

Probing the Druggability of the Notch1 Ankyrin Domain Using a Fragment-Based Approach



Noha Mansour Hassan Abdel-Rahman

Newnham College

University of Cambridge

A thesis submitted for the degree of
Doctor of Philosophy
November 2010

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

I declare that the dissertation submitted is not substantially the same as any submitted for a degree or diploma or other qualifications at any other University.

Noha Abdel-Rahman

November 2010

To my family

Acknowledgments

First and foremost, I offer my sincerest gratitude to my supervisor Professor Sir Tom Blundell for giving me the opportunity to study in his group and working towards my PhD. I would like to thank him for his guidance, support and endless insight he offered. I would also like to thank Professor Alfonso Martinez Arias for his encouragement, useful discussions and constructive suggestions.

I am greatly indebted to Dr Matthias Ehebauer for his patience, guidance, valuable suggestions and supervision. I am grateful to Dr Deepti Gupta for her help and advice. I would like to thank Professor Peter Johnson for providing a licence to use SPROUT. I am thankful to Dr Duncan Scott and Dr Hamid Nasiri for their advice and help. I would like to thank all past and present members of TLB who have always been helpful in the hours of need. I thank Dr Will Pitt for his useful discussions, Dr Rinaldo Montalvao for his time and help, Dr Anja Winter for her advice in Biacore and virtual screening, Dr Victor Bolanos-Garcia for his useful suggestions. I would like to also thank the drug discovery group in particular the TB group: Dr Sachin Surade for his skilful advice, Dr Marcio Dias and Leonardo Silvestre for making synchrotron trips a fun experience. I thank Dr Richard Bickerton and Alicia for their help with PICCOLO. I thank Qian, Alicia, Leonardo, Takashi, Maja, Anna, Vivian, Yi-chun, Jawon, Adrian, Semin, Sung, Bernardo, Ola, Stephanie and all members of the crystallography group. I also thank our lovely lab manager Cynthia Lampert Moore and our cheerful Graham Eliff for his computing support. I would like to thank BP-Egypt, the Cambridge Overseas Trust, Newnham College and Professor Sir Tom Blundell for their financial support.

I would like to extend my thanks to my fellow Newnhamite friends Aisha, Zarina and Solava for the lovely times, continuous support and their friendship.

Last but not least, I thank my family for their immeasurable support and encouragement. I thank my father for inspiring and motivating me to pursue my

degree, my mother for her unconditional love and her constant prayers, my lovely sister and brother for their emotional support.

Abstract

Notch signalling is a highly conserved pathway that is important in the developmental processes that control cell differentiation and cell fates. This canonical pathway involves binding of a transmembrane ligand in one cell to the extracellular domain of a transmembrane Notch receptor in an adjacent cell. Ligand binding triggers two sequential proteolytic cleavages that shed a Notch intracellular domain (NICD). This is followed by translocation of NICD to the nucleus where it interacts with a transcription factor CSL and forms an activated Notch transcription complex, which induces the transcription of Notch target genes.

Abnormal expression or mutations in the different components of the pathway are associated with a number of diseases and cancers. An enhanced activity of Notch signalling resulting from a mutation in the extracellular domain is implicated in the progression of T-acute lymphoblastic leukaemia (T-ALL). Several therapeutic agents have been developed to target the Notch signalling pathway such as, γ -secretase inhibitors, antibodies targeting different regions of the Notch receptor and recently a synthetic stapled peptide, which was found to inhibit the formation of the transcription complex. The current inhibitors have their own disadvantages including lack of selectivity, cost of goods and delivery to the target. Thus, a more selective approach to target downstream protein-protein interactions by small molecules would provide an attractive approach to the design of new therapeutic agents that target this pathway. Here I report a fragment-based approach to target the ankyrin domain, a historically known but challenging, often-considered “undruggable” target.

In this dissertation I describe the application of various biophysical and computational approaches to find, characterise and design compounds. The initial screening of a commercial fragment-library exploited a fluorescent-based thermal shift assay that identified 36 fragment hits. Some of the fragments were kinetically characterised by Surface Plasmon Resonance (SPR) and their affinities were found to be in the millimolar range. Several attempts at soaking

and co-crystallising the fragments in the ankyrin domain crystal resulted in only two successful crystal structures that clearly define the positions of the fragments and their interactions with the ankyrin domain. One fragment binds to a pre-defined hotspot residue at the interface between the ankyrin domain and CSL. The other fragment is located at the interface between the ankyrin domain and Mastermind (MAML). The structural and kinetic data assisted the design of larger compounds with more extensive interactions using drug design software such as SPROUT and a docking program (GOLD). However, the optimised fragments did not show much improvement in affinity underlying the difficulty of flat protein-protein interface. The results reported here show the first structures of small molecules binding to the ankyrin domain of Notch1 receptor

Contents

Declaration	ii
Acknowledgments	iv
Abstract	vi
List of Figures	ix
List of Tables	xi
Abbreviations	xii
1 Introduction	1
1.1 Biological significance of Notch signalling	1
1.2 The components of Notch pathway	1
1.3 The architecture of the Notch receptor	2
1.4 The Notch CSL-dependant signalling pathway	5
1.5 Notch in disease and cancer	7
1.6 Rationalising targeting Notch signalling	7
1.7 Inhibitors of γ -secretase	8
1.8 Targeting the Negative Regulatory Region (NRR)	9
1.9 Development of peptidomimetics	10
1.10 Druggability of protein-protein interfaces	10
1.11 Advantages of targeting a protein-protein interface	13
1.12 Examples of protein-protein inhibitors	13
1.13 Fragment-based approach	15
1.14 Properties of protein-protein inhibitors	17
1.15 Targeting protein-protein interfaces using fragment-based Approach	18
1.16 Fragment screening methods	18
1.17 Objectives	22
2 Preparation and Analysis of the Ankyrin domain	24
2.1 Protein expression and purification	24
2.1.1 Chemicals	24
2.1.2 Protein expression	25
2.1.3 Protein purification	25
2.1.4 Results and discussion	25
2.2 Circular dichroism spectroscopy	27
2.2.1 Material and Methods	27
2.2.2 Results and discussion	27
2.3 Analysis of the ankyrin interactions in Notch transcription complex	30
3 Fragment Library Screening	36
3.1 Introduction to Fragment screening	36
3.1.1 Fragment library design	36
3.1.2 Screening techniques	38
3.2 Fluorescent-based thermal shift assay screening	38
3.2.1 Background of fluorescent-based thermal shift assays	38

3.2.2 Advantages of fluorescent-based thermal shift assays.....	38
3.2.3 Limitations of fluorescent-based thermal assay screening.....	39
3.2.4 Buffer Screening	40
3.2.4.1 Materials and methods.....	40
3.2.4.2 Results.....	41
3.2.4.3 Discussion.....	42
3.2.5 Screening of Fragment Library.....	44
3.2.5.1 Materials and methods.....	44
3.2.5.2 Results and discussion.....	45
3.3 Conclusions.....	60
4 Structural Characterisation of Ankyrin- Fragment complexes.....	61
4.1 Introduction.....	61
4.2 Optimising a crystallographic system for SBDD.....	61
4.3 Materials and methods.....	64
4.3.1 Crystallisation and optimisation.....	64
4.3.2 Soaking and co-crystallisation.....	64
4.3.3 Data collection and processing.....	66
4.4 Results and discussion.....	66
4.4.1 Crystallisation.....	66
4.4.2 Soaking and co-crystallisation.....	68
4.4.3 The first crystal structure.....	72
4.4.3.1 Implications of complex formation.....	75
4.4.3.2 Implication of dimerisation.....	76
4.4.4 The second crystal structure.....	79
4.4.4.1 Implication of the fragment 9F07 on complex formation.....	80
4.4.4.2 Implication of the fragment 9F07 on dimerisation.....	81
4.4.5 The relative position of the two fragments to each other.....	82
4.5 Conclusions.....	83
5 Kinetic Studies Using Surface Plasmon Resonance.....	84
5.1 Introduction.....	84
5.1.1 Surface Plasmon Resonance.....	85
5.1.2 Advantages of Surface Plasmon Resonance.....	86
5.1.3 Limitations of Surface Plasmon Resonance.....	87
5.2 Characterisation of fragments and small molecules.....	88
5.2.1 Characterisation of some of the fragment hits.....	88
5.2.1.1 Materials and methods.....	88
5.2.1.2 Results and discussion.....	90
5.2.2 Characterisation of optimised fragment.....	97
5.2.2.1 Materials and methods.....	97
5.2.2.2 Results and discussion.....	98
5.2.3 Characterisation of Indirubin-3'-monomxime.....	102
5.2.3.1 Results and discussion.....	102
5.3 Conclusions.....	104
6 Computational Methods for Fragment Elaboration.....	105
6.1 Introduction.....	105
6.2 <i>De novo</i> design of molecules using SPROUT.....	106
6.2.1 Fragment rowing and fragment linking.....	106
6.2.2 Materials and methods.....	107
6.2.3 Results and Discussion.....	110

6.2.3.1 Fragment linking.....	110
6.2.3.2 Fragment growinusing SPROUT.....	112
6.3 Virtual screening.....	117
6.3.1 Materials and Methods.....	118
6.3.1.1 Sample preparation.....	118
6.3.1.2 Protein preparation.....	118
6.2.1.3 The docking configuration.....	118
6.2.2 Results and discussion.....	119
6.2.2.1 Substructure analysis.....	119
6.2.2.2 Protein flexibility.....	119
6.2.2.3 Choice of the target template.....	120
6.2.2.4 Choice of scoring function.....	121
6.2.2.5 Selection of compounds.....	123
6.2.3 Testing compounds experimentally.....	127
6.2.3.1 Detecting binding by Surface Plasmon Resonance.....	127
6.2.3.2 Results and discussion.....	127
6.3 Summary and Conclusions.....	129
7 Conclusions and future directions.....	131
7.1 Conclusions.....	131
7.2 Lessons learnt for future directions.....	134
7.3 Future directions.....	134
7.3.1 Choice of Target.....	134
7.3.2 Library design.....	134
7.3.3 Fragment screening.....	135
References.....	137

