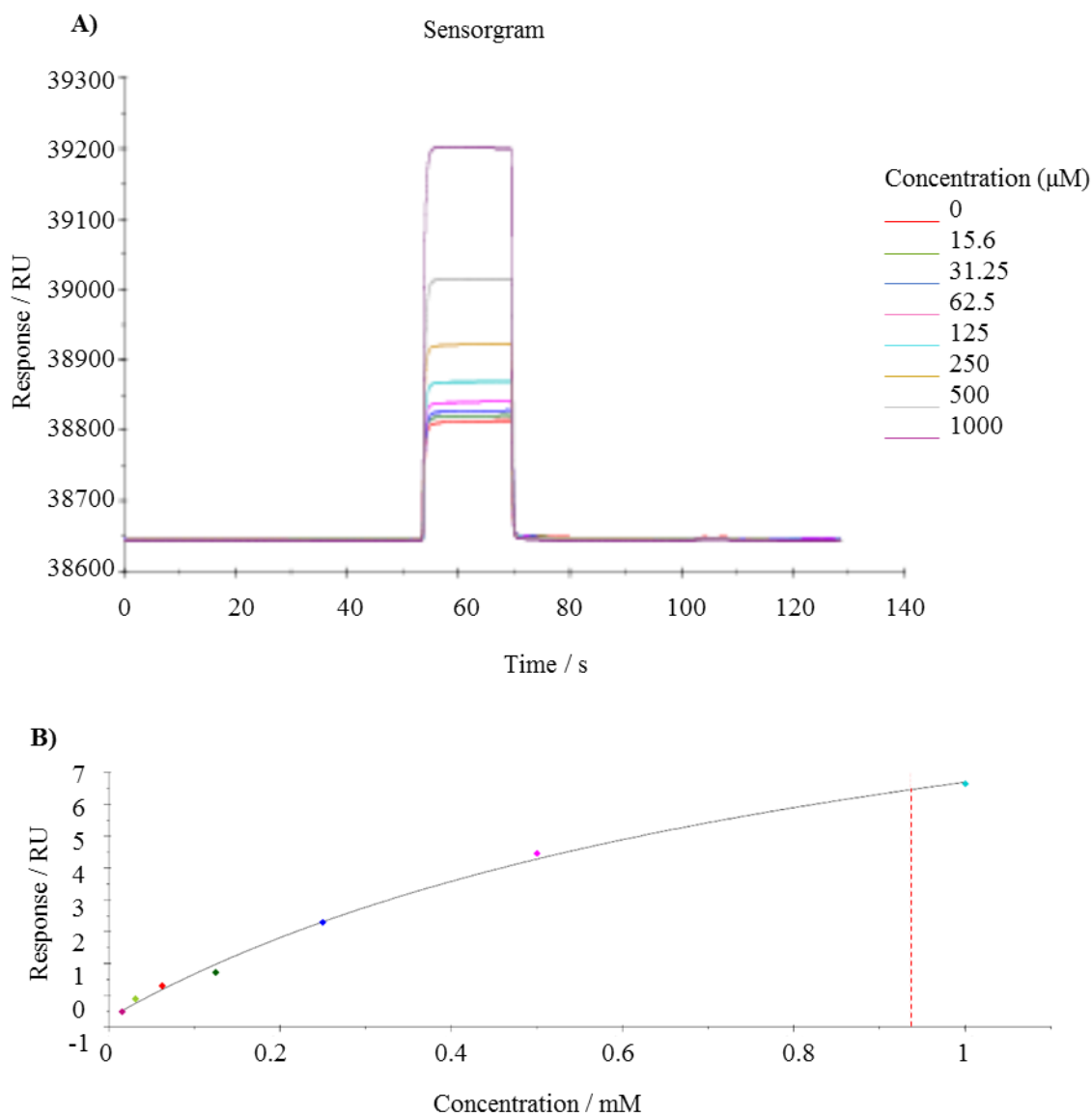


Figure 3-3 Use of steady state fitting for the calculation of binding affinity. A) Multi-cycle kinetic sensorgrams. Each curve shows the sensorgram using a different ligand concentration. **B)** The corresponding steady state response curve and the calculated K_d shown as red dotted line.

3.2 Materials and Methods

3.2.1 Instrumentation and materials used

For these SPR experiments the Biacore T200 instrument was used and 1-ethyl-3-(3-diaminopropyl) carbodiimide hydrochloride (EDC) and N-hydroxysuccinimide (NHS) chemicals were purchased from Bioacore.

A Ni^{2+} -nitrilotriacetic acid (NTA) sensor chip with 4 flow channels in series was purchased from Bioacore and was used to immobilise and covalently stabilise three different Cyp isoforms: 6His-CypA, 6His-CypB and 6His-CypD as described in (Wear et al., 2005). For the purposes of the experiment all 4 different surfaces on the sensor chip were prepared in an identical way, as described in (Wear et al., 2005) but different Cyp isoforms were immobilised on each surface as shown on Figure 3-4. The 1st surface was without any Cyp and was used as reference to see the response that each compound is generating when interacting just with the sensor surface. The 2nd cell was used for the immobilisation of CypA and in total 3200 RU (reference corrected) of CypA were immobilised on the chip. CypB was in the 3rd cell at a density of 3000 RU and CypD in the 4th at the same density as CypB. Chip was prepared by Dr Martin Wear.

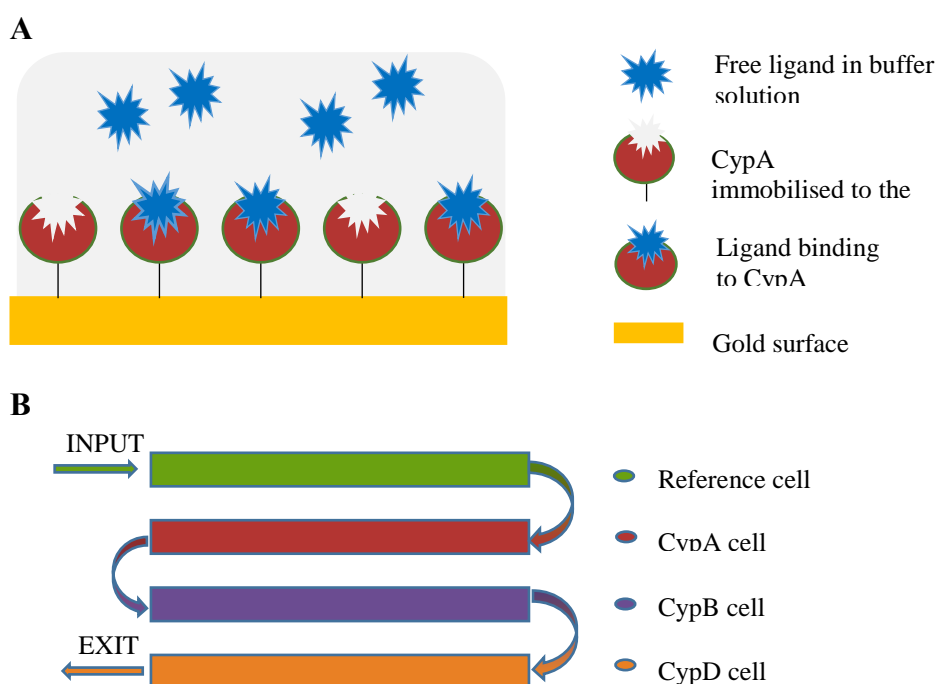


Figure 3-4 Assay and sensor chip set up. **A)** Shows an example of one of the four cells of the chip, with the protein to be immobilised on the gold surface of the sensor chip and the ligands to pass over in the buffer solution. **B)** Shows how the four cells are set up for this experiment. Cell 1 is the reference cell without any protein immobilised, while cells 1, 2 and 3 have CypA, CypB and CypD immobilised respectively.

3.2.2 Dilution and Preparation of fragments

SPR measurements were taken in a buffer solution consisting of: 10 mM PBS, pH 7.4, 137 mM NaCl, 2.7 mM KCl; 0.005 % surfactant P20; 0.5 % DMSO (final concentration) as described. Stock solutions for each compound were prepared under different conditions based on the solubility of each compound. Either 50 - 100mM concentration in 100% ethanol or 5 - 10 mM in 100% PBS solution. Stock solutions from each compound were then diluted with buffer solution for a final compound concentration of 1mM. Stock solution used to solubilise each of the final 48 compounds used in the dilution series SPR measurements can be seen in the Table 7-2 in Chapter 7.3.

3.2.3 Single point measurements

Single point SPR measurements were performed as described (Wear et al., 2005; Wear and Walkinshaw, 2006). A single concentration for each compound 1mM in buffer solution was injected over the surface of the sensor chip at 30 $\mu\text{l}\cdot\text{min}^{-1}$ for 15 s and allowed to dissociate for another 15 s at 25 °C. The surface was regenerated between each measurement and the compounds were washed off the surface of the sensor by running an excess of buffer solution at 40 $\mu\text{l}\cdot\text{min}^{-1}$ for 30 s followed by a further 5 s stabilization period.

Every 30 measurements, 0.2 μM of CsA in buffer solution was passed over the surface at 30 $\mu\text{l}\cdot\text{min}^{-1}$ for 15 s, to verify the activity of Cyp immobilised on the chip and that the surface of the chip remained unchanged. The software provided with the Biacore T200 instrument was used to analyse the results of the single point measurements and select the compounds to be used in the next experiment. Single point measurements were taken with the kind help of Dr Martin Wear.

3.2.4 Dilution series measurements

Two fold dilution series concentrations SPR measurements were performed as described (Wear et al., 2005; Wear and Walkinshaw, 2006). Each compound ranging from 1 mM – 15.62 μ M in buffer solution were injected in cycles over the surface of the sensor at 30 μ l.min⁻¹ for 15 s and allowed to dissociate for another 15 s at 25 °C. As in the previous experiment, surface was regenerated between each measurement and the compounds were washed off the surface of the sensor. Moreover, as in the single point measurements, every 30 measurements, 0.2 μ M of CsA in buffer solution was passed over the surface to verify that the activity of Cyp immobilised remained unchanged. Dilution series measurements were taken with the kind help of Dr Martin Wear.

3.2.5 Data processing and fitting

First all sensorgrams were corrected by subtracting the response on the reference cell and the response from the blank injections. Binding affinity and specificity were determined for each compound by fitting a steady state affinity, 1:1 interaction model using the analysis software provided with the instrument. Standard deviation (std) values are not reported in the results of SPR experiments, neither in figures nor table of results in Chapter 0 or Appendix. This is because std values can only be confidently reported for SPR experiments when measurements are repeated in duplicate or triplicate and the mean and std values are calculated from the repeats. Measurements for this experiment were only taken once, firstly because this experiment is quite expensive and secondly because the expected activity range for these small fragments is on the high micro-molar – low milli-molar range. For the SPR K_d results reported in this thesis a coefficient of variation (CV) of ~ 20 - 40% can be taken into account that is similar to the CV reported in previous studies (Cannon et al., 2004; Katsamba et al., 2006).

3.3 Results and Discussion

3.3.1 Single point measurements

Single point measurements on every soluble compound from the library were taken primarily to identify any possible interaction – binding of the compounds on CypA, CypB or CypD. Also using single point measurements is a fast way to screen all the library and distinguish the active from the less active or inactive compounds for further studies. Doing dilution series for every compound from the 98 in the library would be very time consuming and expensive and wouldn't give much information about the binding affinity, kinetics or specificity because of the very weak binding of these compounds on Cyps, which is very close to the limit of the SPR technique.

As described before, all compounds were tested at 1mM concentration in buffer solution on 4 different surfaces on the sensor chip (in order: reference cell, CypA, CypB and CypD). To analyse the results of the experiment, first we subtract the response of each compound on the 1st reference cell, from the response of the same compound on all the other cells. In this way we exclude any positive response generated due to the interaction of the compound with the gold surface of the sensor chip. Moreover the response of each compound was normalised based on its mass. As described in the SPR introduction (Sub-chapter 3.1.2) the change of angle of the reflected light is proportional to the mass of the molecules interacting with the sensor. So a molecule with higher mass would generate a higher response than a molecule with lower mass. This mass correction was necessary in order to be able to compare the responses from all compounds and select the compounds with the higher response, regardless of their mass, for the next experiment. Mass correction was achieved by dividing the reference corrected response of each compound by the molecular weight of the compound and multiplied by 100 as Equation 13 shows below.

$$R_{corr}^{mass} = \frac{R_{corr}^{ref}}{Mw_{lig}} * 100 \quad \text{Equation 13}$$

Where R_{corr}^{mass} is the mass and reference corrected response for each compound, R_{corr}^{ref} is the reference corrected response and Mw_{lig} is the molecular weight of the ligand.

The results from the single point SPR experiment can be seen in Figure 3-5. Results show the response for the interaction of each compound with each Cyp isoform, after the corrections applied as described above. The positive control in this assay was CsA (with a K_d of ~ 13 nM) and in the figure can be seen in the red circle. Its low response is because of the mass weight correction, as CsA has a M_w of 1202 Da while the M_w of the rest of the compounds is $\sim 95 - 250$ Da. Moreover CsA was tested in a much lower concentration ($0.2\mu\text{M}$) than the rest of the fragments (1mM). The noise of modern SPR instruments is usually under 0.5 RU (Giannetti, 2011). This means that the cut off can be set very low around 2RU. Nevertheless from this experiment we are only interested for the most active compounds, and since most of the compounds have response above the noise ratio then the cut off was set to 5 RU.

From this figure two different trends can be noticed. The first one is the response of each compound in relation to the other compounds and the second one is the response of each compound related to each Cyp isoform. From this graph it seems that the response of each compound in relation to the other compounds is not affected by the Cyp isoform immobilised on the chip. Moreover it seems that the response of all compounds when passed over CypA surface is higher than their response when interacting with CypB or CypD. This is not related with the different masses of the proteins since the response for each compound is reference subtracted so the contribution in response from the protein mass is subtracted. Moreover this cannot be related to the density of each protein on the sensor chip, as explained on Chapter 3.2.1, because the density of all isoforms is very close ($\sim 94\%$) and the difference on response because of it is less than 0.065 per RU.

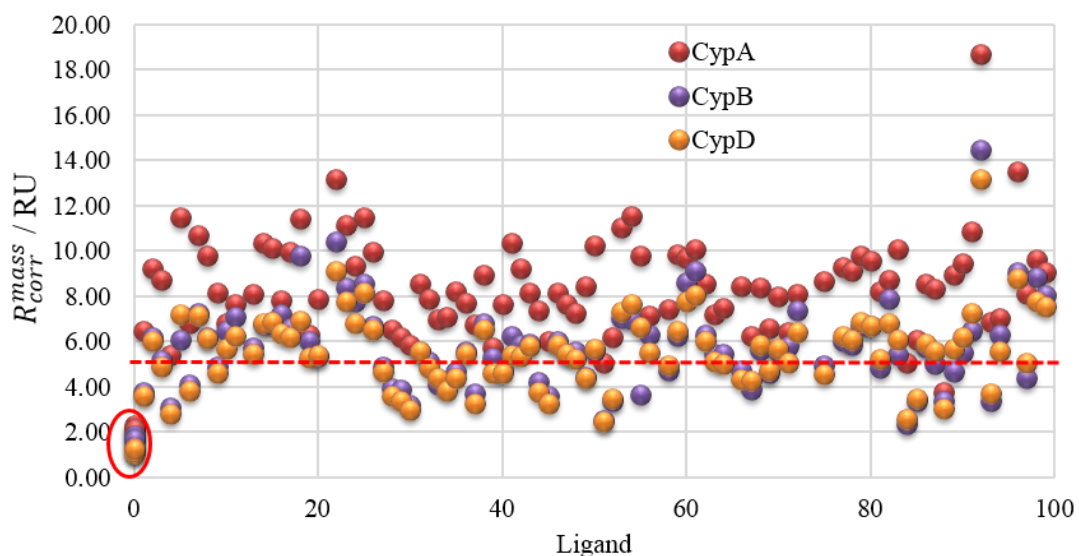


Figure 3-5 Single point measurement SPR results. On the Y axis is the response of each compound (in response units) and on the X axis the 93 compounds tested for activity. The results of activity of the compounds on each Cyp isoform can be seen in different colours on the same figure. CypA results are shown on red, CypB on violet and CypD on orange. The cut off for activity can be seen in a red dotted line, while the positive control can be seen clustered in the red circle.

To further examine the activity of the compounds on CypA and compare it with the activity of the compounds on the other surfaces, the histogram of number of compounds in relation with their response on all Cyp surfaces was plotted, as shown in Figure 3-6. As an example in red line is the histogram of compounds' responses on CypA divided in eleven bins based on their response (bin1 has $0 \leq \text{activity} \leq 2$, bin2 has $2 \leq \text{activity} \leq 4$... bin11 has activity ≤ 20). Comparing the activities on CypA, CypB and CypD it is clear that the activities on CypA are on average 3 response units higher than the activities on the other two isoforms. The activities of compounds on CypB and CypD are almost identical.

This may show that compounds in this library have higher affinity towards CypA isoform but this can be assessed more precisely after the dilution measurements of each selected compound and the determination of their affinity on every surface. At

this stage this results are only used in order to distinguish active compounds from inactive.

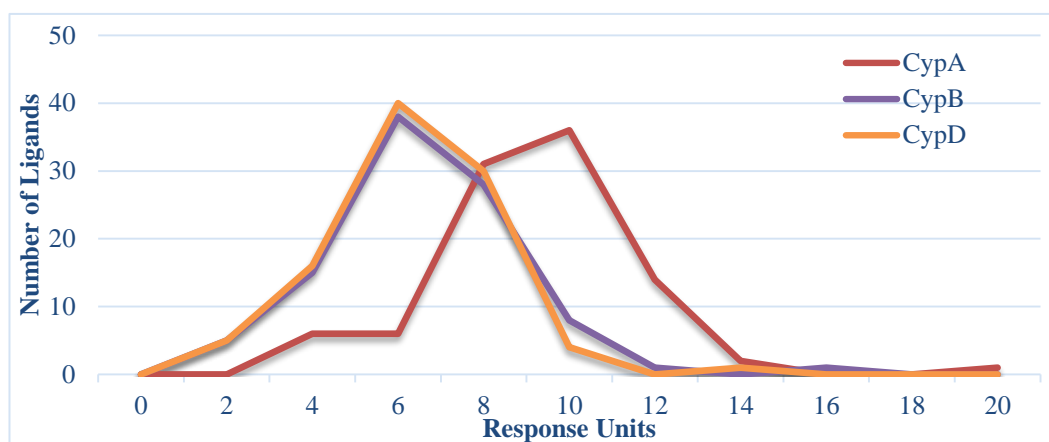


Figure 3-6 Histogram of number of compounds based on their activity on the 3 surfaces of the chip. Compounds are separated in bins based on their activity on CypA (shown in red), CypB (shown in violet) and CypD (shown in orange). Average response on CypA is 8.17, on CypB 5.57 and on CypD 5.38.

To further examine the response of each compound in relation to the other compounds the ranking of compounds, based on their activity, on each surface was plotted. Figure 3-7 shows this correlation between the activity rankings of compounds when passed over CypA, CypB and CypD surfaces. Correlation between compounds ranking on CypA and CypB surfaces, shown in blue, is $R^2 = 0.60$ while the correlation between CypA and CypD, shown in green, is 0.77. It is important to note that the correlation between CypA with CypB and CypD is not as high as expected just by looking Figure 3-7. Nevertheless the 50 most active compounds on CypA are almost the same as the 50 most active compounds on the other Cyp isoforms. Moreover this figure is just to help select the most active compounds for further examination with SPR and is not used to rank reliably compounds based on their activity.

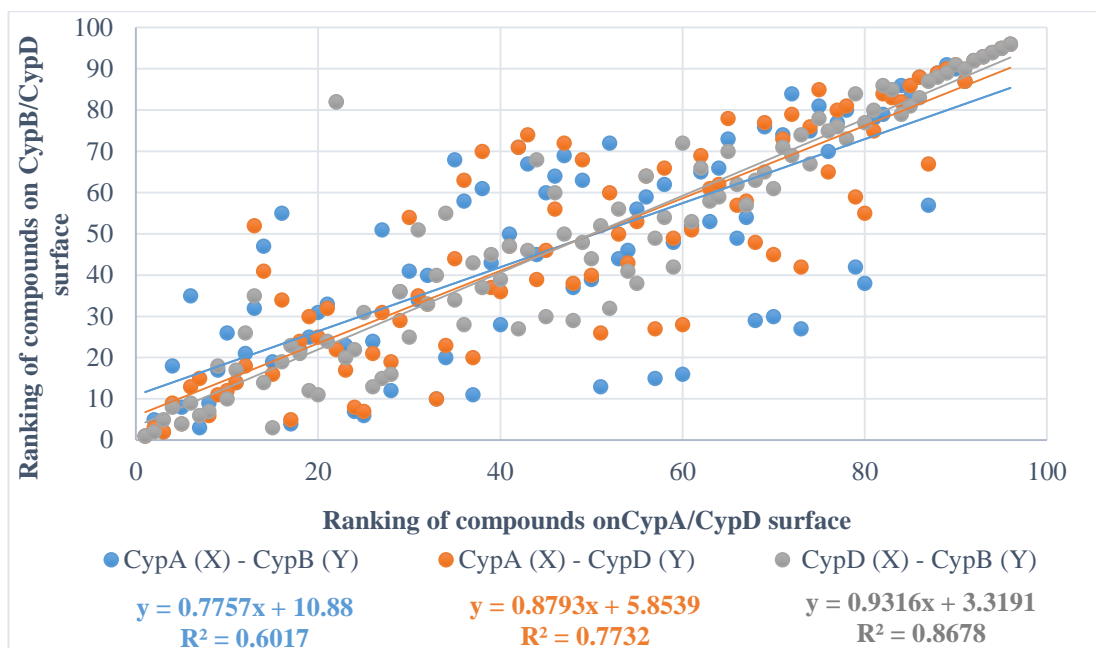


Figure 3-7 Correlation between the activity ranking of compounds upon interaction with CypA, CypB and CypD isoforms. Correlation between CypA-CypB is shown in blue, CypA-CypD in orange and CypD-CypB in grey.

So taking in account the raw data of the single point SPR experiment and the observations from the analysis of the results as shown in the Figure 3-6 and Figure 3-7, the compounds for the dilution series measurements have been selected. These compounds were the 48 compounds with the highest response units on CypA surface. The lowest response for these compounds was 8 on CypA and 6 on CypB and CypD, so that was above the cut-off that we set at the beginning.

3.3.2 Dilution series measurements

3.3.2.1 Active compounds and calculated dissociation constants

Compounds previously selected from the single point experiment were diluted in a two - fold dilution series of concentrations, starting from 1mM and finishing to 15.62 μ M. Measurements were taken as described in Sub-chapter 3.2.4. Dissociation constants were calculated for every compound using the software provided using a steady state affinity model for the analysis of all compounds. The k_d and k_a of all compounds were

Rational design of isoform specific ligands very fast, because of their weak binding, and it was not possible to determine them from the kinetic experiment. Hence they are not reported in this thesis. For the estimation of the K_d values (affinity) the only method be usable was thus the steady state modelling. Steady state affinity modelling was applied as described in Sub-Chapter 3.1.2.2. The steady state response against concentration curves for all compounds measured can be seen in

Figure 7-1 in Chapter 7.3. Based on the calculated binding affinities most compounds can be divided into three main categories: compounds that cannot be described from a steady state model, inactive, and active compounds, as can be seen from the Figure 3-8.

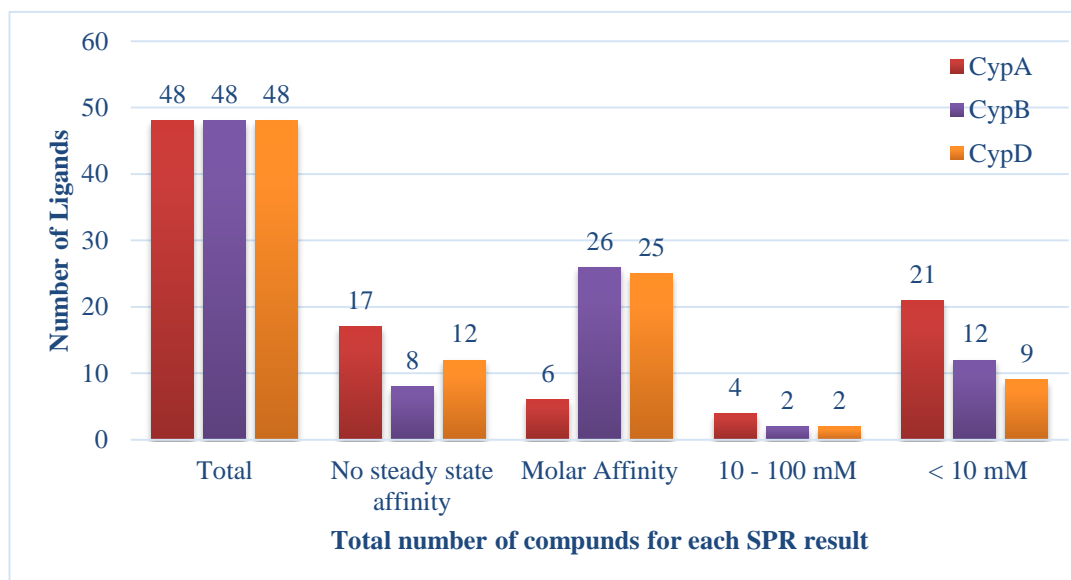


Figure 3-8 Dilution series SPR results. Figure shows the total number of compounds tested using dilution series concentration and their activity range on CypA, CypB and CypD. Compounds are divided into 4 categories based on their activity and the results for each surface can be seen in different colours: red for CypA, violet for CypB and orange for CypD.

In the first category are compounds for which a steady state affinity kinetic model is not able to describe their response versus concentration curves. For these compounds

Rational design of isoform specific ligands

it seems that their interaction with Cyp surfaces is not as simple as the others. Also this could be a reason of poor solubility or unexpected behaviour of the compound during its transfer from the bulk buffer solution to the surface of the chip during SPR measurements. Further studies are needed for the measurement of the dissociation constants of these compounds. In the second category are compounds that are considered as inactive, since their calculated affinity from the steady state affinity model is in the molar range.

In the last category are compounds that show a low milimolar activity on Cyp isoforms. Most of these compounds have a dissociation constant under 10mM and they are thus considered as active in this study. There are a few exceptions where activity is higher than 10mM but lower than 100mM but these compounds are considered as inactive too, as the activity cut off we set is a dissociation constant of 10mM.

As a rule of thumb a K_d prediction with SPR is considered accurate when the predicted K_d value is equal to or less than half of the total analyte concentration (GE Healthcare Life Sciences, 2012). As can be seen from this SPR experiment, the cut off of K_d between active and inactive compounds is 10mM while the average activity is 3.7mM and the maximum concentration of ligands is 1mM. So it is clear that these results are well above the confidence level of SPR. This means that the K_d values for our “active” compounds are approximate. So although we can confidently distinguish active from inactive compounds using a cut off of 10mM, we should refrain from ranking actives according to measured K_d values. Hence affinity numbers have to be treated with some suspicion and more parameters need to be taken in account before making any conclusions. In our case a stoichiometric and specificity analysis was taken into account before we selected our final SPR hit compounds.

3.3.2.2 Binding stoichiometry of compounds on Cyps

For all active compounds (i.e $K_d \leq 10\text{mM}$), the stoichiometry of interaction with each protein was calculated to examine if their activity with Cyps is a simple 1:1, as was

Rational design of isoform specific ligands assumed from the steady state affinity model. To do that the experimental Rmax values for each compound were compared with the theoretical calculated value and the final stoichiometry reported below is the quotient, rounded to the nearest one, of $R_{max}^{exp} / R_{max}^{theo}$. That means that if the quotient of Rmax was 1.51 it was considered as 1:2 activity while if the quotient was 1.50 it was considered as 1:1. Generally for a well fitted model R_{max}^{exp} should not be more than two times the theoretical value (GE Healthcare Life Sciences, 2012).

As was described before in the introduction of SPR, Rmax is the maximum theoretical signal (in RU) between a protein and a ligand pair interaction. The experimental Rmax is calculated from Equation 12 by fitting a steady state affinity curve on the experimental response versus concentration curves. For all compounds this was done automatically from the software provided with the Biacore T200. The theoretical Rmax had to be calculated for each ligand – protein pair of interest using:

$$R_{max}^{theo} = \frac{MW_{lig}}{MW_{pro}} \cdot RU \quad \text{Equation 14}$$

Where MW_{lig} is the molecular weight of the ligand, MW_{pro} is the molecular weight of the protein and RU is the reference corrected response of protein on the surface chip. The RU for reference corrected CypA was ~3200RU while for CypB and CypD was ~3000RU.

A histogram showing the number of compounds per protein and stoichiometry can be seen in Figure 3-9. The histogram shows the total number of active compounds per protein and the compounds are then divided into three categories. Compounds that show 1:1 stoichiometry, compounds that have stoichiometry more than 1:1 and compounds with less than 1:1 stoichiometry.

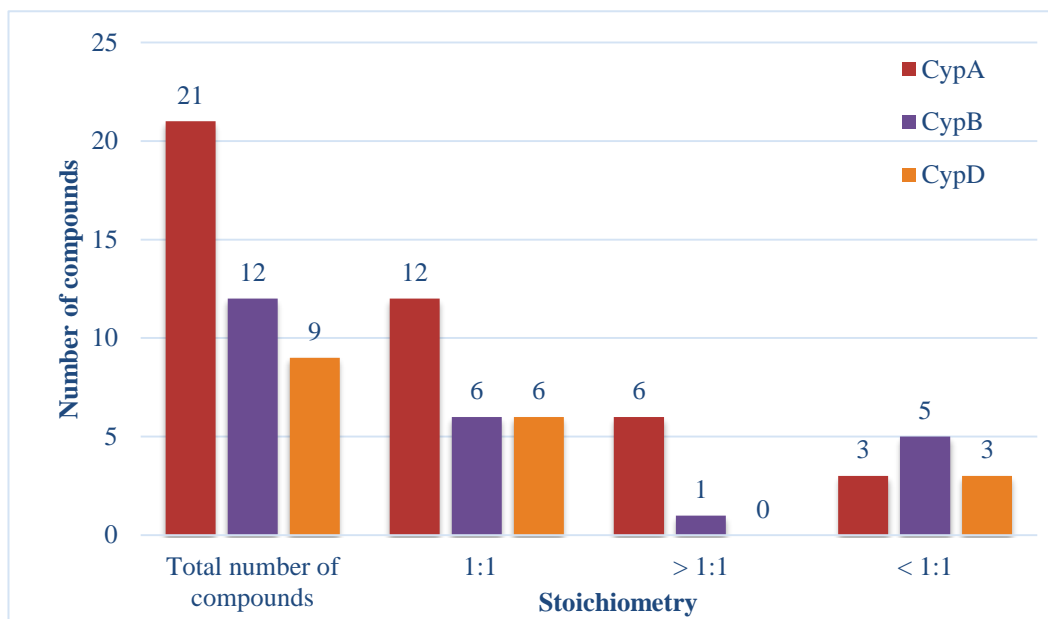


Figure 3-9 Stoichiometry of interaction of active ligands with Cyps. Figure shows the total number of active compounds on each surface. Compounds are then divided into three main categories. Compounds with 1:1 stoichiometry with Cyps, compounds with $\geq 1:1$ stoichiometry (more than one compounds interact with the protein), and compounds with $\leq 1:1$ stoichiometry. The stoichiometry for each surface can be seen in different colours: red for CypA, violet for CypB and orange for CypD.

From the graph it can be seen that 6 compounds have stoichiometry of more than 1:1 on CypA and 1 compound on CypB. 1:1 stoichiometry means that one molecule of ligand is interacting with one molecule of the protein. More than 1:1 stoichiometry means that the ligand is able to interact with the protein but these interactions are unspecific and more than one molecules of the ligand can bind at the same time to the protein. This can also show random binding or a promiscuous binding of ligands to the surface of the protein and not a direct binding into a binding pocket of the protein. But it can also be a combination of the two, random interactions on the surface and also binding into a binding pocket. Promiscuous binding could be identified from the steady state plots because even if the concentration of the ligand increased the protein binding site will still not be saturated, and the plot will never reach steady state. Moreover 3 compounds show less than 1:1 stoichiometry on CypA, 5 on CypB and 3 on CypD. Compounds with stoichiometry less than 1:1 are more difficult to explain. This could occur with a blocked or partially blocked binding site, which shouldn't be the case in

Rational design of isoform specific ligands this experiment as CypA was prepared as previously tested protocols and was tested using CsA. This can happen if something else binds to the bind site and block the binding of a compound or if the protein is not immobilised with the correct orientation. Moreover it can be a result of precipitation of the ligand. All compounds in these two categories must be treated with more care than the others and the main focus should be on compounds that are active and show 1:1 stoichiometry.

On the other hand 12 compounds show 1:1 stoichiometry of interaction with CypA while 6 show the same stoichiometry when interacting with CypB and CypD. These compounds are both active and have the correct or expected stoichiometry (1:1) and they can be a promising starting point as inhibitors of Cyps. There are some compounds that show 1:2 stoichiometry. These compounds are on the edge of reliability or cut off of the experiment. Some of them can be equally good as the one that show 1:1 activity but this can be decided also by looking the steady state affinity fitting. If the fitting is good and the graph is close or tends to reach a maximum on RU, taking in account that their K_d is lower than 10mM then these compounds could considered as acceptable for further studies.

3.3.2.3 Selectivity profiles

After separating the active from inactive compounds and excluding those with stoichiometry of more, or less than 1:1, 16 compounds are left. Some of them, based on the results before, show selectivity for one or more isoforms while others do not show any selectivity. Among these compounds there are 3 that show activity on all isoforms tested (CypA, CypB and CypD), while 7 of them show selectivity for CypA, 2 compounds for CypB and 2 others for CypD. All compounds can be divided into groups based on their selectivity as can be seen in the Figure 3-10. Each and every one of the 15 selected compounds can also be seen in Figure 3-11 in more details. Figure 3-11 shows the molecular structure of each compound, and also the dissociation constant on the Cyp isoform that show activity.

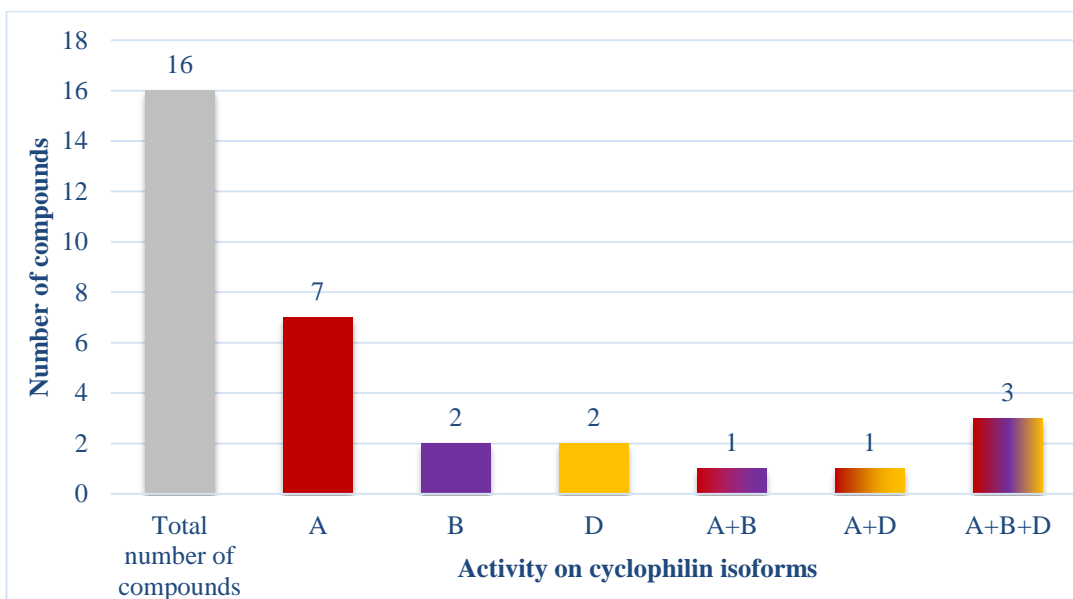


Figure 3-10 Histogram of number of active compounds with 1:1 stoichiometry per selectivity profile. Figure shows a total of 16 compounds that are active and have 1:1 stoichiometry. Compounds are separated into 6 bins based on their selectivity for one or more Cyp isoforms.

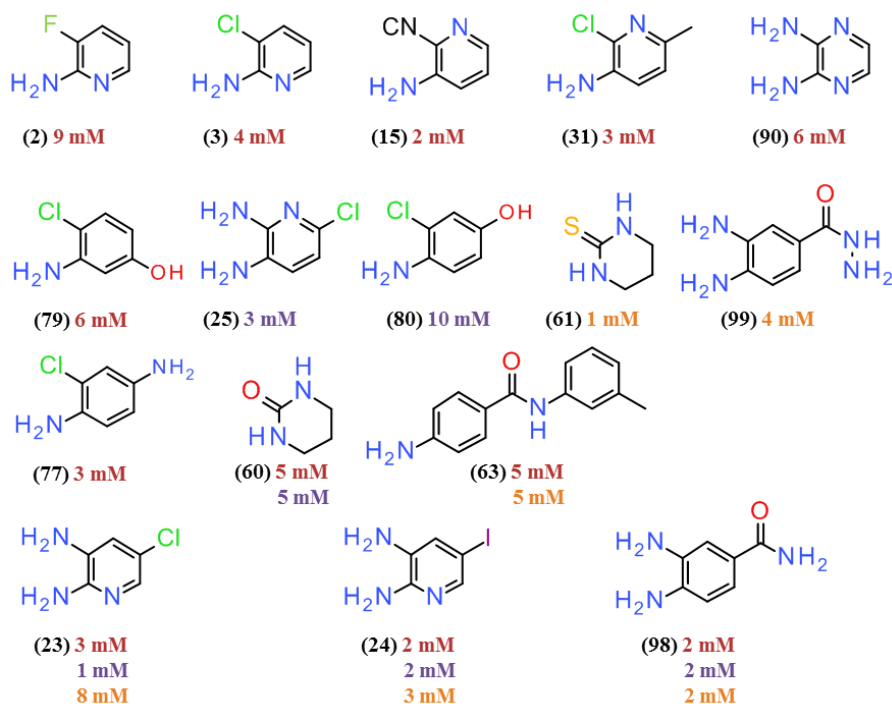


Figure 3-11 Final selected compounds with activity less than 10mM and 1:1 stoichiometry. Compound's number is shown in parenthesis and the K_d for each compound is colour coded based on each Cyp isoform. As before, activities on CypA are shown in red, CypB on violet and CypD in orange.

3.4 Conclusions

In this part of our study we have used SPR as a primary assay to screen almost 100 compounds from a library of small fragments, for activity against three Cyp isoforms, CypA, Cypb and CypD. As a first step single point measurements were taken at 1mM to identify the most active compounds. The 48 compounds with the higher response were selected and were tested in a 2 fold dilutions series in order to measure their binding stoichiometry and K_d values. Compounds with affinity less than 10mM and 1:1 stoichiometry against any Cyp isoforms have been identified and are considered as hit compounds. These compounds, alongside with some other compounds from the library, have been further tested with other techniques such as X-ray crystallography and free energy calculations as described in the next chapters to verify their binding to Cyp, determine their binding pose and identify the interactions they establish in the binding side of Cyps.

3.5 References

- Cannon, M. J.; Papalia, G. A.; Navratilova, I.; Fisher, R. J.; Roberts, L. R.; Worthy, K. M.; Stephen, A. G.; Marchesini, G. R.; Collins, E. J.; Casper, D.; et al. Comparative Analyses of a Small Molecule/enzyme Interaction by Multiple Users of Biacore Technology. *Anal. Biochem.* **2004**, *330* (1), 98–113.
- Davis, T. M.; Wilson, W. D. Determination of the Refractive Index Increments of Small Molecules for Correction of Surface Plasmon Resonance Data. *Anal. Biochem.* **2000**, *284* (2), 348–353.
- GE Healthcare Life Sciences. Biacore Assay Handbook. *GE Healthc. Bio-Sciences AB* **2012**, 1–78.
- GE Healthcare. Biacore Sensor Surface Handbook. **2008**, 8–10.
- Giannetti, A. M. *From Experimental Design to Validated Hits: A Comprehensive Walk-through of Fragment Lead Identification Using Surface Plasmon Resonance*, 1st ed.; Elsevier Inc., 2011; Vol. 493.
- Katsamba, P. S.; Navratilova, I.; Calderon-Cacia, M.; Fan, L.; Thornton, K.; Zhu, M.; Bos, T. Vanden; Forte, C.; Friend, D.; Laird-Offringa, I.; et al. Kinetic Analysis of a High-Affinity Antibody/antigen Interaction Performed by Multiple Biacore Users. *Anal. Biochem.* **2006**, *352* (2), 208–221.
- Markey, F. What Is SPR Anyway. *BIA J.* **1999**, 14–17.
- Müller, K. M.; Arndt, K. M.; Plückthun, a. Model and Simulation of Multivalent Binding to Fixed Ligands. *Anal. Biochem.* **1998**, *261* (2), 149–158.
- Myszka, D. G. Kinetic Analysis of Macromolecular Interactions Using Surface Plasmon Resonance Biosensors. *Current Opinion in Biotechnology.* 1997, pp 50–57.
- Myszka, D. G. Kinetic, Equilibrium, and Thermodynamic Analysis of Macromolecular Interactions with BIACORE. *Methods in Enzymology.* 2000, pp 325–340.
- Myszka, D. G. Analysis of Small-Molecule Interactions Using Biacore S51 Technology. *Anal. Biochem.* **2004**, *329* (2), 316–323.

Myszka, D. G.; Rich, R. L. Implementing Surface Plasmon Resonance Biosensors in Drug Discovery. *Pharmaceutical Science and Technology Today*. 2000, pp 310–317.

Wear, M. A.; Walkinshaw, M. D. Thermodynamics of the Cyclophilin-A/cyclosporin-A Interaction: A Direct Comparison of Parameters Determined by Surface Plasmon Resonance Using Biacore T100 and Isothermal Titration Calorimetry. *Anal. Biochem.* **2006**, *359* (2), 285–287.

Wear, M. A.; Patterson, A.; Malone, K.; Dunsmore, C.; Turner, N. J.; Walkinshaw, M. D. A Surface Plasmon Resonance-Based Assay for Small Molecule Inhibitors of Human Cyclophilin A. *Anal. Biochem.* **2005**, *345* (2), 214–226.

4 X-ray crystallography using Cyp complexes

4.1 *Introduction and Aims*

4.1.1 **Aims of this Chapter**

The aims of this chapter is to study the binding of fragments from our fragment library, on Cyp isoforms using X-ray crystallography. For this study, as in SPR and computational studies, CypA, CypB and CypD were selected, while compounds from the library were selected based on SPR results and research interest for specific compounds. A more specific aim of the chapter is to identify new compounds that can bind Cyps and explain their binding pose and the interaction with protein in the active side. A further aim is to study if some fragments can selectively bind some Cyp isoforms, and if the binding pose of fragments in the active side of Cyp is the same for all isoforms.