List of Figures

1.1	A model describing the events of CSL-dependant Notch signalling. (NICD: Notch Intracellular Domain, MAML: Mastermind). PDB codes for the isolated ankyrin domain (1YYH, 2F8X) and for the complex (2F8Y). CSL is known as Su(H), Lag-1, RBP-jk in <i>Drosophila</i> , <i>C. elegans</i> and in mammals respectively.....	6
1.2	A representation of the structural and functional regions of Notch receptor and ligand. The Notch receptor extracellular region consists of EGF-like repeats whose number varies in different Notch homologues and LNR domain, the intracellular region is composed of RAM, ankyrin domain (ANK) and a terminal PEST region. The extracellular region of Notch ligand always contains a receptor binding DSL domain and EGF-like repeats.....	6
1.3	Examples of peptidomimetics. A) Terphenyl derivatives B) Terphthalamide C) Stapled peptides.....	6
2.1	Purification of the ankyrin domain by Ni-NTA anion exchange chromatography (AEC) and size exclusion chromatography (SEC). A) SDS-PAGE showing Ni-NTA purification. (M=Marker, FT= flowthrough, W(1-2)=wash 1, E(1-2)= elution) B) SDS-PAGE showing AEC and SEC fractions C) anion exchange chromatography D) size exclusion chromatography.....	27
2.2	Far-UV CD spectrum of the ankyrin domain showing two minima at 208 nm and 220 nm.	30
2.3	Denaturation curve showing the ellipticity as a function of temperature to determine the unfolding temperature at the midpoint of the curve.....	30
2.4	A schematic representation of the solvent accessibility of the interacting molecules of the ankyrin domain (PDB 2F8Y) at the interface between ank-CSL and ank-MAML as detected by PICCOLO; red represents the residues at the interface periphery and orange the residues at the interface core. The upper panel shows a sphere representation and in the lower panel a cartoon representation of the residues.....	31
2.5	Cartoon representation of the interactions detected by PICCOLO between the ankyrin domain (maroon) with MAML (light blue) from the complex crystal structure (PDB 2F8Y). Ionic interactions are depicted in dashed red lines, hydrogen-bond interactions in cyan, pi-cation in magenta and hydrophobic interactions in grey. Water molecules are shown as green dots.....	33
2.6	Cartoon representation of the interactions detected by PICCOLO between the ankyrin domain (maroon) with CSL (light orange) from the complex crystal structure (PDB 2F8Y). Ionic interactions are depicted in dashed red lines, hydrogen-bond interactions in cyan, pi-cation in magenta and hydrophobic interactions in grey. Water molecules are shown as green dots.....	35

3.1 The derivative unfolding curves of the protein in different buffers at low salt concentration. Many peaks appeared indicating some instability followed by what seems to be aggregation...	43
3.2 Derivatives of the melting curves for the buffer screens in high salt concentration. The only buffer that showed a reproducible two-state unfolding transition was CHES pH 8.5 as shown which appears as a chevron.....	43
3.3 The thermal unfolding curves of the fragments 11B07, 11B08, 11D02, 11E06 and 11C02 . Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.....	48
3.4 The thermal unfolding curves of 12B2, 12B5, 12C3, 12C5, 12E3, 12E5, 12E6. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.....	49
3.5 The thermal unfolding curves of 12F2, 12F3, 12F7, 12G4, 12G5, 12H5, 13A8, 13A9 and 13A11. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.....	50
3.6 The thermal unfolding curves of 13G10, 13H07, 1406, 14B04, 14C04. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.....	51
3.7 The thermal unfolding curves of 1A02, 1B02, 1C02, 2A08, 2B11. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.....	52
3.8 The thermal unfolding curves of 3B11, 3G02, 3H03, 4D03, 4E03 and 9F07. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.....	53
3.9 Classification of fragment hits according to distinctive chemical structure. Five main classes were found to bind by fluorescent-based thermal shift screening. These include: (a) benzyl derivatives, (b) fused bicyclic rings, (c) biaryl compounds, (d) phenyl derivatives, and (e) 5-membered heterocyclic rings.....	54
3.10 The fourth subgroup of phenyl derivatives and the corresponding ΔT_m values.....	60
4.1 Crystal hits of the ankyrin domain found in different screens.....	69

4.2 A cartoon representation of the ankyrin domain showing symmetry related molecules of the crystal structure (PDB 1YYH). The dotted circle highlights the interface which the ankyrin participates in protein-protein interactions in the ternary transcription complex with CSL and MAML. The crystal contacts in this space group with other symmetry related molecules do not interfere with the binding interface. (The symmetry molecules that are only in contact with ankyrin molecule (in orange) are shown for simplicity).....**69**

4.3 Figures showing the two sites the binding fragment 12C05. The interactions are shown in A) with residues D2109, E2076, D2109 and Y2075. B) Interactions of the fragment at the dimer interface N1984, H2019 and H2019. In C) and D) electron density maps $2|F_o-F_c|$ at a contour level of 1 σ showing the fragments at the two sites in A) and B) respectively.....**77**

4.4 Implications of fragment 12C05 for complex formation. A) Crystal structure of the ternary complex shows a salt bridge interaction between R40 of MAML and D2109 of the ankyrin domain (PDB 2F8X) B) Superposed structure of the ankyrin domain with the bound fragment 12C05 shows a salt bridge interaction between the fragment and the same D2109 of ankyrin domain. The ankyrin domain is in teal, MAML in orange and CSL in violet.....**79**

4.5 Crystal structures of the ankyrin domain both in the isolated structures (PDB 1YYH and PDB 2F8Y) and in complex (PDB 2F8X) show conserved head-to-head interactions through conserved residues of the ankyrin domain. In A) Crystal contact residues between molecules in the asymmetric unit show head-to-head interaction (PDB 1YYH) B) Residue R1985 in one chain of one molecule interacts with the other chain of another molecule in the asymmetric unit (PDB 1YYH). C) Crystal contacts of the ternary complex are involved through the ankyrin domain pseudo-twofold symmetry axis of the crystal (PDB 2F8X). D) The head-to-head interaction of the ankyrin domain in the ternary complex is involved again through residue R1985. E) Another crystal structure of higher resolution of the isolated ankyrin domain (PDB 2F8Y) shows contact residues between the two chains of the same molecule. F) Residue R1985 is involved in hydrogen-bond interactions with backbone carbonyl groups (PDB 2F8Y).....**81**

4.6 A) The crystal structure of the ankyrin domain with fragment 9F07 showing weak hydrogen-bond interactions with residues E2072, G2073, A2038 and N2039 (ankyrin domain in teal and 9F07 in orange) C) The crystal structure of 9F07 at the second site between the two ankyrin chains (Chain A in teal and chain B in violet). B) and D) electron density maps $2|F_o-F_c|$ at a contour level at 1 σ showing the fragments at the two sites in A) and C) respectively.....**82**

4.7 Implications of the binding of fragment 9F07 for complex formation. A) Residue E2072 of ankyrin forms a hydrogen-bond interaction with Y381 of CSL (PDB 2F8X) B) Superposing the structure of the fragment complex with the complex with MAML and CSL, showing the Fragment 9F07 interacting with the residue 2072 through the backbone carbonyl group. C) Displacement of carbonyl group of E2072 by a distance 1.2 Å and the carboxyl group by a distance 1.9 Å. (Ankyrin in teal, MAML in orange, CSL in violet).....**85**

4.8 Superposing structures of the ankyrin domain with the two fragments. A) A surface view of the ankyrin domain showing the two fragments at a close distance of about 6 Å. B) Superposed structures with MAML and CSL showing the location of two fragments at the interfaces. C) Surface view of the ankyrin domain showing the 2 fragments located between the two chains.....**86**

5.1 Sensorgrams of fragments shown to bind using thermal shift and estimated K_d values. In A) the fragment 12C05 and in B) fragment 12D05, the sensorgrams were fitted to a 1:1 model and

kinetic parameters were derived. In C) fragment 14B04 , D)fragment 12B05, E)fragment 14H04 and F)fragment 1605, K_d values were only measured by steady state affinity.....99

5.2 A summary of the optimisation approaches designed to optimise the fragment 12C05. The chemical moieties in dashed red circles were identified as pharmacophoric points involved in important interactions with ankyrin amino acid residues. These moieties were replaced by other rings or substitutions that were thought likely to enhance these interactions.....101

5.3 The effect of replacing a phenyl ring with a naphthalene ring and removing the methyl amino side chain on the affinity. In A) the sensorgram of the compound showing fast on- and off-rates, and in B) a steady state affinity measurement of a K_d value of about 21 mM.....102

5.4 Sensorgram showing the effect of replacing the pyrroline ring with a morpholine ring. Aggregation is observed at higher concentration (18.75 mM).....103

5.5 Replacing the methyl amino side chain by a methyl pyrroline inhibited the binding and aggregation was observed in sensorgram of two high concentration values (37.5 and 18.75 mM).....103

5.6 Adding a methoxy group did not have a significant affect on the affinity which was measured by steady-state which showed that the new molecule was binding at a K_d 15.23.....104

5.7 Sensorgrams of indirubin-3'-monoxime at various concentrations showing no binding responses to the ankyrin domain.....107

6.1 A surface view of the ankyrin domain with the two structures generated by SPROUT after attempting to link the two fragments together. The conformations of both fragments were maintained as in the original crystal structure and skeleton of linkers (within red circles) were built to connect them together without causing any clashes with the protein residues.....114

6.2 A surface view of the ankyrin domain binding to the fragment 12C05 showing two potential sites that could be targeted for fragment optimisation. Site 1 provides a deep small pocket that could enhance the affinity. On the other hand, Site 2 lies close to the protein-protein interface with MAML and could provide an enhanced inhibitory effect. The residues that line both sites can aid in selecting the potential groups that can be added to introduce a new interaction....117

6.3 An example of the structures generated by docking selected functional groups to the adjacent pocket at (Site 1). Most of the output structures added an aliphatic branched side chain. The cyclic side chains were scored less as they might introduce some steric constraints.....118

6.4 An example of the output structures that scored higher in optimising the fragment by growing to Site 2. The functional groups selected were restricted to aliphatic side chains that appeared more suitable for targeting a flatter surface. The sulphonamide group-containing side chain appeared at top of the list.....119

6.5 The docking results of the fragment 12C05 as a control molecule using different docking configurations. In A) the fragment was re-docked using the protein-ligand structure as a template and GoldScore docking result molecule shown in magenta. In B) the same configuration was used after including a water molecule that may assist and improve in docking; the docking result in green. In C) a high-resolution structure was used as the target template and ChemScore as scoring function resulting in fragment conformation shown in pink. In the three figures the reference fragment molecule is represented in a bright orange.....125

6.6 A surface view of the ankyrin domain with the docked structures of A) compound 1 and B) compound 2. The structures are then compared with the original fragment in C) and D) showing the common phenyl ring at an approximate same orientation compared to the fragment12C05.....	129
6.7 The sensorgrams of A) compound 1 and B) compound 2 showing no binding at a series of concentrations. The blank subtracted sensorgrams are solvent corrected to compensate for the effect of DMSO. The spikes observed mark the start and end of injection event.....	131

List of Tables

Table 1.1 The different parameters of Lipinski rule of 5 and rule of.....	16
Table 2.1 A list of the properties of the interactions between the ankyrin	31
Table 2.2 A list of the properties of the interactions between the ankyrin	34
Table 3.1 Reagents used in buffer screening.....	40
Table 3.2 Buffer list used for screening, the pH values and the concentration..	41
Table 3.3 The composition of the fragment and control wells showing.....	45
Table 3.4 A table of the third subgroup, ring (B) is a phenyl.....	55
Table 3.5 First subgroup of benzyl derivatives.....	56
Table 3.6 Second group of bicyclic derivatives.....	57
Table 4.1 Data collection and refinement statistics ankyrin-fragment.....	71
Table 4.2 Summary of the interactions the fragment 12C05 forms at the first..	73
Table 4.3 Summary of the interactions that fragment 12C05 forms at the.....	75
Table 4.4 Summary of the interactions the fragment 9F07 forms.....	80
Table 5.1 Preparing the aliquots for solvent correction solutions.....	90
Table 5.2 A table list of the fragments tested and their corresponding.....	95
Table 5.3 List K_d and pKa values of fragments 12C5 and its analogous.....	97
Table 6.1 A list of selected programs for docking and <i>de novo</i> design	106
Table 6.2 Summary of scores of compound 1.....	125
Table 6.2 Summary of scores of compound 2.....	125

Abbreviations

ALS	Advanced Light Source
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CCD	Closed-circuit device
CHES	2-(N-Cyclohexylamino)ethane Sulfonic Acid
CSD	Cambridge Structural Database
DDT	Dithiothreitol
DMSO	Dimethyl Sulfoxide
EDC	1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide
ESRF	European Synchrotron Radiation Facility
GOLD	Genetic Optimisation for Ligand Docking
GSK-3 β	Glycogen synthase-3 β
HB	Hydrogen bond
HBS	HEPES buffer saline
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HTS	High-throughput Screening
K _d	Binding constant
K _{off}	Dissociation constant
K _{on}	Association constant
MES	2-(N-morpholino)ethanesulfonic acid
MC	Monte Carlo
MD	Molecular dynamics
NHS	N-hydroxysuccinimide
NMR	Nuclear Magnetic Resonance
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction

PDB	Protein Data Bank
PIPES	Piperazine-N,N'-bis(2-ethanesulfonic acid)
RBP-Jk	Recombination Binding Protein
RMSD	Relative mean-square-deviation
RU	Resonance Units
SAR	Structure activity relationship
SPR	Surface Plasmon Resonance
Tris	Tris(hydroxymethyl)aminomethane
T _m	Melting temperature
Vdw	Van der waal

Chapter 1

Introduction

The development of a single cell into a whole organism involves multiple rounds of cell division. The Notch pathway is one of several that enable cell communication during this process. The cell responses are influenced by the intensity of the signal and crosstalk with other signalling pathways.

A notch in the wings of *Drosophila melanogaster* was first noticed by John Dexter in 1914 (Dexter, 1914). Soon after Thomas Morgan identified alleles of the gene in 1917 (Morgan, 1917) but the Notch gene was not itself identified in *Drosophila* until the 1980's (Wharton *et al.*, 1985; Kidd *et al.*, 1986). Since then it has become evident that the Notch pathway is highly conserved and regulates a wide range of developmental processes, as demonstrated by extensive loss- and gain-of- function mutational experiments in various organisms. Indeed, the Notch pathway has provided a good area to investigate various functional, genetic, and structural aspects of a major signalling pathway. Understanding the defects and abnormalities associated with the signalling events has assisted in rationalising different treatment regimens.

1.1 Biological significance of Notch signalling

There are different mechanisms by which Notch can regulate cell fates; an important mechanism is lateral inhibition, in which equivalent cells can equally express both Notch receptors and ligands (Bray *et al.*, 1998). However, subtle changes which are amplified by feedback loops will activate the expression of Notch receptor more than the ligand in one cell. This cell becomes the cell signalling cell whereas neighbouring cells become the receiving cells. Notch signalling cells remain undifferentiated and the neighbouring receiving cells adopt a different fate (Bray *et al.*, 1998).

Notch signalling modulates many biological processes such as apoptosis, cell regulation, and lineage decisions during embryonic development. Notch plays an important role in vascular development in the human embryo. In endothelial cells, Notch receptors (Notch1, Notch2, and Notch4) are expressed, whereas in vascular smooth cells only Notch3 is expressed (Joutel *et al.*, 2000; Lindner *et al.*, 2001; Villa *et al.*, 2001). Notch signalling is also involved in CNS development by inhibiting neuronal differentiation. Knockout studies of Notch components cause precocious neuronal differentiation (Ishibashi *et al.*, 1995; de la Pompa *et al.*, 1997; Hatakeyama *et al.*, 2004).

1.2 The components of Notch pathway

a) Notch receptor: There are four Notch receptors in the mammalian signalling pathway: Notch1, Notch2, Notch3, Notch4.

b) Notch ligands: The ligands are single transmembrane proteins and they are members of the (Delta, Serrate, Lag2) DSL family. The five mammalian ligands include two Jagged (Jagged1 and Jagged2) and three Delta (Delta-like-1, Delta-like-3 and Delta-like-4) family proteins.

c) DNA-binding proteins: These are the core of the transcriptional activation complex. They are known as CSL which is a collective name for C-promoter-binding factor in mammals (known as CBF-1 or RBP-J), Suppressor of Hairless Su(H) in *Drosophila melanogaster* and Lag-1 in *Caenorhabditis elegans*.

d) Target genes: Some Notch genes have been identified which participate in developmental processes. In *Drosophila* the Hairy/Enhancer of Split family genes were described as the direct Notch target genes. In mammals, Hes1, Hes5, Hes7 and a subfamily of Hes (He1, Hey2, HeyL) were activated by Notch signalling. These are basic helix-loop-helix (bHLH) protein transcription factors that are involved in cell fate suppression.

1.3 The architecture of the Notch receptor

The Notch receptor is a transmembrane receptor formed of extracellular and intracellular regions (Figure 1.2).

a) The extracellular region:

The extracellular region is composed mainly of 36 EGF-like (Epidermal Growth Factor-like) tandem repeats, a characteristic to the Notch family. Each repeat is a globular domain, consisting of 40 amino acids containing six cysteine residues that form three conserved disulphide bonds (Wharton *et al.*, 1985). The role of the EGF-like repeats has been investigated by extensive deletion mutagenesis studies in *Drosophila melanogaster*. These studies revealed that the repeats 11 and 12 are both important and sufficient to mediate interactions with the ligand *Delta* (Rebay *et al.*, 1991). Similar studies were conducted on human Notch1 that concluded that the EGF-like repeat 12, and not repeats 11 or 13, was important for ligand binding (Cordle *et al.*, 2008).

The Notch receptor function is modulated through its EGF-like repeats by posttranslational modification. Fucose is transferred to serine and threonine residues in the conserved region of EGF-like repeats by O-fucosyl transferase1 (Okajima & Irvine, 2002). This is followed by adding *N*-acetylglucosamine by fucose β -1,3-*N*-acetylglucosaminyltransferase (Panin *et al.*, 1997; Bruckner *et al.*, 2000; Moloney *et al.*, 2000)

b) The negative regulatory region

The negative regulatory region (NRR) consists of 3 cysteine-rich Lin12-Notch (LN) repeats followed by a hetero-dimerisation region.

The intracellular region

The intracellular domain is composed of RAM region (RBPJk-associated molecule region), 7 characteristic ankyrin repeats, transactivating domain and PEST domain.

The RAM region is natively unstructured in solution as was shown by biophysical studies (Nam *et al.*, 2003). There have been speculations about the role of RAM in human. However, it is thought to help in docking the ankyrin domain correctly, by interacting first with the CSL and so increasing the local concentration of the ankyrin domain to bind to CSL (Friedman *et al.*, 2008; Gordon *et al.*, 2008).

The ankyrin domain

The ankyrin repeats have a conserved secondary and tertiary structure. Each repeat is composed of 33 amino acid residues, arranged in two antiparallel α -helices connected by a short loop. The α -helices in each repeat are connected to helices of the adjacent repeat in a head-to-tail manner by a β -hairpin structure which is oriented perpendicular to the helices. This gives the ankyrin domain an L-shape in cross section. The ankyrin domain is a modular repeat protein which makes it impossible to be stabilised by direct interactions far apart in the sequence – so called long-range interactions. Instead, the repeats are stabilised by intra- and inter-repeat hydrophobic and hydrogen-bond interactions. The repeats are packed against each other with the inner helices being shorter than the outer helices. The terminal or capping repeats are polar and solvent accessible.