4.1.2 **Introduction to X-ray crystallography**

X-ray crystallography is an experimental technique that is used for the identification of the molecular structure of a crystal. This is determined by analysis of diffraction patterns of a beam of incident X-ray light shone on the crystal. This experimental technique has three important steps. The first step (crystallisation) is the most critical and difficult one. It involves the formation of a good quality crystal of the protein of interest. This crystal should grow in 3D and be large (typically larger than 0.1 mm), and the surface of the crystal should be without any cracks, crystal defects or twinning (multiple separate crystals sharing the same or some crystal lattice points). In the second step (data collection) the crystal is hit by an X-ray beam, which is reflected in multiple directions upon interaction with crystal atoms. The reflections and intensities of the scattered beams are collected as the crystal is rotated, for every orientation of the crystal. In the third and last step (data processing, analysis and structure refinement) the data collected is combined with structural data (if, as in the case of cyclophilins, a structure is available), refined previously from X-ray or another experiment for the generation of the final refined crystal structure.

4.1.2.1 Crystallisation

As was highlighted above the crystallisation step is the most important and may be the most difficult step in X-ray crystallography. For a good quality crystal that is critical for the quality of the final X-ray structure, the crystal should be large and single, grown in all 3 directions, have no cracks, defects or twinning. For these result to be achieved there are several factors to be taken into account. These factors include the protein sample to be crystallised, the crystallisation additives or precipitants, and the technique used. Most importantly the protein of interest should be pure and in high concentration. Special care should be taken to prevent any impurities or additives such as detergent or cofactors, from previous steps to remain in the protein sample for crystallisation.

Moreover the crystallisation conditions and precipitants like pH, temperature, salts, organic additives, and polymers can affect the formation of the crystal as well as the quality of the crystals formed. Crystallisation of a protein usually happens in two sequential steps, nucleation and growth steps. Nucleation is the most difficult one as it is the step where the protein should change phase, from the disordered phase that exist in solution to the ordered “nucleation” phase. Nucleation is key for the development of the crystal, and after its formation then the growth steps takes place until the nucleus becomes a fully grown crystal. For both nucleation and growth steps the supersaturation of the drop is the most important factor. Supersaturation is achieved by adding precipitants in the crystallisation buffer. Precipitants can be divided in four main categories: (a) salts, (b) organic solvents, (c) high molecular weight polymers and (d) low molecular weight polymers and not volatile organic compounds (McPherson and Gavira, 2014). The principal effect of salts, such as sulfates, phosphates, acetates, or citrates, are to dehydrate the proteins by sequestering water molecules through H-bond interactions. This promotes the interaction between proteins, their clustering together in an ordered phase and the crystal formation. Organic solvents reduce the dielectric of the medium which causes the reduction in the interactions between solute and protein molecules and the increase in the attraction

Rational design of isoform specific ligands between proteins. Some common organic solvents includes low molecular weight alcohols such as methanol, ethanol, propanol and butanol, dioxane or acetonitrile. Arguably the most common precipitant used of variable chain size is polyethylene glycol (PEG) polymers. This polymer promotes the separation of the protein molecules from the solution through volume exclusion effect. This polymer, in contrast with proteins, has no ordered structure. Since they are very flexible in solution they occupy a much larger volume space than similarly sized ordered structures. Because of this, the protein access to solvent is limited and the protein molecules are forced to aggregate with each other and form a more stable structure that leads to the crystal formation. Lastly pH can affect the solubility of a protein by changing the protonation state of protein residues. This can decrease interactions between water molecules and protein atoms and promote their separation.

Finding the best crystallisation conditions is not an easy process but there are a two major techniques that can be followed (McPherson and Gavira, 2014). If crystallisation conditions are known for similar protein or some crystallisation variables are thought to be particularly important, then a systematic approach can be tried. In a systematic approach one or two crystallisation variables, eg pH and salt concentration, are selected and are systematically changed until the best conditions for the protein crystallisation are found. The second one, the so called “shotgun approach”, is useful if nothing is known about the protein and no previous data exists (Jancarik and Kim, 1991). In this approach a mix and match technique is applied where different crystallisation variables are tested, randomly or based on previous works. Upon identification of crystal formation then the crystallisation conditions can be further optimised using the first approach. For the second approach a number of companies are also offering crystallisation kits with which someone can test hundreds of different conditions.

The last factor that should be taken into account for protein crystallisation is the technique to be used. The major techniques for protein crystallisation are vapour diffusion by hanging drop, vapour diffusion by sitting drop, microbatch under oil and

Rational design of isoform specific ligands microdialysis crystallization. The two most common techniques used, and the two techniques used in this research are the hanging and sitting vapour diffusion methods, Figure 4-1. Both techniques are based on the same principle where water is transferred over time from the drop to the well solution. Over time the well solution will attract water from the drop, using vapour diffusion, until the concentration of reagents in both well and drop reach an equilibrium. An effect of this equilibration process, is the increase in the concentration of the protein in the drop, enabling the supersaturation of the sample.

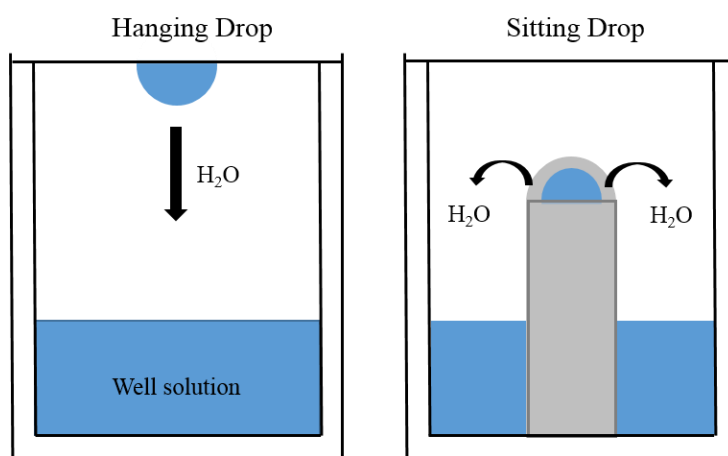


Figure 4-1 Hanging and sitting vapour diffusion crystallisation techniques

4.1.2.2 Data collection

Protein crystals are loaded onto loops, flash frozen in liquid nitrogen and sent for X-ray screening and data collection. The diffracted X-ray beams are collected as the crystal is rotated by 180° and are recorded as black spots on the detector, as can be seen in Figure 4-2. One image of these spots is not enough to reconstruct the crystal structure. Many images from different angles are necessary to ensure that there is enough information for the structure to be defined, so usually a few thousand images are taken per crystal.

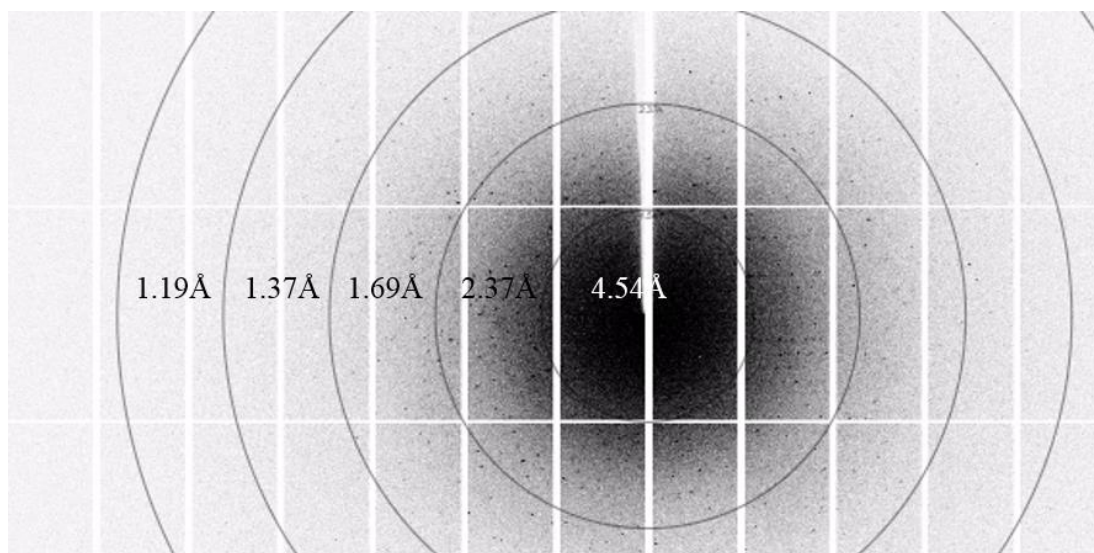


Figure 4-2 Data diffraction pattern of a protein crystal. The spots closer to the centre represents data of low resolution, while the further from the centre is the spot the higher its resolution.

The diffraction of X-rays can be explained using Bragg's law that was first introduced by William Lawrence Bragg and his father William Henry Bragg (Bragg, W. H. Bragg, 1913). In Bragg's law, as can be seen also in Figure 4-3, the crystal can be represented as a set of parallel planes separated by a constant parameter d_{hkl} . According to the law an incident X-ray wave, with wavelength λ , is diffracted from an array of atoms in a primitive lattice producing spherical waves. In most cases diffracted waves interfere destructively and cannot be recorded except in the case where two or more parallel X-ray waves are diffracted from two parallel lattice planes, with distance d_{hkl} , through an angle 2θ and they remain in phase with a phase shift as an integer multiple of their wavelength. This condition can be expressed as:

$$2d_{hkl}\sin\theta = n\lambda \quad \text{Equation 15}$$

Where d_{hkl} is the distance between two parallel planes and hkl are the Miller indices (Miller, 1839), θ is the angle between the incident wave and the lattice plane, λ is the wavelength of the wave and n is an integer, the order of the reflection.

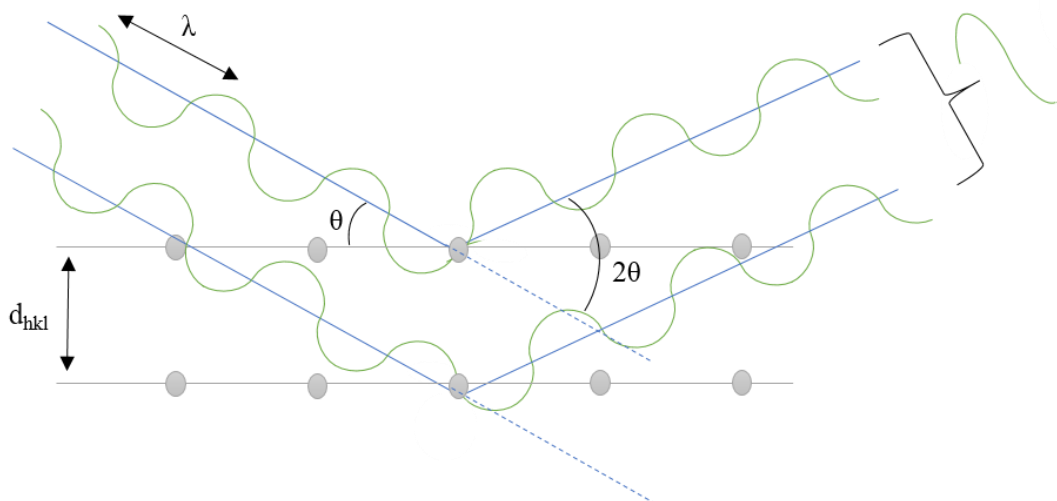


Figure 4-3 Bragg's law with just one atom at each lattice point

If two or more atoms (1, 2 ... n) are present in the unit cell of the lattice then waves will be scattered from all of these atoms in the unit cell. Scattered waves are described from their amplitudes (atomic scattering factors) $f_0, f_1, f_2 \dots f_n$, and their phases (angles) $\varphi_1, \varphi_2 \dots \varphi_n$. Constructive combination of these component waves results in a net (resultant) wave that can be described from amplitude F_{hkl} and phase a_{hkl} . F_{hkl} is obtained from the linear superposition of the scattering factors of each atom and phase a_{hkl} arise from the difference in position of the different atoms (Figure 4-4).

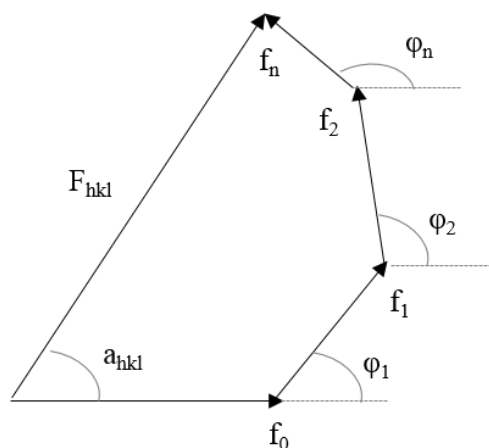


Figure 4-4 Constructive combination of waves incident to different atoms

Combining amplitude F_{hkl} and phase a_{hkl} , the structure factor F_{hkl} of every diffracted beam can be defined. F_{hkl} (the “structure factor”) is a very important quantity as it expresses both the phase and the amplitude of the reflected waves and it has a central role in the refinement of crystal structures (Wallwork, 1980). The mathematical equation that is used for the determination of F_{hkl} is:

$$F_{hkl} = F_{hkl} \exp(ia_{hkl}) = \sum_i f_i \exp[2\pi i(hx_j + ky_j + lz_j)] \quad \text{Equation 16}$$

Where F_{hkl} is the amplitude and a_{hkl} is the phase of the diffracted wave. The sum is over all the atoms of the crystal and x_j , y_j and z_j are the coordinates of the j^{th} atom.

The structure factors of all reflections hkl should be taken into account and Fourier transform methods are used for the determination of the three dimensional electron densities of the crystal structure. But for the different scattered beams to be combined we should know their amplitudes and phases. The direction of the beam can be specified from the Miller indices of the crystal. The amplitude can be calculated from the intensity of the beam that is related to the blackness of the spot. The phase of the beam though is lost during data collection and needs to be determined for the calculation of the structure factors.

4.1.2.3 Structural Refinement

As highlighted above, the ability to determine the crystal structure of a protein depends on the ability to define the structure factors for the atoms in the protein. But as explained before this cannot be done directly from the data collected after crystallography since the phase of the diffracted waves cannot be measured. For that deferent methods exist, such as molecular replacement, in the case where a previously solved structure or homology models exist. Otherwise methods such as anomalous dispersion or isomorphous replacement can reduce the phase problem with the determination of the position of a few heavy atoms socked in the crystal. Model structure can be a previously refined crystal structure of the same protein, or of an

Rational design of isoform specific ligands isoform of that protein, or even a structure modelled *de novo* using different computational tools. Using the model structure then the amplitude of the calculated structure factor $|Fc|$ from the proposed model, is compared with the observed amplitude from the X-ray experiment $|Fo|$ using the Equation17:

$$R = \frac{\sum ||Fo| - |Fc||}{\sum |Fo|} \quad \text{Equation17}$$

Where the summation is for all the reflections of the crystal that give significant intensities.

For a fully refined and solved perfect structure R should be 0, while for a random atom configuration it is ~ 0.6 . The usual R value for a well refined structure is ~ 0.2 . Achieving a perfect model, $R=0$, is basically impossible because of water molecules and disordered protein regions that cannot be accurately modeled. Moreover, this process is inherently biased, since the calculated modeled amplitudes $|Fc|$ are used along with the observed experimental amplitudes $|Fo|$ for the calculation of the electron density and the improvement of the model. So for this reason usually a small subset of experimental reflections, about 10%, is removed from the refinement process and the R value is calculated using the remaining 90%. The rest 10% of the reflections are used for the calculation of the free R (R_{free}). R_{free} is calculated exactly the same way as R and shows how well the model predicts the structure using only the 10% of the reflections that were not used in the refinement process. For a well refined model R_{free} is usually just a little bit higher than R , around ~ 0.26 . So in the refinement process the crystallographer has to do multiple rounds of structure refinement and model validation (R/R_{free} calculation) until the model agrees as much as possible with experimental data.

After the refinement, the coefficients for two types of electron density maps are reported. The first one, $2|Fo|-|Fc|$, is the map that represents the data that exist in the crystal structure, i.e. electron densities of atoms that were correctly modelled in the structure. This map shows the electron densities $\rho(x,y,z)$ at every point in the 3D map,

Rational design of isoform specific ligands and the mean and standard deviation (std) of the entire map is calculated. The intensity of each point, is expressed in standard deviation units (σ) away from the mean intensity of the whole map. Usually the intensity of the $2|Fo|-|Fc|$ map is presented at 1.0σ . A map at the 1.0σ level shows all the points that have a std higher than 1.0 from the mean intensity of the map. Lower σ value show more density at each point of the map but can include noise, while higher values can exclude real data.

The second map that is usually calculated, $|Fo|-|Fc|$, is the map that shows electron densities in the regions of the structure where there are differences between observed and calculated values. Both maps are inspected and should be taken into account during the refinement process.

4.2 *Materials and Methods*

4.2.1 *Materials*

TRIS hydrochloride, PBS tablets, polyethylene glycol (PEG) 8000, PEG 3350, Ammonium Nitrate, Sodium Citrate, Sodium Acetate and Sodium Chloride were obtained from Sigma-Aldrich.

4.2.2 *Methods*

4.2.2.1 *Crystallization of CypA*

Previously purified and his-tag cleaved CypA was concentrated to 30.42 mg ml^{-1} in PBS buffer. The vapor diffusion by hanging drop method in $6 \text{ }^\circ\text{C}$ was used for the crystallization of CypA. The total volume of the well was 1ml and the precipitation solution consisted of 100mM Tris-HCl pH 8.0 and 21 – 24% PEG 8000 (final concentrations). The drop consisted of $1.5 \text{ }\mu\text{L}$ of CypA sample in PBS and $1.5 \text{ }\mu\text{L}$ of well solution.

4.2.2.2 Crystallization of CypB

Previously purified and his-tag cleaved CypB was concentrated to 40 mg ml⁻¹ in PBS buffer. The vapor diffusion by hanging drop method in 6 °C was used for the crystallization of CypB. The total volume of the well was 1 ml and the precipitation solution consisted of 250mM Ammonium Nitrate and 22 – 24% PEG 3350 (final concentrations). The drop consisted of 1.5 µL of CypA sample in PBS and 1.5 µL of well solution. After 24 hours, if no crystals were observed, drops were seeded using a micro-seeding (streak seeding) technique using as seed stock a drop with a previously grown CypB crystal.

4.2.2.3 Crystallization of K133I CypD

Wild type CypD is not crystallizing easily so a K133I mutant was used. Previously purified and his-tag cleaved K133I CypD was concentrated to 42 mg ml⁻¹ in PBS buffer. The vapor diffusion by hanging drop method in 18 °C was used for the crystallization of K133I CypD. The total volume of the well was 1 ml and the precipitation solution consisted of 100mM Sodium Cacodylate pH 6.5, 200mM Sodium Acetate and 32% PEG 8000 (final concentrations). The drop consisted of 1.5 µL of K133I CypD sample in PBS and 1.5 µL of well solution. The small and fine needles observed were only useful for seeding. So 2µL of the drop were taken and diluted with 98 µL of well solution. The solution was spun to crush the needles and the solution was saved as seeding solution for future drops.

4.2.2.4 Co-crystallization of K133I CypD

Previously purified and his-tag cleaved K133I CypD was concentrated to 40mg ml⁻¹ in PBS buffer. The vapor diffusion by hanging drop method in 18 °C was used for the crystallization of K133I CypD. Ligand solution of **compound 1** from (Daum et al., 2009b) was prepared using 5% of 25mM **compound 1** in 100% EtOH, and 95% of precipitation buffer. The total volume of the well was 1ml and the precipitation solution consisted of 45mM Sodium Citrate pH 2.9 and 19% PEG 3350 (final

Rational design of isoform specific ligands concentrations). The drop consisted of 0.9 μ L of K133I CypD sample in PBS, 1.0 μ L of well solution, 0.2 μ L of seeding solution 0.1 μ L of ligand solution.

4.2.2.5 Soaking of crystals and data collection

Each ligand was introduced into an *apo* Cyp crystal using a one-step soaking procedure. For the crystal-soaking experiments, a crystal of Cyp was placed into a 2 μ L drop of soaking solution composed of 30% PEG 8000, 100mM Tris-HCl, buffer pH 8.0 and saturating concentrations of ligands (50mM - 100mM of ligands) and then flash-frozen in liquid nitrogen. For *apo* Cyp, a crystal of Cyp was placed into a 2 μ L drop of soaking solution composed of the same buffer as before but without any ligand concentration.

X-ray intensity data were collected from crystals at the Diamond synchrotron-radiation facility in Oxfordshire, England. Intensity data for *apo* Cyp and Cyp-ligand complexes were collected from single crystals flash cooled in liquid nitrogen at 100K. Data were processed with MOSFLM (Battye et al., 2011) and scaled with SCALA (Evans, 2006). The data-collection and processing statistics are summarized in Appendix Table 7-3.

4.2.2.6 Structure determination

All structures were originally processed with DIMPLE (Difference Map PipeLinE), a CCP4 supported software (Winn et al., 2011), which is an automated software pipeline that rapidly process crystals of known proteins. DIMPLE was followed by multiple cycles of restrained refinement using REFMAC 9 (Murshudov et al., 1997, 2011). The side chains of the models were manually adjusted while ligands and water molecules were added where appropriate using Coot 10 (Emsley et al., 2010). After further cycles of restrained refinements and manual adjustments of ligands, side chains and water molecules, the overall quality of the electron density maps were improved. The refinement statistics are summarized in Appendix Table 7-3.

4.3 Results and Discussion

4.3.1 Crystallization of CypA

Crystallisation of CypA was achieved as described in the methods section above. More than 40 ligands from our fragment library were soaked or co-crystallised with CypA, out of which 36 ligands yielded good diffraction data. Among these compounds that were successfully crystallised and yielded good diffraction data but were not observed in the CypA crystal are: **4, 7, 38, 49, 53, 55, 58, 59, 63, 64, 65, 66, 85, 94, 95** and **96** (Figure 4-6). Encouragingly the structure of 16 fragments have been successfully been solved in the active site of CypA. Fragments that were found to bind CypA are: **1, 2, 3, 5, 9, 16, 49, 56, 60, 61, 81, 89, 91, 97, 98** and **99**. The structure for all of these compounds can be seen in Figure 4-5. These 16 compounds are binding in either the Abu or Pro CypA pockets, depending on their H-bond donors and acceptors, and also the hydrophobicity of each ligands.

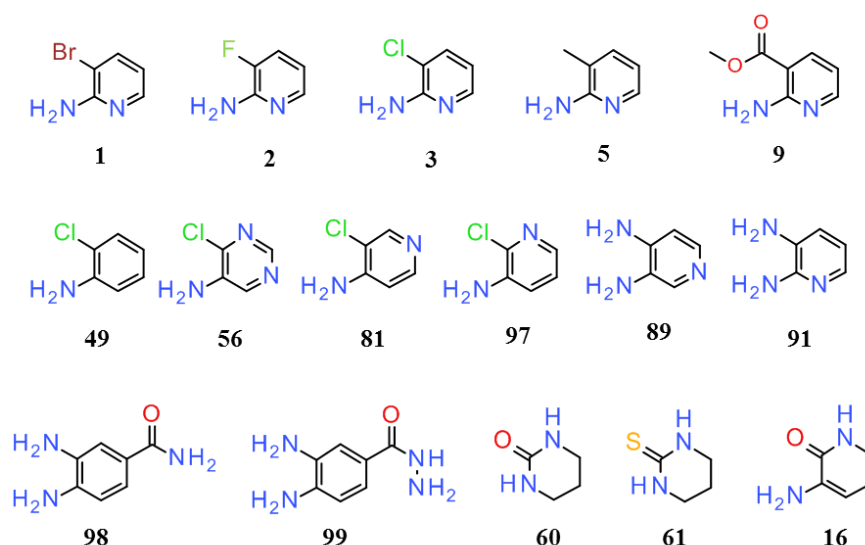


Figure 4-5 Fragments that bind CypA based on X-ray crystallography. Data collection, refinement and Ramachandran plot statistics for one crystal structure for each of these complexes are reported in the Appendix 7.7 in Table 7-3.

Rational design of isoform specific ligands

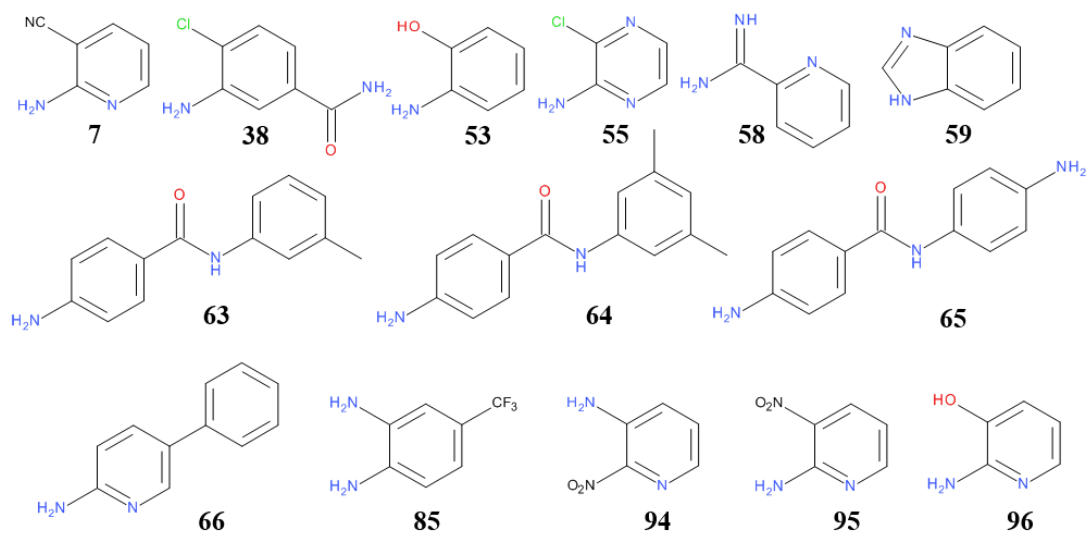


Figure 4-6 Fragments that do not bind CypA based on X-ray crystallography

4.3.2 Binding of compounds in the Abu pocket of CypA

Ligands with hydrogen bond donors and acceptors such as **1**, **2**, **3**, **5**, **9**, **49**, **56**, **81**, **89**, **91**, **97**, **98** and **99** are binding in the more hydrophilic Abu pocket of CypA. All of these compounds bind with the same pose in the protein pocket and they establish similar interactions with CypA. An example can be seen in Figure 4-7 using fragment **98**.

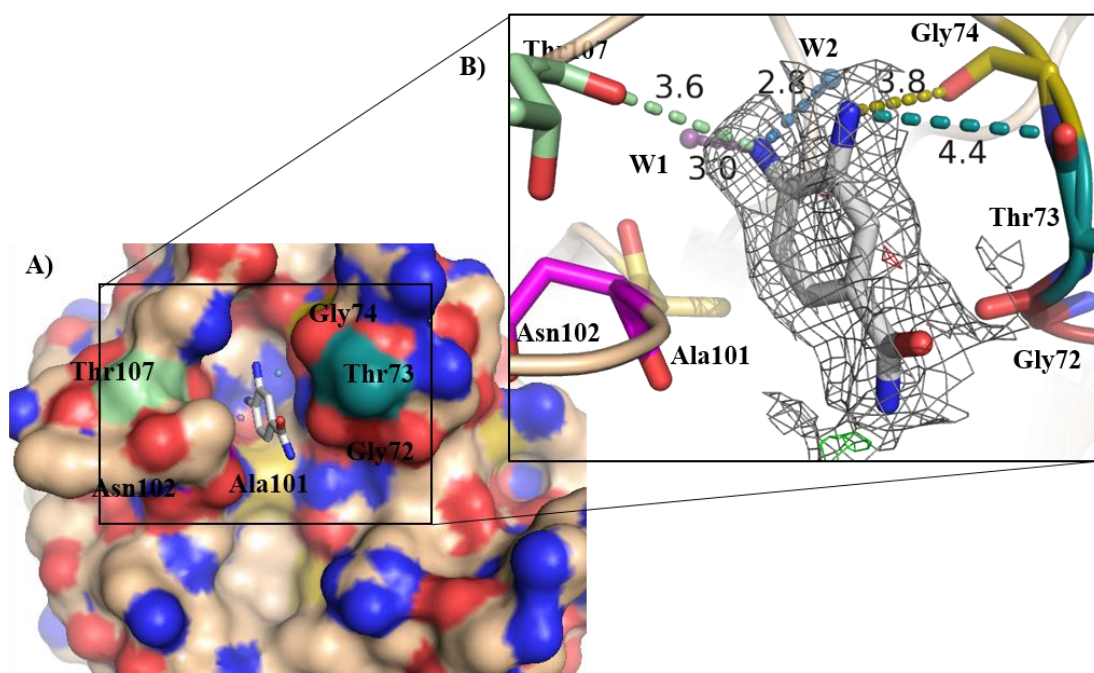


Figure 4-7 Binding of 98 in the Abu pocket of CypA. A) 3D surface structure of **98** binding in the Abu pocket of CypA. B) $2F_o - F_c$ electron density and $F_o - F_c$ maps of **98** bonded to the Abu pocket CypA, at 1.0σ and 3.0σ are shown in black and red/green respectively. Distances between ligand with water and CypA molecules are highlighted and colour coded based residue colour.

In this figure we can see that the amine substituent of the ring of **98** is able to establish H-bond interactions with the two tightly bonded water molecules (W1 and W2) in the Abu pocket. In turn these water molecules are acting as H-bond bridges between the ligands and CypA. The ligand do not seem to establish direct H-bonds with the protein because the distance and angle between ligands and protein residues, such as Thr107 and Gly74, is slightly bigger than the optimal distance for H-bonding. In contrast, these distances allow the van der waals and/or Lennard-Jones interactions between protein and ligand. The other fragments, **1**, **2**, **3**, **5**, **9**, **49**, **56**, **81**, **89**, **91**, **97** and **99** have a very similar pose to **98**, in the Abu pocket of CypA, establishing similar H-bonds as **98**. Structure of each of these compounds in the pocket as well as their electron density and distances between fragments and protein/water molecules can be seen in Appendix 7.5 Figure 7-2 and Figure 7-3

Rational design of isoform specific ligands

By visual examination of the structure of all fragments that bind in the Abu pocket and also from their distances with key water and protein residues as shown in Figure 4-7 and in Figure 7-2 and Figure 7-3 in Appendix 7.5, it seems that the amine substituent of the aromatic ring, as well as the aromatic the ring itself, play an import role for the binding of these fragments on the protein. In this common binding pose this amine substituent is hidden in the pocket, to facilitate the H-bond interactions with the conserved water molecules, leaving the other side chain of the ring, in the ortho position, more water exposed.

In this ortho position of the ring, and based on the structures of all fragments, Appendix 7.5, it seems that most of the small functional groups, with 1 heavy atom can be facilitated, Table 4-1. This includes all halogens (F, Cl, Br and I), H-bond donors such as amines, and hydrophobic groups (methyl). Only in one case, **9**, a functional group with more than 1 heavy atoms in this position of the ring is able to bind. This can be an effect from the two neighbouring loops, 70's and 80's loops, that are very close to the water exposed part of the fragments. Figure 4-8 and Table 4-1, show the distances between the R₂ site chain of the ligands and the neighbouring residues Thr73, Gly74 and Lys82, on the 70's and 80's loop respectively.

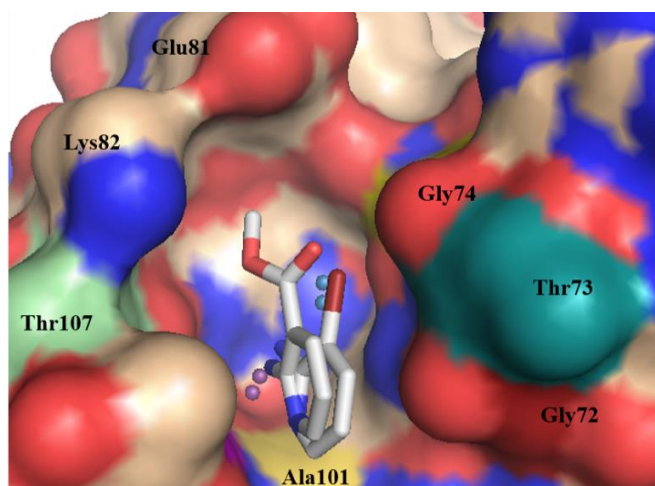
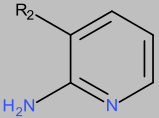
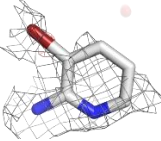
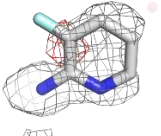
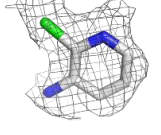
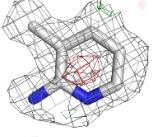
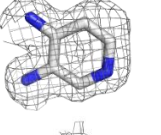
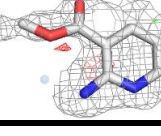


Figure 4-8 Overlay between **1** and **9** in the Abu pocket of CypA. Carbon atoms are colour coded in grey and oxygen, nitrogen and bromine atoms in bright red, blue and dark red respectively.

Rational design of isoform specific ligands

Table 4-1 Distances between ligands' side chain (R₂) and protein residues in the 70's and 80's loops. 2Fo – Fc and Fo – Fc electron density maps of all ligands, at 1.0 and 3.0 σ, are shown in black and red/green respectively. SPR affinity in mM and stoichiometry can also be seen (Not determined are compounds that were not tested on a 2 fold SPR dilution series).

Ligand number	Ligand structure and e- density map	R ₂	Distance (Å)			SPR affinity (mM) / stoichiometry
			Distance (Å)			
			R ₂ -Thr73	(carbonyl oxygen)-Gly74	(carbonyl oxygen)-Lys82	
						
1		Br	3.4	3.5	5.6	Not determined
2		F	4.5	4.0	4.8	9 / 1:1
97		Cl	3.8	3.6	5.0	1 / <1:1
5		Me	3.6	3.8	4.8	Molar affinity
91		NH ₂	4.5	3.9	4.7	Molar affinity
9		COCH ₃	3.3	3.7	4.0	Not determined

The distances between the R₂ side chain and the 70's and 80's loops are 3.5 - 4.5 Å and 4.5 - 5.5 Å respectively. In the case of **9** those distances are 3.5 - 4.0 Å respectively. These close proximity between the fragments the 70's and 80's loops, indicating that these loops are able to facilitate or impede fragment binding in the Abu pocket through non-bonded, electrostatic or van der Waals interactions.

A crystallographic artefact that was observed during the refinement of structures of **1**, **2**, **3**, **5**, **56** and **81** is the binding of these fragments not only in the Abu pocket of CypA but also out of the pocket close to the 80's loop (Figure 4-9 and Figure 7-4 in Appendix 7.6). This second binding position is totally water exposed and is in-between CypA and a symmetric copy of the protein. Binding in that position seems to be facilitated from a H-bond between the amine substituent of the fragments and the side chain of Glu84 as well as from a halogen bond between the halogens of these compounds and the backbone carbonyl of Asn106, Figure 4-9 and Figure 7-4. Moreover Trp121 of a symmetric CypA is in very close proximity, and establishes short van der Waals interaction with the aromatic ring of the compounds. The second molecule of these fragments can be seen in the refined structures and in the Appendix Table 7-3 but is not considered as a second binding pose but as a crystallographic artefact.

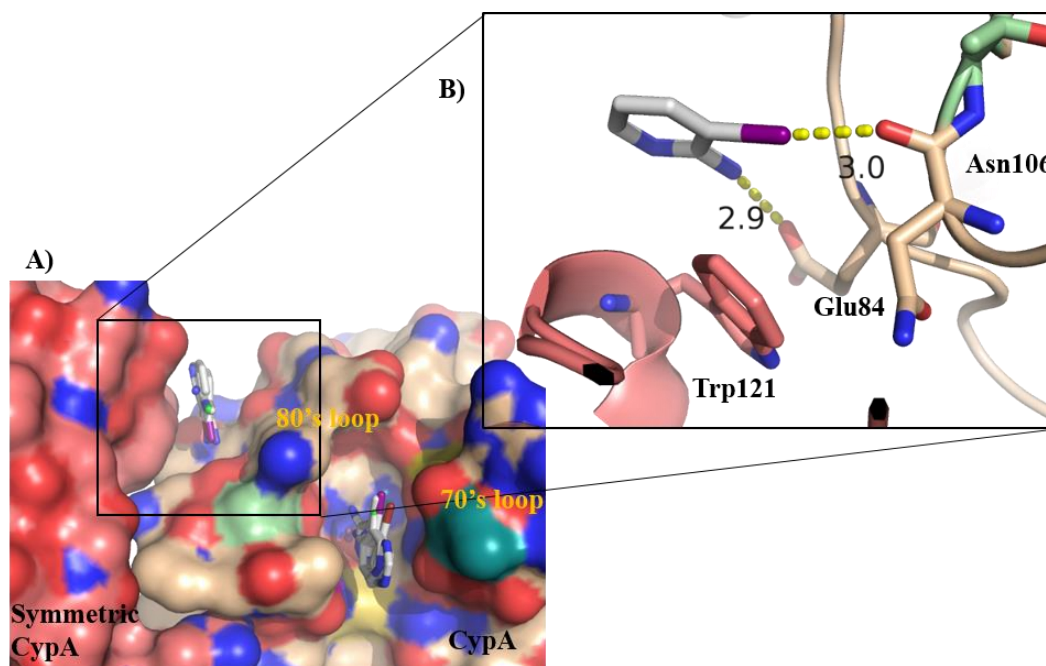


Figure 4-9 Binding of fragments out of the 80's loop of CypA. **A)** Overlay of **1**, **2**, **3**, **5**, **56** and **81** binding on a 3D surface structure of CypA. A symmetric CypA structure can be seen in pink. **B)** shows interactions between **4** with Glu84 and Asn106 residues of CypA and a close van der Waals interactions to Trp121 of the symmetric CypA.

4.3.2.1 Pyridine, benzene or pyrimidine

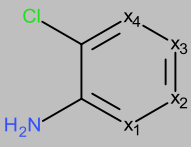
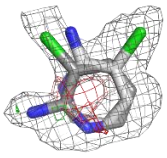
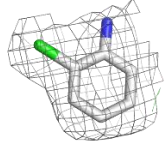
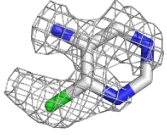
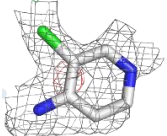
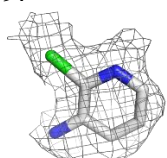
Looking at the chemical structures of these fragments, Figure 4-5, we can notice that many of them are almost identical with the only change between them being an atom in their aromatic ring. For example compounds **3**, **97**, **56**, **49**, **81** have one chloride and one amide substituents in ortho positions, and what changes is the ring which can be a benzene, pyridine or pyrimidine.

All these ligands are able to bind in the Abu pocket with the position of their aromatic ring to be mostly conserved. On the other hand the position of the substituents on the ring can be changed. This can be observed from the X-ray structures of each fragment binding in the Abu pocket of CypA, Figure 7-2 and Figure 7-3 in Appendix 7.5 and Table 4-2. Ligands **97** and **81** have the same binding mode, as discussed before in paragraph 4.3.2, where the amine substituent of the ring is hidden in the pocket and the chlorine substituent is more water exposed. Ligands **56** and **49** adopt a slightly different binding mode. **49** is rotated almost by 90° in the pocket while **56** has the chlorine substituent hidden in the pocket and the amine substituent more water exposed (exactly the opposite compared to the other compounds i.e. **97** and **81**). Ligand **3** seems to be able to partially adopt more than one binding modes. Table 4-2 shows the structure of each of these fragments and the distances of their side chains (-NH₂ and -Cl) from water and protein residues. These distances are very similar between all these compounds and are always $\leq 4.0\text{\AA}$ which is ideal for electrostatic interactions. These figure and the data on the table show that all pyridine, benzene and pyrimidine rings are tolerated in the Abu pocket of CypA and the lipophilic interactions between the aromatic ring and the protein, in combination with the electrostatic interactions of the substituents on the ring, play in important role for their binding.

All these fragments bind exclusively in the Abu pocket, except **49** that shows binding in the Abu and Pro pocket too. This double binding behaviour suggests that the lack of a nitrogen atom in the ring can have an effect on the binding of **49** on CypA, allowing it to bind both Abu and Pro pockets. Also SPR results from the previous

chapter verifies the double binding of **49** on CypA, which shows 1:2 stoichiometry, Appendix 7.3 Table 7-2, and a more than ten-fold reduction in dissociation constant of **49** compared to **3**, **97** or **56**, Appendix 7.3 Table 7-2.

Table 4-2 Distances between fragments' side chains (–NH₂ and –Cl) and protein/water residues. 2Fo – Fc and Fo – Fc electron density maps of all ligands, at 1.0 and 3.0 σ , are shown in black and red/green respectively. SPR affinity in mM and stoichiometry can also be seen (Not determined are compounds that were not tested on a 2 fold SPR dilution series).

Ligand number, structure and e- density map	Distance (Å)								SPR affinity (mM) / stoichiometry	
	X ₁	X ₂	X ₃	X ₄	Distance (Å)					
					NH ₂ -W ₁	NH ₂ -W ₂	Cl-Thr73	Cl-Gly74		
										
3		N	C	C	C	3.6	2.6	3.8	3.6	4 / 1:1
49		C	C	C	C	-	-	-	-	24 / 1:2
56		C	N	C	N	4.0	4.0	4.3	3.5	1 / <1:1
81		C	C	N	C	3.5	3.2	3.9	3.6	Not determined
97		C	C	C	N	3.0	2.7	3.8	3.6	1 / <1:1

Another fragment that binds slightly different from the others is **3**, Figure 4-10. Refinement of **3** in the same pose as previous compounds, as shown in Figure 4-10 (A), $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ density maps show differences between the observed and calculated values. Further refinement of **3** using a different pose in the Abu pocket of CypA, such as pose B, Figure 4-10 (B), gives slightly less differences between the observed and calculated values. Refining the structure of **3** using both poses together, as shown in Figure 4-10 (A+B) gives similar results to (B) but still not perfect. This difficulty to reduce the differences between observed and calculated value, as will be seen later in the free energy chapter of this thesis, is because **3** adopts both poses in Abu pocket, and not just one of them.

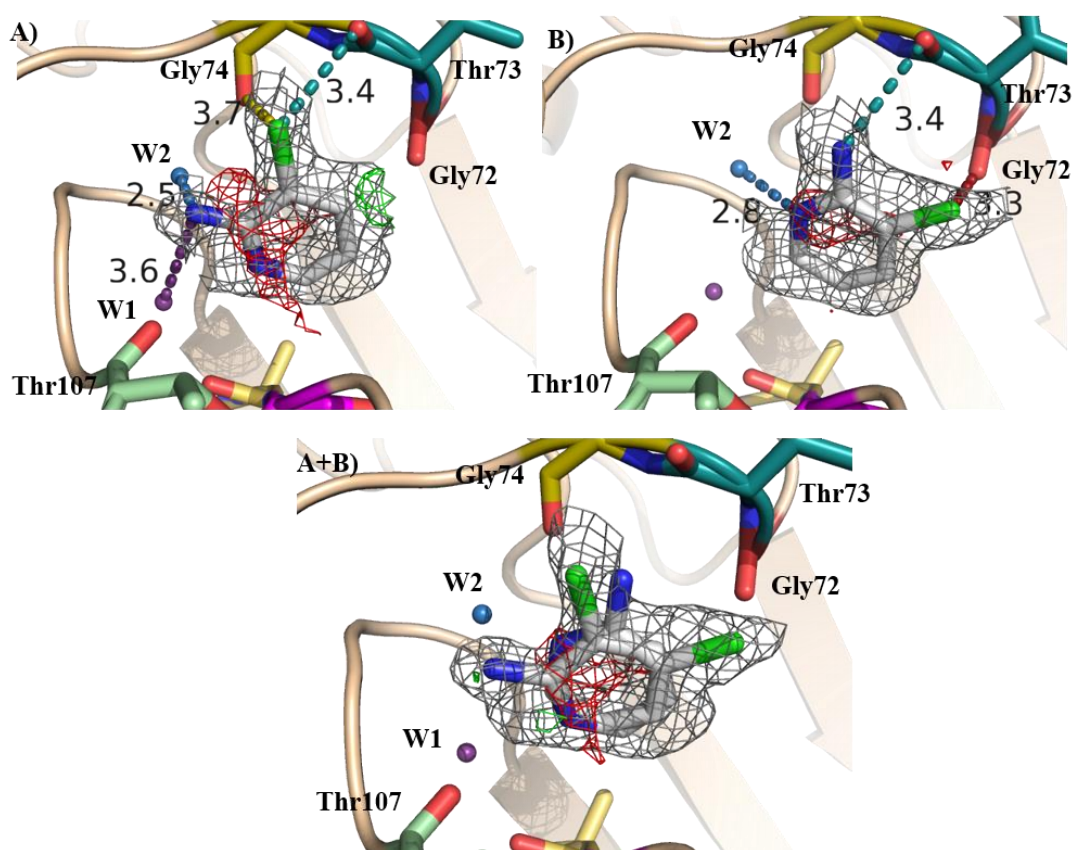


Figure 4-10 Possible binding poses of **3 in the Abu pocket of CypA.** $2F_o - F_c$ and $F_o - F_c$ electron density maps, at 1.0 and 3.0 σ , are shown in black and red/green respectively. Distances between ligand with water and CypA molecules are highlighted and colour coded based residue colour.

Rational design of isoform specific ligands

Two other fragments, with a very similar structure, are compounds **91** and **89**. Both are diaminopyridines but they have the same side chains in different positions on the ring. If we rotate these compounds by 180° in the Abu pocket the only change would be the nitrogen atom of the ring. This can make the refinement of their crystals and the decision of where the nitrogen of the ring is more difficult. Nevertheless the high quality of the crystals, in some cases, allow the determination of the exact position of the nitrogen atom.

This can be seen from Figure 4-11. In this figure we can see that by rotating **89** by 180° in the pocket, from (A) to (B), $|F_o| - |F_c|$ density map shows a positive difference between the observed and calculated values suggesting that density is missing in the area highlighted in green. When **89** is rotated from (B) to (A) this difference disappears. That indicates that the correct pose of **89** in the pocket is (A). This is supported from the H-bond interaction between **89** and W3 water molecule in pose (A). The same is not happening though with **91**. Both $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ density maps in both poses (C) and (D) are very similar, Figure 4-11. In this case is almost impossible to decide where the nitrogen atom of the ring should be.

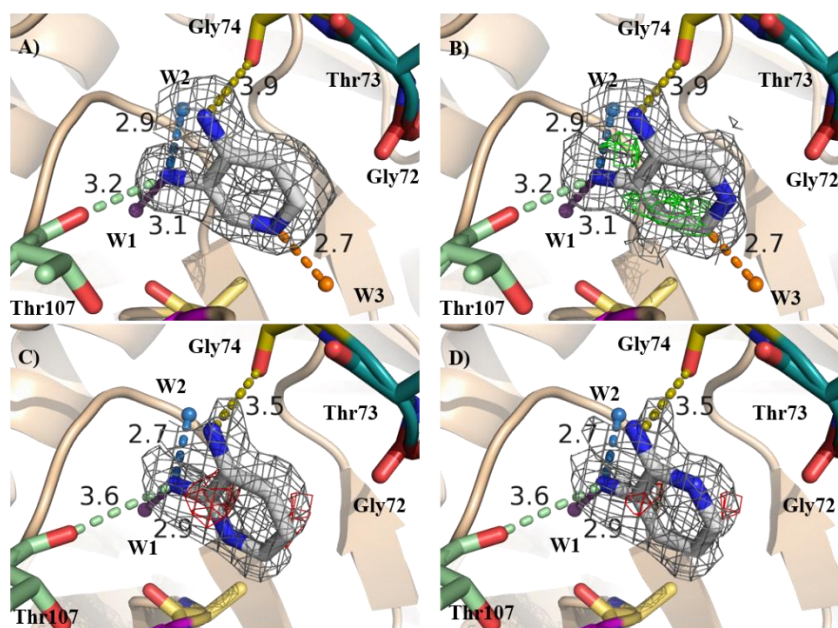


Figure 4-11 Two possible poses of binding of **89** and **91** in the Abu pocket of CypA. **A** - **B**) Electron density maps, $2|Fo| - |Fc|$ in grey at 1.0σ and $|Fo| - |Fc|$ in green/red at $3.0/3.0\sigma$, of two possible binding poses of **89** in the Abu pocket. The only difference between them is the nitrogen atom of the ring, which is rotated by 180° between the two poses. **C**) and **D**) show the two different possible poses for **91**. Distances between ligands with water and CypA molecules are highlighted and colour coded based residue colour.

A difference between **89** and **91** that plays an important role in the binding pose of these fragments, is a water molecule (W3). W3 as can be seen in Figure 4-11 (A), is in H-bond distance (2.7 \AA) to the nitrogen atom of **89** ring. When the ring flips from (A) to (B) this H-bond is broken. In contrast to **89**, **91** does not seem to form any H-bond interaction with any protein or water molecules.

4.3.3 Binding of compounds in the Pro pocket of CypA

The last three fragments, **16**, **60** and **61**, in contrast to the other fragments that bind in the Abu pocket, show a clear binding in the Pro pocket of the protein. All of them establish H-bond interactions in the pocket with the conserved active side Cyp residues (Arg55, Gln63 and Asn102) as can be seen in Figure 4-12 below.

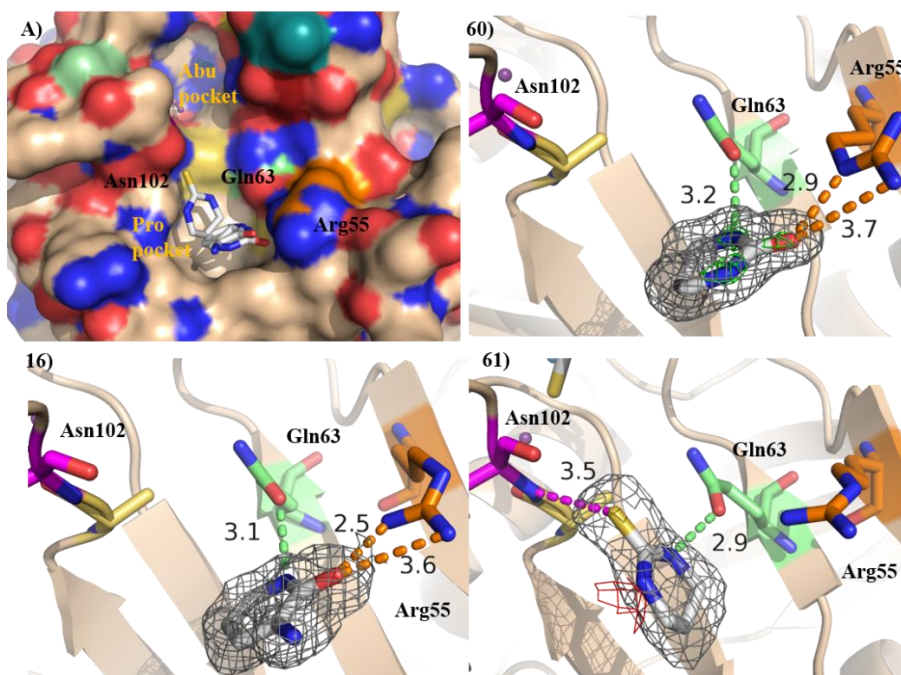


Figure 4-12 Binding of 16, 60 and 61 in the Pro pocket of CypA. A) Overlay of **16**, **60** and **16** binding on a 3D surface structure of CypA. **16** – **61**) Interactions between **16**, **60** and **61** with Arg55, Gln63 and Asn102 residues of CypA in the Pro pocket. $2F_o - F_c$ electron density and $F_o - F_c$ maps, at 1.0σ and 3.0 are shown in grey and red/green respectively. Distances between ligands and CypA residues are highlighted and colour coded based residue colour.

60 and **61** have a non-aromatic six membered ring in contrast to the fragments that bind in the Abu pocket. The loss of aromaticity of the ring seems to give some extra hydrophobicity to the fragments that favours their binding in the Pro pocket of the protein. It is surprising though that with just an atom change between **60** and **61**, from oxygen to sulphur respectively, the binding pose is changing, as can be seen from Figure 4-12. Moreover the activity and stoichiometry of binding seems to be different between **60** and **61**. **60** is binding exclusively to the Pro pocket, while **61** binds in the Pro (Figure 4-12) pocket but also may bind partially in the Abu pocket. This difference between **60** and **61** is also confirmed from SPR measurements. Table 7-2 in Appendix 7.3 show that **60** has a binding affinity and stoichiometry of 4.63 mM and 1:1 respectively, while **61** has 5.34 mM and 1:6 respectively. So the oxygen atom of **60** seems to be key for its binding position and its activity.

4.3.4 Crystallization of CypB

CypB crystallization was performed as described in the methods section above. Crystallization of this isoform was facilitated using the micro-seeding – streak seeding technique. The resulting crystals had the form of needles that are not usually the ideal for crystallography. Nevertheless those needles were not fragile and were big enough (> 0.2mm) and could be used in soaking experiments, Figure 4-13.

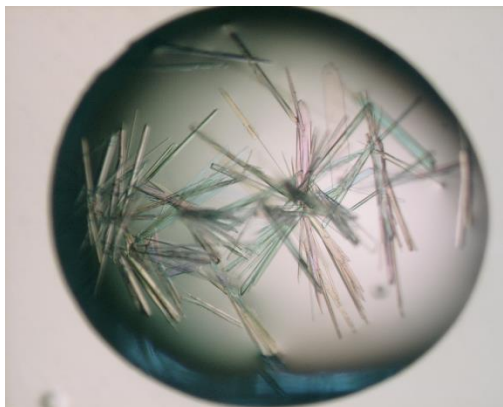


Figure 4-13 CypB crystals

Diffraction data was collected and crystal structures were successfully refined for the *apo* CypB crystal as explained in the methods section above. Moreover CypB crystals were soaked using many small fragments from our library including **3**, **5**, **25**, **29**, **56**, **60**, **63**, **77**, **89**, **91**, **97** and **98**. For all of these compounds crystals yielded good diffraction data, with resolution (< 2.0 Å) and the data were refined using the *apo* CypB structure solved before. From the solved structures no ligand was found to bind in the active side of CypB, in contrast to CypA. The reason for this, as was found during the first step of the structure refinement of the soaked crystals, was the presence of two chains of CypB in a unit cell, Figure 4-14. Close inspection and further refinement of the crystals showed that the N-terminal of the second chain of CypB is binding in the active side of the first chain. Figure 4-14 shows the crystal structure of CypB refined from a crystal of *apo* CypB soaked with ligand **97** from our fragment library.

As can be seen from Figure 4-14 the N-terminal region of chain B establishes a lot of H-bonds in the active side of chain A, especially in the proline pocket and the saddle point of the active side. Furthermore water molecules and chain intramolecular interactions facilitate these binding. Although chain B cannot be seen to extend until the Abu pocket of chain A, the Abu pocket is also closed. This is because the side chain of Arg122 bent over the Abu pocket, establishing H-bonds with the 70's loop of the protein. So naturally the binding of these fragments in the active side of CypB is prevented.

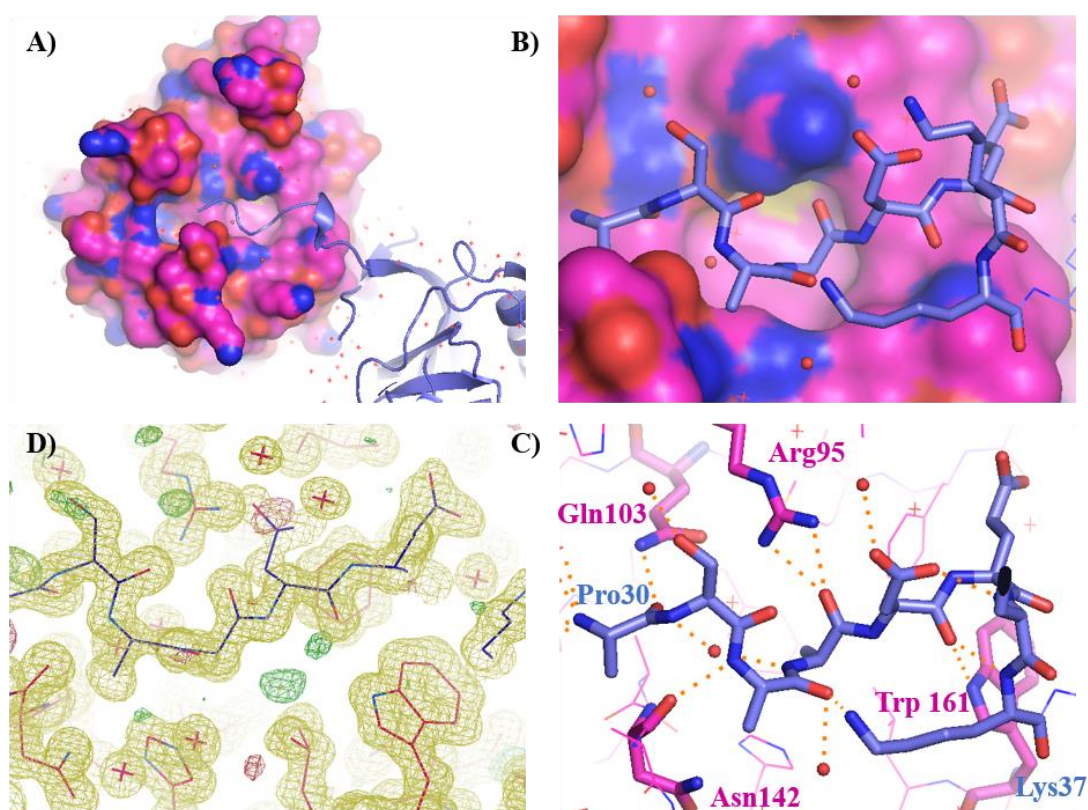


Figure 4-14 Crystal structure of *apo* CypB solved from an *apo* crystal of CypB soaked with 97. **A)** - **B)** The N-terminal of chain B CypB (shown in blue) binds in the active side of chain A CypB (shown in magenta). **C)** H-bonds between chain A and chain B that includes most of the Cyp catalytic residues, Trp161, Asn142, Arg95 and Gln103. **D)** Electron density map $2|Fo| - |Fc|$ (shown in yellow mesh), at 1.0σ , shows the binding of the chain B N-terminal in the active side of chain A CypB.

It seems that for further studies of binding of these fragments in the active side of CypB, a shorter sequence of CypB is needed. A new CypB construct should exclude the first ~10 - 12 amino acids of the N-terminal that binds in the active site.

An interesting observation though, during crystallisation and structure refinement of CypB, is that the presence of organic compounds, such as these small fragments facilitate and favoured these interaction between chain A and chain B. After examination of the electron density maps of the *apo* and fragment soaked CypB crystals, the two maps seems to have some differences. In the crystals that were soaked with small fragments, the electron densities of the N-terminal atoms of chain B are clearly highlighted in the active side of chain A as Figure 4-14 shows. In the *apo* crystal though these electron densities are much weaker and the structure of the N-terminal cannot fully refined. The resolution and refinement statistics are also slightly better for the crystals soaked with different small compounds.

4.3.5 Crystallization of CypD

Crystallisation of CypD was the most challenging among the three isoforms tested in this thesis. First of all as was previously published from Schlatter *et al.*, wild type (WT) CypD is difficult to crystallise (Schlatter *et al.*, 2005). For this reason a mutant is required to facilitate the protein crystallising and this mutation is K133I. A lot of researches reported the use of this CypD mutant in soaking or co-crystallisation experiments with different ligands (Kajitani *et al.*, 2008; Valasani *et al.*, 2014a; Gelin *et al.*, 2015).

For the crystallisation experiments in this thesis different crystallising conditions reported from the previously published articles were tested. In more details two different conditions reported from Valasani *et al.* have been tried using both hanging and sitting drop techniques. Unfortunately none of them worked for the crystallisation of the *apo* K133I CypD. Finally suggested conditions, chapter 4.2.2.3, from Dr Jacqueline Dornan, yielded some very small and fragile needles. These conditions are

Rational design of isoform specific ligands similar to the previously reported conditions from (Kajitani et al., 2008). The resulting drop with the fragile needles was used as a seeding solution and not for soaking experiments as the crystals were not good enough for this experiment. Co-crystallisation conditions of K133I CypD with a ligand from (Daum et al., 2009b) seemed to facilitate the crystallisation of the protein. Although, the resulting needles were almost as fine and fragile as before.

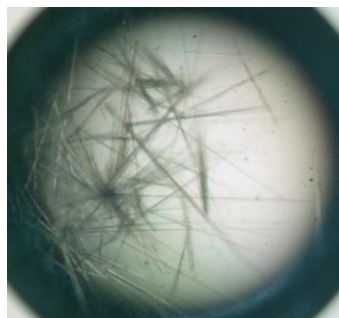


Figure 4-15 K133I CypD co-crystallised with a ligand from Daum et al., 2009

A crystal was successfully used for data collection but the refined crystal structure revealed only the *apo* CypD and no ligand binding in the active side. In their paper Daum et al. have tested analogues of compound **1** for activity on CypA, CypB and CypD. According to their enzymatic activity results, **1** has shown sub micromolar activity on CypA ($0.52 \pm 0.15 \mu\text{M}$) and CypD ($2.42 \pm 0.76 \mu\text{M}$) but no activity on CypB ($>100 \mu\text{M}$). Surprisingly from our crystallographic results no binding was observed of **1** on a CypD crystal, while data collection from CypA and CypB co-crystallised with this compound was not successful.

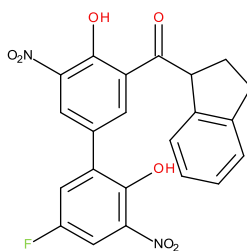


Figure 4-16 Chemical structure of compound 1 from Daum et al., 2009

Details of crystal and refinement statistics can be seen in the Table 7-3. Soaking experiments of the mutant CypD with fragments from our library were not possible. This is because of the nature of the *apo* K133I CypD crystals as explained before, and further effort is needed for the optimisation of the crystallisation conditions of the mutant CypD.

4.4 Conclusion

In this chapter the binding of ligands from our fragment library on Cyp isoforms was examined using X-ray crystallography. In this experiments three Cyp isoforms have been used: CypA, CypB and CypD. More than forty compounds were tested for activity on CypA out of which seventeen found to bind in the active site of CypA. Twelve compounds (**1**, **2**, **3**, **5**, **9**, **56**, **81**, **89**, **91**, **97**, **98** and **99**) found to bind exclusively in the Abu pocket, two (**16** and **60**) in the Pro pocket while two of them (**49** and **61**) may bind partially in both Abu and Pro pockets. The driving force for binding of compounds in the Abu pocket seems to be the amine substituent of the ring. This amine substituent makes H-bond interactions with tightly bonded water molecules in the pocket. Ortho to that amine substituent, halogens, methyl or amines groups can be facilitated while any ring type, e.g. pyridine, pyrimidine or benzene can bind in the pocket. Increasing the hydrophobicity of the compounds, favours the binding in the Pro pocket, as 16, 60 and 61 which stabilise their binding with H-bond interactions with Arg55, Asn102 and Gln63, Figure 4-12.

Rational design of isoform specific ligands

Binding of compounds was also examined in CypB and CypD isoforms. Unfortunately it was not able to identify any fragments binding in these proteins for different reasons. CypB has a different space group and a longer N-terminal chain, which binds in the active side of another copy of CypB and prevents the binding of any fragments. On the other hand crystallisation of CypD was found to be more difficult than anticipated and more efforts are needed for optimisation of the crystallisation conditions of CypD.

Last but not least a compound reported in the literature from Daum et al., as a selective inhibitor of CypA and CypD, has been tested. Although crystallisation attempts of this compound with CypD were successful unfortunately no binding was observed into the active side of CypD. Despite its strong enzymatic inhibition, based on Daum et al., the structure of these compounds are not observed crystallographically.

4.5 References

- Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. W. iMOSFLM: A New Graphical Interface for Diffraction-Image Processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 271–281.
- Bragg, W. H. Bragg, W. L. The Reflection of X-Rays by Crystals. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1913**, 428–438.
- Daum, S.; Schumann, M.; Mathea, S.; Aumuller, T.; Balsley, M. A.; Constant, S. L.; De Lacroix, B. F.; Kruska, F.; Braun, M.; Schiene-Fischer, C. Isoform-Specific Inhibition of Cyclophilins. *Biochemistry* **2009**, *48* (26), 6268–6277.
- Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (4), 486–501.
- Evans, P. Scaling and Assessment of Data Quality. In *Acta Crystallographica Section D: Biological Crystallography*; 2006; Vol. 62, pp 72–82.
- Gelin, M.; Delfosse, V.; Allemand, F.; Hoh, F.; Sallaz-Damaz, Y.; Pirocchi, M.; Bourguet, W.; Ferrer, J.-L.; Labesse, G.; Guichou, J.-F. Combining 'dry' Co-Crystallization and In Situ Diffraction to Facilitate Ligand Screening by X-Ray Crystallography. *Acta Crystallogr. Sect. D* **2015**, *71* (8), 1777–1787.
- Jancarik, J.; Kim, S. H. Sparse Matrix Sampling. A Screening Method for Crystallization of Proteins. *J. Appl. Crystallogr.* **1991**, *24* (pt 4), 409–411.
- Kajitani, K.; Fujihashi, M.; Kobayashi, Y.; Shimizu, S.; Tsujimoto, Y.; Miki, K. Crystal Structure of Human Cyclophilin D in Complex with Its Inhibitor, Cyclosporin A at 0.96-Å Resolution. *Proteins* **2008**, *70* (4), 1635–1639.
- Laskowski, R. A.; Swindells, M. B. LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **2011**, *51* (10), 2778–2786.
- McPherson, A.; Gavira, J. A. Introduction to Protein Crystallization. *Acta Crystallogr. Sect. F Structural Biol. Commun.* **2014**, *70* (1), 2–20.
- Miller, W. H. *A Treatise on Crystallography*; J. & J. J. Deighton: Cambridge, 1839.
- Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular

Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D: Biological Crystallography*. 1997, pp 240–255.

Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 355–367.

Schlatter, D.; Thoma, R.; Kung, E.; Stihle, M.; Muller, F.; Borroni, E.; Cesura, A.; Hennig, M. Crystal Engineering Yields Crystals of Cyclophilin D Diffracting to 1.7 Å Resolution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2005**, *61* (5), 513–519.

Valasani, K. R.; Carlson, E. A.; Battaile, K. P.; Bisson, A.; Wang, C.; Lovell, S.; Yan, S. S. High-Resolution Crystal Structures of Two Crystal Forms of Human Cyclophilin D in Complex with PEG 400 Molecules. *Acta Crystallogr. Sect. F Structural Biol. Commun.* **2014**, *70* (6), 717–722.

Wallwork, S. C. *Introduction to the Calculation of Structure Factors*; University College Cardiff Press: Cardiff, Wales, 1980.

Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; et al. Overview of the CCP4 Suite and Current Developments. *Acta Crystallographica Section D: Biological Crystallography*. 2011, pp 235–242.

5 Free energy calculations of Cyp complexes

5.1 Introduction and Aims

5.1.1 Aims of this Chapter

The main objective of this chapter was the study of Cyp isoforms in complex with small molecules from our fragment library using computational chemistry techniques and more specifically alchemical binding free energy calculations. From this study we aim to discriminate active compounds from inactives, and specify whether a fragment likes to bind in the Abu or Pro pocket of cyclophilins. A specific objective is to analyse our simulations to identify preferred interactions of fragments in Abu or Pro pockets, and any potential functional groups that are favoured in these pockets. Last but not least, we aim to replicate and understand observed selectivity profiles from SPR data for binding of these fragments to different Cyp isoforms. As in SPR experiments, the work is focused on the binding of different fragments to CypA, CypB and CypD.

5.1.2 Introduction to Molecular Dynamics (MD)

5.1.2.1 Molecular Dynamics

Molecular Dynamic (MD) simulations compute the positions and momenta of a collection of particles as a function of time, and can be used to calculate thermodynamic, and structural properties of a system. To perform a MD simulation of a protein or a protein – ligand complex two steps are necessary. The first one is to establish the initial configuration of the system, usually obtained from X-ray crystallography, or another experimental technique, or a theoretical model. The second step is to assign initial velocities to all atoms of the system. This can be done by drawing random values from the Maxwell-Boltzmann distribution:

$$p(\mathbf{V}_i) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{1}{2} \frac{m_i \mathbf{V}_i^2}{k_B T}\right] \quad \text{Equation 18}$$

The Maxwell-Boltzmann equation calculates the probability that an atom i of mass m_i and at temperature T has a velocity vector $\mathbf{V}_i = (V_{ix}, V_{iy}, V_{iz})$. After setting up the initial

Rational design of isoform specific ligands coordinates and velocities of the system the MD simulation can start and the trajectory is obtained by differentiating the Newton's Second Law of motion, Equation19.

$$\mathbf{F}_i = m_i \mathbf{a}_i \quad \text{Equation19}$$

$$\mathbf{F}_i = m_i \frac{d\mathbf{V}_i}{dt} = m_i \frac{d^2\mathbf{r}_i}{dt^2} \quad \text{Equation20}$$

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} \quad \text{Equation21}$$

$$\mathbf{F}_i = - \frac{dU}{dr_i} \quad \text{Equation22}$$

The result of the differentiation of Newton's Second Law, Equation21, describes the motion of an atom i of mass m_i , depending on the force \mathbf{F}_i that acts on the atom. The force \mathbf{F}_i that acts on atom i at each time step can be calculated by differentiating the potential energy function of the system U , as Equation22 shows, as described from the appropriate force field that is used in the MD simulation, described below in subchapter 5.1.2.2.

5.1.2.2 Molecular Mechanics

Molecular Mechanics (MM) or Force Fields (FF) are based on the Born–Oppenheimer (BO) approximation that enables the electronic and nuclear motions to be separated. MM methods estimate the energy of a system as a function of the nuclear positions only, while Quantum Mechanical (QM) methods deal explicitly with the electrons in the system. MM methods are the method of choice to perform calculations on systems with a significant number of atoms (>1000s). A typical biomolecular FF splits the potential energy U into four components: bond stretching, angle bending, bond rotation (torsion) and non-bonded interactions (electrostatic and van der Waals).

$$\begin{aligned}
U = & \sum_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 \\
& + \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\
& + \sum_{i=1}^N \sum_{j=i+1}^N \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}} \right)
\end{aligned}
\tag{Equation23}$$

U as described before, is the potential energy of the system that is calculated from four different components. Angle and bond terms are usually modelled by a harmonic potential that increase the energy as the angle θ_i or bond length l_i parameters deviate from their reference values $\theta_{i,0}$ and $l_{i,0}$, respectively. The higher the k_i value, the more difficult it is for the bond or the angle to deviate from its equilibrium position. The third term is the bond rotation - torsion term that models the change in energy as bonds rotate. A torsion term is usually modelled from a simple periodic function and is controlled by the V_n , n and γ parameters that depends on the atoms that define the specific torsion, while ω is the rotational angle axis. The last term in the equation describes the non-bonded interactions in the system, between all pairs of atoms, using the Coulomb and Lennard-Jones potentials for the electrostatic and van der Waals interactions respectively. In this term σ_{ij} (the finite distance at which the inter-particle potential is zero) and ε_{ij} (the depth of the potential well) are the Lennard-Jones parameters while r_{ij} (distance between i and j atoms), ε_0 (the electrical permittivity of space or electric constant) and q_i and q_j (the charges of the i and j atoms respectively) are the Coulomb parameters.

5.1.3 Introduction to free energy calculations

Free energy is a very useful thermodynamic quantity from which we can understand how two chemical species associate, interact or recognize each other. As an example this can be the association between two proteins, the conformational equilibrium of a protein or the interaction between a protein and a drug molecule. In our case computation of free energies can give us insights into how different ligands from our fragments library interact with different Cyp isoforms. The first scientist who laid the

Rational design of isoform specific ligands basis for the free energy calculations was John Kirkwood using the thermodynamic integration (TI) methods (Kirkwood, 1935), while a few years later Zwanzig developed the now famous free energy perturbation (FEP) method (Zwanzig 1954). All free energy calculation techniques, including TI and FEP are based on the assumption that the states of a system $\pi(\mathbf{p}, \mathbf{r})$, where \mathbf{p} is the system's momenta vector and \mathbf{r} is its position vector, have a constant number of particles N , pressure p and temperature T , and can be described by the Boltzmann probability distribution function:

$$\pi(\mathbf{p}, \mathbf{r}) = \frac{\exp(-\beta H(\mathbf{p}, \mathbf{r}))}{Q(N, P, T)} \quad \text{Equation24}$$

$$Q(N, P, T) = \int \exp(-\beta(H(\mathbf{p}, \mathbf{r}) + pV)) d\mathbf{p} d\mathbf{r} dV \quad \text{Equation25}$$

Where $H(\mathbf{p}, \mathbf{r})$ is the Hamiltonian of the system, which is a function of the particles' positions \mathbf{r} and momenta \mathbf{p} , $\beta = 1/K_b T$, K_b is the Boltzmann's constant, T is the absolute temperature, p the pressure and $Q(N, P, T)$ is the isothermal – isobaric partition function. The absolute Gibbs free energy G of a system can be described as a function of Q , as described in Equation26. The direct calculation of Q is impossible because configuration space is very large, but the change (ΔG) in absolute Gibbs free energy between two related states, A and B, can be calculated instead (Equation27), using MD, described in the previous sub-chapter 5.1.2.1 or Metropolis Monte Carlo (MMC) algorithms. This is easier because for ΔG we only need to calculate the ratio of Q of the two different states.

$$G = -\beta^{-1} \ln Q(N, p, T) \quad \text{Equation26}$$

$$\Delta G_{AB} = -\beta^{-1} \ln \left(\frac{Q_A}{Q_B} \right) \quad \text{Equation27}$$

There is thus a need for comparing the free energy changes (ΔG). In this study we want to compare the standard free energy of binding, ΔG_1^o and ΔG_2^o , of two different fragments, L1 and L2, to protein P (Equation28 and Equation29).





This can be done in two different ways (Michel and Essex, 2010) as can be seen in Figure 5-1. The first one is by calculating the standard free energy of binding, ΔG_1^o and ΔG_2^o of these ligands, in water and in complex with a protein. The second route is to compute relative free energy of binding $\Delta\Delta G_{\text{bind}}$. The calculation of relative free energy can be done by evaluating two quantities (ΔG_L and ΔG_{PL}). The first one ΔG_L , is the free energy change associated with the mutation of ligand L_1 to ligand L_2 in a water box. The second step, ΔG_{PL} , is the free energy change associated with the mutation of L_1 to L_2 while both are bound to protein P, again in a water box. This mutation process is sometimes called alchemical transformation as it often proceeds via intermediate states that have no physical meaning. Calculation of relative binding free energies is allowed as the free energy change is a state function, and so can be calculated using different reaction pathways to evolve the system.

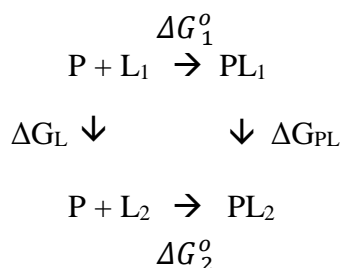


Figure 5-1 Thermodynamic cycle of free energy. The relative free energy of binding ($\Delta\Delta G$) of fragments L_1 and L_2 can be calculated from the thermodynamic cycle in two ways. Either as the difference of ΔG_1^o and ΔG_2^o ($\Delta\Delta G = \Delta G_1^o - \Delta G_2^o$) or as the difference of ΔG_{PL} and ΔG_L ($\Delta\Delta G = \Delta G_{PL} - \Delta G_L$).

Calculation of relative free energies of binding are often preferred to absolute free energies of binding. This is because calculation of relative binding free energies can often be estimated more accurately and more precisely than absolute binding free energy. Relative free energy calculations require less sampling of conformational space for both protein and ligand as the ligand and the protein are always in complex,

Rational design of isoform specific ligands so there is no need to simulate conformational change in the protein due to the absence of ligand. On the other hand the conformational changes of ligand are prevented or minimised at least since ligand is always in complex to the protein or is mutated from one ligand to another in a water box, so there is no need to sample the whole conformational space of one ligand.

5.1.3.1 Coupling parameters

Both TI and FEP need a parameter in order to link computationally the two end states L_1 and L_2 . Usually this parameter is called the coupling parameter and is denoted with the Greek letter λ . This parameter takes values between 0 (that corresponds to L_1) and 1 (that corresponds to L_2) and there are two ways to couple this parameter to the free energy calculations. The first one is called single topology (Jorgensen and Ravimohan, 1985) where the λ parameter is introduced directly to the atom parameters such as charge, Lenard-Jones parameters, bond length and dihedrals. Atom parameters are linearly interpolated ($n = 1$) from one state to another. For example for the atom charge q of atom i :

$$q_i = \lambda^n q_{L2} + (1 - \lambda)^n q_{L1} \quad \text{Equation30}$$

The second way, dual topology does not interpolate the atom properties but the potential energy instead (Kollman, 1993).

$$U(\lambda) = \lambda^n U(L2) + (1 - \lambda)^n U(L1) \quad \text{Equation31}$$

So at $\lambda=0$ the total potential energy of the system is equal to the potential energy of the $L1$ and when the $\lambda=1$ the total potential energy of the system is equal to the potential energy of the $L2$. In-between these two values the total potential energy is calculated from the equation above as a function of the coupling parameter.

If the $L1$ and $L2$ molecules have different numbers of atoms, then some atoms have to be added and/or removed, by turning on/off their potential energy. Making a ligand or some atoms disappear during a simulation leaves an empty space and causes a volume change in the place where the atoms were before. The empty space becomes in turn

Rational design of isoform specific ligands occupied by other protein or water atoms that overlap with the now “invisible” perturbed atoms. This volume change and overlap of atoms cause numerical instabilities in the calculation of the free energy changes and there are different ways to overcome them. One of the most common way is the use of a soft core potential energy function for the perturbed atoms (Zacharias et al., 1994).

$$U_{soft} = (1 - \lambda)4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}^{12}}{(\lambda\delta\sigma_{ij} + r_{ij}^2)^6} \right) - \left(\frac{\sigma_{ij}^6}{(\lambda\delta\sigma_{ij} + r_{ij}^2)^3} \right) \right] + \frac{(1 - \lambda)^n q_i q_j}{4\pi\varepsilon_0 \sqrt{(\lambda + r_{ij}^2)}} \quad \text{Equation32}$$

Where σ_{ij} and ε_{ij} are the Lennard-Jones parameters of atoms i and j , whose distance is r_{ij} and their atomic charges q_i and q_j . So using these soft core potential the repulsion of atoms is finite even if they are very close and the fluctuations in the free energy gradient of the potential energy differences are considerably reduced.

5.1.3.2 Calculation of relative binding free energy

For the analysis of the simulations and the calculation of the free energies (either ΔG_{PL} or ΔG_L) different post-processing techniques can be used. Such techniques are the finite difference thermodynamic integration technique (Mezei, 1987), the Bennet’s Acceptance Ratio analysis (BAR) (Bennett, 1976) and the Multiscale-BAR estimator (MBAR) (Shirts and Chodera, 2008).

Using TI, the free energy change is obtained from Equation33:

$$\Delta G(\lambda_0, \lambda_1) = \int_{\lambda=0}^{\lambda=1} \frac{\partial G}{\partial \lambda} d\lambda \quad \text{Equation33}$$

In this equation, the finite difference thermodynamic integration technique first evaluates the free energy gradient at every λ value between 0 and 1, and then the integral is numerically approximated. The result is the free energy change from one ligand to another in complex (ΔG_{PL}) or in solution (ΔG_L).

On the other hand, BAR method is a reweighting technique. BAR assumes that the two states λ_0 and λ_1 share some common configurational space, but differ in their energy and hence probabilities, because of their difference in a parameter, in this case the λ parameter. To calculate the free energy change between these two states BAR is using the following equation:

$$\Delta G(\lambda_0, \lambda_1) = -\beta^{-1} \ln \frac{\langle f(\beta(U(\lambda_1) - U(\lambda_0) - C)) \rangle_{\lambda_0}}{\langle f(\beta(U(\lambda_0) - U(\lambda_1) + C)) \rangle_{\lambda_1}} + C \quad \text{Equation 34}$$

Where $\Delta G(\lambda_0, \lambda_1)$ is the free energy change between states 0 and 1 while $U(\lambda_1)$ and $U(\lambda_0)$ are the potential energies at $\lambda=1$ and $\lambda=0$ respectively. The difference between the potential energies at the two states are calculated using the same configuration. Either the configuration at λ_0 , in the numerator, or λ_1 , in the denominator.

To estimate free energies in our simulations we have used the last technique, MBAR analysis, which is a generalisation of BAR analysis. It is used for the calculation of free energies with more than two λ states, as in our calculations. It is essentially reduced to BAR when the system has only two states ($\lambda=0$ and $\lambda=1$). MBAR has a direct way of calculating the free energies from the simulations and the errors as described in the article of (Shirts and Chodera, 2008).

5.2 *Materials and Methods*

5.2.1 **Preparation of proteins and ligands**

For this computational study CypA, CypB and CypD proteins were used. Proteins' structures were taken from the X-ray crystal structures with PDB IDs: 1CWA, 3ICH and 2BIT respectively. All water molecules were removed from the structures except the three tightly bound water molecules in the Abu pocket of cyclophilins (Stegmann et al., 2009). All proteins were capped at the C-terminal and N-terminal with an N-methyl and acetyl groups respectively. Protonation of the histidine residues, was done according to the prediction of the Maestro software from Schrödinger package, using

Rational design of isoform specific ligands the Protein Preparation Wizard tool (Madhavi Sastry et al., 2013; Schrödinger LLC, 2013). Specifically the protonation states of histidine's used in CypA are: HIE54, HIE70, HID92, HID126; in CypB: HIE54, HID92, HID126 and in CypD: HIE54, HIE70, HID92, HID126 and HID131.

Ligands used in this free energy calculations study are all ligands purchased as denoted in (Appendix 7.2 Table 7-1) except ligands **18, 36, 39, 43, 47, 48, 68, 69, 71, 72**. These ligands were not identified as hits neither in SPR nor in X-ray experiments and also these ligands can be charged in the pH range where the experiments were performed. Net charge changes in free energy calculations are challenging and can lead to large discrepancies between calculated and experimental free energies (Rocklin et al., 2013; Reif and Oostenbrink, 2014). To correct these discrepancies from the so called finite size effects, additional correction terms are needed that are complicated to evaluate (Rocklin et al., 2013; Reif and Oostenbrink, 2014). So charged ligands were excluded from the study. The rest of the ligands were prepared using Maestro software from Schrödinger package (Schrödinger LLC, 2013).

5.2.2 Docking calculations for complex set up

All ligands were docked into the Abu and Pro pockets of CypA, CypB and CypD using the autodock/vina plugin of Pymol (Schrödinger LLC; Trott and Olson, 2010). All proteins were aligned to a CypA X-ray structure solved in this study. For the docking calculations in the Abu pocket, the grid was centered in the Abu pocket of the protein. The box size of the grid was set at $x = y = z = 13.12 \text{ \AA}$ and between the grid points ($x = y = z = 35$) the spacing was 0.375 \AA . The whole protein was used as a receptor and only one pose was generated for each ligand. Moreover for the docking in the Abu pocket three water molecules that are very tightly bound in the pocket (Stegmann et al., 2009), were kept in the pocket and were treated as part of the protein. The final docking pose for each ligand was compared and make sure that was similar to the crystallographic one, if available. In a similar way, all ligands were docked in the Pro pocket of CypA, CypB and CypD. The grid was centered in the Pro pocket of the protein while the box size of the grid was set at $x = y = z = 7.50 \text{ \AA}$ and between the

Rational design of isoform specific ligands grid points ($x = y = z = 20$) the spacing was 0.375 Å. As before, the whole protein was used as a receptor and only one pose was generated for each ligand. The grid box in the Pro pocket was smaller than the one in the Abu pocket as these fragments prefer to bind in the Abu pocket. So a smaller grid box prevents fragments to be docked out of the Pro pocket.

5.2.3 Perturbation map and free energy files set up

For the calculation of relative binding free energies, ligands should be perturbed from one ligand to another in complex with the protein and alone in a water box. For this reason a perturbation map was generated by manually connecting all the ligands via multiple transformations, as can be seen in Appendix 7.9 Perturbation maps for free energy calculations. The centre of the map was ligand **3** because of its high chemical similarity with the rest of the ligands, because its K_d was determined from SPR experiments, and was also found to bind in the Abu pocket of CypA using X-ray crystallography. All ligands were connected by trying to keep the number of perturbation steps away from the central compound to less than three. Also the chemical similarity of ligands was qualitatively taken into account. The compounds with higher structure similarity were connected between them, in order to reduce the number of perturbed atoms in each simulation for a more accurate relative binding free energy calculation.

Input files for free energy simulations were set up using FESetup software (Loeffler et al., 2015), which is a python software for automated setup that uses AmberTools (Case et al., 2012) for the parametrization of ligands and proteins. Proteins parametrization was done using the latest Amber force field for protein residues ff14SB (Maier et al., 2015), while ligands were parametrized using the general Amber force field, GAFF, for small organic molecules (Wang et al., 2004, 2006) that uses AM1-BCC charges (Jakalian et al., 2000, 2002). All systems were solubilized in a rectangular box of TIP3P water molecules, with a box length of 10 Å away from the edge of the solute, and Na⁺ or Cl⁻ ions were added to neutralize the net charge of the system box. The systems were energy minimized for 200 steps to release any non-favored interactions,

Rational design of isoform specific ligands followed by a heating step to 300 K for 200ps, with harmonic potential restraints on all non-solvent atoms of $10 \text{ kcal mol}^{-1} \text{ \AA}^2$ force constant. Systems were then equilibrated using an NVT ensemble (with standard volume and temperature) for 200 ps with the same restraints as the previous step. Finally systems were equilibrated in an NPT (with standard pressure and temperature) ensemble at 1 atm for 5ns to equilibrate the density of the solvent. The final output snapshot was used as the starting point for the free energy calculations.