The ankyrin domain was thought to consist of six repeats, but recent studies confirmed the presence of an additional seventh repeat (Zweifel & Barrick, 2001; Ehebauer, 2005). The first repeat is partially folded and it assumes a regular ankyrin fold when it forms a complex with CSL and MAML (Zweifel 2003; Ehebauer 2005; Nam 2006). As a result of the repeat architecture of the ankyrin domain, it adopts a curved and concave structure.

It has been reported recently that the ankyrin domain undergoes hydroxylation at residue Asn1945 by factor inhibiting hypoxia-inducing factor (FIH) (Coleman *et al.*, 2007). Post-translational hydroxylation of intracellular proteins is a rare event. Hydroxylation of the ankyrin domain does not affect the formation of the

Notch ternary complex as hydroxylation involves residues which are distal to any protein-protein interfaces. This hydroxylation may not affect the Notch signalling directly but it is thought that the hydroxylated ankyrin domain may regulate (hypoxia-inducible factor) HIF signalling (Coleman *et al.*, 2007). Crystallographic analysis suggested that the ankyrin domain might undergo a major conformational change to bind to FIH. However, this was based on the structure of a short ankyrin peptide in an ankyrin-FIH complex, which does not correctly reflect the exact interaction in solution (Coleman *et al.*, 2007). On the other hand, biophysical studies using circular dichroism (CD) spectroscopy and differential scanning calorimetry experiments showed that some ankyrin domain repeats may appear more flexible than revealed in a frozen crystallographic state (Bradley & Barrick, 2005).

The repeat structure of the ankyrin domain provides a scaffold for protein-protein interactions with various protein partners. It interacts with a number of proteins involved in the Notch pathway, including Deltex through the first five repeats in *Drosophila*. Deltex appears to be a positive regulator of Notch signalling (Matsuno *et al.*, 1995). In mammals on the other hand, the mammalian Deltex acts as an antagonist of Notch signalling (Diederich *et al.*, 1994). The ankyrin domain also interacts with p300 (Oswald *et al.*, 2001). The histone acetyltransferases, PCAF and GCN5 (Kurooka *et al.*, 2000), interact with the ankyrin domain, possibly playing a role in the RBP-J-mediated transactivation of the intracellular domain of Notch. The ability of the ankyrin domain to bind to different targets could be due to the fact that it does not recognise specific sequences of targets or motifs, but instead binding involves discontinuous epitopes across the ankyrin molecule.

The stability and folding of the ankyrin domain have been studied extensively. Although it was originally thought that the ankyrin domain would unfold in a multistate manner involving a population of intermediates, energetic studies revealed that it folds cooperatively in a two-state manner (Bradley & Barrick, 2006). Thermodynamic studies on *Drosophila* Notch ankyrin domain revealed

that the seventh repeat contributes more to the stability of the whole ankyrin domain than repeats six and five (Bradely & Barrick, 2002).

1.4 The Notch CSL-dependant signalling pathway

The Notch receptor is synthesised in the endoplasmic reticulum as a single-pass transmembrane protein. The single polypeptide is cleaved at S1 site by proteases of the furin family after being transported to the Golgi network (Logeat *et al.*, 1998) and forms a “heterodimer” receptor composed of extracellular region involved in ligand binding and an intracellular region responsible for membrane-tethered signal transduction.

Ligands of the DSL (Delta, Serrate, Lag2) family in one cell interact with the extracellular domain of Notch receptor in an adjacent cell. This results in two sequential cleavage processes; the first occurs at the S₂ site by a member of the ADAM family of metalloproteases (Brou *et al.*, 2000). This is followed by a second cleavage at the S₃ site by the γ -secretase activity of the Presenilin-Nicastrin-Aph1-Pen2 protein complex (Struhl *et al.*, 1999) releasing the intracellular domain. The intracellular domain is then translocated to the nucleus to interact with members of CSL (CBF1/ RBPjk, Su (H), Lag-1) family of transcription factors replacing the corepressors. The CSL forms a complex with Skip, SMRT (Silencing Mediator of Retinoid and Thyroid receptors)/ N-coR (nuclear repressor co-repressor), CIR (CBF-1-interacting repressor) and histone deacetylases. Mastermind (MAML) protein is then recruited where it regulates the turnover of the intracellular by hyper-phosphorylation domain. The activated Notch transcription complex induces the transcription of target genes, including members of the *Enhancer-of-split* family of basic helix-loop-helix (bHLH) transcription factors that suppress cell fates (Mumm and Kopan, 2000; Schweisguth, 2004).

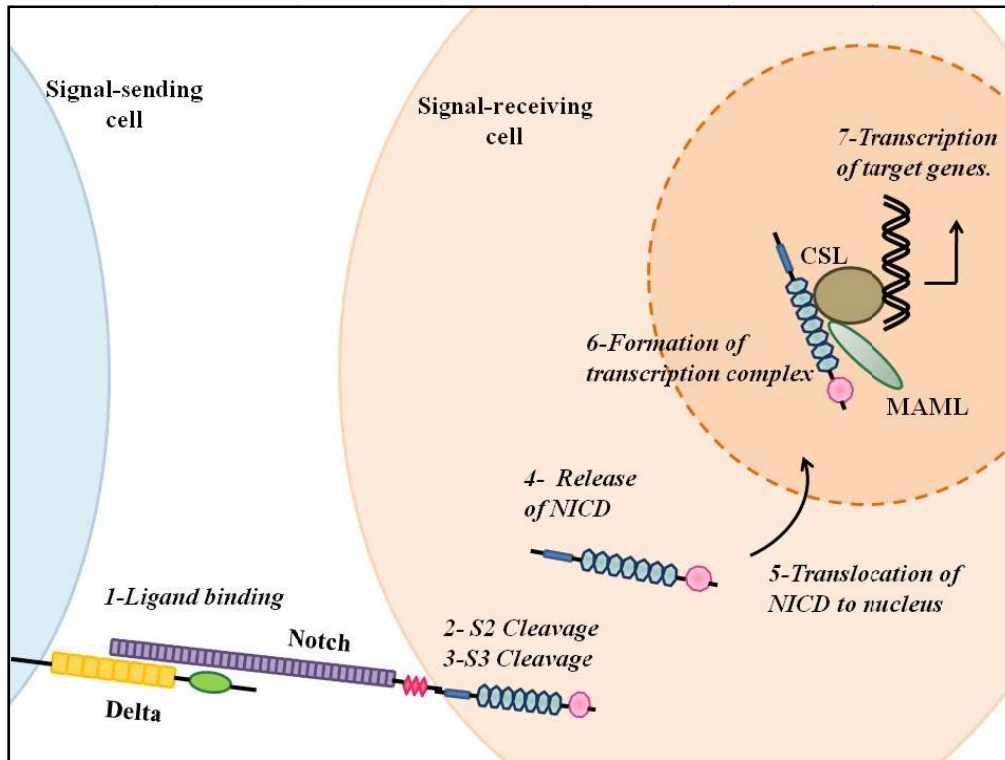


Figure 1.1 A model describing the events of CSL-dependant Notch signalling. (NICD: Notch Intracellular Domain, MAML: Mastermind). PDB codes for the isolated ankyrin domain (1YYH, 2F8X) and for the complex (2F8Y). CSL is known as Su(H), Lag-1, RBP-jk in *Drosophila*, *C. elegans* and in mammals respectively.

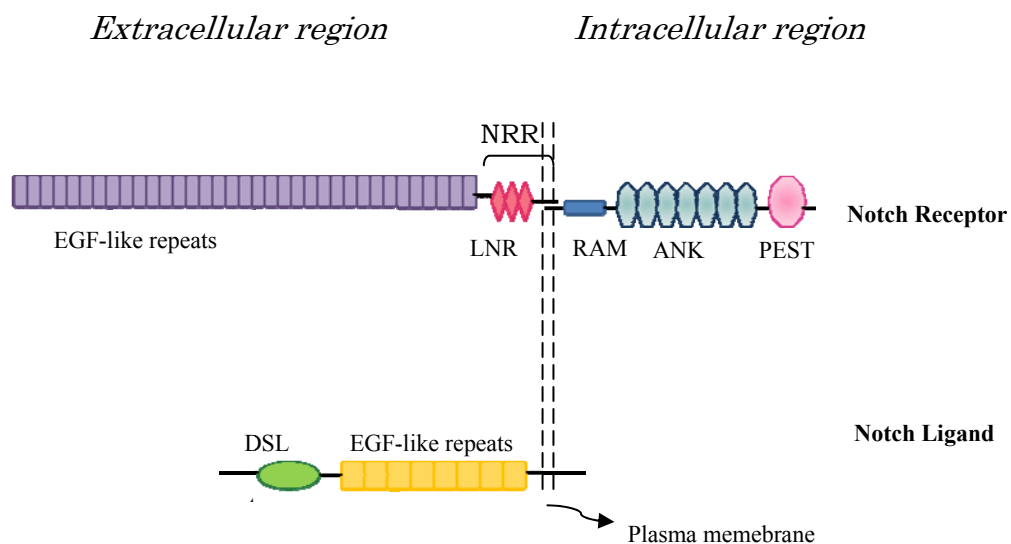


Figure 1.2 A representation of the structural and functional regions of Notch receptor and ligand. The Notch receptor extracellular region consists of EGF-like repeats whose number varies in different Notch homologues and LNR domain, the intracellular region is composed of RAM, ankyrin domain (ANK) and a terminal PEST region. The extracellular region of Notch ligand always contains a receptor binding DSL domain and EGF-like repeats

1.5 Notch in disease and cancer

Abnormalities in expression of different Notch receptor components are implicated in progression of various diseases and cancers. Missense mutations or microdeletions in the EGF-repeats in Notch3 particularly in the cysteine residues are associated with the developmental vascular disorder known as CADASIL (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy; (Joutel & Tournier-Lasserre, 1998). Another developmental disorder known as Alagille syndrome is caused by a mutation of a cysteine residue in the 11th EGF-like repeat in Notch2. This syndrome which is characterised by clinical cardiac, ocular and facial defects is also caused by mutations in ligand Serrate 1 (McDaniell *et al.*, 2006).