5.2.4 Alchemical free energy simulations

Free energy simulations were performed using TI technique using the Sire – OpenMM framework (Woods, C.; Calabro, C.; Michel; Eastman and Pande, 2010) on GPUs (GeForce GTX465 and Tesla/M2090/K20 cards). The free energy gradients were calculated using 12 equidistant λ windows (0.0000, 0.0909, 0.1818, 0.2727, 0.3636, 0.4545, 0.5455, 0.6364, 0.7372, 0.8182, 0.9091, 1.00). Systems were energy minimised for 1000 steps and then equilibrated at the appropriate λ value for 2 ps. The total length of each λ simulations was 5ns, and the energies were saved every 200 fs, to be used from MBAR post-processing. A softcore potential was used in all simulations as explained before (Michel et al., 2007).

The hydrogen mass repartitioning (HMR) technique (Hopkins et al., 2015) was used to increase the timestep of the simulation and decrease the time needed to run each simulation. A 4 fs timestep and a repartitioning factor of 4 was used and all bonds were constrained. Simulations were performed in the NPT ensemble, where temperature control was achieved with the Andersen thermostat (Andersen, 1980) and pressure control was achieved using a Monte Carlo Barostat. Periodic boundary conditions were used with a 10 \AA cut off for the non-bonded interactions.

In relative binding free energy simulations atomic distance restraints of $7.5 \pm 3.0 \text{ \AA}$ in the Abu and Pro pockets were applied with a force constant of 5 kcal mol^{-1} . One atom of each ligand was selected and was restrained to one atom of the protein in the Abu

Rational design of isoform specific ligands pocket (GLN111/backbone O atom) when ligands were bonded in the Abu pocket. For simulations of ligands in the Pro pocket the protein atom was (Met61/backbone CA atom) and the ligand atom was the same as in the Abu pocket. Using these restraints the ligand was penalized with the respective force constant if the distance between the selected atoms was > 10.5 or < 4.5 Å. For a distance of $4.5 - 10.5$ Å ligands were not penalized. Although restraints are usually not needed in relative binding free energy simulations, they were considered necessary in our simulations. This is because fragments binding in a shallow pocket of a protein, with low millimolar affinity, tend to drift away from their starting binding position into the bulk water of the system box. The selected atomic distance restraints and the force constant were enough to restrain our fragments in the binding pocket, and prevent them drifting away from the protein, while the same time fragments are free to move in the pocket and explore all possible interactions with the protein residues in the pocket.

For ligand **3** that was the central compound for the free energy calculations, absolute binding free energy calculations were performed with **3** binding to the Abu and Pro pockets of CypA, CypB and CypD. In this simulations $\lambda=0$ corresponds to ligand **3** and the final state $\lambda=1$ corresponds to a ligand with dummy atoms (atoms without Lennard-Jones or coulombic parameters). Absolute binding free energy calculations for this ligand in water and in complex with the protein were performed in two steps. In the first step only the Coulombic parameters were turned off. The final frame of this step was then used as the starting position of the second step with the Lennard-Jones parameters of the ligand atoms turned off. As in the relative binding free energy calculations, ligand **3** was restrained in the pocket of the protein using a force constant of 5 kcal mol^{-1} . One atom of ligand **3** was distance restrained to two atoms of the protein in the Abu pocket (GLN111 and ASN102 CA backbone atoms) and two atoms of the protein in the Pro pocket (HID126 and ILE57 CB atoms). In the Abu pocket the distance restraint tolerance was set to 5.5 ± 1.5 Å and in the Pro pocket to 7.5 ± 1.0 Å, and no restraints were applied while the distance was within the given ranges.

5.2.5 Data processing and fitting

As explained in the Sub-chapter 5.1.3.2 and the Equation35 shown below, the relative free energy change for transforming ligand L1 to ligand L2 ($\Delta\Delta G(L1 \rightarrow L2)$) was calculated as the difference in the free energy change of transforming L1 to L2 in a water box ($\Delta G_w(L1 \rightarrow L2)$) and in complex with the protein ($\Delta G_p(L1 \rightarrow L2)$).

$$\Delta\Delta G(L1 \rightarrow L2) = \Delta G_p(L1 \rightarrow L2) - \Delta G_w(L1 \rightarrow L2) \quad \text{Equation35}$$

MBAR implementation in python (pymbar) was used for the calculation of both $\Delta G_w(L1 \rightarrow L2)$ and $\Delta G_p(L1 \rightarrow L2)$ free energy changes (Shirts et al., 2014). Relative binding free energy calculations were only performed, so errors were propagated from each step and the final reported error was calculated using the equation below.

$$\begin{aligned} \text{err}(\Delta\Delta G(L1 \rightarrow L2)) \\ = \text{sqrt}((\Delta G_w^{err}(L1 \rightarrow L2))^2 + (\Delta G_p^{err}(L1 \rightarrow L2))^2) \end{aligned} \quad \text{Equation36}$$

In absolute binding free calculations the absolute binding free energies were calculated as Equation37 :

$$\Delta G^o(L3) = \Delta G_w^C + \Delta G_w^{LJ} - \Delta G_p^C - \Delta G_p^{LJ} - \Delta G_p^{o\text{ restr}} \quad \text{Equation37}$$

Where $\Delta G^o(L3)$ is the absolute binding free energy of ligand **3** (L3). ΔG_w^C and ΔG_p^C are the free energies of turning off the coulombic parameters of the L3 in water and in complex respectively while ΔG_w^{LJ} and ΔG_p^{LJ} are the free energies of turning off the Lenard Jones parameters of the ligand. $\Delta G_p^{o\text{ restr}}$ is the free energy cost of releasing the restraints and getting standard state conditions, as explained in (Mobley et al., 2006) and it was calculated using a custom script. Each absolute binding free energy was repeated in quadruplicate and the $\Delta G^o(L3)$ was calculated as the average of the quadruplicate. The final error reported for the absolute binding free energy was calculated as the standard error of the mean, from the four repeats, as shown below:

$$\text{err}(\Delta G^o(L3)) = \frac{s}{\sqrt{n}} \quad \text{Equation38}$$

Where s is the standard deviation of the average absolute binding free energy from the 4 repeats and n is the number of repeats.

Based on absolute binding free energy of **3**, and because the absolute binding free energies of all ligands were not determined (rather relative energies were calculated), the absolute binding free energies for the rest of the ligands were estimated using the equation below:

$$\Delta G^{\circ}(\mathbf{L}) = \Delta G^{\circ}(\mathbf{L3}) + \Delta\Delta G(\mathbf{L} \rightarrow \mathbf{L}_i) + \sum_{i=m}^n \sum_{j=i+1} \Delta\Delta G(\mathbf{L}_i \rightarrow \mathbf{L}_j) + \Delta\Delta G(\mathbf{L}_j \rightarrow \mathbf{L3}) \quad \text{Equation39}$$

Where \mathbf{L}_i and \mathbf{L}_j are all the ligands in the perturbation path between ligand **3** ($\mathbf{L3}$ that is the centre of the perturbation map) and \mathbf{L} (that is the target ligand).

The experimental free energies were calculated using the experimental dissociation constants (K_d) determined from SPR experiments using the Equation40 below:

$$\Delta G^{\circ} = RT \ln\left(\frac{K_d}{C^{\circ}}\right) \quad \text{Equation40}$$

Where R is the ideal gas constant ($0.00198588 \text{ kcal mol}^{-1} \text{ K}^{-1}$), T is the temperature (298.15 K) and C° is the standard state concentration (1 mol L^{-1}). Standard deviation (std) values are not reported in the results of SPR experiments but as explained a coefficient of variation (CV) of $\sim 20 - 40\%$ can be taken into account. So taken into account an average CV of 30% , three experimental values (K_d , $K_d + \text{CV } K_d$, $K_d - \text{CV } K_d$) were converted to binding free energies using Equation40 and the standard error of mean of these values is reported as the experimental error.

5.3 Results and Discussion

5.3.1 Absolute binding free energies of ligand **3**

As was explained in the methods before, ligand **3** was selected as the central compound for this free energy calculation study on Cyp. In order to determine computationally the binding free energy of **3** to the Abu and Pro pockets of CypA, CypB and CypD,

absolute binding free energies were performed as explained in the sub-chapter 4.2. The results from these calculations can be seen in the figure below.

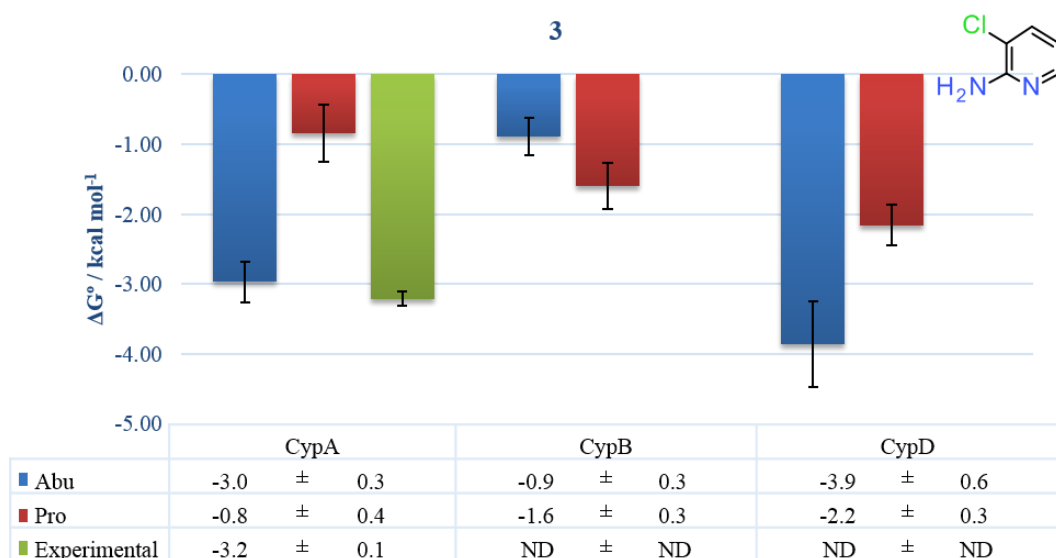


Figure 5-2 Absolute binding free energy of lig3 to Abu and Pro pockets of CypA, CypB and CypD

SPR experiments have shown a binding energy of **3** to CypA of -3.2 ± 0.1 kcal mol⁻¹. Free energy calculations in Abu and Pro pocket of CypA show that **3** has binding free energy for these two pockets of -3.0 ± 0.3 and -0.8 ± 0.4 kcal mol⁻¹ respectively. First of all we can see that there is a difference in the free energy of binding, of ~ 2 kcal mol⁻¹, for **3** when it binds to Abu or Pro pockets of CypA. Secondly the experimental value is very close to the calculated value in the Abu pocket, while the value in the Pro pocket is much lower (2.4 kcal mol⁻¹) than the experimental one. So from the calculated values of free energy of binding of **3** to CypA we can conclude that there is a clear preference for binding of **3** into the Abu pocket of CypA. Moreover this in agreement with X-ray results were **3** was identified to bind in the Abu pocket of CypA.

In CypD, **3** seems to have the same preference of binding between Abu and Pro pocket. The calculated binding free energy of **3** in the Abu and Pro pocket of CypD are -3.86 ± 0.61 and -2.16 ± 0.30 kcal mol⁻¹ respectively. Although the binding free energy in

Rational design of isoform specific ligands the Abu pocket is higher than the binding free energy in the Pro pocket, the binding free energy in the Pro pocket in CypD is much higher than the binding free energy in Pro pocket of CypA. This is a surprising result, since the binding pocket of Cyp is very highly conserved and especially the Pro pocket. The calculated binding free energies of **3** to the Abu and Pro pocket of CypB are both under $-1.60 \text{ kcal mol}^{-1}$ and this shows that for some reason, **3** does not like to bind CypB, as well as the other isoforms. Experimentally no binding free energies were measured by SPR for **3** on CypD or CypB as their response versus concentration curves cannot be described from a steady state affinity kinetic model (

Figure 7-1in Appendix 7.4).

Although we cannot be sure why experimentally we cannot measure binding free energies of **3** on CypD or CypB, the calculated binding free energies maybe can give us some possible explanation for this different behaviour in their response curves. **3** seems to bind weakly CypB, but it seems that it can bind into both pockets of CypD. This weak binding on CypB and unspecific binding on CypD, could be the reason why response versus concentration curves of **3** on CypB and CypD cannot be described from a steady state affinity kinetic model.

5.3.2 Relative binding free energy calculations on CypA

As explained in the methods section before, the relative binding free energies of 86 fragments from our library were calculated, binding in the Abu and Pro pockets of CypA, using free energy calculations. Two of our main goals are first to see if we can differentiate active from inactive compounds and second if we can see a preference of binding of these fragments in the Abu or Pro pocket of CypA.

A histogram summarizing the results from these calculations can be seen in Figure 5-3. In this histogram the compounds are separated in bins based on their K_d (bottom x-axis) and ΔG° values (top x-axis). The total number of compounds in each bin is presented in the figure. From this graph is clear that there is preference of binding of

these fragments in the Abu pocket of CypA. The absolute binding free energy range of these fragments in the Abu pocket of CypA is 0.5 – 30 mM while in the Pro pocket is 30mM – 2 mol L⁻¹. Although in these free energy calculations there is a clear separation between the binding free energies of fragments in the Abu or Pro pockets the same is not happening with the separation of active and inactive compounds.

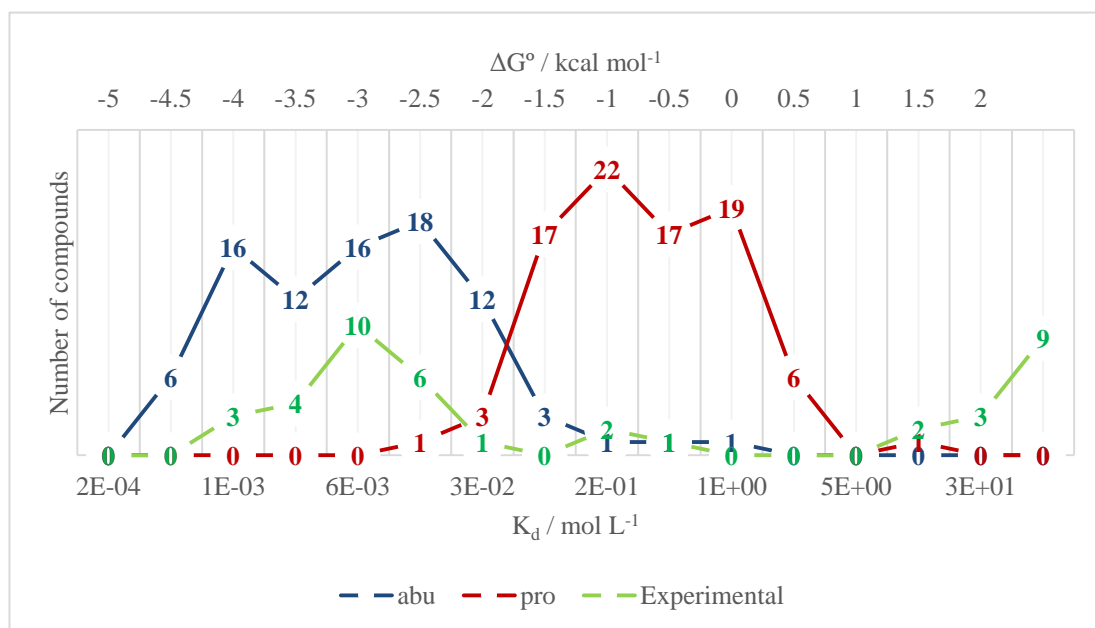


Figure 5-3 Histogram of calculated and experimental absolute binding free energy of fragments on CypA. In the bottom X axis is K_d in Molar and in the top X-axis the ΔG in kcal mol⁻¹ while in the Y-axis is the number of compounds in each bin of the histogram. The experimental and calculated results, in Abu and Pro pockets, can be seen in green, blue and red lines respectively.

As can be seen from the graph, the experimental results range from 1 mM to more than 1 M while the calculated values in the Abu pocket are all under 30 mM and in the Pro pocket above 30 mM. In the SPR chapter in this thesis, the active/inactive cut off that we set was 10mM. If we use the same cut off for the free energy calculations, it means that almost all compounds are active in the Abu pocket and inactive in the Pro pocket, and this is not true (based also on the X-ray crystallographic results). This behavior is of course affected from many other parameters that we should take in account when we compare the SPR and the free energy calculations results. First of all the K_d that is

Rational design of isoform specific ligands reported as experimental value is the result of SPR experiments, in which fragments are free to diffuse in/out of the binding site of the protein or just interact randomly on the surface of the protein or both. In our simulations fragments are binding only, and they are restrained, in the Abu or in the Pro pocket. The expectation is that the experimental and computational values are similar if the experimental conditions match the computational i.e. the fragment binds specifically or predominately one of the two pockets of CypA. In this way the interaction between protein and ligands would be the same experimentally and computationally and the free energy values from the two methods should be similar. In the cases where ligands have a more complex behavior on the surface of the protein, especially when we are dealing with fragments, the experimental value cannot be easily described computationally.

Moreover in the figure, the size of the experimental and the computational libraries are not the same. In the SPR measurements, from 98 compounds, 50 compounds were not tested while for 17 compounds the steady state affinity kinetic model was able to describe their response versus concentration curves. In the free energy calculations 86 compounds were tested and we have a relative binding free energy for all of them.

5.3.3 Insights in the selectivity profiles of 5 compounds from free energy calculations

One of the aims of the chapter is to study the binding free energy of different fragments from our library, which show different selectivity profiles in SPR experiments. For this purpose 5 fragments were selected, **4**, **60**, **61**, **90** and **98**. These compounds have different binding profiles between them and some of them show some selectivity for one Cyp isoform, while others had the same profile over all Cyps. So the first goal is to see if free energy calculations agree with the SPR results, and the second goal is to try to explain any observed selectivity.

For these calculations relative binding free energies were performed as explained in the methods sub-chapter 4.2. The absolute binding free energies of these compounds

Rational design of isoform specific ligands were then estimated using the calculated relative binding free energies of these compounds (relative to **3**) and the calculated absolute binding free energy of **3**.

5.3.3.1 Ligand **61**

SPR results for ligand **61**, show that it has a small preference over CypD (1.1 kcal mol⁻¹) compared to CypA, while for CypB, its response versus concentration curves cannot be described from a steady state affinity kinetic model (

Figure 7-1 in Appendix 7.4). Free energy calculations, as can be seen in Figure 5-4, agree with SPR experiments. **61** has a very small binding free energy to CypB, under 1.0 kcal mol⁻¹ for both pockets, while at the same time it shows a much higher activity on CypD than CypA, 0.6 and 1.9 kcal mol⁻¹ when bound to Abu or Pro pocket respectively. From Figure 5-4 we can notice also that the binding affinities of this compound to the Abu and Pro pockets of CypA or CypD are very close, with their difference within statistical error. This indicates that **61** may be able to bind the Pro pocket of CypA or CypD as well as the Abu pocket of these proteins. This trend can also be confirmed from the X-ray studies of **61** with CypA. The structure of **61** was solved in the Pro pocket of the protein (Figure 4-12) but there was also a second blob of electron density in the Abu pocket, which could be a second molecule of the same ligand binding at the same time the other pocket of the protein.

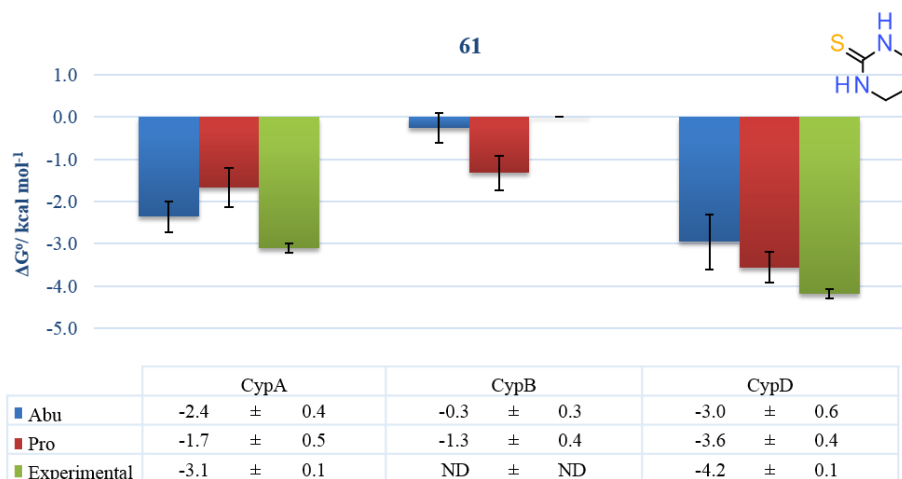


Figure 5-4 Absolute binding free energies of 61 to the Abu and Pro pockets of Cyp

5.3.3.2 Ligand 60 and Ligand 98

60 and **98** are two compounds that experimentally show no selectivity over one Cyp isoform. Instead they bind all three isoforms with similar binding free energy. Free energy calculations on **60** agree with the experimental observations but show different binding profiles of **60** to the Abu or the Pro pockets of the proteins. On CypA, **60** prefers to bind to the Abu pocket, on CypB it prefers to bind to the Pro pocket, while on CypD it can bind to both pockets equally. Further examination is needed to understand why **60** shows preference over different pocket upon binding to different isoforms, especially on CypA. X-ray studies of **60** with CypA, revealed that this compound is binding in the Pro pocket of CypA, and this is in contrast with what the free energy calculations suggests.

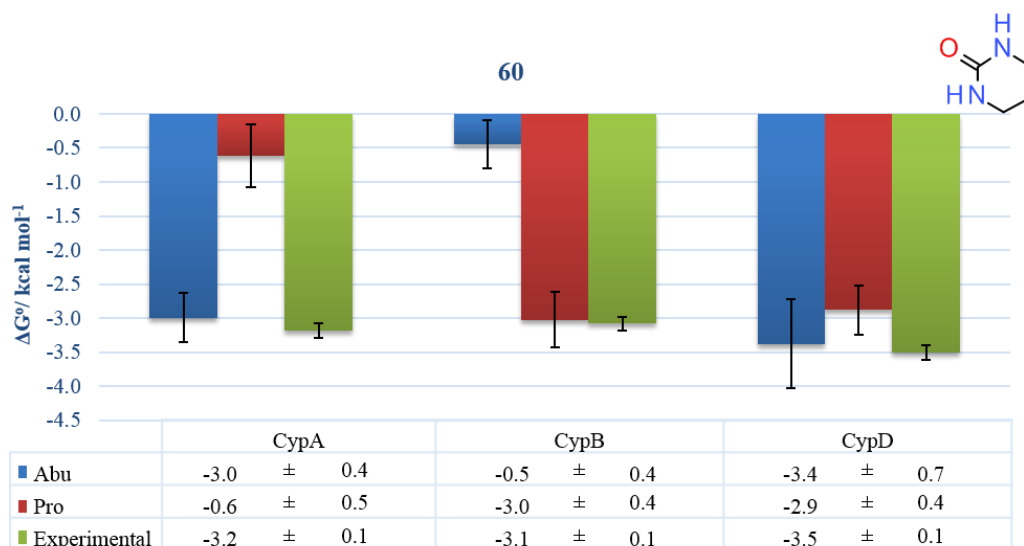


Figure 5-5 Absolute binding free energies of **60** to the Abu and Pro pockets of Cyp

On the other hand **98** both experimentally and computationally show the same binding profile over CypA and CypD, with the calculated and experimental absolute binding free energies to be similar within error bars. Moreover there is a clear preference over the Abu pocket instead of the Pro pocket and this is in agreement with X-ray studies that show **98** binds in the Abu pocket of CypA. Experimental binding free energy of **98** on CypB though is not the same as the computational binding free energy. Experimental binding free energy on CypB, -3.7 ± 0.1 kcal mol⁻¹, is similar to the ones on CypA (-3.6 ± 0.1 kcal mol⁻¹) and CypD (-3.8 ± 0.1 kcal mol⁻¹) while the computational binding free energy on CypB is positive in the Abu pocket (1.3 ± 0.4 kcal mol⁻¹) and -1.3 ± 0.5 kcal mol⁻¹ in the Pro pocket.

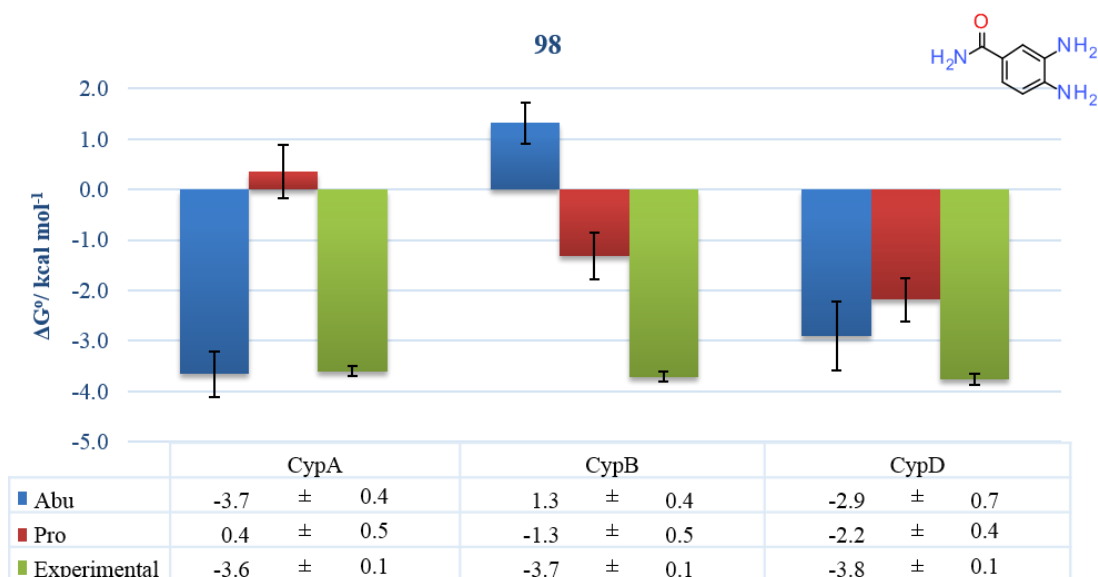


Figure 5-6 Absolute binding free energies of 98 to the Abu and Pro pockets of Cyp

5.3.3.3 Ligand 90

Experimental binding free energies of **90** were reported only for CypA as the response versus concentration curves of **90** on CypB and CypD cannot be described from a steady state affinity kinetic model. Computationally, and as can be seen from Figure 5-7 below, the calculated binding free energy of this compound (-2.3 ± 0.6 kcal mol⁻¹) in the Abu pocket of CypA is very close to the experimental value (-3.0 kcal mol⁻¹). For CypB the binding is very weak and in the Abu pocket of CypD the calculated binding free energy is -0.4 kcal mol⁻¹ more negative than in CypA. This is in contrast with the experimental results and further examination is needed to fully understand why computationally we can see binding of **90** to CypD, but experimentally something more complex is happening. Nevertheless the experimental and calculated binding free energies of **90** on Cyp look similar to **3**. And maybe, as **3**, a weak binding on CypB and an unspecific binding on CypD, could explain why response versus concentration curves of **90** on CypB and CypD cannot be described from a steady state affinity kinetic model.

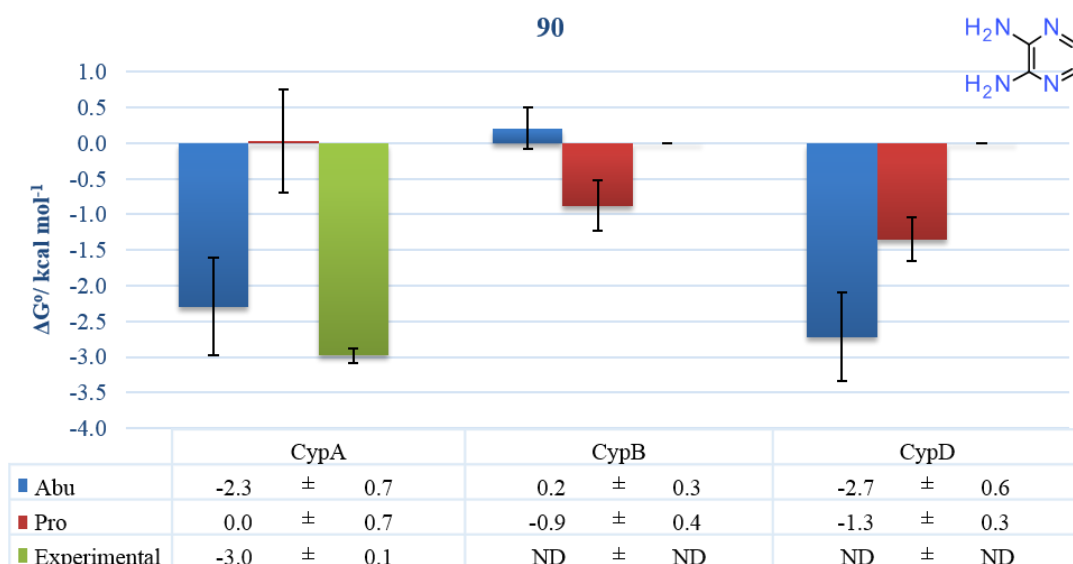


Figure 5-7 Absolute binding free energies of 90 to the Abu and Pro pockets of Cyp

5.3.3.4 Ligand 4

Dissociation constants for **4** were not determined experimentally for any Cyp isoform. This is because for the two fold dilution series measurements of SPR only the 48 most active compounds from the single point concentration measurements were selected. Nevertheless from the results of the free energy calculations, as can be seen in Figure 5-8, **4** seems to be able to bind both CypA and CypD in the Abu pocket with free energy of binding of -3.5 ± 0.3 and -4.3 ± 0.6 kcal mol⁻¹ respectively. This result supports the fact that compounds that were not selected and were not tested using a dilution series measurements on SPR, cannot be described *a priori* as inactive. Within those compounds, there may be some that can be equally active or more active than compounds that have been tested.

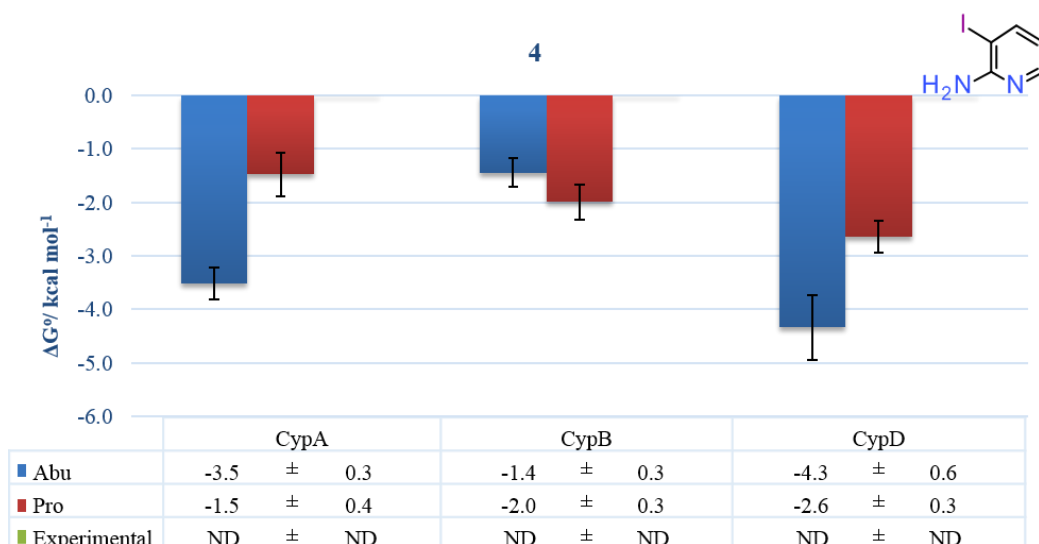


Figure 5-8 Absolute binding free energies of 4 to the Abu and Pro pockets of Cyp

5.4 Interaction of fragments in the Abu and Pro pockets of CypA

In the Pro pocket, as can be seen also from Figure 5-3, the binding free energy of the fragments is much less negative (binding is much weaker) than binding of these fragments in the Abu pocket. This is a result of two parameters, firstly because the pro pocket is slightly shallower than the Abu pocket and secondly, and maybe more important, the Pro pocket is more hydrophobic than the Abu pocket. The result of these parameters is that the fragments that are forced to stay in the Pro pocket are unable to form a stable H-bond(s) that will keep them in place for the whole length of the simulation. Nevertheless from the simulations we can see that fragments in the Pro pocket can form H-bonds with side chain of Arg55 and the backbone of Asn102, as described also in the X-ray chapter of this thesis (Figure 4-12), but also with side chain of residues Gln63 and His126.

On the other site, from the MD simulations of the fragments in our library in complex with Cyp isoforms, and specifically in the Abu pocket of CypA, many important observations arise. The first one is the binding poses with which these fragments can bind in the pocket and the population of these poses as were defined from our MD

Rational design of isoform specific ligands simulations. The second observation, are the interactions between fragments and protein or water molecules in the Abu pocket of CypA. The third and last observation, is a functional group that is common in most fragments and seems to be important for their binding in the Abu pocket of CypA.

Analysis of the X-ray crystallographic structures of fragments, in the active site of CypA, revealed a common binding pose that is believed to be the major pose for all fragments binding in the Abu pocket (Figure 4-7). In some cases though, as for example ligand **3**, refinement of the structures show that these fragments may have more than one pose in the Abu pocket (Figure 4-10). Analysis of all the trajectories (20 ns in total) obtained from the absolute free energy calculation of **3**, in the Abu pocket of CypA, have shown an agreement with the X-ray experiments. The RMSD of **3** from the X-ray crystallographic pose, was calculated for all four independent MD simulations and from this analysis, as can be seen in the Figure 5-9, three different RMSD clusters have been identified.

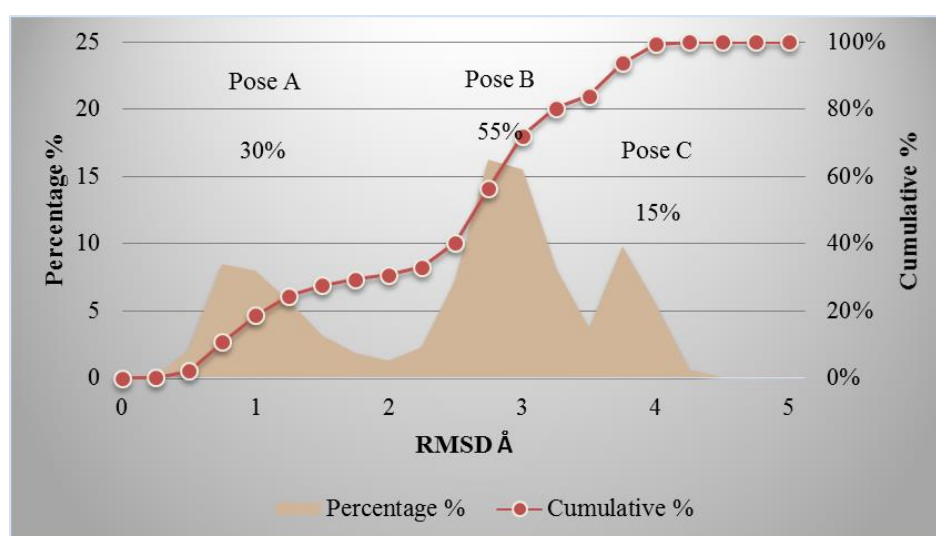


Figure 5-9 Histogram of RMSD of **3** from the X-ray crystallographic pose

The first cluster, named pose A, with 1 ± 1 Å RMSD, represents the cluster at which **3** adopts a pose very similar the X-ray structure. Interestingly pose A was not the only

Rational design of isoform specific ligands pose in the simulation but also was not the major one too. Two more clusters of RMSD were calculated, pose B and pose C, with the RMSD of 2.8 ± 0.7 and 3.75 ± 0.50 from pose A respectively. The population of all poses were also calculated. Pose B found to be the most populated one, with **3** to adopt this pose in 55% of the simulation time, pose A was less populated (30%) and pose C was the least populated, only 15%.

Moreover by plotting the RMSD of each trajectory versus time, Figure 5-10, we notice two important things. The first one is that the starting pose of the fragment does not affect the pose that the fragment adopts during the simulation. In all repeat simulations, **3** was docked in the same conformation and that was similar to pose B. In some simulations, **3** was instantly changing pose from pose B to pose A (Figure 5-10 Repeat 2), while in some others pose B was retained (Figure 5-10 Repeat 3). Furthermore the second observation is that the restraints used in this free energy calculations were just enough to keep the ligands in the pocket and were not preventing the ligand from freely moving in the pocket. These can be rationalised by comparing repeat 1 with repeats 2 and 3 from Figure 5-10. In repeat 2, one pose (pose B) was retained throughout the simulations. In repeat 3, **3** change pose from B to A from the first steps of the simulation and retain this pose for the rest of the simulation. On the other hand in repeat 1, **3** was able to adopt all three different poses for some time each (Figure 5-10).

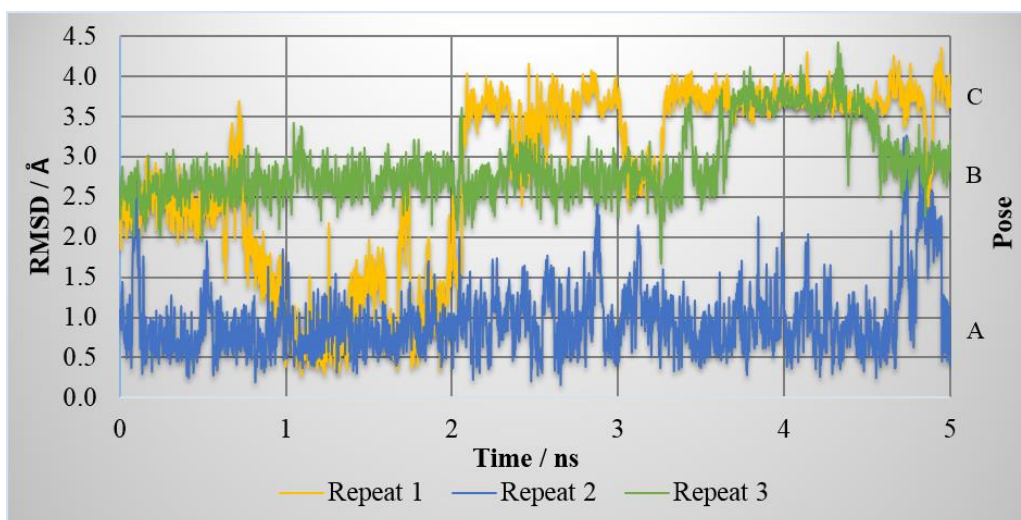


Figure 5-10 RMSD versus time of 3 from the X-ray crystallographic pose. Calculations are shown from three independent CypA:3 simulations. Repeat 1 is shown in yellow, repeat 2 in blue and repeat 3 in green.

Visual examination of the trajectories of CypA:3 complex revealed the structural details of each different pose that 3 adopts in the Abu pocket of CypA. Moreover the H-bonds that 3 establishes with the different protein or water molecules in each pose were identified. Characteristic pose A structures and HBA – HBD distance histograms can be seen in Figure 5-11.

In pose A, which is similar to the observed X-ray crystallographic data, the primary amine of 3 is hidden in the Abu pocket and H-bonded to Thr107, Ala101, W1 and W2, Figure 5-11. Column B. In all cases the primary amine of 3 acts as a hydrogen bond donor (HBD) and the oxygen atom of other residues as hydrogen bond acceptor (HBA). Exception is the case of W2, which can act as both HBD and HBA. Column A shows the H-bond distance histograms. Bond distances (Å) are separated in bins (X-axis) and y-axis shows the simulation time percentage at which each bond distance is adopted. Almost in all H-bond distance histograms we can see that these distances are separated in three different clusters, which roughly represent the three different poses (A, B and C). The population of each cluster can be calculated, ~30% for pose A, ~50% for pose B and ~15% for pose C. For all distances these populations are almost

the same as previously calculated from the RMSD of **3** conformations. Exception again is W2 and this is because this molecule is also involved in H-bonding with **3** in pose B.

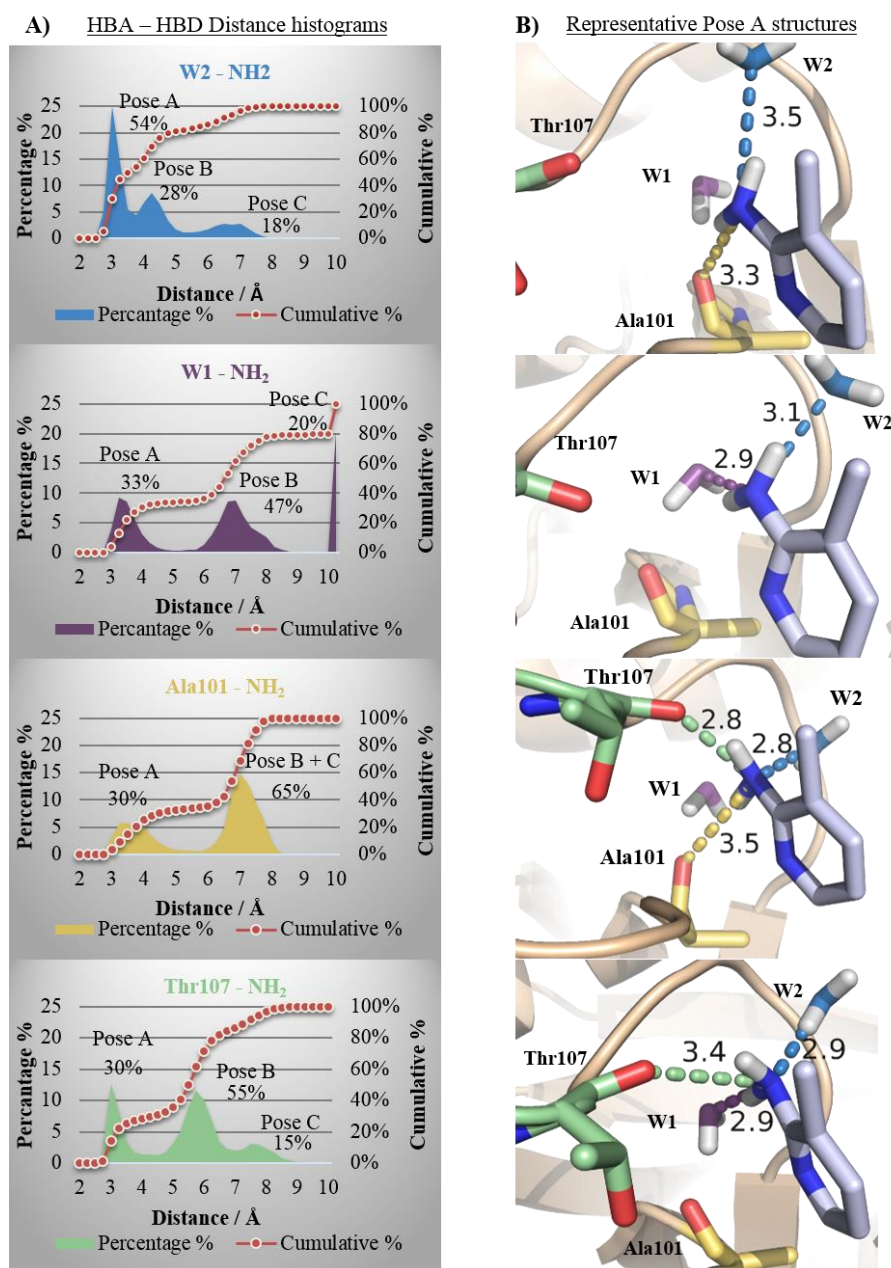


Figure 5-11 Representative structures of pose A and H-bond distances with key protein and water molecules. Each protein and water molecule that is involved in H-bonding with **3** is colour coded separately and is the same colour in both structures (column B) and H-bond distance histograms (column A). The distance between HBA and HBD is also highlighted in the structures using the same colour coding.

In the second pose (pose B), **3** adopts a different conformation in the pocket, Figure 5-12 column B. In this pose, **3** is rotated by 90° in the Abu pocket and it is H-bonded to Thr73 and W2. The primary amine of **3** acts as a HBD and Thr73 acts as the HBA, while the same time W2 can act as a HBD and the nitrogen atom of the pyridine ring acts as an HBA. Also, as before in Figure 5-11, from the distance histograms the three different poses can be identified and their populations was calculated to be similar to before (~35% for pose A, ~50% for pose B and ~15% for pose C).

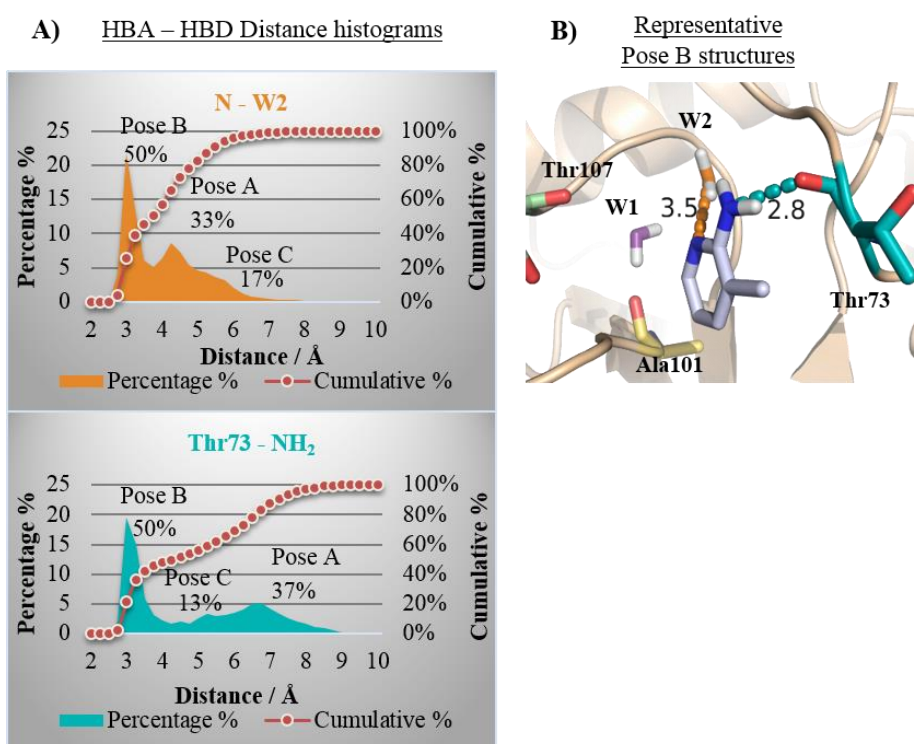


Figure 5-12 Representative structures of pose B and H-bond distances with key protein and water molecules. Each protein and water molecule that is involved in H-bonding with **3** is colour coded separately and is the same colour in both structures (column A) and H-bond distance histograms (column B). The distance between HBA and HBD is also highlighted in the structures using the same colour coding.

In pose C, which is the least populated pose, **3** is rotated another 90° degrees, and now its primary amine is positioned towards the saddle point of CypA active site and it is H-bonded with the backbone carbonyl oxygen of Gly72, Figure 5-13.

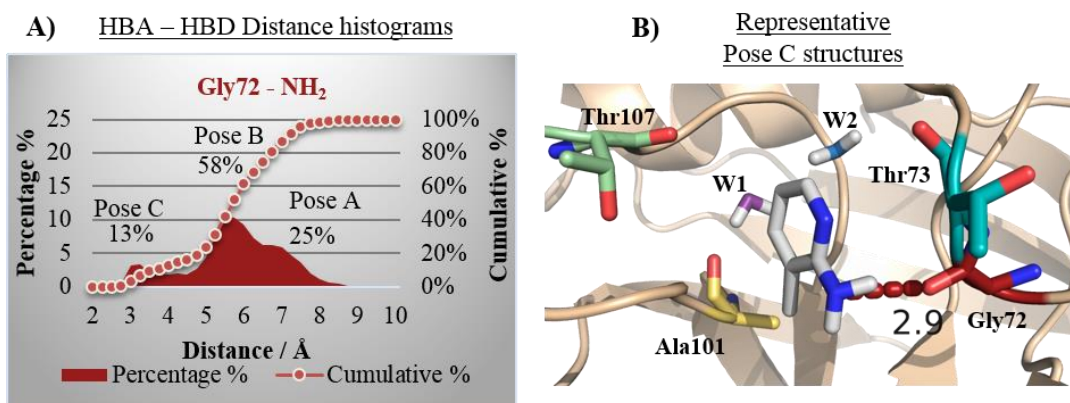


Figure 5-13 Representative structure of pose C and H-bond distance with a key protein molecule. Each protein and water molecule that is involved in H-bonding with **3** is colour coded separately and is the same colour in both structures (column A) and H-bond distance histograms (column B). The distance between HBA and HBD is also highlighted in the structures using the same colour coding.

The population of each H-bond in pose A, pose B or pose C, is very similar to the population of the states as were defined before from RMSD of **3**. This shows that actually these H-bonds are very stable and helps to stabilise the fragment in the Abu pocket. In all three poses that **3** can adopt in the Abu pocket, all its H-bonds are established with one functional group. This conserved functional group that is always involved in H-bonding is the primary amine substituent of the pyridine ring. This functional group is a very good hydrogen bond donor (H-donor), a property that is important for the binding of these fragments. The primary amine group can interact and establish hydrogen bonds with the backbone carbonyl oxygen of four different CypA residues (Gly72, Thr73, Thr107 and Ala101) and the two tightly bonded water molecules (W1, W2) in the Abu pocket.

Rational design of isoform specific ligands

In a recently published article from (Gelin et al., 2015), a number of compounds have been identified as inhibitors of Cyp and co-crystallised in the active side of CypD. These compounds have a very similar substructure with fragments from our library. In their research Gelin *et al* have shown that the primary amine substituent of a benzyl moiety can bind in the Abu pocket of CypD and establish H-bond interactions with Thr107. X-ray structure refined from this group, with PDB ID 4zsc, was aligned with **3**, in pose A, from our fragment library (Figure 5-14). In this figure we can see that the two compounds are aligned completely and both can establish the same interactions with the protein.

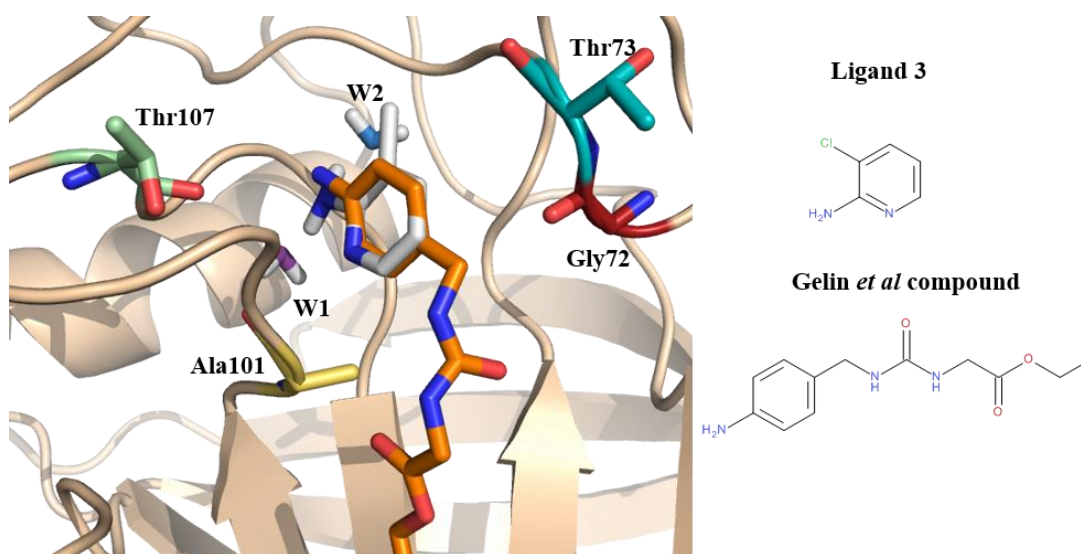


Figure 5-14 Alignment of **3** with compound from (Gelin et al., 2015) in complex with CypA. Compound from Gelin *et al* is shown as orange while **3** is shown in light grey.

5.5 Conclusion

In this part of our research, we have examined the binding free energies of 96 compounds from our fragment library, in the Abu and Pro pockets for three Cyp isoforms, CypA, CypB and CypD. For this purpose alchemical free energy calculations have been used to calculate the binding free energies of these fragments and compare them with experimental results. Results have shown that binding free energy calculations tend to overestimate the binding free energy of the compounds compared to SPR experiments. This is the reason why from these results it is difficult to differentiate between active and inactive compounds. Nevertheless in some cases absolute binding free energy calculations on selected compounds, eg **3**, were in close agreement with the experimental binding free energy.

Moreover in this study molecular dynamics/free energy calculations were able to predict the preferred binding of these fragments and specifically their preference to bind in the more hydrophilic Abu pocket of Cyp instead the hydrophobic Pro pocket. Also the computational study of these fragments when bound to different Cyp isoforms have shown that, free energy calculations sometimes can give very similar results to the SPR data and in some cases it can give some insights in the experimental measurements.

Furthermore from the simulations we could identify the three different poses and their populations that **3** adopts in the Abu pocket as well as the main interactions between the protein and the fragment in all three poses. These interactions include the H-bonds between the nitrogen atom of the primary amine and the nitrogen atom of the pyridine ring of the fragments with the backbone carbonyl oxygen of CypA residues (Gly72, Thr73, Thr107 and Ala101) and two tightly bonded water molecules. This preference on binding and the interactions between protein and ligand that was identified from our computational studies completes and confirms our X-ray results and previously published data from other groups.

5.6 References

- Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure And/or Temperature. *J. Chem. Phys.* **1980**, 72 (1980), 2384.
- Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, 22 (2), 245–268.
- Case, D. A.; Darden, T.; Cheatham, T. E.; Simmerling, C.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. Amber 12. *Univ. California, San Fr.* **2012**, 350.
- Eastman, P.; Pande, V. OpenMM: A Hardware-Independent Framework for Molecular Simulations. *Comput. Sci. Eng.* **2010**, 12 (4), 34–39.
- Gelin, M.; Delfosse, V.; Allemand, F.; Hoh, F.; Sallaz-Damaz, Y.; Pirocchi, M.; Bourguet, W.; Ferrer, J.-L.; Labesse, G.; Guichou, J.-F. Combining ‘dry’ Co-Crystallization and In Situ Diffraction to Facilitate Ligand Screening by X-Ray Crystallography. *Acta Crystallogr. Sect. D* **2015**, 71 (8), 1777–1787.
- Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, 11 (4), 1864–1874.
- Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, 21 (2), 132–146.
- Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, 23 (16), 1623–1641.
- Jorgensen, W.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J Chem Phys* **1985**, 83 (6), 3050–3054.
- Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, 3 (1935), 300–313.
- Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, 93, 2395–2417.

Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model.* **2015**, *55* (12), 2485–2490.

Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27* (3), 221–234.

Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

Mezei, M. The Finite Difference Thermodynamic Integration, Tested on Calculating the Hydration Free Energy Difference Between Acetone and Dimethylamine in Water. *J. Chem. Phys.* **1987**, *86* (12), 7084–7088.

Michel, J.; Essex, J. W. Prediction of Protein-Ligand Binding Affinity by Free Energy Simulations: Assumptions, Pitfalls and Expectations. *J. Comput. Aided. Mol. Des.* **2010**, *24* (8), 639–658.

Michel, J.; Verdonk, M. L.; Essex, J. W. Protein-Ligand Complexes: Computation of the Relative Free Energy of Different Scaffolds and Binding Modes. *J. Chem. Theory Comput.* **2007**, *3* (5), 1645–1655.

Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the Use of Orientational Restraints and Symmetry Corrections in Alchemical Free Energy Calculations. *J. Chem. Phys.* **2006**, *125* (8), 084902.

Reif, M. M.; Oostenbrink, C. Net Charge Changes in the Calculation of Relative Ligand-Binding Free Energies via Classical Atomistic Molecular Dynamics Simulation. *J. Comput. Chem.* **2014**, *35* (3), 227–243.

Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hunenberger, P. H. Calculating the Binding Free Energies of Charged Species Based on Explicit-Solvent Simulations Employing Lattice-Sum Methods: An Accurate Correction Scheme for Electrostatic Finite-Size Effects. *J. Chem. Phys.* **2013**, *139* (18).

Schrödinger LLC. The PyMOL Molecular Graphics System, Version 1.4.1.

Schrödinger LLC. Schrödinger Release 2013-3: Maestro, Version 9.6. New York, NY

2013.

Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129* (12).

Shirts, M. R.; Chodera, J. D.; Beauchamp, K. R. pymbar v2.1.0-beta
<https://github.com/choderalab/pymbar>.

Stegmann, C. M.; Seeliger, D.; Sheldrick, G. M.; De Groot, B. L.; Wahl, M. C. The Thermodynamic Influence of Trapped Water Molecules on a Protein-Ligand Interaction. *Angew. Chemie - Int. Ed.* **2009**, *48* (28), 5207–5210.

Trott, O.; Olson, A. J. AutoDock Vina. *J. Comput. Chem.* **2010**, *31*, 445–461.

Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.

Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260.

Woods, C.; Calabro, C.; Michel, J. www.siremol.org www.siremol.org (accessed Apr 29, 2016).

Zacharias, M.; Straatsma, T. P.; McCammon, J. a. Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100* (12), 9025.

Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. II. Polar Gases. *J. Chem. Phys.* **1955**, *23* (10), 1915.

6 Discussion, conclusions and future work

6.1 Aims of this Chapter

This chapter summarises the results obtained in this study and discusses whether these results answer the questions that were set at the beginning of this research. Furthermore in this chapter, the experimental and computational techniques used in this project are compared, the differences and agreements between them are highlighted and emphasis is given into how such techniques can complement each other for the identification of new Cyp inhibitors.

The aims of this study, as were set in the introduction chapter of this thesis, can be divided into three main areas. The first one is the identification of a new class of Cyp inhibitor, the second one is the analysis of the binding of these inhibitors in the active site of Cyps and the last one is to examine the selectivity profile of their binding to Cyps. Each of these three areas will be described separately in the following subchapters. For each one, the individual aims set at the beginning, and the results obtained will be discussed. Possible future work that would give new insights will be reported too.

6.2 Identification of a new class of Cyp inhibitors

6.2.1 Aims

The main aim of this section of the study, was the identification of a new class of Cyp inhibitors and their use for the further study of protein dynamics and thermodynamics. In extension to this we aimed to determine the binding affinity of these compounds as well as obtain structural data to confirm their binding in the active site of the Cyp isoforms. To do that we proposed the use of SPR, to determine binding affinities (Chapter 3) and X-ray crystallography, to provide structural data (Chapter 4), as the primary screening techniques. Moreover free energy calculations (Chapter 5) were

Rational design of isoform specific ligands used as a complementary technique to see if there is an agreement between computational and experimental results.

6.2.2 Results and discussion

As explained in Chapter 3, an SPR experiment with a single point measurement for all soluble compounds, from our fragment library, was performed to identify the compounds with the higher response on Cyps. 48 compounds with the higher response were selected and tested in 2-fold dilution series, to determine their binding affinity on Cyp isoforms. To separate actives from inactive compounds, two cut-offs, affinity \leq 10mM and 1:1 stoichiometry, were used. Based on these cut offs, 12 compounds were found to be active on CypA and 6 compounds on CypB and CypD (Figure 3-11).

Following up SPR results, X-ray experiments were performed by soaking CypA, CypB and CypD crystals with different ligand solutions. From the X-ray results, as described in Chapter 4, 16 ligands were found to bind in the active site of CypA (Figure 4-5), while crystallisation of ligands in the active site of CypB and CypD was not possible because of different issues discussed in Chapter 4.

Furthermore, as described in Chapter 5, free energy calculations were performed, using 86 compounds from our fragment library, to calculate their binding free energy to the Abu and Pro pockets of CypA, and differentiate active from inactive compounds. The calculated binding affinity in the Abu pocket of CypA was calculated to be between 1 - 10 mM, for 72% of the compounds. In the Pro pocket the calculated affinity was \geq 100 mM, for 71% of the compounds. From these results is clear that free energy calculations successfully predicts that the preferred binding pocket of these fragments on CypA is the Abu pocket. This result is in agreement with our X-ray data where 13 out of 16 active compounds bind in the Abu pocket of CypA (Figure 7-2 and Figure 7-3) Furthermore free energy calculations were able to estimate the binding affinity range (1 – 10mM) of these fragments in the Abu pocket,

a result that is also in agreement with our SPR measurements that show the binding affinity of these fragments to be in the low mM scale (Figure 3-8).

Compared to the successful predictions of the range of the binding free energies and the preferred binding site of these fragments, relative free energy calculations were not so successful on differentiating “active” from “inactive” compounds. Also were not so successful on ranking compounds based on their binding affinity. To identify “hit” or “active” compounds from our free energy calculations, if the same cut-off to SPR is used that is 10mM, then the percentage of “active” compounds from the whole dataset is 79%. If free energy calculations were used as a primary screening technique, 79% “active” compounds would be too high and not very helpful for the selection of the best fragments. This highlights the difficulty of accurate free energy calculations when we are modelling fragments with low binding affinity. In our free energy calculations, and for comparison with the other experimental techniques, the “active”/“inactive” cut-off could be set to 3mM, which gives 33 “active” compounds (38% from the whole dataset used in this study).

Looking now at the results from all three techniques used, is clear that several fragments, from this in-house bespoke library, are active against Cyp isoforms. Their binding affinity was confirmed from both SPR and free energy calculations, and it was calculated to low mM range ($\leq 10\text{mM}$). Moreover X-ray crystallography confirmed the ability of these fragments to bind in CypA active sites, by successful crystallisation of 17 of them.

Figure 6-1 below shows the number of active compounds, on CypA, as identified from the three different techniques (SPR, X-ray and MD). Also it shows compounds that were found to be active according to two out of three techniques, or in all three techniques.

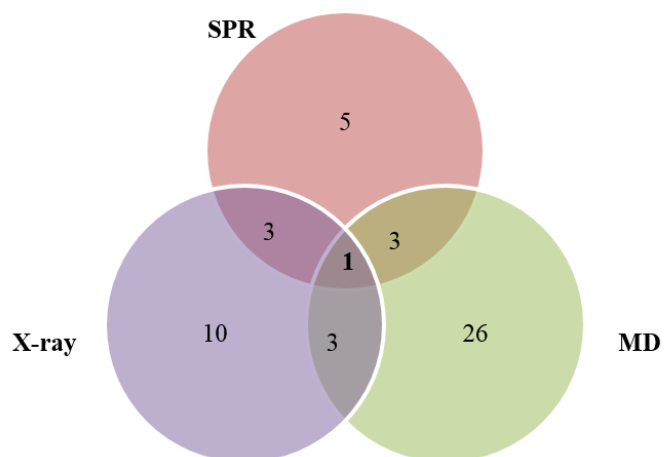
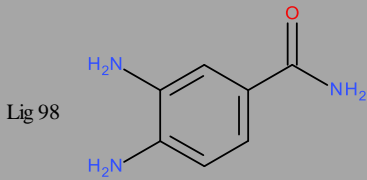
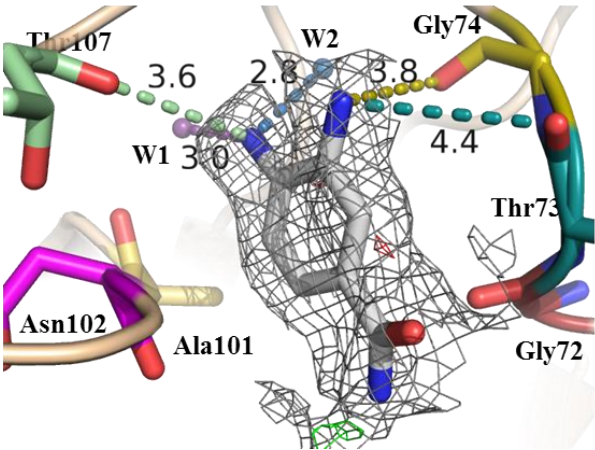


Figure 6-1 Venn diagram showing the hit compounds from the three different screening techniques, SPR, X-ray and MD

Table 6-1 One compound that was “active” in all three techniques

Technique		Ligand	
X-ray			
SPR	K_d (kcal mol ⁻¹)	-3.6 ± 0.1	
	Stoichiometry	1:1	
MD (kcal mol ⁻¹)		-3.7 ± 0.4	
Ligand efficiency		0.33	

Although the binding free energy of each fragment varies between SPR and MD, and the number of “active” compounds varies between the different techniques, there are cases where fragments are active in all SPR, X-ray and MD. Ligand **98** has a binding free energy of -3.6 ± 0.1 kcal mol⁻¹ based on SPR measurements, and -3.7 ± 0.4 kcal mol⁻¹ based on free energy calculations, Table 6-1. Moreover **98** was crystallised in the Abu pocket of CypA, establishing H-bond interactions with water molecules (W1 and W2) and protein residues (Thr107 and Gly74) in the pocket, Table 6-1 and Figure 4-7.

6.2.3 Future Work

Free energy calculations and SPR experiments were performed using almost all the compounds in the library (98 small fragments). X-ray, in contrast to the previous two techniques, was only performed on a small subset of the library and only using CypA isoform. That was due to experimental constraints that prevent the use of the full library on all isoforms in the given timeline of this study. Nevertheless if the screening of all fragments on different Cyp isoforms was possible, this would give more data and more information about potential CypA, CypB and CypD binders. Moreover expansion of our fragment library with fragments of larger size and higher molecular weight would allow the identification of more potent Cyp inhibitors.

6.3 *Analysis of the binding of these fragments in the active site of Cyps*

6.3.1 Aims

This section of the study was focused on the analysis of the binding of the hit compounds in the active site of Cyps. Initially, this analysis was performed on the crystal structures obtained from the X-ray crystallography (Chapter 4). The goal was to further understand the preference of binding of fragments into the Abu or Pro pockets of Cyps, to identify preferred interactions in the binding pockets, and any potential functional groups that were favoured in these pockets. The analysis also included trajectories obtained from the free energy calculations (Chapter 5).

6.3.2 Results and discussion

X-ray studies revealed the binding of 17 fragments, from our small compounds library in the active site of CypA. Fourteen fragments were found to bind in the Abu pocket of CypA, while three others prefer to bind to the Pro pocket. This preference for fragments binding to the Abu pocket of CypA, was also observed in the free energy calculation studies (Figure 5-3). The calculations showed a clear preference of fragments binding in the Abu pocket, with an average binding affinity of -3.25 ± 0.90 kcal mol⁻¹, compared to -0.94 ± 0.70 kcal mol⁻¹ for compounds binding to the Pro pocket.

The ability of these fragments to establish H-bond interactions with two tightly bonded water molecules in the Abu pocket of CypA and their close proximity (≤ 4 Å) to several CypA residues was revealed from X-ray crystallography (Figure 7-2 and Figure 7-3). Fragments are able to establish these interactions with the amine substituent of their aromatic ring that acts as H-bond donor. The water molecules in the pocket are acting as H-bond bridges between the fragments and the protein. This experimental result was in agreement not only with our MD results but also with previously published data (Stegman et al., 2009).

Moreover, in Chapters 4 and 5 the preference of aromatic rings (pyridine, pyrimidine or benzene) to bind in the Abu pocket and non-aromatic rings to bind in the Pro pocket was discussed (Table 4-2 and Figure 4-12).

Further study of the binding of these fragments in the Abu pocket, using MD simulations, revealed additional details. The MD trajectories have shown the ability of several fragments to establish H-bonds with residues Thr73, Ala101 and Thr107 (Figure 5-11) as well as the two conserved water molecules. The binding of these fragments in the pocket was found to be very similar to the binding of compounds previously published (Gelin et al., 2015) that have similar substructures with fragments from our library (Figure 5-14). This shows that fragments' binding modes are able to

Rational design of isoform specific ligands predict potential moieties or side chains orientations of lead compounds and their interactions with protein residues.

Moreover new data for the dynamics of the fragments in the pocket were revealed (Figure 5-9). This data have shown that the fragments are able to move in the pocket and adopt three different major conformations. Conformation A, which is similar to the crystallographic one, and is populated ~30% during the MD trajectories, has the amino group hidden in the pocket and H-bonded to water and protein residues (Ala101 and Thr107) (Figure 5-11). In conformation B, which is the most populated one, ~55% of MD time, the fragment is rotated by 90° and the amino group is more exposed and H-bonded to Thr73 (Figure 5-12). The last conformation, pose C, which is the least populated from the three (~15%), the fragment is further rotated by 90° and the amino group this time establishes H-bond with Gly72 (Figure 5-13).

Although this analysis was focused only on fragment **3**, the same may apply to other fragments too. The $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ density maps of other fragments (e.g. **5**), as the density maps of **3**, show small differences between the observed and calculated values (Figure 7-2 and Figure 7-3). This is a good example where MD can assist X-ray crystallography during the refinement process for the better modelling of the fragment's binding pose.

The ability of MD to identify different binding conformations of fragments, and characterise the interactions between fragments and protein at each one, can be very useful for the next stage of FBDD. This stage as discussed previously in the introduction, Figure 1-4 is to grow one fragment, or connect many of them together, to optimise their binding affinity and selectivity. In this project until now, our data suggest that pose A is a good pose to start growing the fragments out of the Abu pocket. This is because pose A, facilitates the H-bond interactions between fragments and protein and water residues as previously shown in Figure 5-11, Figure 7-2 and Figure 7-3. The interactions that fragments establish in pose B and C, as also as the

Rational design of isoform specific ligands interactions that fragments establish in the Pro pocket, show positions of potential protein-ligand interaction sites that can be taken into account when growing out the fragments.

6.3.3 Future Work

As explained previously, X-ray crystallography gave a lot of useful information on the binding of these fragments to the binding site of CypA, but crystal structures of these fragments in the binding site of CypD and CypB were not obtained. This would give insights in the binding of these fragments on the specific Cyp isoforms. Moreover the use of computational techniques, such as energy decomposition analysis (Woods et al., 2014), principal components analysis (PCA) or other techniques would give more information on the binding of these fragments and complement the results obtained from the free energy calculations. In particular energy decomposition analysis, would help us understand the contribution of different water and/or protein residues to the fragments binding affinities. Also this would help us understand if the binding of several fragments is driven from their hydrophobicity or from their interactions with the protein.

6.4 Selectivity profile of fragments binding on Cyps

6.4.1 Aims

In the last section of this project the aim was to study the binding selectivity profiles of fragments to three different Cyp isoforms CypA, CypB and CypD. To do that we proposed the use of SPR as the first screening technique, in order to measure the binding affinity of fragments on different Cyp surfaces. From the analysis of the SPR data we aimed to identify fragments that show preferences over one or two CypA surfaces, and use these fragments to further study their selectivity profiles. Based on SPR results, and after selection of fragments of particular interest, free energy calculations were performed. Firstly to examine if we can replicate the experimental results and secondly trying to give rational to the observed selectivity profiles.

Furthermore the use of X-ray crystallography was proposed to give some structural insights to the binding selectivity profile data of the fragments.

6.4.2 Results and discussion

SPR studies have shown that fragments from this library of compounds have different binding profiles when interacting with CypA, CypB or CypD surfaces. Differences were observed from the first SPR experiment, where a single point measurement, at 1 mM, was taken for all fragments. From this experiment we observed that fragments behave differently on the three Cyp surfaces and there is a clear preference of interaction with CypA instead of CypD or CypB (Figure 3-6). On average, the measured activities on CypA are 3 response units higher than the activities on the other two isoforms while the activities of fragments on CypB and CypD are almost identical.

After selection of the 48 most active fragments, 2-fold dilution series measurements were performed to measure the binding affinity of these fragments on all three surfaces. The results from these measurements were consistent with the previous SPR measurements and a similar trend as before was observed (Figure 3-8).

From these measurements we can see that ~44% of the fragments have a low mM, ≤ 10 mM, binding affinity to CypA compared to 25% and 19% on CypB and CypD respectively. Moreover the fragments that have molar activity on CypB and CypD are 54% and 52% compared to only 12.5% on CypA. Taking out fragments with stoichiometry $\neq 1:1$, 16 fragments have activity ≤ 10 mM and stoichiometry of 1:1 (Figure 3-11) for at least one of the three isoforms. From those 61 fragments, the percentage of them that bind selectively on CypA surface is still 44% compared to 12.5% on CypB and CypD (Figure 3-10).

As explained before in the SPR chapter the observed binding affinities for these fragments, are well above the confidence level of SPR, which means that the K_d values for our “active” compounds are approximate. Nevertheless we can confidently report

Rational design of isoform specific ligands “active” from “inactive” compounds using a cut off of 10mM. Moreover the binding selectivity profile data of these fragments are consistent throughout the SPR measurements. Furthermore the selectivity profile of these fragments, to our knowledge, are very different compared to the selectivity profiles of lead molecules. Lead molecules, especially on Cyps, usually have high affinity on all Cyps with very small or no selectivity at all. In comparison, these small fragments have a clear preference over CypA compared to the other two isoforms. These profiles are an interesting observation and if we are able to verify them with another technique and understand their rational, may be a very powerful tool for the selective inhibition of Cyps.

From the 16 compounds in total that had activity ≤ 10 mM and stoichiometry of 1:1 (Figure 3-11), six of them were selected to be further examined with free energy calculations. These fragments were **3**, **4**, **60**, **61**, **90** and **98** and most of them have different selectivity profiles. **3** and **90** have a measured binding affinity on CypA of 4 and 6 mM respectively, but no measured binding affinity on CypB or CypD. **60** and **98** have a low mM affinity for all three isoforms and do not show any selectivity. **61** have a low sub mM affinity and 1:1 stoichiometry only on CypD, and low mM affinity on CypA but not 1:1 stoichiometry. In contrast, **4** was not used in the 2-fold dilution series, as its response after the single point SPR measurements was too low. So the first objective of the free energy calculations was to replicate the SPR selectivity profile data of these five fragments and the second aim to give some insights to why some of those bind selectively to one Cyp isoform while some others bind equally well to all of them.

In some cases, free energy calculations were able to replicate the measured binding selectivity profiles and provide some rationale for them. An example is the study of fragment **3** (Figure 5-2). The computational and experimental binding free energies of **3** on CypA are -2.97 ± 0.29 kcal mol⁻¹ and -3.21 ± 0.11 kcal mol⁻¹ respectively, which are very close and within statistical error. The same time the computational binding free energies of **3**, in both Abu and Pro pockets of CypB are very low, while in both

Rational design of isoform specific ligands pockets of CypD, its binding free energies, are much higher than the respective ones on CypA. Experimentally no binding free energies were measured by SPR for **3** on CypD or CypB as their response versus concentration curves cannot be described from a steady state affinity kinetic model (

Figure 7-1).

Although we cannot be sure why SPR experiment cannot measure binding free energies of **3** on CypD or CypB, free energy results, show that for some reason, **3** binds very weakly on CypB, but it seems that it can bind into both pockets of CypD. This weak binding on CypB and unspecific binding on CypD, could be the reason why response versus concentration curves of **3** on CypB and CypD cannot be described from a steady state affinity kinetic model.

Free energy calculations agree with SPR experiments also in the study of **61**. As was observed from MD, **61** has a very small binding free energy to CypB, under $1.0 \text{ kcal mol}^{-1}$ for both pockets, while at the same time it shows higher activity for both pockets of CypA and an even higher affinity for both pockets of CypD (Figure 1-4). This can be again an indication that **61** is active on both CypA and CypD, but in both cases the fragment binds in more than one position. That was also observed from X-ray crystallography where the structure of **61** was solved in the Pro pocket of CypA (Figure 4-12) but a second possible partial binding position was observed in the Abu pocket.

For the other fragments though, **61**, **90** and **98**, the results from the two techniques partially agree, and there are differences on the binding free energies of these fragments on some isoforms and pockets (Figure 5-5, Figure 5-7 and Figure 5-6 respectively). Differences between experimental and computational results are expected though, especially in this study where we are using fragments with low mM affinity. Moreover in the SPR experiment, fragments are free to diffuse in/out of the binding site of the protein or just interact randomly on the surface. In our simulations fragments they are restrained to bind only in the Abu or in the Pro pocket of the protein.

In total, free energy calculations have proven to be a useful complementary technique to SPR and were able to give possible explanations as to why experimental SPR measurements were not possible for some fragments. Furthermore where experimental measurements were possible, in some cases the selectivity profile of compounds was very similar in both free energy and SPR data.

An interesting observation from X-ray experiments on CypA-fragments complexes, is the close proximity of the fragments to the 80's loop of the protein (Figure 4-8 and Table 4-1). As discussed previously in the Introduction chapter, this loop and specifically amino acids 81 and 82 are not conserved across Cyp family, Figure 1-7 and Table 1-2. In contrast these amino acids can change significantly from one Cyp to another. The close proximity of the fragments to these amino acids may be able to introduce some specificity on Cyp inhibition, and further examination is needed to prove it, by successful crystallization of CypB and CypD complexes.

6.4.3 Future Work

Unfortunately crystallisation of fragments with CypB and CypD was not successful and crystal structure data are not available, so X-ray crystallography was not able to help us understand the observed selectivity profiles at this stage. Successful crystallization though of fragments in the active site of CypB and CypD would be very useful and would probably give insights to the observed selectivity profiles.

6.5 Conclusions

Overall this study has shown that the combination of experimental and computational techniques can be useful to identify new fragments, predict and measure binding affinities, and characterise the dynamics of protein-fragment complexes. Because it is challenging to detect weak binding affinities associated with small fragments, it is preferable to focus attention on these fragments that can be characterised as “active” by different techniques. Compound **98** is an example that could be used to generate more potent Cyp binders by growing out of the Abu pocket.

6.6 References

Gelin, M.; Delfosse, V.; Allemand, F.; Hoh, F.; Sallaz-Damaz, Y.; Pirocchi, M.; Bourguet, W.; Ferrer, J.-L.; Labesse, G.; Guichou, J.-F. Combining ‘dry’ Co-Crystallization and In Situ Diffraction to Facilitate Ligand Screening by X-Ray Crystallography. *Acta Crystallogr. Sect. D* **2015**, *71* (8), 1777–1787.

Stegmann, C. M.; Seeliger, D.; Sheldrick, G. M.; De Groot, B. L.; Wahl, M. C. The Thermodynamic Influence of Trapped Water Molecules on a Protein-Ligand Interaction. *Angew. Chemie - Int. Ed.* **2009**, *48* (28), 5207–5210.

Woods, C. J.; Malaisree, M.; Michel, J.; Long, B.; McIntosh-Smith, S.; Mulholland, A. J. Rapid Decomposition and Visualisation of Protein-Ligand Binding Free Energies by Residue and by Water. *Faraday Discuss.* **2014**, *169*, 477–499.

7 Appendix

7.1 Sequences of recombinant human Cyps

The sequences of all recombinant human CypA, CypB and CypD used in the experimental part of this study can be seen below. The first 24 amino acids in red represent the hexa-histidine tag plus the Tev protease binding sites, used for the Ni²⁺ affinity chromatography purification of the proteins and the cleavage of the extra amino acids respectively. For CypD, the amino acid highlighted in blue is the lysine 133 (K133) which is mutated to isoleucine for the X-ray crystallographic studies of the protein.

CypA / Residues: V2 – E164

MSYYHHHHHHDYDIPTTENLYFQGVNPTVFFDIAVDGEPLGRVSELFADKV
PKTAENFRALSTGEKGFGYKGS CFHRIIPGFCQGGDFTRHNGTGGKSIYGE
KFEDENFILKHTGPGILSMANAGPNTNGSQFFICTAKTEWLDGKHVVFGKVK
EGMNIVEAMERFGSRNGKTSKKITIADCGQLE

CypB / Residues: S21 – E215

MSYYHHHHHHDYDIPTTENLYFQGSVFFLLLPGPSAADEKKKGPKVTVKVY
FDLRIGDEDVGRVIFGLFGKTVPKTVDNFVALATGEKGFGYKNSKFHRVIKD
FMIQGGDFTRGDGTGGKSIYGERFPDENFKLKHYGPGWVSMANAGKDTNG
SQFFITTVKTAWLDGKHVVFGKVLEGMEVVRKVESTKTDSRDKPLKDVIIA
DCGKIEVEKPFIAIAKE

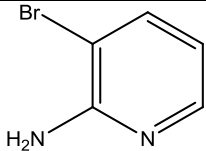
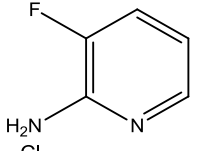
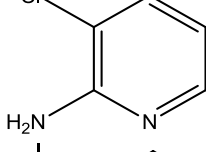
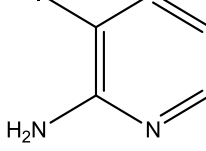
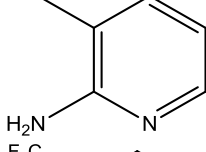
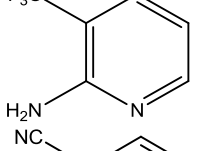
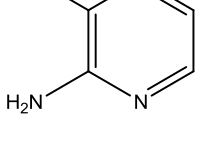
CypD / Residues: S43 – S207

MSYYHHHHHHDYDIPTTENLYFQGS GNPLVYLDVDANGKPLGRVVLELKA
DVVPKTAENFRALCTGEKGFGYKGSTFHRVIPSFMCQAGDFTNHNGTGGKSI
YGSRFPDENFTLKHVGPVLSMANAGPNTNGSQFFICTIKTDWLDGKHVVV
GHVKEGMDVVKKIESFGSKSGRTSKKIVITDCGQLS

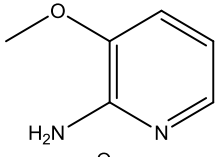
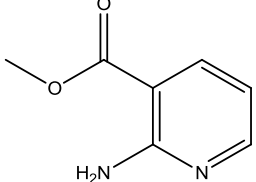
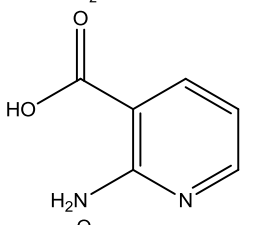
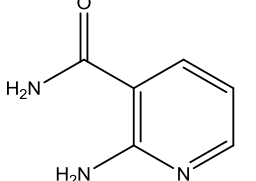
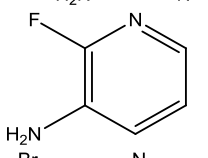
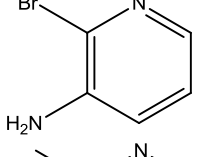
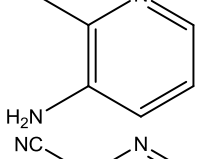
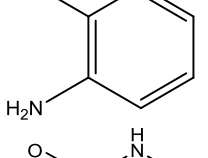
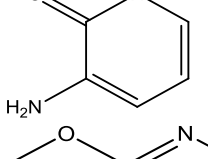
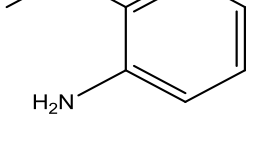
7.2 Purchased compounds

Table 7-1 below shows the compounds purchased to be test for activity on CypA, CypB and CypD isoforms using X-ray crystallography, SPR and free energy calculation studies. Table shows the ligand number for each compound that is used in this study as also as the chemical structure, IUPAC name, MW, manufacturer and code number for each compound.