The first evidence of a link of the mutations in Notch to cancer was found in T-cell acute lymphoblastic leukaemia (T-ALL) patients. Chromosomal translocation was found in about 10% of the patients resulting in expression of Notch receptor where the extracellular domain was removed (Ellisen *et al.*, 1991). In humans, it was found that 50% of (T-ALL) patients have activating mutations in the extracellular hetero-dimerisation domain and C-terminal PEST domain of Notch1. These mutations activated Notch1 signalling and showed an increase in transcriptional activity (Weng *et al.*, 2004).

1.6 Rationalising targeting Notch signalling

Targeting the Notch signalling pathway has recently been reviewed and some unique features have been identified that could assist in rationalising the design of Notch inhibitors (Rizzo *et al.*, 2008). Complete inhibition of the Notch pathway may not be required since Notch activation is dose-dependant. An intermittent inhibition of the Notch signal may be sufficient as the half-life of the active form is very short. Another essential feature that should be considered during the design of the treatment regimen is that Notch activity is expressed in different cells with different responses. There should be a means to modulate the Notch signalling without causing undesirable side effects. This may require a more

selective approach to targeting certain stages or steps in a more context-specific manner. Theoretically, the Notch pathway can be targeted at different levels, including ligand binding, fucosylation of Notch receptor, cleavage by ADAM proteases and γ -secretase, and protein-protein interactions of the Notch transcriptional complex.

It is probably better to consider combination therapeutic regimens in targeting developmental pathways such as Notch. Such regimens are often developed by clinical trial and error when limited information of the mechanism of these pathways is available. A better understanding of the pathways and how they crosstalk with each other would help in designing these therapeutic regimens. The best combination regimen can be designed only after performing studies that can investigate and detect the type of cancers that are targeted by Notch inhibitors, the role of different components of the Notch pathway in cancer progression, and the pathways that crosstalk with Notch in certain cancers. However, this could be difficult as Notch is involved in an extensive cross talk network with other pathways.

1.7 Inhibitors of γ -secretase

The γ -secretase inhibitors found their ways into early stage clinical trials having the advantage of relatively easy oral administration and higher bioavailability (Shih & Wang, 2007). The main drawback of using γ -secretase inhibitors is the lack of specificity and selectivity. This is because they act on many substrates, including CD44, which is an important adhesion molecule for the extracellular matrix components and the intramembraneous cleavage product becomes a signal transduction molecule (Pelletier *et al.*, 2006). E-cadherin is another substrate for γ -secretase that is a major cell-cell adhesion receptor important for different cellular behaviours. In order to disassemble the E-cadherin–catenin complex it is cleaved, releasing β -catenin, which is an essential modulator in the Wnt signaling pathway (Marambaud *et al.*, 2002). γ -Secretase activates the release of the intracellular domain of ERBB4, a receptor tyrosine kinase (Vidal *et*

al., 2005). In addition, proteases participating in other cellular functions may be targeted by γ -secretase inhibitors. Another disadvantage of the γ -secretase inhibitors is their gastrointestinal toxicity; diarrhoea was observed in pre-clinical models caused by goblet cell metaplasia (Milano *et al.*, 2004). This suggested the use of a combination therapy to reduce the gut toxicity, which includes adding glucocorticoids in addition to the antileukemic agents.

1.8 Targeting the Negative Regulatory Region (NRR)

The crystal structure of NRR of Notch2 revealed a possible mechanism by which the S₂ cleavage is induced to activate Notch signalling (Gordon *et al.*, 2007). It is proposed that after ligand binding the S₂ becomes exposed through a significant conformational change that is caused by either a small allosteric change or a mechanical force. This dissociates the LNR from the HD (Heterodimerisation Domain) where the S₂ site is buried in a hydrophobic groove (Gordon *et al.*, 2007). This could be a general mechanism in all Notch receptors, which share a high degree of sequence identity. The auto-inhibition conformation that protects the Notch receptor from ligand-independent activation suggests the possibility of targeting this region for developing potential therapeutic agents that modulate the Notch signalling. This is of particular interest especially in the treatment of T-ALL where mutations in the hydrophobic core of the HD are implicated in the development of T-ALL (Malecki *et al.*, 2006). Antibodies have been developed to target the NRR of Notch3 (Li *et al.*, 2008). Some antibodies were found to bind at one face of the NRR confirming the proposed model (Li *et al.*, 2008)). Similarly, antibodies to target the NRR in order to antagonise Notch1 and Notch2 were generated to stabilise the off-conformation of NRR (Wu *et al.*, 2010). In another study, two classes of highly potent antibodies were characterised *in vitro* (Aste-Amézaga *et al.*, 2010). The first group was ligand-competitive and targeted the ligand-binding site in the extracellular EGF-like repeats. The second group comprised the allosteric NRR-binding antibodies.

1.9 Development of peptidomimetics

The assembly of a ternary complex composed of the intracellular domain, CSL and MAML is instrumental in switching on the transcription of Notch target genes. As described previously, the intracellular domain interacts with CSL to form a long groove to allow for the binding of MAML. The role of MAML is crucial in stabilising the Notch complexes. MAML is an α -helical polypeptide that was shown to adopt an unexpected kinked structure in the ternary-DNA complex as revealed in the crystal structure (Nam *et al.*, 2006) (PDB 2F8X). The crystal structure showed that neither the ankyrin domain nor CSL undergo major conformational change on complexation. This suggested that MAML could be considered as a recognition motif in the Notch transcriptional activation complex. In fact, Notch signalling has been shown to be antagonised by a dominant fragment of MAML (residues 13-74) (Maillard *et al.*, 2004). This fragment, known as dnMAML, suggested that helix-mimetics such as hydrocarbon-stapled α -helical peptides might be useful.

A number of peptides were designed that scan the contact surface with ankyrin and CSL (Moellering *et al.*, 2009). Synthetic peptides that showed higher helical content bound with increased affinity to a binary RAMAnk-CSL complex. Such peptides were shown to suppress Notch1 gene expression in reporter gene assays, globally suppress Notch signalling in gene expression profiling experiments, reduce the proliferative capacity in T-ALL cell lines, and proved effective *in vivo* by inhibiting leukaemic progression (Moellering *et al.*, 2009). These observations provide promising evidence of the possibility of a therapeutic agent acting as a direct transcriptional antagonist. However, the complexity of the system may lead to undesirable off-target activity.

1.10 Druggability of protein-protein interfaces

The possibility of small molecule drugs, so called new chemical entities (NCEs), modulating protein functions is referred to as the “druggability” of the target (Hajduk *et al.*, 2005b). It has been used as a means for target identification and target validation. Several approaches have been devised to assess the

druggability of proteins. These methods rely on the application of geometry-based or energy-based algorithms to 3D structures of proteins. Geometry-based methods were designed to predict concave pockets or calculate molecular surface complexity, whereas energy-based algorithms calculate the binding potentials or energies. However, most protein-protein interactions differ from protein-ligand interactions; it has been observed that protein-ligand interactions involve fewer and larger pockets whereas protein-protein interactions use numerous smaller pockets. Geometry-based methods were unable to predict small cavities on flat protein-protein interfaces. Energy-based methods on the other hand may be able to predict binding pockets, which are formed by adaptive or conformational changes on binding, using a dynamic model of the protein structure.

Protein-protein interfaces have not evolved to bind to small molecules. Most are flat shallow surfaces that are devoid of deep pockets or cavities that can accommodate the binding of small molecules. Many protein-protein interactions appear to be achieved through the additive effect of numerous but weak interactions across the interface. It is quite difficult for a small molecule to mimic these widely-spaced interactions as it can participate only in a limited number of interactions. However, protein-protein interactions are important in many biological processes and in cases where the biology is understood have been implicated in disease progression; these targets have emerged as attractive drug targets despite their high chemical risk (Wells & McClendon, 2007). This has been encouraged by the discovery of hotspot residues that contribute a very large proportion of the binding energy at the interface. It was found that the energy of binding is not equally distributed across the interface and that the hotspot residues are usually enriched with tryptophan, tyrosine and arginine (Bogan & Thorn, 1998). Hotspot residues are investigated using alanine scanning mutagenesis. In this method, an amino acid is substituted with alanine and the change in the energetic contribution of the substituted residue to protein binding is recorded. Changes in the binding free energy of at least 2 Kcal/mol indicated a hot spot residue (Bogan & Thorn, 1998). Interestingly, these residues are usually surrounded by hydrophobic residues forming a water exclusion “O ring”. Protection from bulk solvent is necessary to strengthen the polar

interactions between complementary hot spot residues across the interface (Bogan & Thorn, 1998). Alanine scanning can be performed either by computational prediction or experimental site-directed mutagenesis which is rather tedious. The same hotspot region can bind to several targets suggesting a promiscuous binding behaviour of proteins involved in protein-protein interactions (DeLano *et al.*, 2000). This adaptivity of proteins suggests that targeting these specific interactions by small molecules can possibly inhibit the interactions across the interface. A recent review by Wells and McClendon (Wells & McClendon, 2007) has described six successful stories of targeting protein-protein interactions by discovering small molecules binding to hotspot residues. The discovery of nanomolar range inhibitors that bind to cytokine interleukin-2 (IL-2) hotspot residues is an interesting example that has employed a fragment-based approach. It was also shown in some cases that binding to a small molecule could trigger substantial conformational change revealing a potential binding cavity that was not seen either in the free protein or in complex. This could not be predicted computationally by virtual screening and requires experimental screening of molecules or inhibitors. Predicting the druggability of protein-protein interfaces still requires more knowledge and investigation of protein-small molecule recognition and its dependence on protein folds, the nature of the amino acids in the binding sites and the structural adaptivity of protein-protein interfaces to binding small molecules by inferring a certain degree of flexibility and mobility.