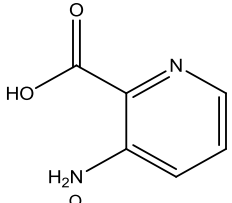
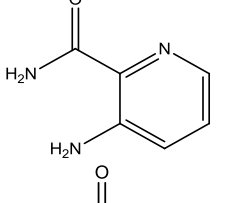
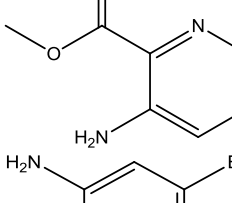
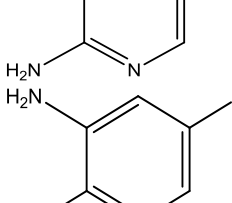
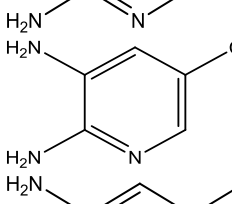
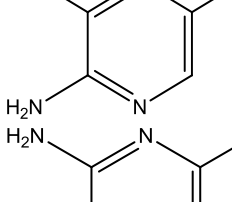
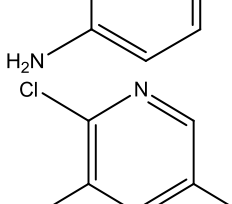
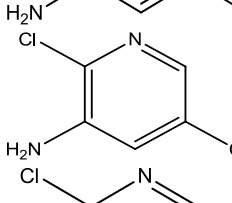
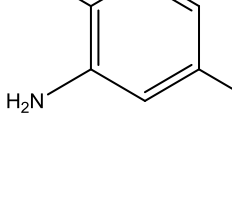


Table 7-1 Table of Purchased compounds

Ligand Number	Structure	Name	M _w	Company	Code Number
1		2-Amino-3-bromopyridine	173.01	Fluorochem	049777
2		2-Amino-3-fluoropyridine	112.10	Fluorochem	045834
3		2-Amino-3-chloropyridine	128.56	Fluorochem	032705
4		2-Amino-3-iodopyridine	220.01	Fluorochem	043682
5		2-Amino-3-methylpyridine	108.14	Fluorochem	032515
6		2-Amino-3-trifluoromethylpyridine	162.12	Fluorochem	023329
7		2-Amino-nicotinonitrile	119.13	Fluorochem	033252

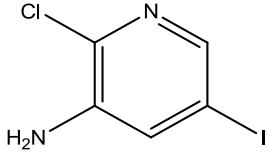
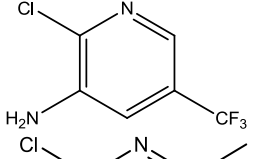
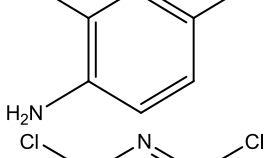
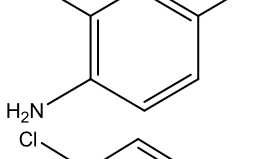
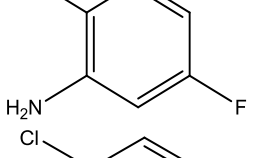
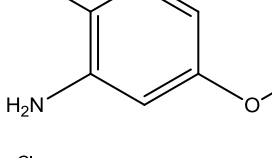
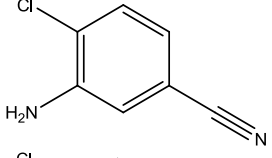
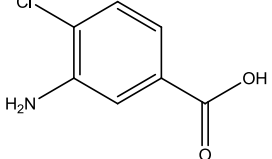
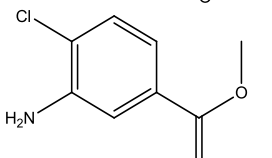
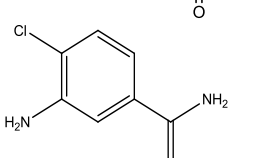
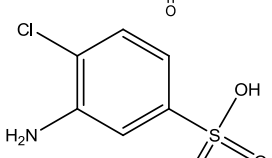
Rational design of isoform specific ligands

8		2-Amino-3-methoxypyridine	124.14	Fluorochem	050813
9		Methyl 2-Aminonicotinate	152.15	Fluorochem	066915
10		2-Aminonicotinic acid	138.13	Fluorochem	033046
11		2-Aminonicotinamide	137.14	Fluorochem	068372
12		3-Amino-2-fluoropyridine	112.11	Fluorochem	033249
13		2-Bromopyridin-3-amine	173.01	Fluorochem	064879
14		3-Amino-2-methylpyridine	108.14	Fluorochem	036315
15		3-Amino-2-pyridinecarbonitrile	119.12	sigma-aldrich	CDS020 253
16		3-Aminopyridin-2(1H)-one	110.11	Fluorochem	318699
17		3-Amino-2-methoxypyridine	124.14	Fluorochem	017404

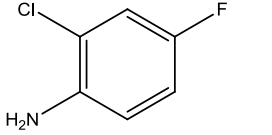
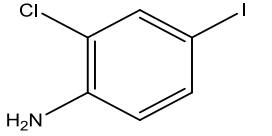
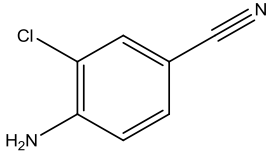
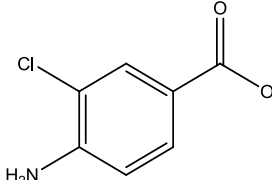
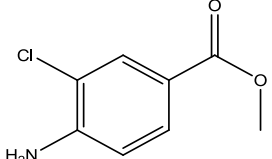
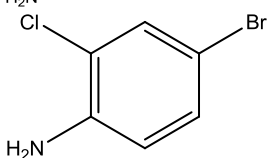
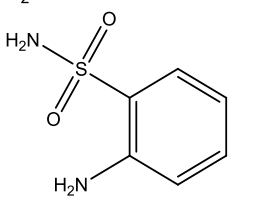
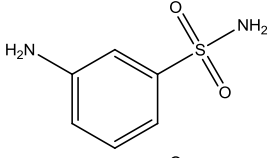
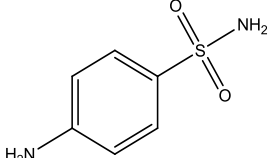
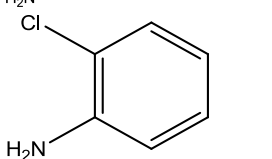
Rational design of isoform specific ligands

18		3-Amino-2-pyridinecarboxylic acid	138.12	Fluorochem	046510
19		3-Aminopyridine-2-carboxamide	137.14	Fluorochem	048148
20		Methyl 3-Aminopyridine-2-carboxylate	152.15	Fluorochem	066881
21		2,3-Diamino-5-bromopyridine	188.03	Fluorochem	043711
22		5-Fluoropyridine-2,3-diamine	127.12	Fluorochem	223387
23		2,3-Diamino-5-chloropyridine	143.57	sigma-aldrich	714623
24		2-amino-5-iodo-3-pyridinylamine	235.02	Key Organics	DC-0722
25		2,3-Diamino-6-chloropyridine	143.57	Fluorochem	049431
26		3-Amino-2-chloro-5-picoline	142.59	Fluorochem	033670
27		3-Amino-2,5-dichloropyridine	163.01	Fluorochem	017241
28		3-Amino-5-bromo-2-chloropyridine	207.46	Fluorochem	033563

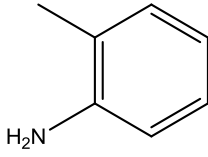
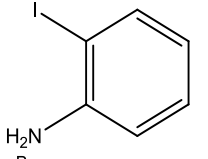
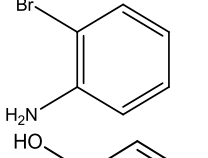
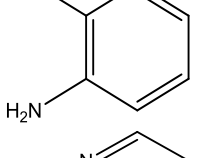
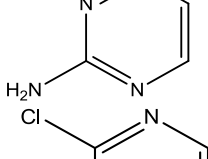
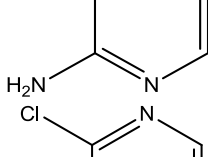
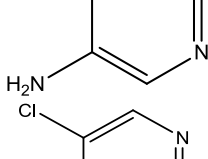
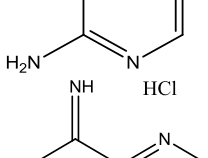
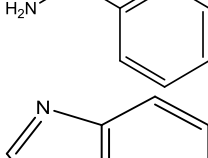
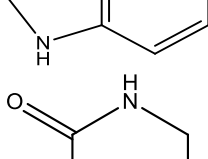
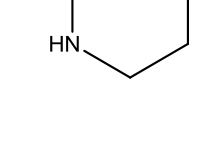
Rational design of isoform specific ligands

29		2-Chloro-5-iodopyridin-3-amine	254.46	Fluorochem	240335
30		2-Chloro-5-(trifluoromethyl)pyridin-3-amine	196.56	Fluorochem	091558
31		2-Chloro-6-methylpyridin-3-amine	142.59	Fluorochem	064787
32		3-Amino-2,6-dichloropyridine	163.01	Fluorochem	017208
33		2-Chloro-5-fluoroaniline	145.56	Fluorochem	004416
34		2-chloro-5-methoxyaniline	157.60	Fluorochem	358387
35		3-Amino-4-chlorobenzonitrile	152.58	Fluorochem	077955
36		3-Amino-4-chlorobenzoic acid	171.58	Fluorochem	033910
37		Methyl 3-amino-4-chlorobenzoate	185.61	Fluorochem	040397
38		3-Amino-4-chlorobenzamide	170.60	Fluorochem	068290
39		3-Amino-4-chlorobenzenesulfonic acid	207.63	Fluorochem	226610

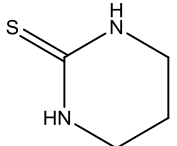
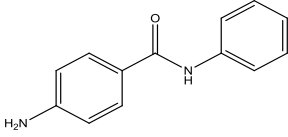
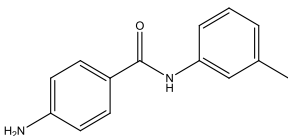
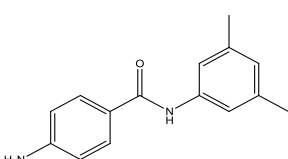
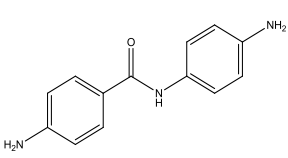
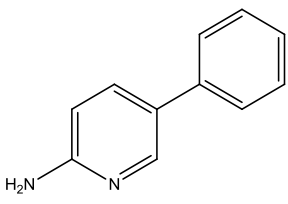
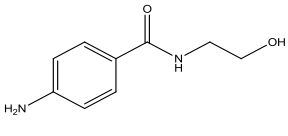
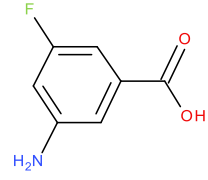
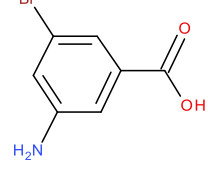
Rational design of isoform specific ligands

40		2-Chloro-4-fluoroaniline	145.56	Fluorochem	001691
41		2-Chloro-4-iodoaniline	126.57	Fluorochem	005436
42		4-Amino-3-chlorobenzonitrile	152.58	Fluorochem	035185
43		4-Amino-3-chlorobenzoic acid	171.58	Fluorochem	021888
44		Methyl 4-amino-3-chlorobenzoate	185.61	Fluorochem	012690
45		4-Bromo-2-chloroaniline	206.48	Fluorochem	003692
46		2-Aminobenzenesulfonamide	172.21	Fluorochem	209596
47		3-Aminobenzenesulfonamide	172.21	Fluorochem	068167
48		4-Aminobenzenesulfonamide	172.21	Fluorochem	037364
49		2-Chloroaniline	127.57	sigma-aldrich	23300

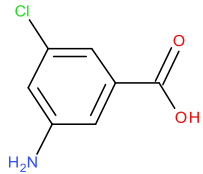
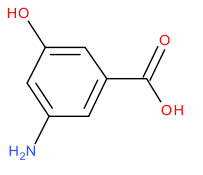
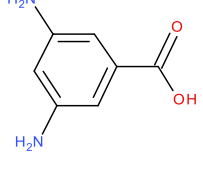
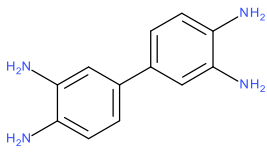
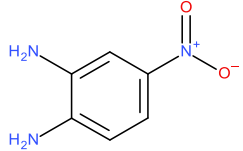
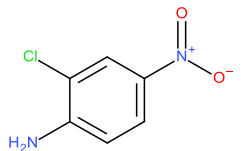
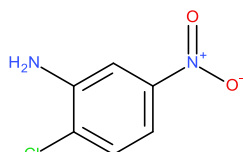
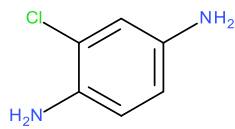
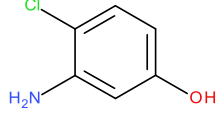
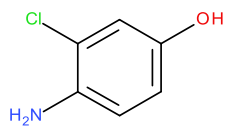
Rational design of isoform specific ligands

50		o-Toluidine	107.16	Fluorochem	043316
51		2-Iodoaniline	219.03	Fluorochem	001038
52		2-Bromoaniline	172.03	Fluorochem	005347
53		2-Aminophenol	109.05	Fluorochem	094977
54		2-Aminopyrimidine	95.10	Fluorochem	036344
55		2-Amino-3-chloropyrazine	129.55	Fluorochem	035839
56		5-Amino-4-chloropyrimidine	129.55	Fluorochem	077408
57		4-Amino-5-chloropyrimidine	129.55	Fluorochem	076283
58		Pyridine-2-carboximidamide hydrochloride	157.60	Fluorochem	044914
59		1H-Benzo[d]imidazole	118.14	Fluorochem	235007
60		Tetrahydro-2(1H)-pyrimidinone	100.12	Fluorochem	208604

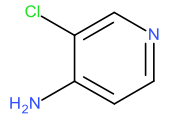
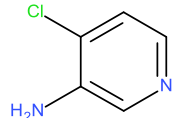
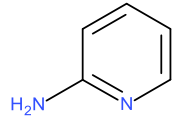
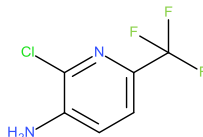
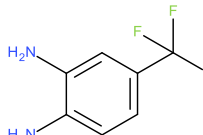
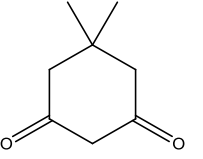
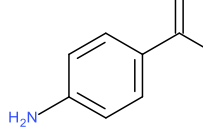
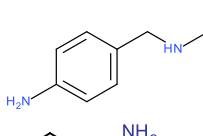
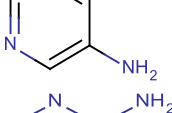
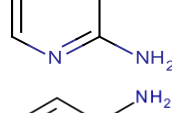
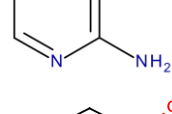
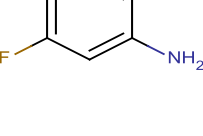
Rational design of isoform specific ligands

61		2-(1H)- Tetrahydropyrimidi nethione	116.19	Fluorochem	018995
62		4-amino-N- phenylbenzamide	212.25	Fluorochem	355603
63		N-(4-aminophenyl)- 4-methylbenzamide	226.28	Fluorochem	310000
64		4-amino-N-(3,5- dimethylphenyl)ben zamide	240.30	Fluorochem	315797
65		4,4'- Diaminobenzanilide	227.27	Fluorochem	125000
66		2-Amino-5- phenylpyridine	170.21	Fluorochem	013115
67		4-Amino-N-(2- hydroxyethyl)benza mide	180.21	Fluorochem	020193
68		3-Amino-5- fluorobenzoic acid	155.13	Fluorochem	078123
69		3-Amino-5- bromobenzoic acid	216.04	Fluorochem	210956

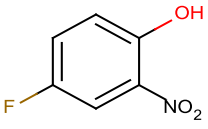
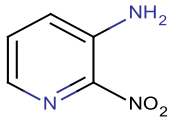
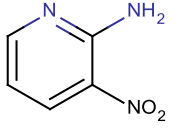
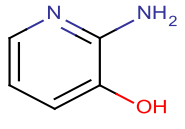
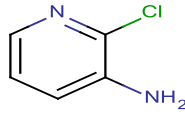
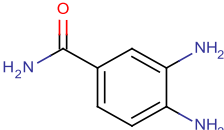
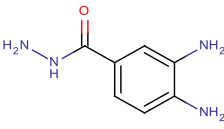
Rational design of isoform specific ligands

70		3-Amino-5-chlorobenzoic acid	171.58	Fluorochem	036195
71		3-Amino-5-hydroxybenzoic acid	153.13	Fluorochem	223494
72		3,5-Diaminobenzoic acid	152.15	Fluorochem	209099
73		3,3'-Diaminobenzidine	214.27	Fluorochem	342164
74		4-Nitro-O-phenylenediamine	153.14	Fluorochem	077317
75		2-Chloro-4-nitroaniline	172.57	Fluorochem	075157
76		2-Chloro-5-nitroaniline	172.57	Fluorochem	078002
77		4-Chloro-1,3-diaminobenzene	142.59	sigma-aldrich	125172
79		3-Amino-4-chlorophenol	143.57	sigma-aldrich	552488
80		4-Amino-3-chlorophenol	143.57	Fluorochem	223449

Rational design of isoform specific ligands

81		4-Amino-3-Chloropyridine	128.56	Fluorochem	012738
82		3-Amino-4-Chloropyridine	128.56	Fluorochem	035847
83		2-Aminopyridine	94.12	Fluorochem	050259
84		3-Amino-2-chloro-6-(trifluoromethyl)pyridine	196.56	Fluorochem	009005
85		3,4-Diaminobenzotrifluoride	176.14	Fluorochem	003179
86		Dimezone	140.18	Fluorochem	011151
87		4-Aminobenzamide	136.15	sigma-aldrich	284572
88		4-Amino-benzyl-urea	165.19	Fluorochem	040246
89		3,4-Pyridinediamine	109.13	sigma-aldrich	D7148
90		4,5-diaminopyrimidine	110.12	Fluorochem	080010
91		2,3-Diaminopyridine	109.13	sigma-aldrich	125857
92		Phenol, 2-amino-4-fluoro-	127.12	Fluorochem	13048

Rational design of isoform specific ligands

93		Phenol, 4-fluoro-2-nitro-	157.1	Fluorochem	2388
94		3-Pyridinamine, 2-nitro-	139.11	sigma-aldrich	L510165
95		2-Pyridinamine, 3-nitro-	139.11	sigma-aldrich	113514
96		3-Pyridinol, 2-amino-	109.11	sigma-aldrich	122513
97		2-Chloropyridine, 3-amino-	128.56	sigma-aldrich	A46900
98		3,4-diaminobenzamide	151.17	BIONET	KH-0709
99		3,4-diaminobenzhydrazide	166.18	Alfa Aesar	L05263

7.3 SPR dissociation constants and stoichiometry results

Table 7-2 Table of dissociation constants and stoichiometry results from the dilution series SPR experiment

Table lists the binding affinities (K_d), as were calculated from a steady state fit, stoichiometry (Stoich.) for every compound upon interaction with CypA, CypB and CypD and their solubility.

K_d :

Molar = Compounds that show binding affinity in the Molar range.

No fitting = Compound's kinetic curve does not fit the steady state affinity model used.

Stoichiometry:

X:Y = X protein molecules interact with Y ligand molecule

X = Stoichiometry was not determined

Solubility

‡ Ligands' stock solutions for SPR measurements have been prepared using 100% Ethanol

† Ligands' stock solutions for SPR measurements have been prepared using 100% PBS solution

Lig. Number	Solubility	SPR code	CypA		CypB		CypD	
			K_d (mM)	Stoich.	K_d (mM)	Stoich.	K_d (mM)	Stoich.
2	‡	A2	8.53	1:1	No fitting	X	No fitting	X
3	‡	A3	4.44	1:1	No fitting	X	No fitting	X
5	‡	A5	Molar	X	No fitting	X	No fitting	X
7	‡	A7	5.76	1:2	2.53	<1:1	Molar	X
8	‡	A8	Molar	X	No fitting	X	No fitting	X
14	‡	B1	Molar	X	No fitting	X	No fitting	X
15	‡	B2	2.27	1:1	No fitting	X	No fitting	X

Rational design of isoform specific ligands

16	†	B3	No fitting	X	No fitting	X	No fitting	X
17	‡	B4	No fitting	X	No fitting	X	No fitting	X
18	†	B5	Molar	X	No fitting	X	No fitting	X
22	‡	B8	98.17	>> 1:10	52.94	>> 1:10	146	>> 1:10
23	‡	B9	2.71	1:1	1.31	1:1	7.97	1:1
24	‡	B10	2.25	1:1	1.89	1:1	3.42	1:1
25	‡	B11	10.28	1:4	3.5	1:1	Molar	X
26	‡	B12	Molar	X	Molar	X	Molar	X
29	‡	C3	8.32	1:2	1.47	< 1:1	12.89	1:1
31	‡	C5	3.52	1:1	No fitting	X	No fitting	X
39	†	C12	No fitting	X	No fitting	X	No fitting	X
41	‡	D3	9.7	1:2	No fitting	X	No fitting	X
42	‡	D4	91.41	>> 1:10	No fitting	X	No fitting	X
49	‡	D11	27.49	1:2	No fitting	X	No fitting	X
50	‡	D12	Molar	X	No fitting	X	No fitting	X
53	‡	E3	Molar	X	No fitting	X	No fitting	X
54	‡	E4	No fitting	X	No fitting	X	No fitting	X
55	‡	E5	Molar	X	No fitting	X	No fitting	X
56	†	E6	0.875	<1:1	5.62	<1:1	6.52	<1:1
59	‡	E8	Molar	X	Molar	X	Molar	X
60	†	E9	4.63	1:1	5.48	1:1	2.69	< 1:1
61	†	E10	5.34	1:6	No fitting	X	0.85	1:1
62	‡	E11	Molar	X	Molar	X	Molar	X

Rational design of isoform specific ligands

63	‡	E12	3.12	1:1	1.73	<1:1	3.64	1:1
64	‡	F1	0.936	<1:1	0.457	<1:1	0.872	<1:1
66	‡	F2	Molar	X	Molar	X	Molar	X
75	‡	F9	Molar	X	Molar	X	Molar	X
77	‡	F10	3.26	1:1	No fitting	X	No fitting	X
79	‡	F12	5.74	1:1	No fitting	X	No fitting	X
80	‡	G1	8.4	1:2	9.82	1:1	Molar	X
82	†	G3	Molar	X	Molar	X	Molar	X
83	‡	G4	No fitting	X	No fitting	X	No fitting	X
86	‡	G7	Molar	X	Molar	X	Molar	X
89	‡	G10	No fitting	X	No fitting	X	No fitting	X
90	‡	G11	6.49	1:1	No fitting	X	No fitting	X
91	‡	G12	Molar	X	No fitting	X	No fitting	X
92	‡	H1	Molar	X	Molar	X	Molar	X
96	‡	H4	Molar	X	82.37	>>1:1	Molar	X
97	‡	H5	1.11	<<1:1	No fitting	X	No fitting	X
98	†	H6	2.28	1:1	1.89	1:1	1.72	1:1
99	†	H7	6.04	1:2	5.41	1:2	4.2	1:1

7.4 Steady state response curves

Figure 7-1 below, pages 178 -197, show the steady state curves for all 48 compounds and three different Cyp isoforms, screened with SPR using a 2 two fold dilution series.

Equation 12 is used to fit the steady state affinity curves and for the calculation of the K_D values as, shown and explained in page 73.

Explanatory features for each graph:

X Axis: Concentration of ligand in mol L⁻¹

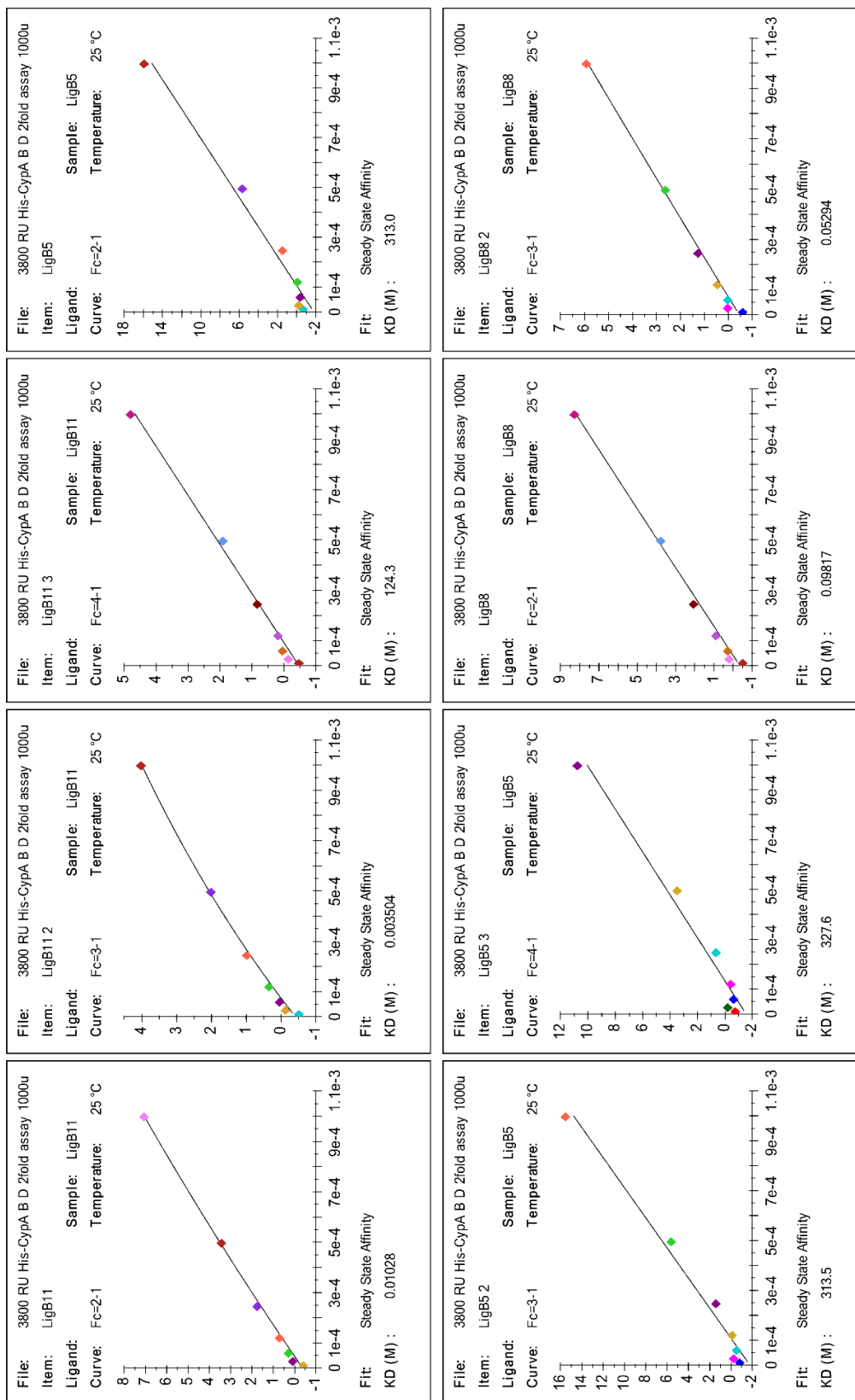
Y Axis: Reference corrected response in response units (RU)

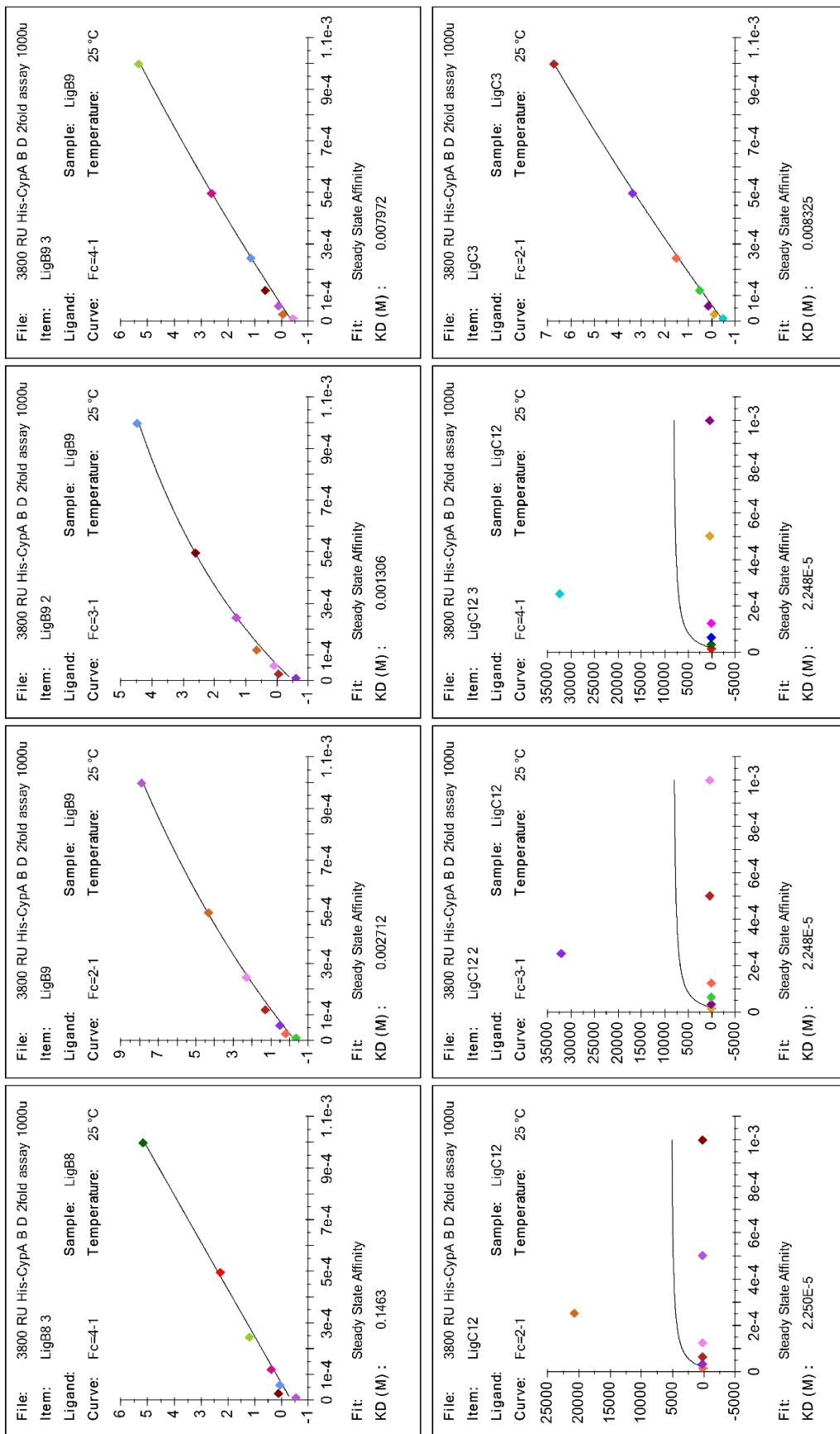
Sample: Ligand tested, coded based on the number of the well of the SPR plate. The ligand number matching this code can be found from Table 7-2 on SPR dissociation constants and stoichiometry results.

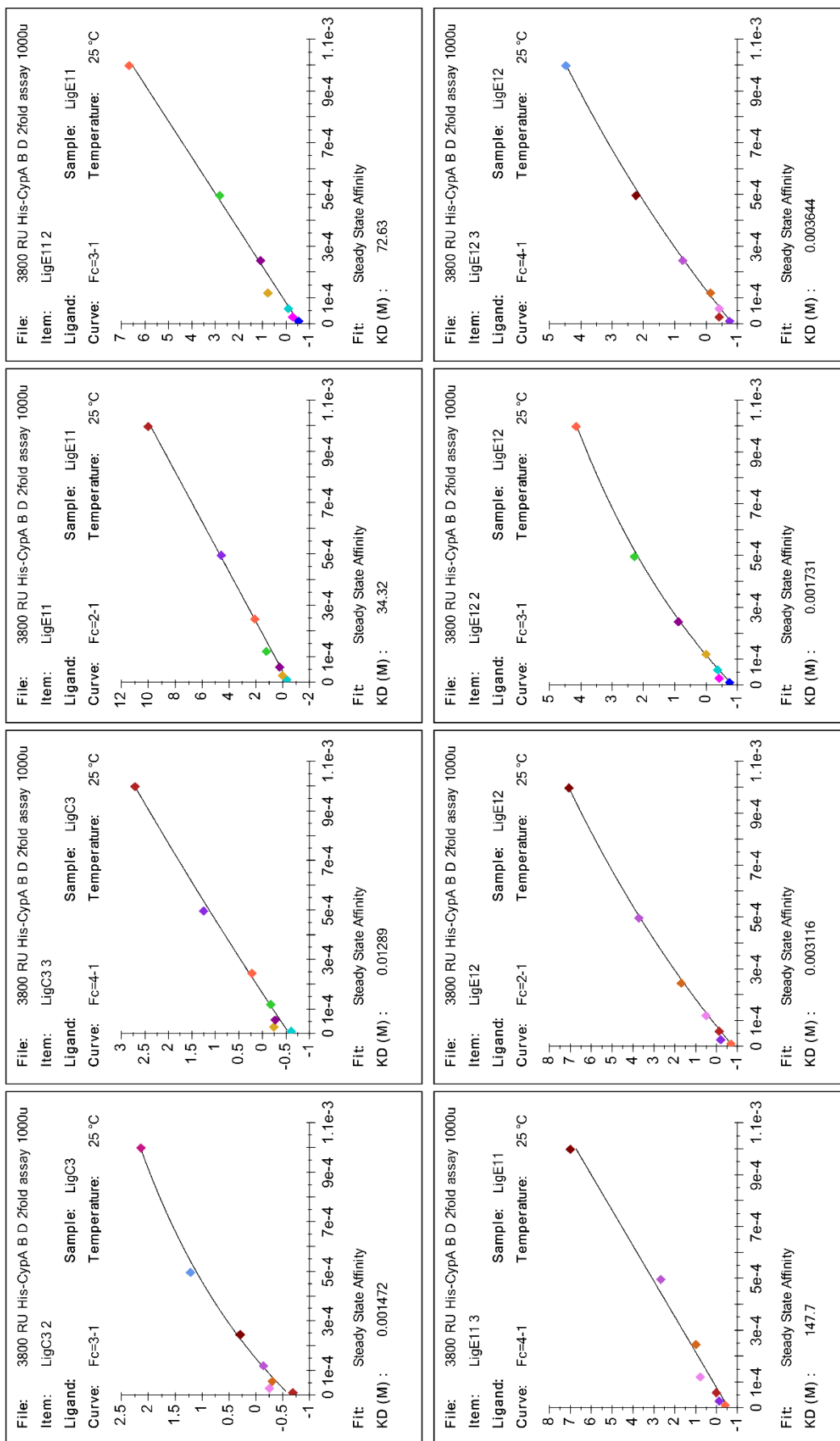
Curve: The cell surface at which this result was measured. As described before cell 1 was the reference cell, CypA, CypB and CypD were immobilised in cell 2, 3 and 4 respectively. So curve 2-1, as an example, shows the steady state affinity curve of the sample when passed over CypA surface, after the reference curve was subtracted.

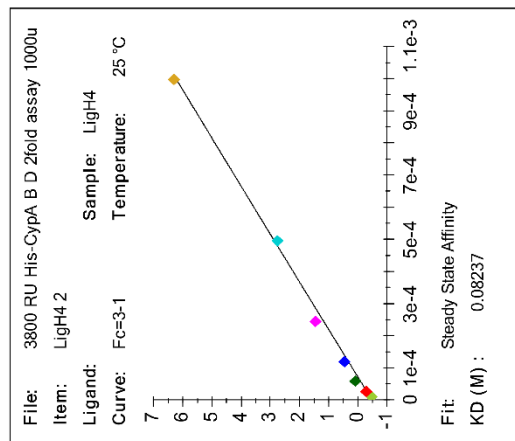
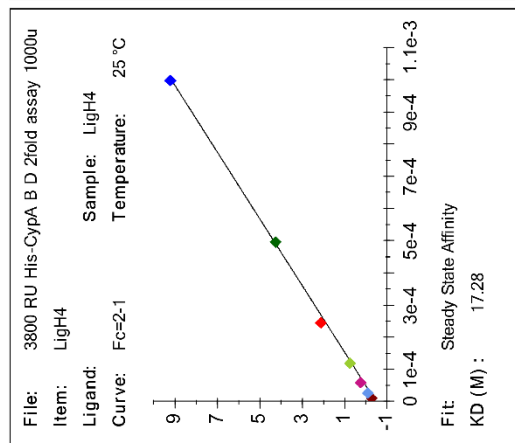
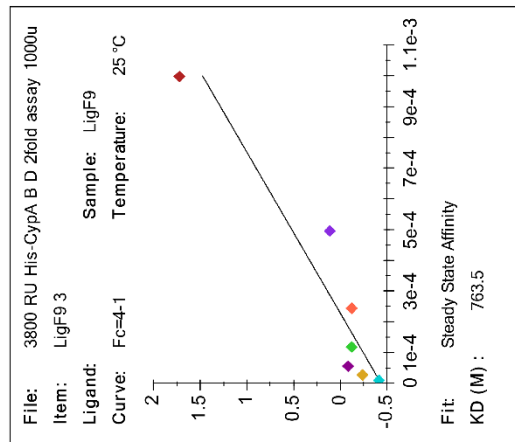
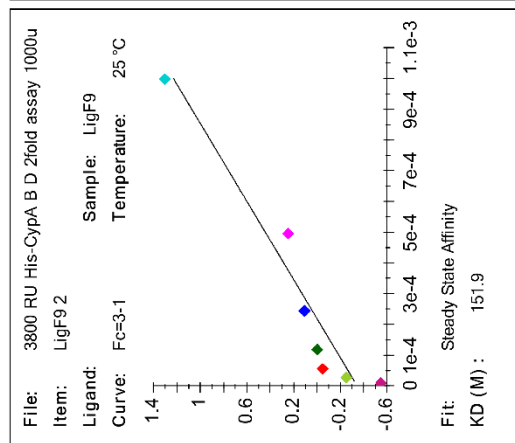
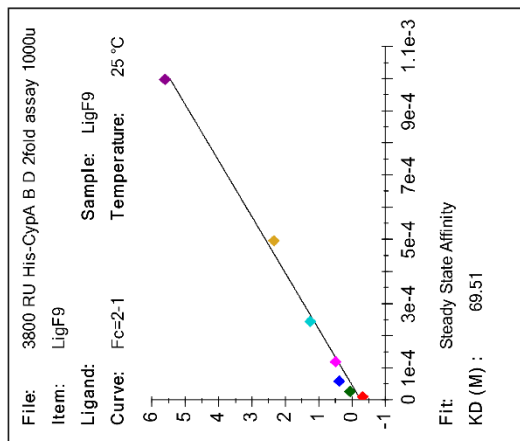
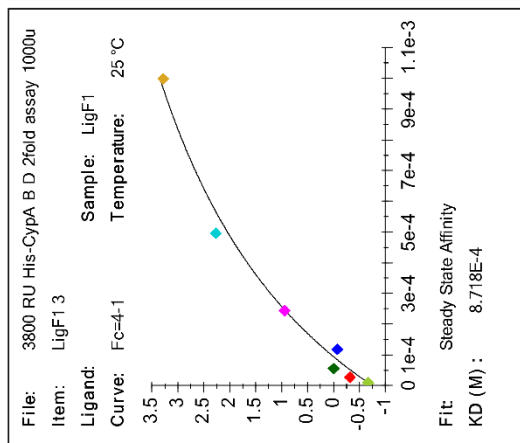
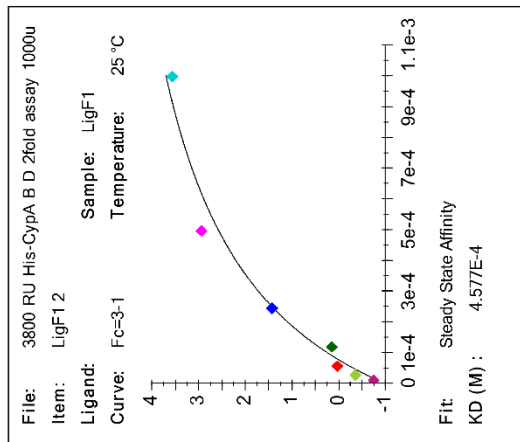
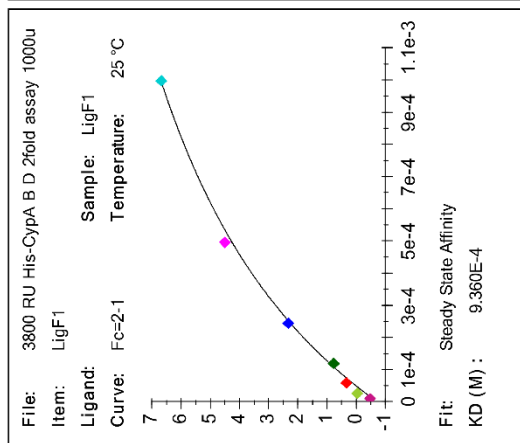
K_D : is the dissociation constant calculated after the steady state fitting

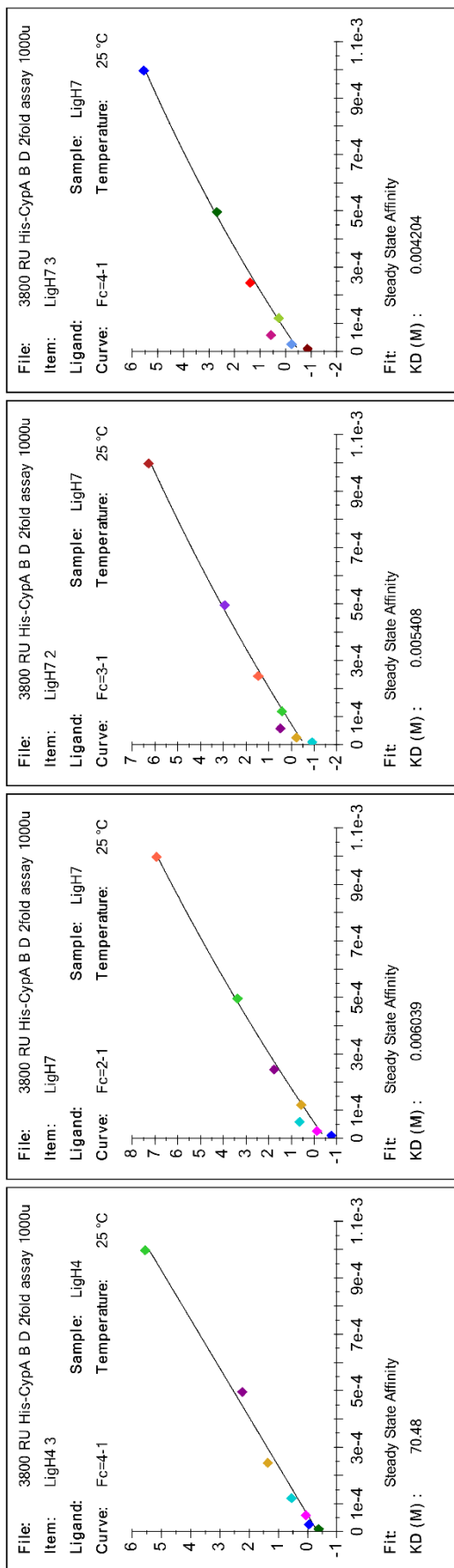
Figure 7-1 Steady state response curves

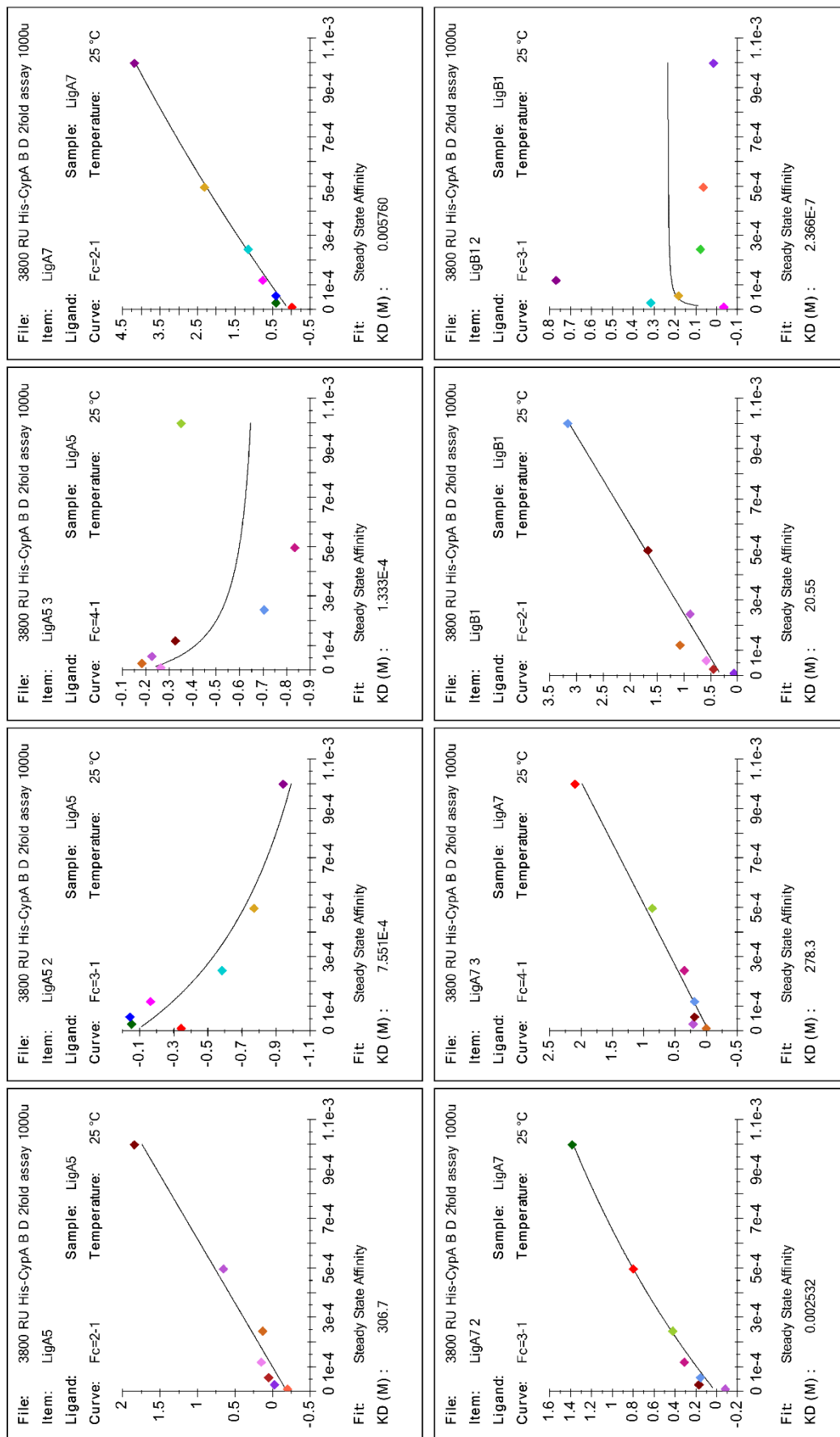


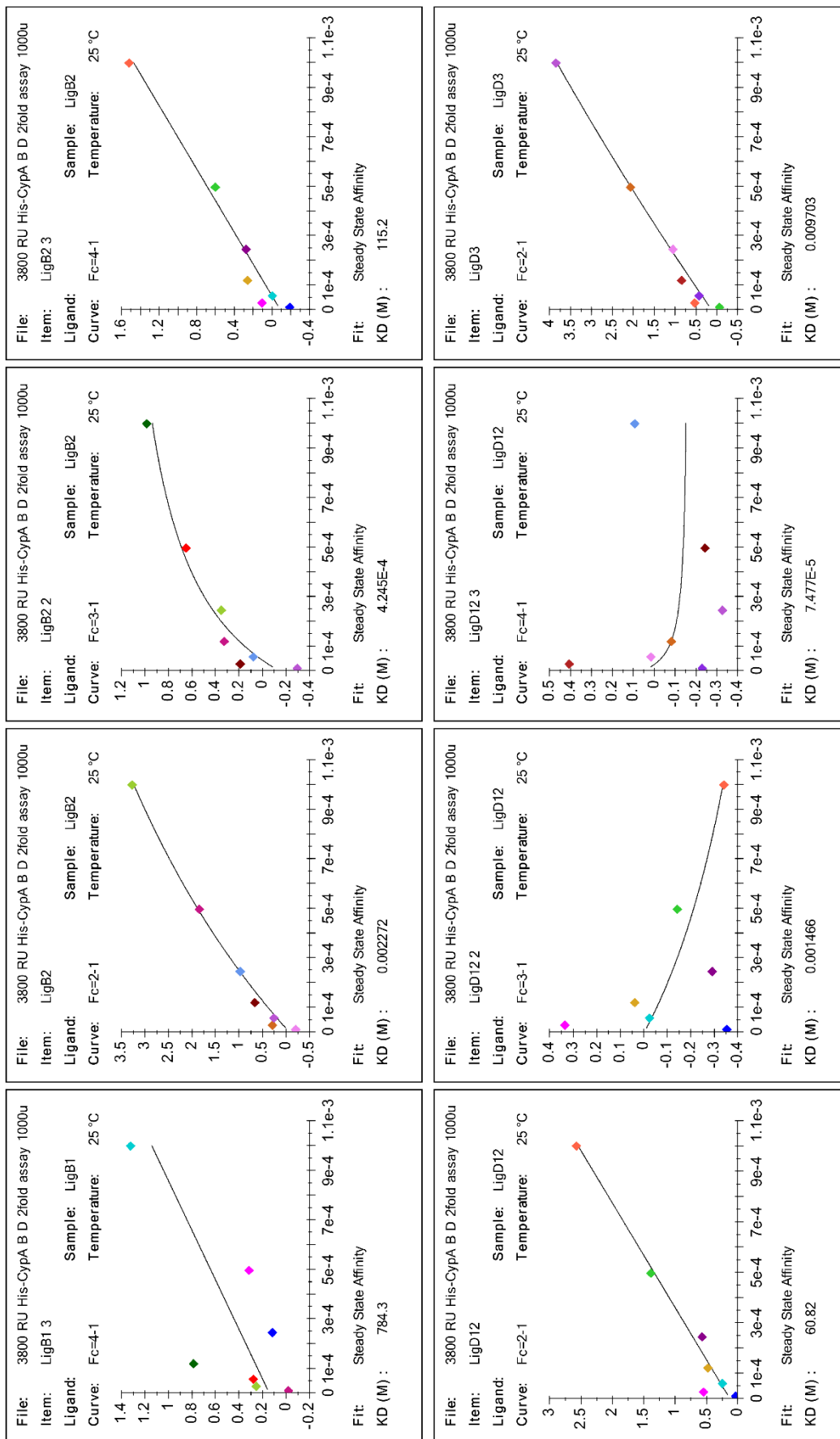


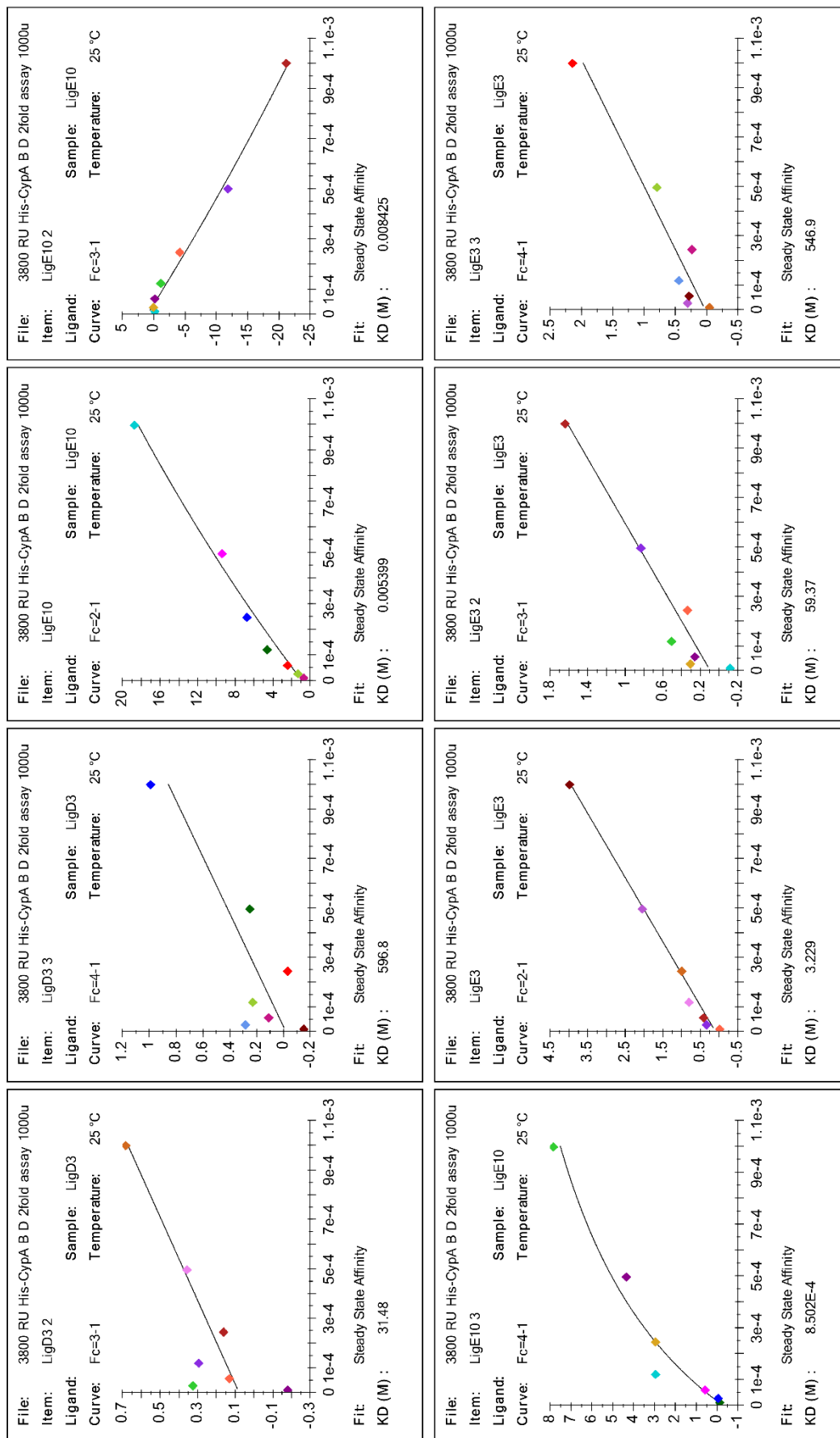


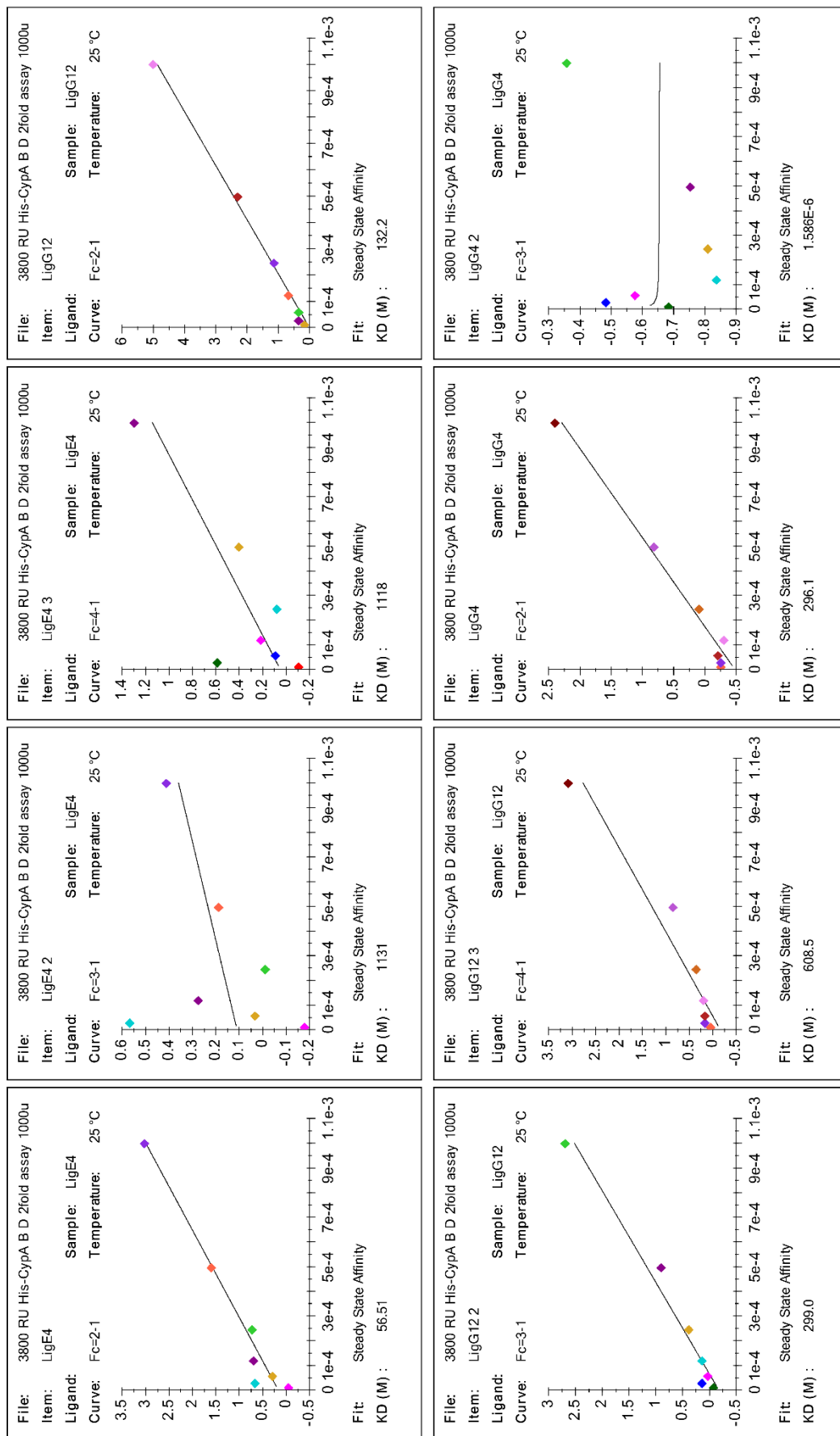


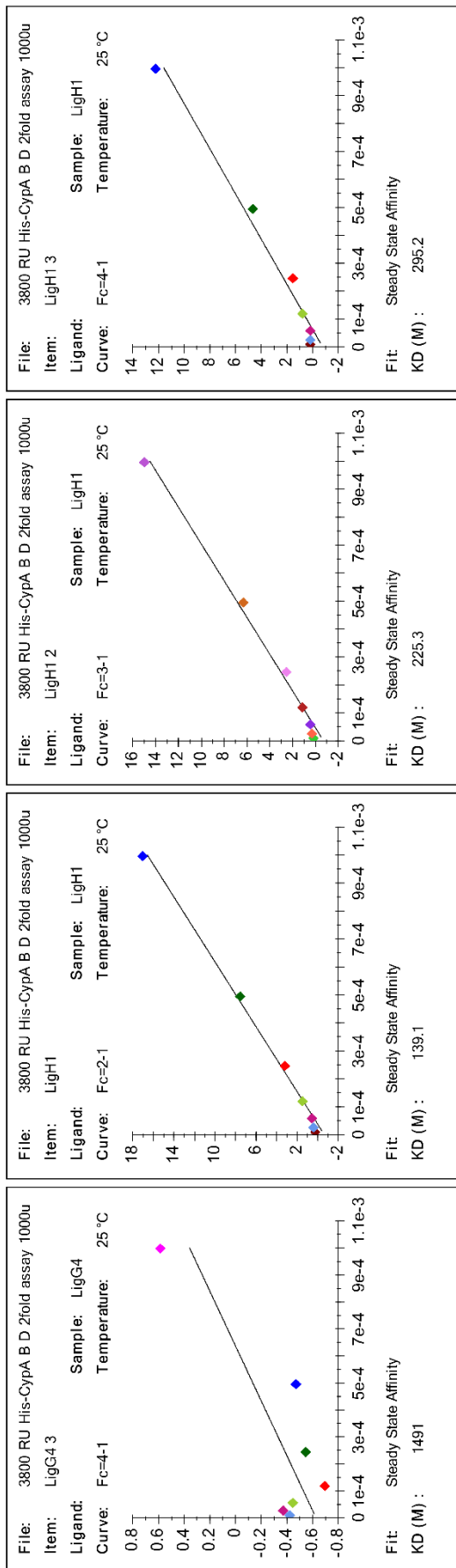


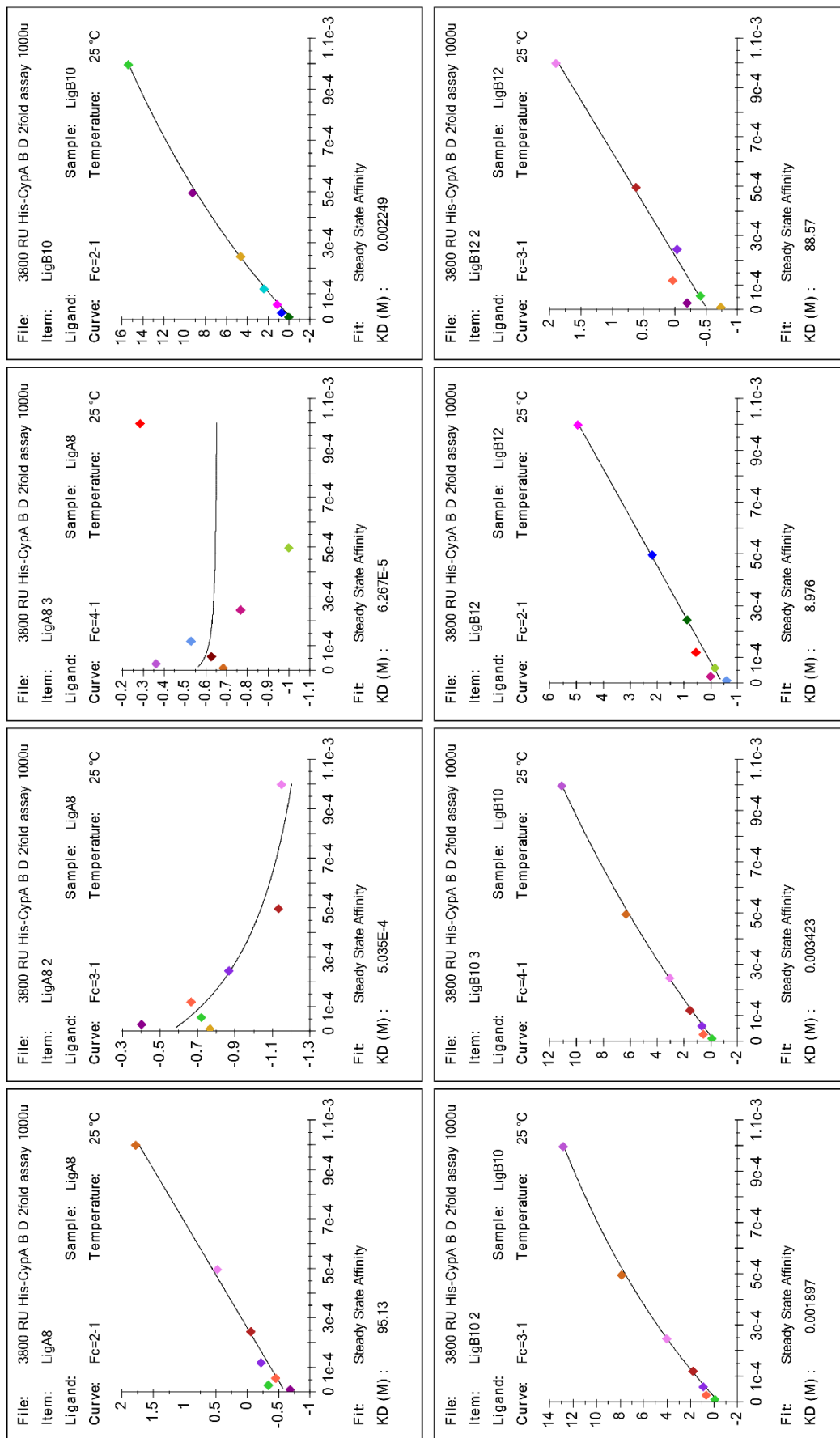


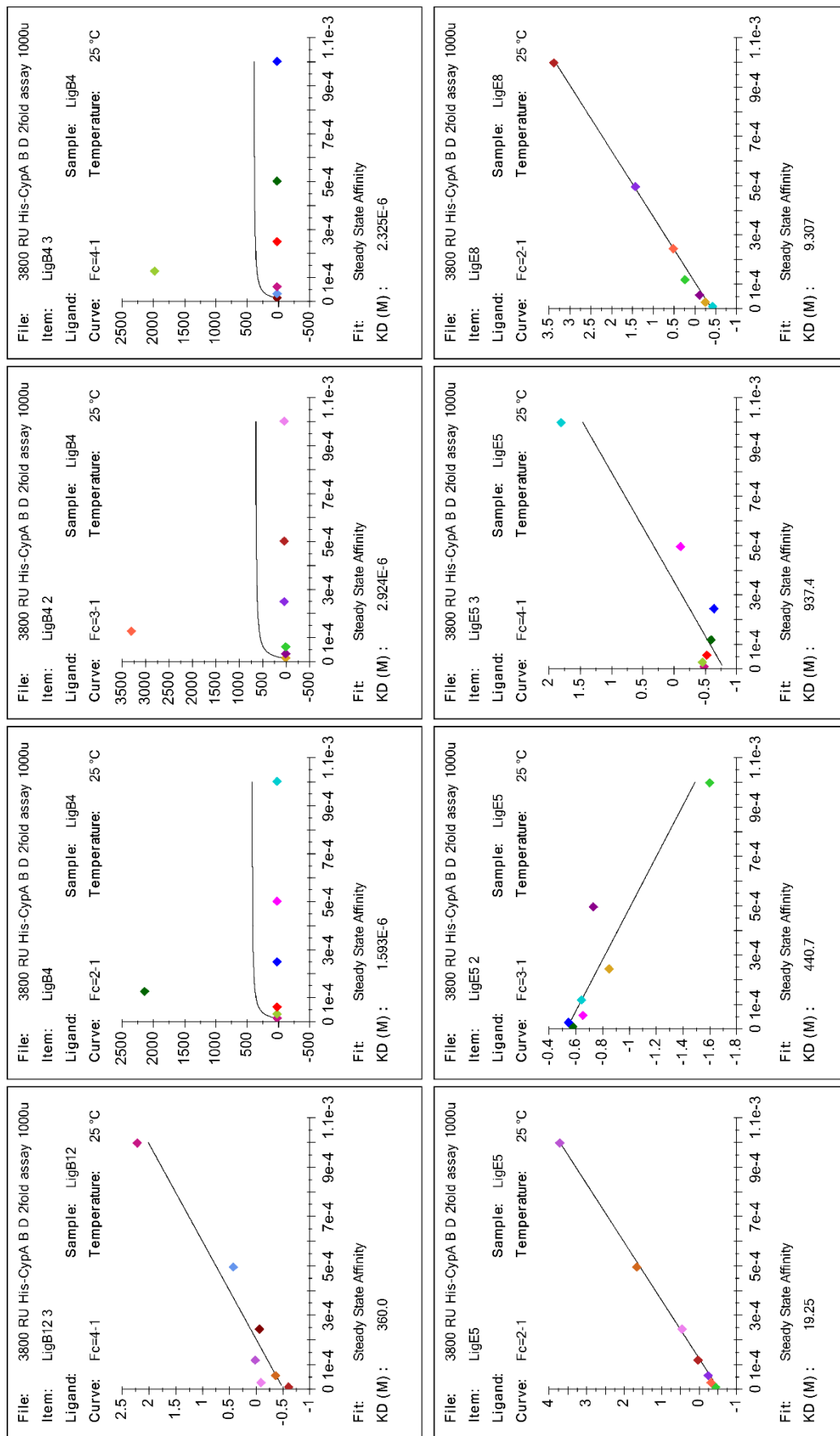


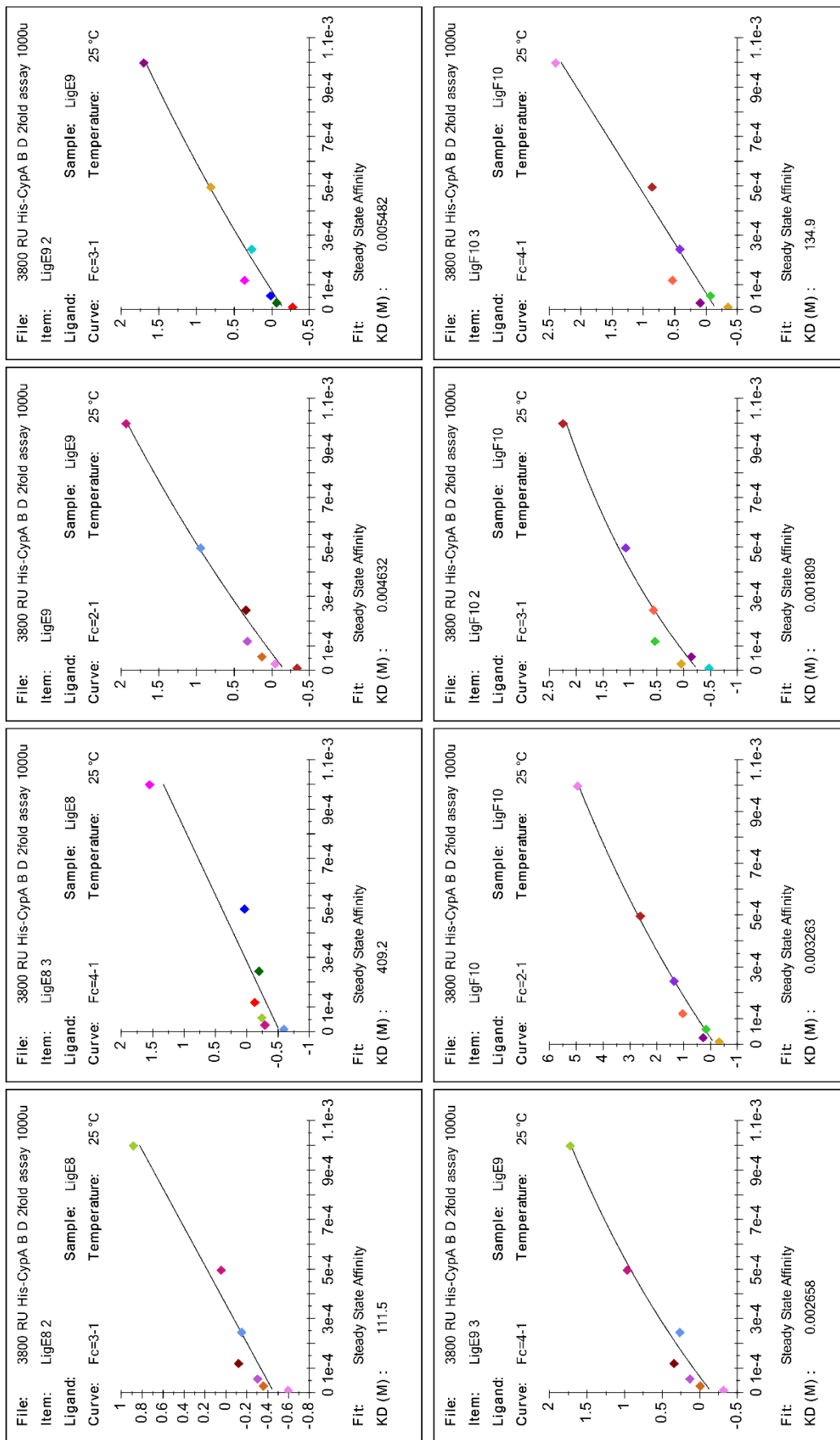


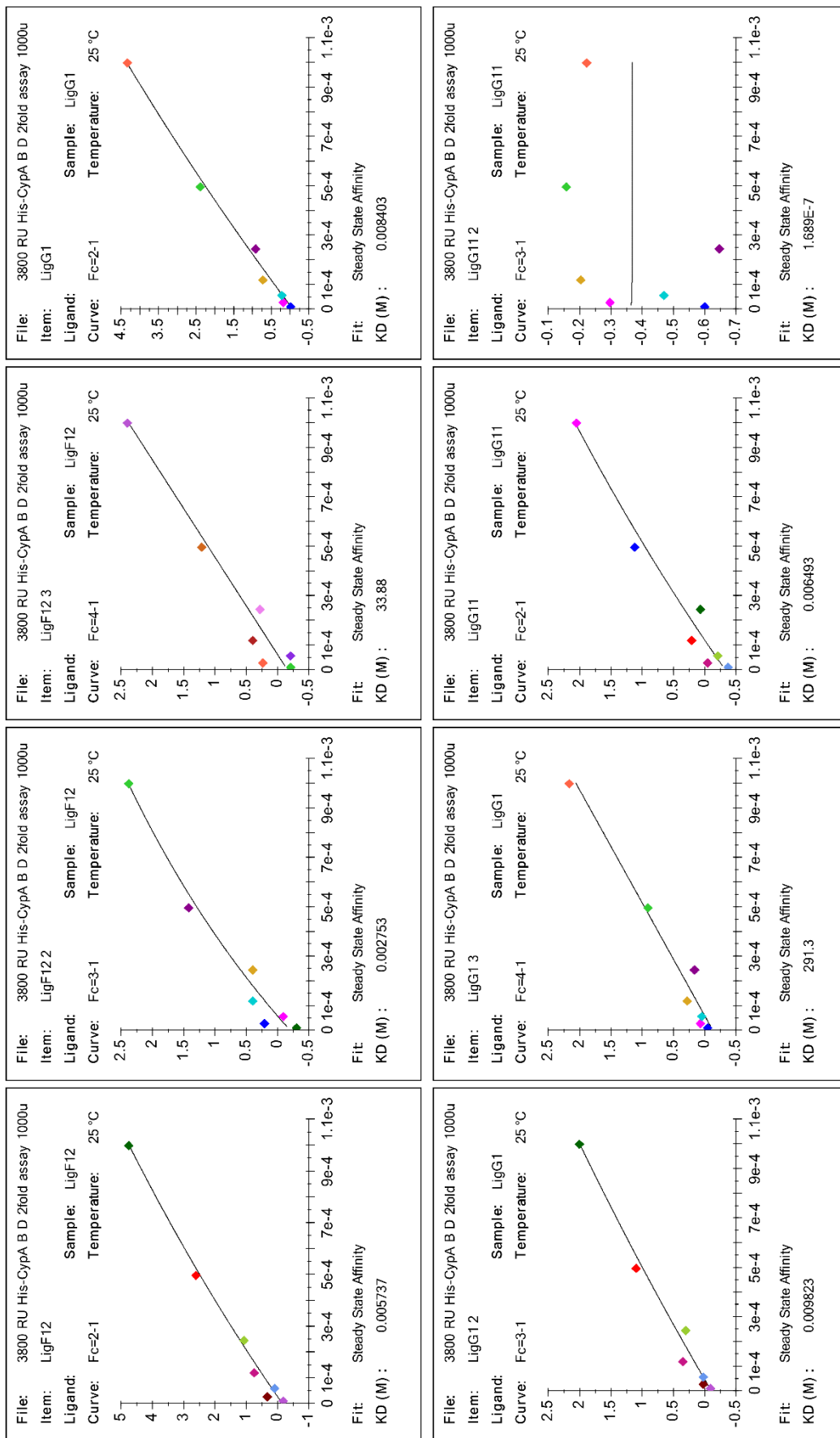


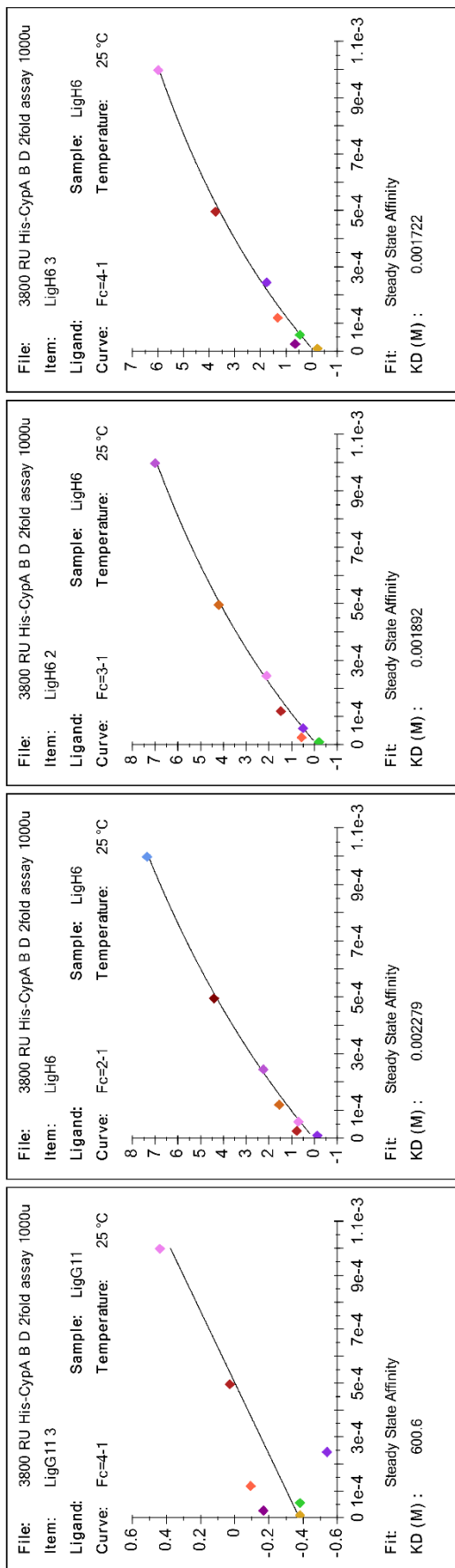


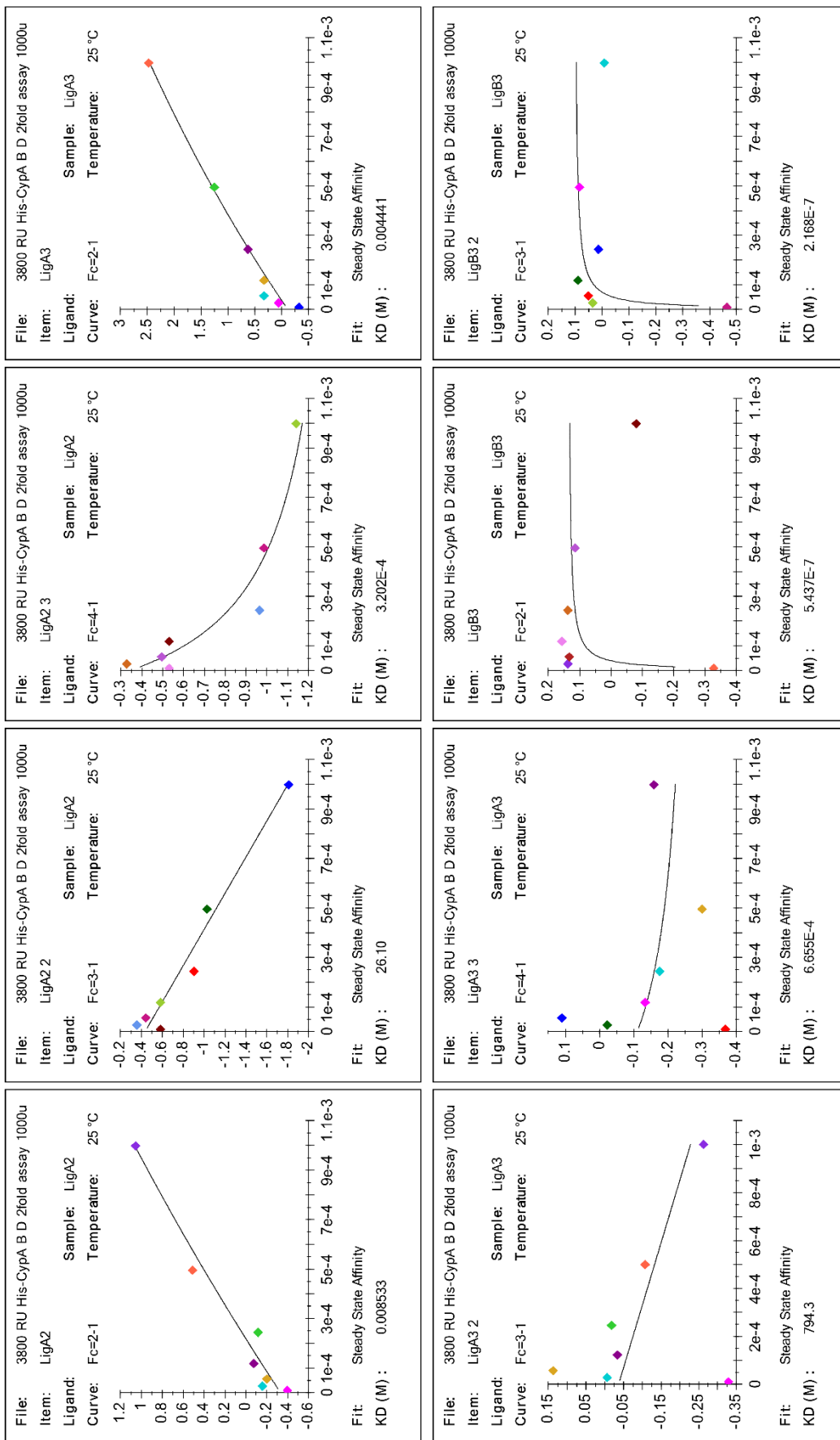


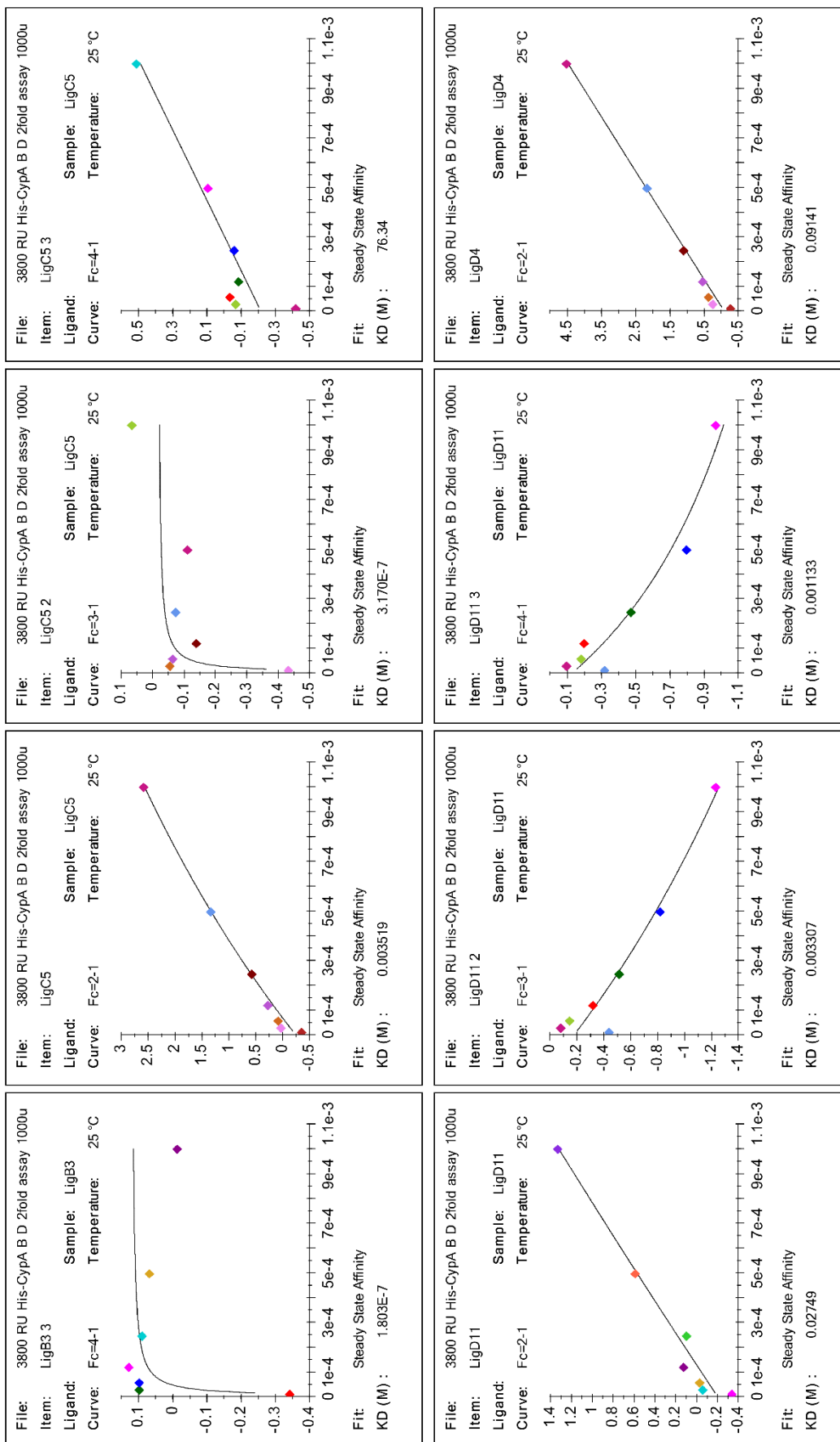


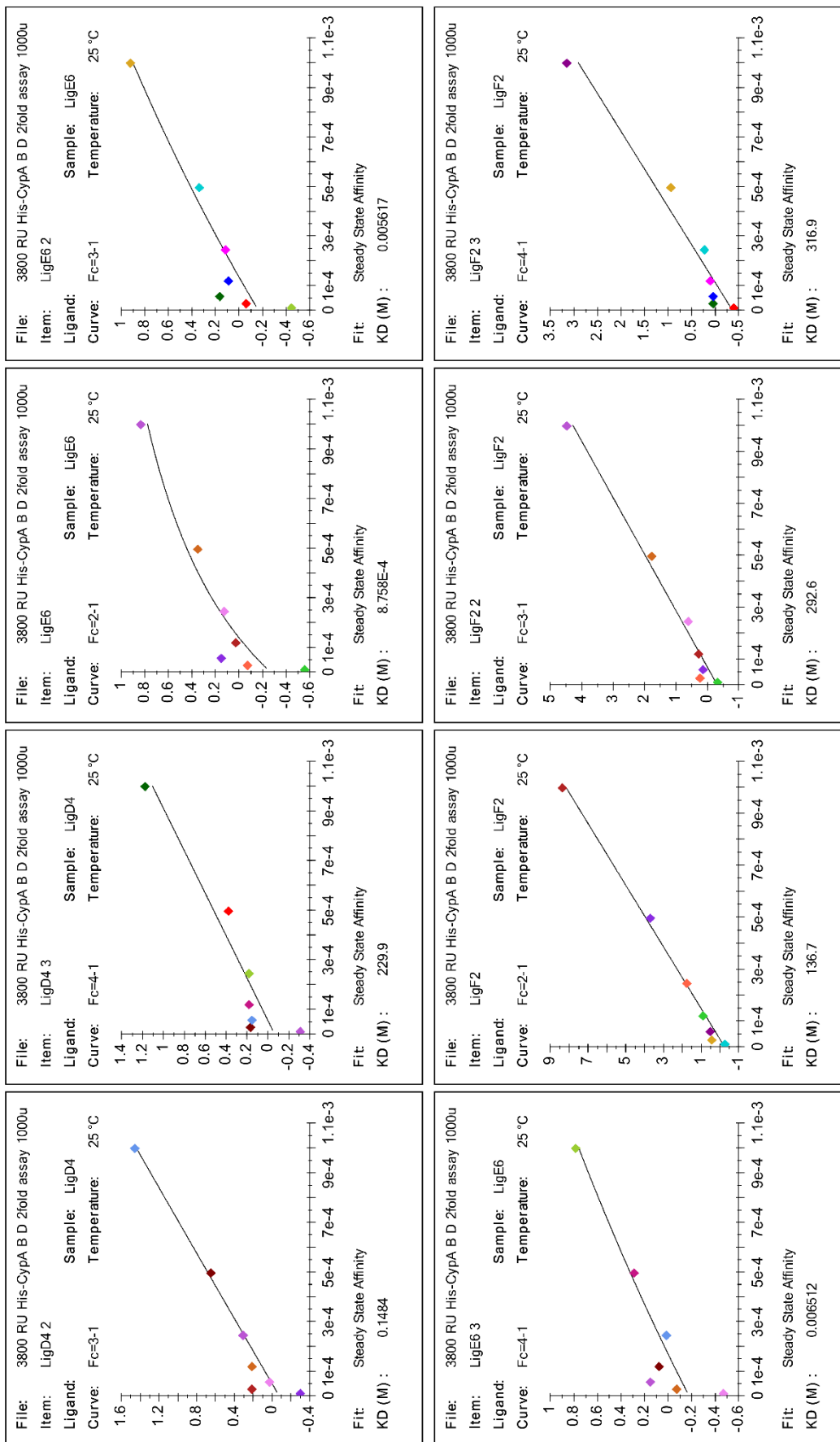


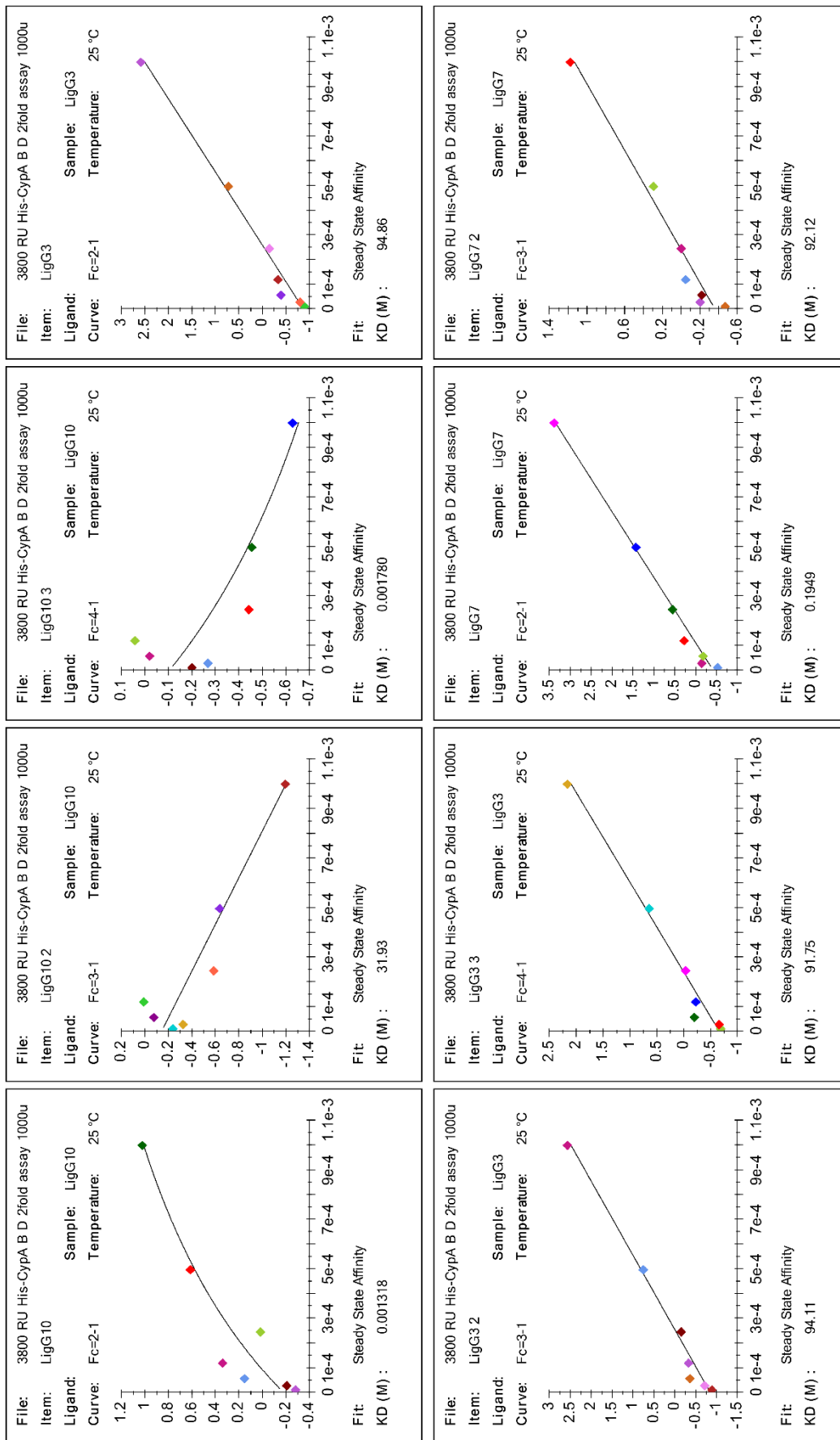


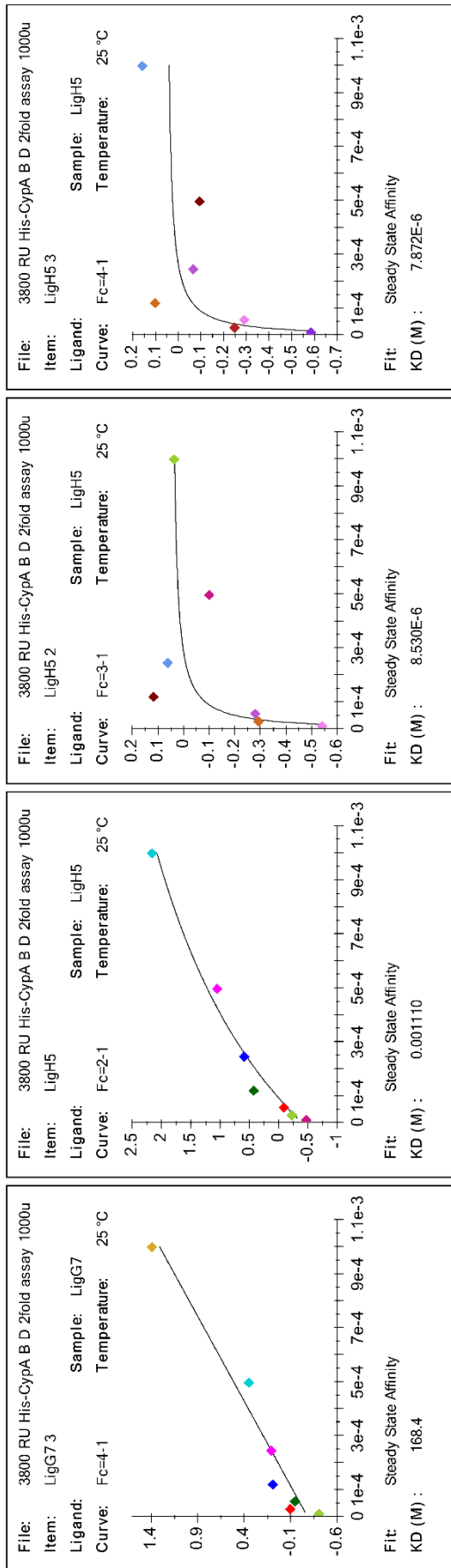












7.5 Binding of fragments in the active site of CypA

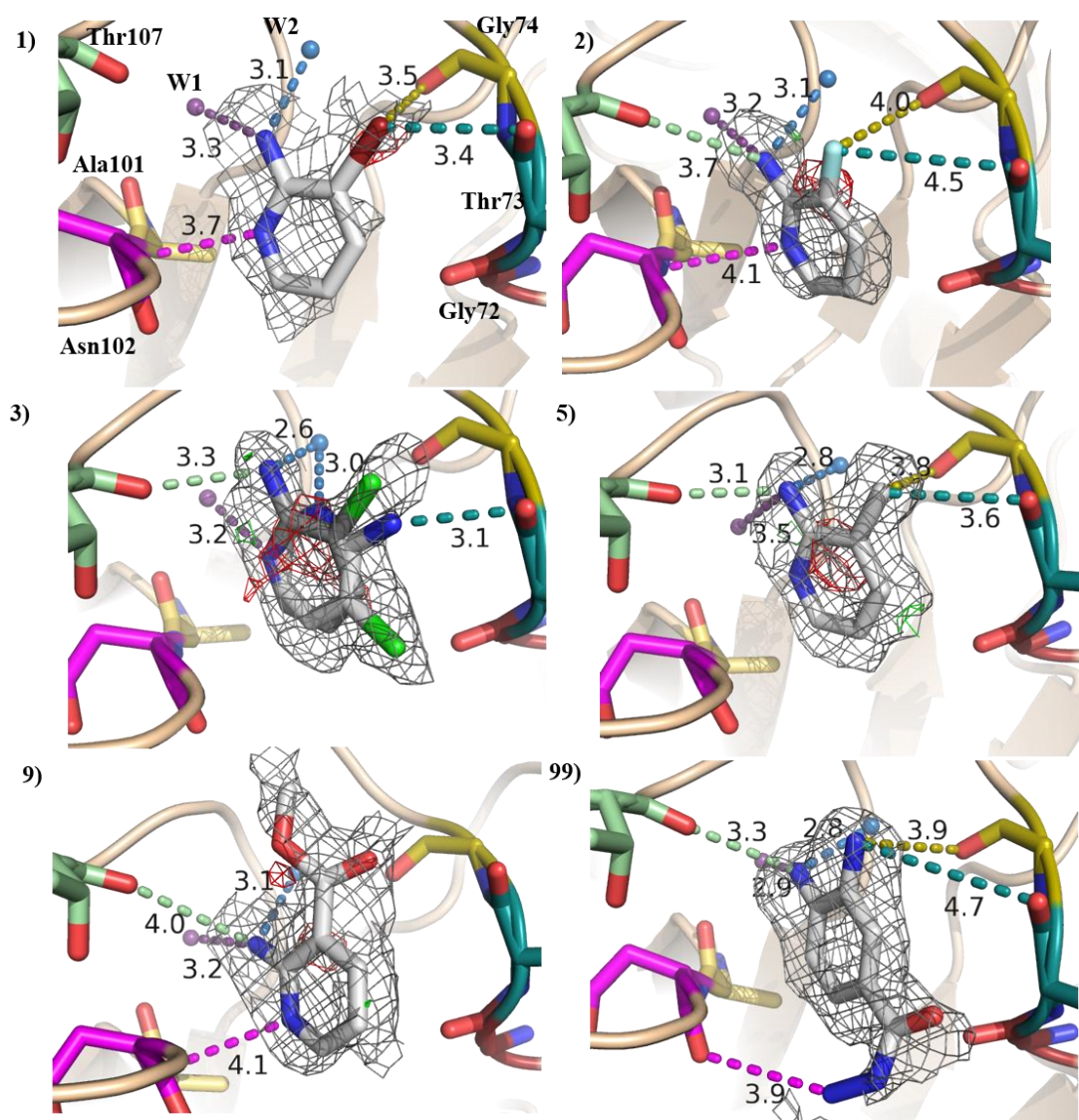


Figure 7-2 Binding of compounds 1, 2, 3, 5, 9 and 99 on the Abu pocket of CypA. $2F_o - F_c$ electron density and $F_o - F_c$ maps, at 1.0σ and 3.0 are shown in grey and red/green respectively. Distances between ligand and water or CypA molecules are highlighted and colour coded based on residue colour.

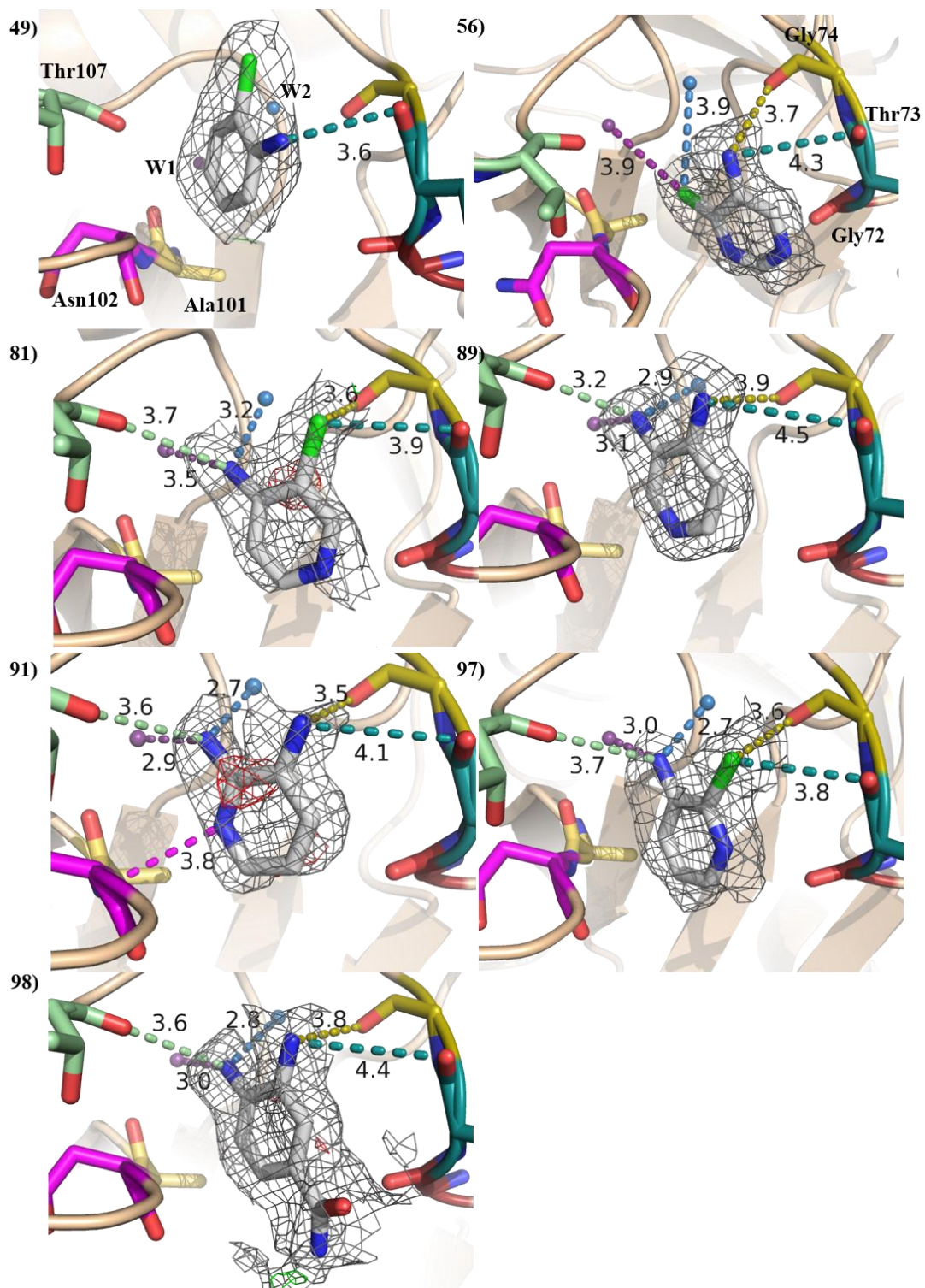


Figure 7-3 Binding of compounds 49, 56, 81, 89, 91, 97 and 98 in the Abu pocket of CypA. $2F_o - F_c$ electron density and $F_o - F_c$ maps, at 1.0σ and 3.0σ are shown in grey and red/green respectively. Distances between ligand and water or CypA molecules are highlighted and colour coded based on residue colour.

7.6 Binding of compounds out of the 80's loop of CypA

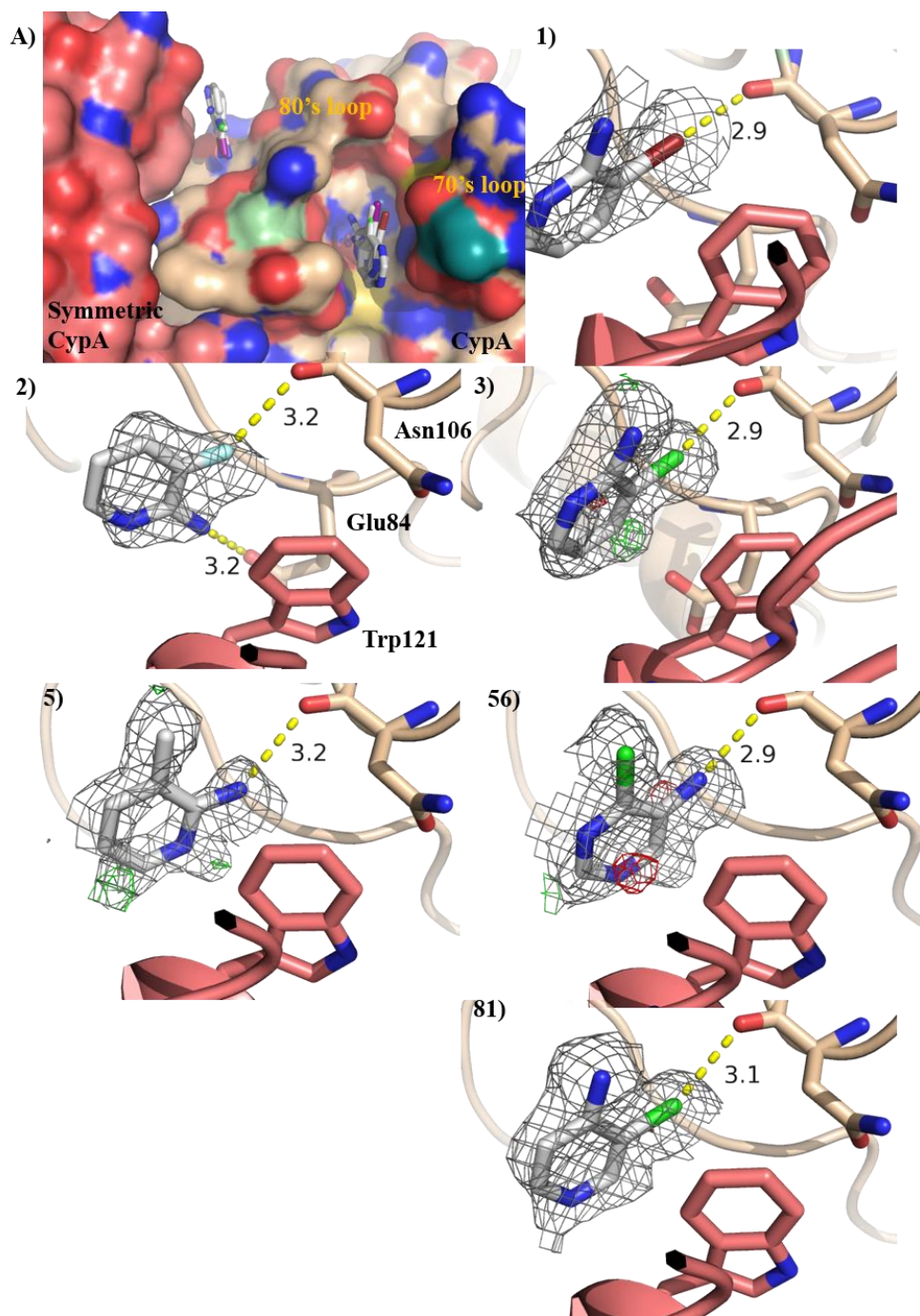


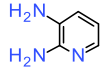
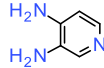

Figure 7-4 Binding of fragments out of the 80's loop of CypA. A) Overlay of **1**, **2**, **3**, **5**, **56** and **81** binding on a 3D surface structure of CypA. A symmetric CypA structure can be seen in pink. **1 – 81** Interactions between ligands with Glu84 and Asn106 residues of CypA and a close van der Waals interactions with Trp121 of the symmetric CypA. Distances between ligand with water and CypA molecules are highlighted and coloured as yellow dotted lines. $2F_o - F_c$ electron density and $F_o - F_c$ maps, at 1.0σ and 3.0σ are shown in grey and red/green respectively

7.7 Data collection, refinement and Ramachandran plot statistics for Cyp crystals

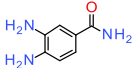
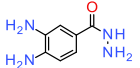
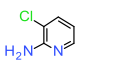
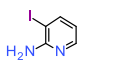
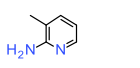
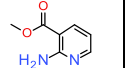
Table 7-3 in the next pages summarises all the statistics and details from the data collection and refinement of *apo* crystal structures of CypA, CypB and CypD as also as the complexes of CypA with ligands: **1, 2, 3, 5, 9, 16, 49, 56, 60, 61, 81, 89, 91, 97, 98 and 99.**

Table 7-3 Data collection, refinement and Ramachandran plot statistics for *apo* CypA CypB and CypD and for CypA in complex with 17 ligands







Rational design of isoform specific ligands



Protein	CypB	CypD	CypA	CypA	CypA	CypA
Ligand	-	-	-	91 	89 	97 
PDB ID	TBD	TBD	TBD	TBD	TBD	TBD
Data collection and processing						
Wavelength (Å)	0.928	0.928	0.976	0.970	0.920	0.920
High resolution limit	55.56	41.04	1.55	1.54	1.48	1.49
Low resolution limit	1.11	1.82	40.50	46.17	23.14	42.78
Completeness	86.26	96.99	99.90	98.90	89.80	99.10
Multiplicity	1.80	6.10	6.20	6.00	6.00	7.10
I/sigma	6.50	7.80	11.30	15.20	12.80	12.10
Rmerge	0.07	0.09	0.09	0.06	0.08	0.08
Anomalous completeness	48.20	98.60	98.50	93.90	81.80	97.90
Anomalous multiplicity	0.60	3.20	3.20	2.90	3.00	3.70
Unit cell dimensions:						
a (Å)	42.16	57.64	40.50	42.61	42.77	42.78
b (Å)	44.55	58.45	52.41	54.37	54.67	54.33
c (Å)	55.64	38.62	88.38	87.40	86.97	87.62
α (°)	90.46	90.00	90.00	90.00	90.00	90.00
β (°)	92.54	90.00	90.00	90.00	90.00	90.00
γ (°)	115.70	90.00	90.00	90.00	90.00	90.00
Space group	P1	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Total observations	218883	12206	173223	181525	185506	238763
Total unique	124598	11241	28038	30421	31129	33803
Refinement statistics						
No. of residues per chain (No. of chains)	187 (2)	165	165	165	165	165
No. of ligands	0	0	0	1	1	1
No. of waters	344	48	234	172	221	381
Number of used reflections	118684	11241	26578	28830	29517	32035
Percentage observed	86.26	96.99	99.81	98.61	89.46	98.90
Percentage of free reflections	4.75	5.05	5.02	5.06	5.04	5.08
R factor	0.16	0.22	0.14	0.15	0.21	0.14
R free	0.19	0.28	0.20	0.19	0.22	0.19
Rms BondLength	0.03	0.02	0.02	0.02	0.02	0.02
Rms BondAngle	2.40	1.84	1.78	1.92	1.96	1.83
Ramachandran plot						
Favoured (%)	95.95	93.71	96.25	94.97	93.75	95.48
Allowed (%)	3.51	5.66	3.75	5.03	5.63	3.87
No. of outliers*	2 (F68B, F68A)	1 (F60)	0	0	1 (F60)	1 (F60)

Rational design of isoform specific ligands

Protein	CypA	CypA	CypA	CypA	CypA	CypA
Ligand	98	99	3	4	5	9
						
PDB ID	TBD	TBD	TBD	TBD	TBD	TBD
Data collection and processing						
Wavelength (Å)	0.920	0.970	0.979	0.979	0.979	0.979
High resolution limit	1.43	1.61	1.60	1.55	1.80	1.60
Low resolution limit	38.49	38.31	45.65	45.83	45.72	45.65
Completeness	99.20	97.30	99.9	99.90	99.50	97.40
Multiplicity	6.80	6.30	5.4	6.30	4.20	3.70
I/sigma	18.80	13.20	7.9	11.20	7.70	9.20