Another hurdle that faces those who wish to modulate protein-protein systems is topological complexity of the interfaces; the two proteins that contribute to the interface both participate in protrusions and sub-pockets. One protein could contribute the most to the protrusions and becomes less likely to bind to small molecules. In addition, in some complexes the protein-protein interaction is regulated by phosphorylation. Developing a small molecule that mimics these charged species would result in a molecule that is too polar to be bio-available. Another important issue to consider is the functionality of the target: to consider whether the aim is to develop a molecule that inhibits the interaction or rather

restores the activity and function of the complex. In cases where the disease is caused by a loss of activity as a result of mutation, developing therapeutic agents to mimic and initiate a response becomes even more challenging.

1.11 Advantages of targeting a protein-protein interface

Protein-protein interfaces can provide an opportunity to develop selective inhibitors. Protein-protein interface inhibitors target hotspot residues, which are usually highly conserved compared to the other residues lying at the interface. This makes the proteins more resistant to spontaneous mutations at the binding site, probably based on the fact that a simultaneous complementary double mutation is uncommon and unfavourable in formation of a functional stable complex. A mutation or a change introduced in one amino acid in one protein at the protein-protein interface would adversely affect the affinity of one protein to the other.

1.12 Examples for protein-protein inhibitors

Peptidomimetics

The first obvious approach to target a protein-protein interface is to mimic the natural binding ligand. However, peptidomimetics may not be the best choice, as they are susceptible to proteolysis, and thus less stable. Furthermore they have lower bioavailability due to their poor absorption, which is caused by their relatively higher molecular weight. They sometimes lack target specificity owing to their flexible conformations that enable them to interact with various receptors. However, they have been tested and developed as potential therapeutic agents that can block the function of some receptors. Different strategies aimed at overcoming these limitations have been developed, including modification of amino acids to stabilise the peptide against metabolism. This has been achieved by replacing the peptide bonds by structures that are not proteolysed, such as $-\text{CH}_2\text{NH}$ or by modifying the peptide backbone by introducing retro-inverse modifications using D- amino acids. In the case of

target proteases, peptide bonds have been replaced by structures that mimic or are isosteric with the enzyme transition state or intermediate, such as $-\text{CH}(\text{OH})-\text{NH}-$.

Mimicking a secondary structure requires maintenance of both conformation and interaction. The α -helical structures can be mimicked either by using stapled peptides that enforce synthetic peptides to acquire a helical conformation, or by using a proteomimetic strategy that employs different scaffolds to reproduce critical contacts with the target protein.

Stapled peptides are designed to enhance helical stability through incorporation of two α -methyl, α -alkenyl amino acids at positions separated by (i and $i+4$) to form one helical turn followed by cyclisation to form a macrocyclic hydrocarbon crosslink (Kim & Verdine, 2009). A staple can also be formed at a position (i and $i+7$) to form two helical turns (Schafmeister *et al.*, 2000). The activity of these stapled peptides is dependant on the stereochemistry of the cyclic hydrocarbon as helical content can influence cellular uptake. The helix stability also depends on the sequence of the peptide and the position of the crosslink. As it is difficult to predict the best conditions for helix stability, it is often necessary to synthesise and screen a small library in order to find the most active candidate. The other strategy is to design small molecules that mimic the residues at positions i , $i+3$ or $i+4$, and $i+7$ of a ten-residue length α -helix.

Numerous approaches have been employed to use scaffolds. For example, the template terphenyl has been substituted at three *ortho*-positions of the terphenyl scaffold by alkyl or aryl substituents to simulate an α -helix (Kim & Hamilton, 2006). However, this resulted in a highly hydrophobic compound that has poor water solubility. The phenyl ring has been replaced by a heterocyclic ring to increase the hydrophilicity and this has improved solubility. The aqueous solubility was even improved when the terphenyl scaffold was replaced by the terphthalamide scaffold (Yin *et al.*, 2005). An intramolecular hydrogen bond maintains a planar geometry keeping the central phenyl core. Adding an additional phenyl ring in the core allows for a higher degree of specificity as it mimics a four-residue α -helix.

Peptidomimetics tend to be most successful where the protein-protein interface involves continuous binding peptides that contribute significantly to the overall affinity between the protein units (Arkin & Wells, 2004).

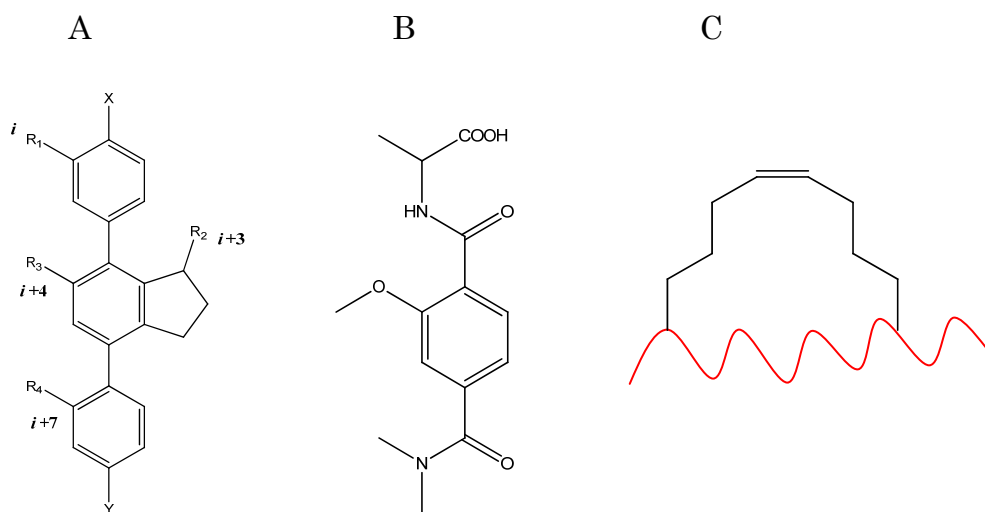


Figure 1.3 Examples of peptidomimetics. A) Terphenyl derivatives B) Terphthalamide C) Stapled peptides

1.13 Fragment-based approach

The “Lipinski rule of 5” has been devised to describe orally active drugs (Lipinski *et al.*, 2001). Although many drugs that reached the clinic have deviated from one or two rules (Wenlock *et al.*, 2003), they are still useful as a guideline for identifying drug-like molecules. However, these rules need some qualification for assessing lead-like molecules during drug development. Lead compounds usually have lower molecular weights, lower lipophilicity and fewer hydrogen-bond acceptors. This is expected, as starting with compounds of drug-like properties would lead to compounds of poorer physical properties during optimisation. This suggested that starting with smaller molecules would result in drug candidates with more favourable properties. The “Lipinski rule of 5” was then replaced by a “rule of three” (Congreve *et al.*, 2003) that helped in establishing screening libraries known as fragments and referred to also as needles (Boehm *et al.*, 2000) seeds (Liebeschuetz *et al.*, 2002) and shapes (Fejzo *et al.*, 1999).

Table 1.1 The different parameters of Lipinski rule of 5 and rule of three

	Lipinski Rule of 5	Rule of three
Molecular Wgt	<500	<300
H-bond donors	≤ 5	≤ 3
H-bond acceptors	≤ 10	≤ 3
Rotatable bonds	≤ 5	≤ 3
LogP	≤ 5	≤ 3

The advantages of using a fragment-based approach are many. First, smaller screening libraries can cover an acceptably diverse chemical space. The estimated 10^{63} of small drug-like compounds (Bohacek *et al.*, 1996) makes it impossible to assemble HTS (High-throughput screening) libraries that cover more than a very tiny portion of the chemical space. On the other hand, smaller fragment libraries can explore the same chemical space as a larger library of drug-sized molecules. This not only reduces the screening time and synthesis, but it is advantageous in data analysis and management. Second, the successful hits identified are likely to have a higher ligand efficiency compared to those obtained by other methods. Ligand efficiency is a measure of the potency of a compound, measured by calculating the average free energy of binding per heavy atom (Hopkins *et al.*, 2003). It is a useful means to prioritise initial hits and assess the compounds throughout the optimisation process as it eliminates any bias introduced by an increase in molecular weight. Small molecules such as fragments form fewer but often better binding interactions. Starting with small fragments of high ligand efficiency would lead to compounds of better pharmacokinetics. This could also be due to the fact that fragments libraries can be selected to avoid unfavourable functional groups. Third, fragment libraries are expected to have higher hit rates than conventional screening libraries of more complex and larger compounds, although this will be dependant on the sensitivity of the screening method used to detect the weakly-binding fragments. Simple fragments could have high complementarity to protein targets. The identified hits can then be optimised by adding chemical groups, in order to increase complementarity between protein and ligand.

Fragments can be developed into lead compounds by fragment evolution or growing, fragment linking, fragment optimisation and fragment self-assembly.

Fragment evolution can be achieved by adding new functionalities to the fragment in order to introduce new interactions to adjacent regions in the binding site. Fragment linking seems to be an attractive approach when two fragments bind in proximal positions in the active site enabling them to be linked together. In fragment self-assembly, the protein acts a template for the reactive fragments that link together to form an active compound. On the other hand, fragment optimisation involves a modification in only certain functionalities to improve properties or solve problems.

These approaches may seem daunting when attempting to transform a millimolar fragment hit to a nanomolar lead drug. However, it can be facilitated by strategic structure-based drug design. Structural information about the binding modes of fragments can be obtained by NMR or X-ray crystallography. In fact, it is thought that structure-based drug design aided in tripling the success of producing potent inhibitors (Hadjuk & Greer, 2007).

1.14 Properties of protein-protein inhibitors

Analyses of the physical and chemical properties of protein-protein inhibitors show they have significantly higher molecular weights than those that target conventional receptor binding or enzyme active sites (Higueruelo *et al.*, 2009; Sperandio *et al.*, 2010). Larger molecules tend to be necessary to duplicate even a small subset of the widely spaced interactions across the interface. They are also observed to be more rigid and have restricted orientations; they tend to have more rings and less rotatable bonds. This rigidity can be advantageous in reducing energetic penalties associated with the loss of entropy. In addition, rigid molecules can create new sub-pockets in the target protein by displacing flexible

protein regions. They are also likely to be more lipophilic with fewer hydrogen-bond acceptors and donors.

Early attempts to develop protein-protein inhibitors were focused on mimicking protein ligands including replacing chemical moieties and side chains with similar functional groups. However, some inhibitors that displayed non-exact matches were effective as long as they form interactions in the binding site.

1.15 Targeting protein-protein interfaces using fragment-based approaches

The compounds used in HTS screening are mainly derived from well-known chemical phenotypes that have been characterised from extensive research on traditional drug targets such as G-protein-coupled receptors and enzyme targets. These compounds may not be the most appropriate for targeting protein-protein interfaces. The observations recorded in the previous section suggest that enriching the standard libraries with compounds that comply with the properties of successful protein-protein modulators may increase the hit rate, enhance potency and improve specificity for protein-protein interactions. Indeed inhibitors of protein-protein interfaces may require new scaffolds or new classes of compounds. Fragment screening could then be the preferred approach for exploring a larger chemical space with higher ligand efficiency.

1.16 Fragment screening methods

Screening weakly binding fragments can be carried out by either biochemical assays or biophysical methods.

i) Biochemical method

Biochemical assays are often referred to as high concentration screening (HCS) as they are performed at high fragment concentrations that can often reach millimolar range. These functional assays can be configured in two different ways. Enzymatic activity can be monitored by the presence or absence of

substrate or by accumulation of product or by-product. Alternatively, fluorescent-based molecules can be used to trace the displacement of a known ligand. In the presence of a correct compound template, the allosteric site binders can be identified. HCS offers the advantages of a fast, quantitative, and a highly scalable method once the assay setup is established. However, false positives can occur by compound aggregation at high concentrations or lack of effective solubility. Cell toxicity caused by high concentration compounds in cell-assays makes this assay unsuitable for some targets.

ii) Biophysical methods

Biophysical methods are the more popular approach to screening fragment libraries; they include direct biophysical assays or direct structure-based screening.

a) X-ray crystallography screening:

Automation of data collection, analysis and interpretation has made it possible to use X-ray crystallography as a screening method (Mooij *et al.*, 2006). Cocktails of fragments dissolved in an organic solvent are used to soak crystals. The cocktail size can be up to eight compounds depending on the ability of the crystals to tolerate higher concentrations of the organic solvent. False positives can be greatly reduced as the bound compound can be visualised and decisions made for improving the binding in a structure-based manner. However, this technique is time and resource consuming, requiring production of significant amounts of protein, usually in milligram quantities, in order to grow crystals of reasonable quality. False negatives are quite frequent for either kinetic or crystallographic reasons or when the fragments are poorly soluble in the crystallisation medium.

b) NMR screening:

The identification of small molecules binding to a protein and linking them together using NMR-based method was first carried out in Abbott laboratories (Shuker *et al.*, 1996). This approach, which has been known as “SAR by NMR”,

can detect weakly binding molecules in the millimolar range (Shuker *et al.*, 1996). Since then, it has been applied for screening fragment libraries. In fact, it has been used as an experimental druggability assay (Hajduk *et al.*, 2005a). False positives or negatives are difficult to identify in this screen. Some structural information can be obtained and unlike X-ray crystallography binding affinities can be measured. NMR screening can be configured in two ways. A protein-based NMR assay makes use of sensitive ^1H - ^{15}N or ^1H - ^{13}C correlation NMR for detecting small molecules binding. Local chemical shift perturbations can indicate a ligand binding event. However, screening is restricted to proteins amenable to isotopic labelling which is time consuming and not suitable for all targets as it requires a large amount of protein which makes it difficult to screen proteins with limited solubility or larger proteins (larger than 50KD). The alternative is to carry out ligand-detected NMR assays which do not need any protein labelling allowing for a rapid assay setup. Water ligand optimised gradient spectroscopy (WaterLOGSY) technique depends on the fast ligand exchange between the bound and unbound forms (Dalvit *et al.*, 2000). This method exploits water magnetisation and bound ligands show positive signals. It can detect affinity values between 10 μM and 10mM. Unlabelled protein without any size limitations can be used in this case.

c) Fluorescence-based thermal shift assay:

This rapid screening method exploits the stabilisation of the protein in the presence of the compound by monitoring protein unfolding in the presence of a fluorescent dye. This method is discussed in detail in Chapter 3.

d) Surface plasmon resonance

The enhanced sensitivity of surface plasmon resonance methods has allowed the study of weakly binding compounds such as fragments. Once the setup is optimised, it can be used as a high throughput method. This method is described in more details in Chapter 5.

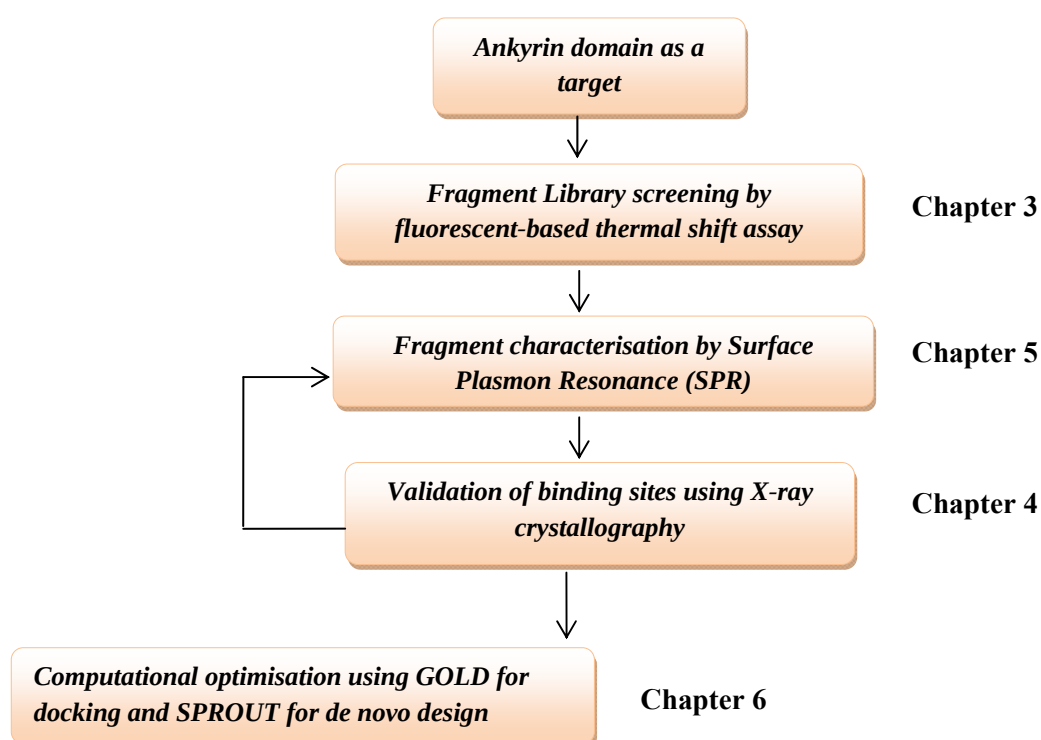
e) Isothermal calorimetric titration

This method is not considered a high-throughput method for screening fragment libraries as it requires large quantities of proteins. In addition, it sometimes can not detect very weakly binding fragments. However, it is a powerful method for deriving important thermodynamic parameters such as enthalpy and entropy of binding of ligands. A recent review by John Ladbury has emphasised the importance of measuring these parameters during the drug discovery and optimisation process (Ladbury *et al.*, 2010). Such data is complementary for hit prioritisation and hit-to-lead optimisation. Future development of the instrumentation may transform this technique to a popular method in fragment screening.

1.17 Objectives

The main objective of this study was to probe the druggability of the ankyrin domain of human Notch1 receptor. The architecture of the ankyrin domain as a repeat protein facilitates its role as a scaffold for binding other proteins. It is involved in many protein-protein interactions but so far there have been no reports of any small molecules that can bind to the ankyrin domain. Finding small molecules would be advantageous in developing compounds that could interfere with the protein-protein interactions involved in the Notch transcriptional complex with CSL and MAML. The ankyrin domain is instrumental in Notch signalling and targeting it with small molecules would be a promising approach to replacing existing therapeutic agents with a lead-like or drug-like chemical entity.

I have embarked on investigating the druggability of the ankyrin domain using a fragment-based approach. The following chart describes the route I followed.



In Chapter 3 and Chapter 5, I test the druggability of the ankyrin domain by screening a commercial fragment library and measuring the hit rate in order to

investigate the likelihood of binding small molecules. The objective here is to identify the most favourable scaffolds or chemical moieties that could bind to the ankyrin domain and the functionalities that could destabilise it. Measuring the affinities of fragments and optimised fragments should assist in developing molecules in a SAR manner.

The objectives of the research described in chapters 4 and 6 are to identify binding sites of the fragment hits and to detect any small cavities on the surface that could accommodate binding to small fragments. A structure-based approach with the assistance of computational tools such as docking and *de novo* drug design software is used to develop and optimise the fragments identified by these methods. The research indicates that a fragment-based approach may provide a route to identifying leads targeting such “undruggable” protein-protein interaction sites.