Rmerge	0.06	0.08	0.14	0.09	0.13	0.10
Anomalous completeness	97.40	95.90	97.60	98.60	91.00	82.20
Anomalous multiplicity	3.50	3.30	5.40	3.20	2.10	1.80
Unit cell dimensions:						
a (Å)	42.87	42.69	42.18	42.56	42.47	42.18
b (Å)	54.58	54.49	53.88	53.82	53.71	53.88
c (Å)	87.44	86.82	85.91	87.42	87.07	85.91
α (°)	90.00	90.00	90.00	90.00	90.00	90.00
β (°)	90.00	90.00	90.00	90.00	90.00	90.00
γ (°)	90.00	90.00	90.00	90.00	90.00	90.00
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Total observations	262523	165888	142673	188585	79680	86815
Total unique	38376	26129	26543	29890	19096	26.564
Refinement statistics						
No. of residues per chain (No. of chains)	165	165	165	165	165	165
No. of ligands	1	1	2	2	2	1
No. of waters	289	209	175	161	193	103
Number of used reflections	36396	24759	25209	28405	18048	25209
Percentage observed	99.10	96.91	99.71	99.88	99.23	99.71
Percentage of free reflections	5.01	5.08	4.80	4.85	4.75	4.83
R factor	0.14	0.16	0.16	0.17	0.16	0.16
R free	0.18	0.21	0.19	0.20	0.21	0.19
Rms BondLength	0.02	0.02	0.02	0.02	0.02	0.02
Rms BondAngle	1.88	1.80	1.93	1.83	1.77	1.93
Ramachandran plot						
Favoured (%)	93.67	95.00	94.84	94.34	95.57	94.97
Allowed (%)	5.70	5.00	4.52	5.03	4.43	4.40
No. of outliers*	1 (F60)	0	1 (F60)	1 (F60)	0	1 (F60)

Rational design of isoform specific ligands

Protein	CypA	CypA	CypA	CypA	CypA	CypA
Ligand	16	56	60	61	49	81
						
PDB ID	TBD	TBD	TBD	TBD	TBD	TBD
Data collection and processing						
Wavelength (Å)	0.979	0.979	0.979	0.979	0.979	0.979
High resolution limit	1.35	1.45	1.30	2.00	2.05	1.60
Low resolution limit	45.25	46.00	45.36	44.95	43.36	45.66
Completeness	99.20	99.30	98.10	98.10	86.30	99.40
Multiplicity	5.10	5.30	5.00	4.70	3.20	4.70
I/sigma	14.50	14.20	12.70	8.20	8.20	9.50
Rmerge	0.06	0.06	0.07	0.13	0.09	0.09
Anomalous completeness	94.60	95.80	93.00	91.00	66.80	94.40
Anomalous multiplicity	2.50	2.70	2.50	2.40	1.50	2.40
Unit cell dimensions:						
a (Å)	42.09	42.48	41.72	40.98	41.71	42.50
b (Å)	52.49	54.40	52.68	52.16	50.28	53.49
c (Å)	89.22	86.13	89.18	88.56	85.58	87.65
α (°)	90.00	90.00	90.00	90.00	90.00	90.00
β (°)	90.00	90.00	90.00	90.00	90.00	90.00
γ (°)	90.00	90.00	90.00	90.00	90.00	90.00
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Total observations	224174	189779	241815	91486	32449	126504
Total unique	43795	35840	48076	13084	10154	26879
Refinement statistics						
No. of residues per chain (No. of chains)	165	165	165	165	165	165
No. of ligands	1	2	1	2	2	2
No. of waters	155	130	170	64	68	96
Number of used reflections	41612	34058	45723	12456	9615	25548
Percentage observed	98.95	98.99	97.71	97.51	85.32	99.02
Percentage of free reflections	4.86	4.85	4.78	4.60	4.78	4.79
R factor	0.16	0.16	0.16	0.21	0.20	0.21
R free	0.18	0.18	0.18	0.27	0.28	0.25
Rms BondLength	0.03	0.03	0.03	0.02	0.02	0.02
Rms BondAngle	2.26	2.41	2.37	1.71	1.80	1.79
Ramachandran plot						
Favoured (%)	96.20	95.54	94.41	93.08	93.08	94.34
Allowed (%)	3.80	3.82	4.35	6.92	6.29	5.66
No. of outliers*	0	1 (F60)	2 (G80, E81)	0	1 (V2)	0

Protein	CypA	CypA
Ligand	2	1
		
PDB ID	TBD	TBD
Data collection and processing		
Wavelength (Å)	0.979	0.979
High resolution limit	1.75	1.85
Low resolution limit	45.91	45.79
Completeness	93.40	98.20
Multiplicity	4.10	3.90
I/sigma	8.90	9.80
Rmerge	0.10	0.08
Anomalous completeness	75.00	85.10
Anomalous multiplicity	1.90	1.90
Unit cell dimensions:		
a (Å)	42.33	42.56
b (Å)	53.91	53.41
c (Å)	87.56	88.94
α (°)	90.00	90.00
β (°)	90.00	90.00
γ (°)	90.00	90.00
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Total observations	78874	68767
Total unique	19335	17584
Refinement statistics		
No. of residues per chain (No. of chains)	165	165
No. of ligands	2	2
No. of waters	104	65
Number of used reflections	18357	16580
Percentage observed	92.39	97.29
Percentage of free reflections	4.80	4.87
R factor	0.20	0.26
R free	0.25	0.36
Rms BondLength	0.02	0.03
Rms BondAngle	2.07	2.23
Ramachandran plot		
Favoured (%)	93.71	90.57
Allowed (%)	5.66	9.43
No. of outliers*	1 (F60)	0

7.8 *Cyp X-ray structures used in free energy calculations*

Details about all Cyp modelled in free energy calculations are given below. For each protein the UniProtKB, PDB ID, Chain and residue numbers are reported, followed by the residues using single letter coding.

> CypA | UniProtKB: P62937 | PDB ID: 1CWA | Chain: A | Residues: 2-164

VNPTVFFDIAVDGEPLGRVSFELFADKVPKTAENFRALSTGEKGFYKGS
 HRIIPGFMCQGGDFTRHNGTGGKSIYGEKFEDENFILKHTGPGILSMANAGPN
 TNGSQFFICTAKTEWLDGKHVVFVGKVKEGMNIVEAMERFGSRNGKTSKKITI
 ADCGQLE

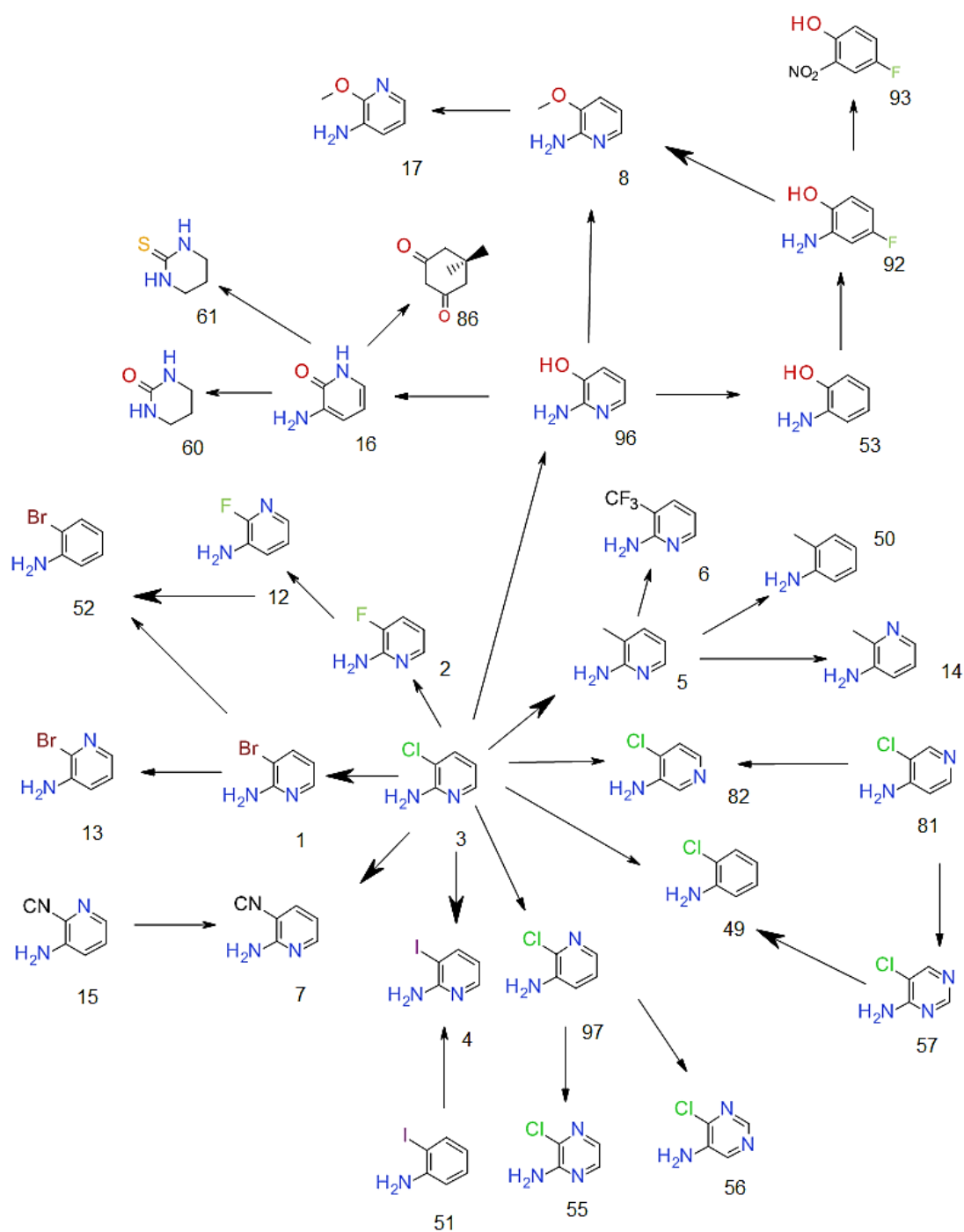
> CypB | UniProtKB: P23284 | PDB ID: 2BIT | Chain: A | Residues: 42-204

VTVKVYFDLRIGDEDVGRVIFGLFGKTVPKTVDNFVALATGEKGFYKNSK
 FHRVIKDFMIQGGDFTRGDGTGGKSIYGERFPDENFKLKHYGPGWVSMANA
 GKDTNGSQFFITTVKTAWLDGKHVVFVGKVLEGMEVVRKVESTKTDSRDKPL
 KDVIIADCGKI

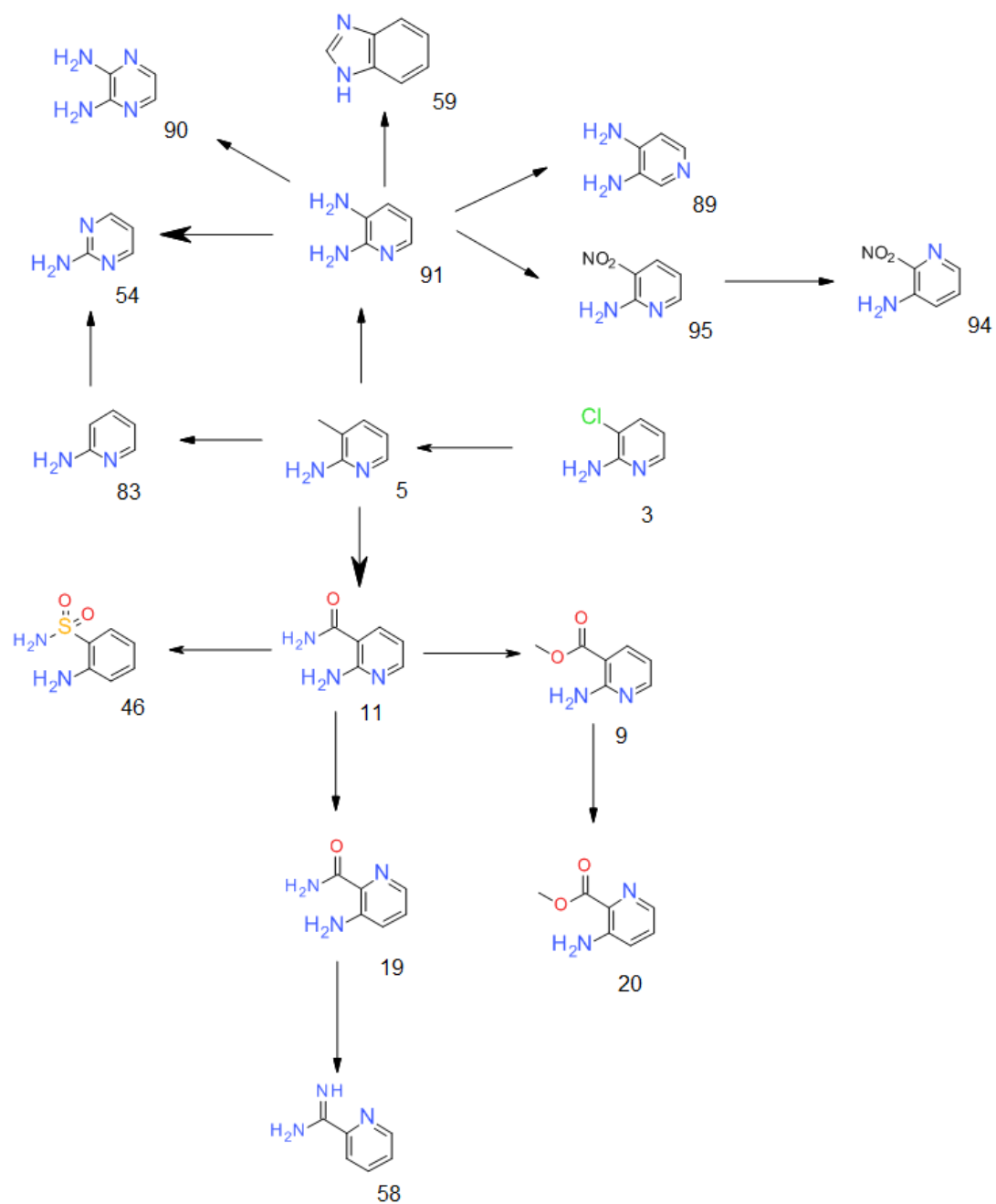
> CypD | UniProtKB: P30405 | PDB ID: 2BIT | Chain: A | Residues: 44-206

GNPLVYLDVDANGKPLGRVVLELKADVVPKTAENFRALCTGEKGFYKGS
 TFHRVIPSFMCQAGDFTNHNGTGGKSIYGSRFPDENFTLKHVGPVLSMANA
 GPNTNGSQFFICTIKTDWLDGKHVVFVGHVIEGMDVVKKIESFGSKSGRTSKKI
 VITDCGQL

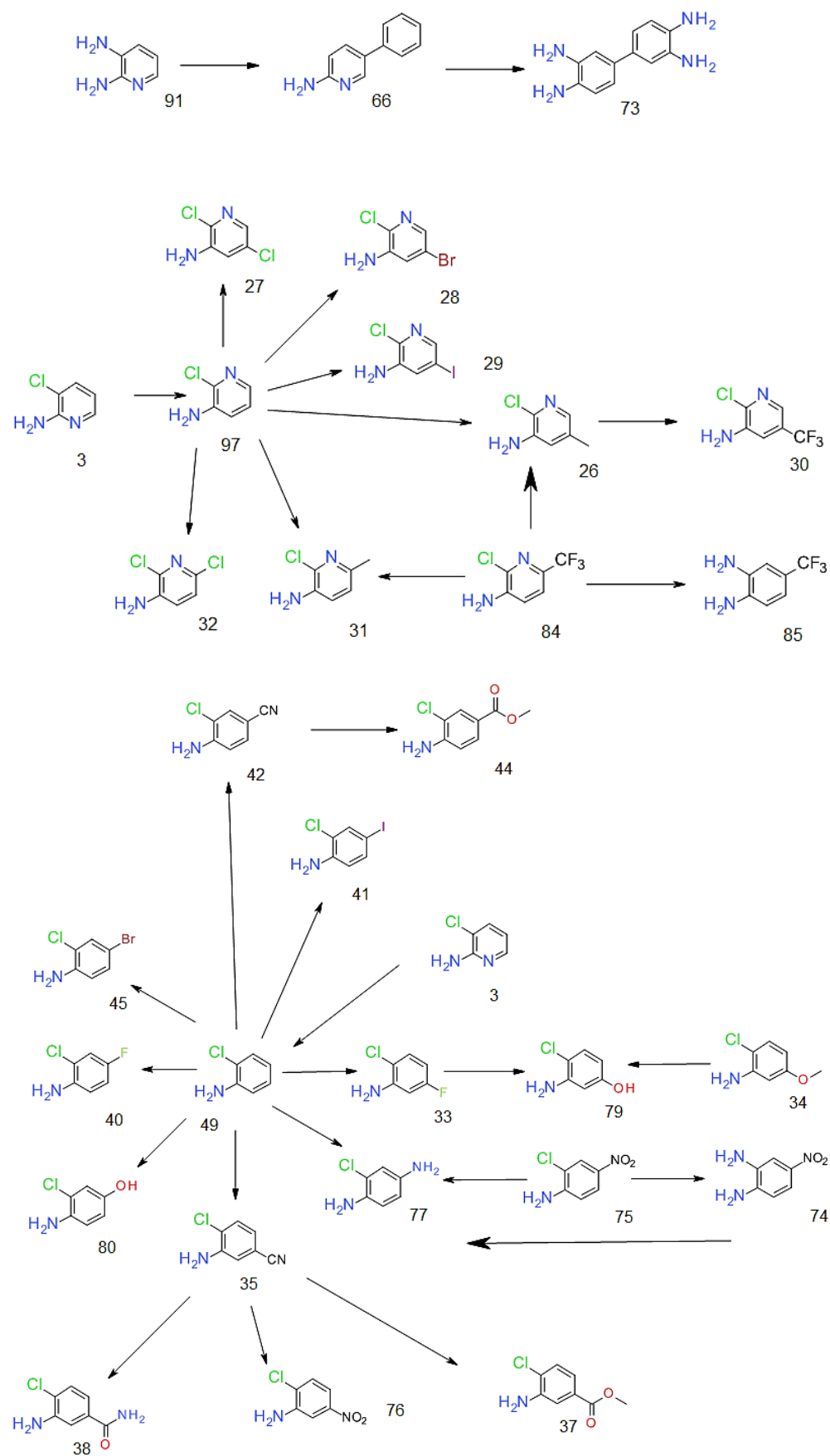
7.9 Perturbation maps for free energy calculations



Rational design of isoform specific ligands



Rational design of isoform specific ligands



Rational design of isoform specific ligands