Chapter 2

Preparation and Analysis of the Ankyrin domain

Many biophysical experiments require highly pure, stable, soluble and homogeneous protein. The degree of purity varies according to the technique and its sensitivity. In NMR studies, two-step purifications are often sufficient to yield protein samples with the required purity level for initial spectra. On the other hand, crystallography needs a higher degree of purity in order to allow reproducible crystal growth in crystallisation trials. An additional purification step often ensures the removal of any impurities that could interfere with crystal packing. In an iterative structure-based drug discovery programme, high protein expression enables the use of different methods for screening compounds, hit identification, hit characterisation and crystallisation to obtain protein-ligand structures. The first step is to establish a systematic scheme for protein expression and purification in a reproducible manner.

In this chapter I describe the expression and purification of the ankyrin domain for structural analysis and its initial biophysical characterisation using circular dichroism. I also describe an assessment of the potential binding sites in the ankyrin domain by examining its interactions in the three-dimensional structure of the Notch transcription complex. This was performed using the relational database PICCOLO and by analysing mutational studies data.

2.1 Protein expression and purification

2.1.1 Chemicals

The chemicals used were from either Sigma-Aldrich (St. Louise, MO), or Melford Laboratories (Ipswich, UK), unless stated otherwise.

2.1.2 Protein expression:

The construct pET41a (+) encoding the human Notch1 ankyrin domain (Ehebauer *et al.*, 2005) was transformed into Rosetta (DE3) *Escherichia coli*. Cells were grown in 1 L 2xYT medium containing 34 µg/ml chloramphenicol and 25 µg/ml kanamycin in a shaking incubator at 37°C. When the OD₆₀₀ reached 0.6, they were induced with 1mM IPTG for 3 hours.

The cells were harvested by centrifuging at 5,000 ×*g* for 20 minutes. The pellet was then resuspended in 25 ml of 20% sucrose; 50 mM Tris-HCl (pH 8.0); 150 mM NaCl containing one protease inhibitor tablet (Complete EDTA free-Roche) and 1mg/ml of lysozyme grade VI chloride.

2.1.3 Protein purification

The resuspended pellets were sonicated (Sonicator XL2020, Misonix) on ice using 30 second pulses for a 2 minute period, centrifuged at 17,000 × *g* for 30 minutes. The supernatant was loaded onto Ni-NTA agarose (QIAGEN) column with a 5 ml bed volume. The column was washed twice with 10 ml of 50 mM Tris-HCL (pH 8.0); 50 mM NaCl; 50 mM imidazole and then eluted twice with 10 ml with 50 mM Tris-HCl (pH 8.0); 300 mM; 150 mM imidazole. The first eluent fraction was then loaded onto a HiTrap Q-HP anion-exchange column (Phenomenex) pre-equilibrated with buffer A: 50 mM Tris (pH 8.0); 50 mM NaCl. Protein was eluted using a gradient of 0-100% of buffer B: 50 mM Tris-HCl (pH 8.0); 1 M NaCl over 20 column volumes. Collected fractions were loaded onto a Superdex 200 HR 10/30 size exclusion column (Phenomenex) and eluted using 50 mM Tris-HCL (pH 8.0) and 50 mM NaCl.

2.1.4 Results and discussion

The ankyrin domain was expressed as a C-terminal-His-tagged protein. The first purification step used affinity chromatography employing a Ni-NTA affinity

column. Analysis of the eluate fractions by SDS-PAGE showed some protein contaminants that were removed by further purification steps. Ion exchange chromatography was used as a second purification step. After visualising the protein on SDS-PAGE, the fractions that contained the protein were pooled together for a final polishing purification step by size exclusion chromatography (Figure 2.1).

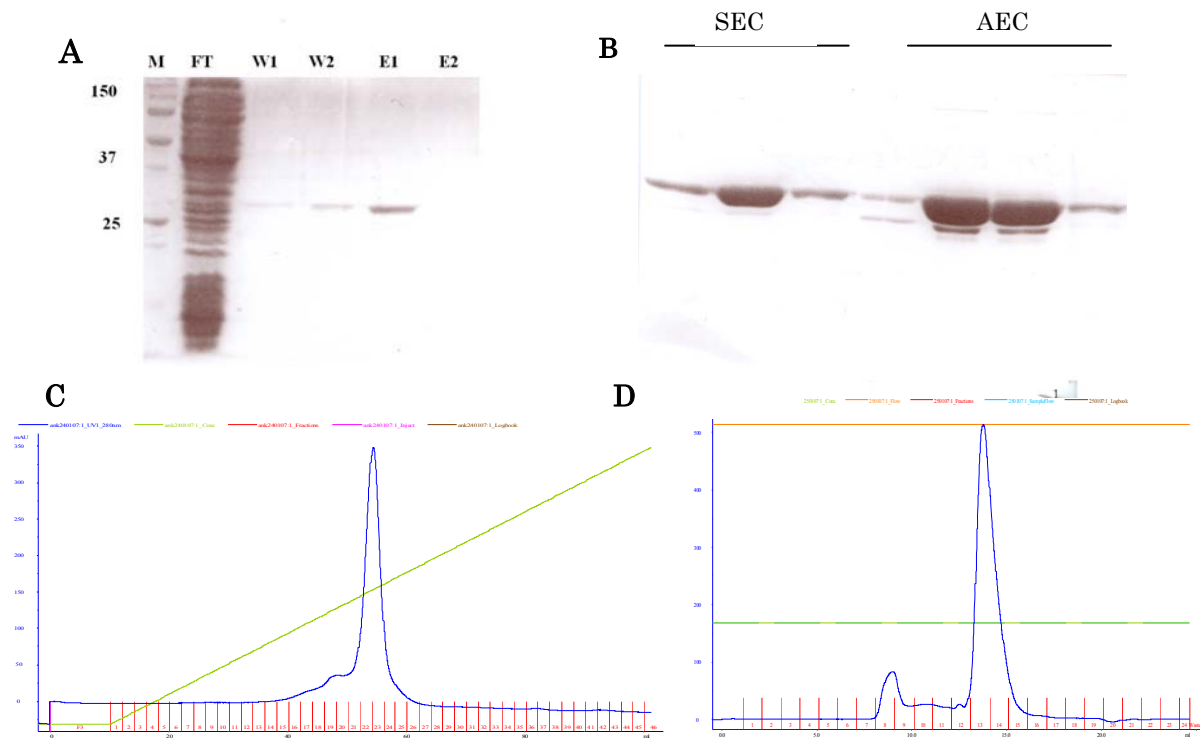


Figure 2.1 Purification of the ankyrin domain by Ni-NTA anion exchange chromatography (AEC) and size exclusion chromatography (SEC). A) SDS-PAGE showing Ni-NTA purification. (M=Marker, FT= flowthrough, W(1-2)=wash, E(1-2)= elution) B) SDS-PAGE showing AEC and SEC fractions C) anion exchange chromatography D) size exclusion chromatography

The single peak indicated a highly pure protein that could be used for further experiments.

This protocol was adapted from Ehebauer *et al.* and assessing its reproducibility was essential before proceeding in any future experiments. Batch-to-batch variations can become problematic in iterative processes. It was possible to purify the protein in one day which makes it very convenient for repetitive experiments. The ankyrin domain was expressed giving a reasonable yield of

about 5 mg/L of media. As shown in Chapter 3, about 20 mgs are required for screening a fragment library. More protein was required to produce crystals to optimise soaking conditions and for co-crystallisation attempts with fragments and ligands as described in Chapter 4. The kinetic characterisation using Surface Plasmon Resonance described in Chapter 5 utilised only small quantities of protein.

2.2 Circular dichroism spectroscopy

2.2.1 Material and Methods

The far-UV circular dichroism spectrum of the ankyrin domain was recorded using an Aviv 215 circular dichroism spectrometer (Aviv Instruments Inc.). The spectrum was recorded at 20°C in a clean 0.1 cm quartz cell, using an average time of 0.5 seconds and a step size of 0.5 nm. The spectra of 15 scans were averaged and subtracted from the baseline of the buffer containing 50 mM Tris pH8 and 50 mM NaCl.

The denaturation of the ankyrin domain was recorded at 222nm due to its high helical content. The temperature range set between 20°C and 90°C and the temperature range was fixed at 0.1°C. The equilibration time at each temperature was adjusted to 0.5 min with heating rate 1°C per minute. Ellipticity (y) is plotted against temperature (x) to examine the unfolding curves.

2.2.2 Results and discussion

Circular dichroism (CD) is a spectroscopic technique used to investigate the folding and unfolding of proteins, to gain clues about the effects of mutations and ligand binding on protein structure, and to estimate the secondary and tertiary structures of proteins. This phenomenon, which is exhibited by optically active molecules, relies on the differential absorption of left and right circularly

polarised light. For proteins, the secondary structure is studied in the far-UV region at wavelengths (180-250 nm) where the peptide bond absorbs. The characteristic CD-spectra have minima at 208 and 222nm when there is α -helix, at 218nm for β -sheet and at 198nm for random coil. The near-UV region (250-350 nm), where aromatic amino acids absorb, provides information on the tertiary structure.

The CD spectrum of the ankyrin domain was recorded in order to investigate its structural stability. The CD spectrum revealed two minima at 208 and 222 nm, indicating a predominantly α -helical protein (Figure 2.2). This confirmed the quality of the protein and its folded nature. Further experiments could then be conducted on the ankyrin domain for drug discovery purposes.

The unfolding temperature of a protein can be measured by recording the CD spectrum as a function of temperature at a fixed wavelength. This assists in determining the presence of any unfolding intermediates, and can be used to identify changes in unfolding caused by variations in salt concentrations or pH. Two-state unfolding is indicated by a sigmoidal curve, from which the unfolding temperature T_m is estimated from the midpoint. The ankyrin domain unfolded in a two-state manner despite its modular architecture (Figure 2.3). The T_m was estimated to be around 45° C and its unfolding behaviour indicated that it would be possible to employ a fluorescent-based thermal shift assay as a screening method. In this method, the protein is required to reproducibly unfold in a two-state manner to measure the difference of T_m in the absence and presence of fragments or compounds.

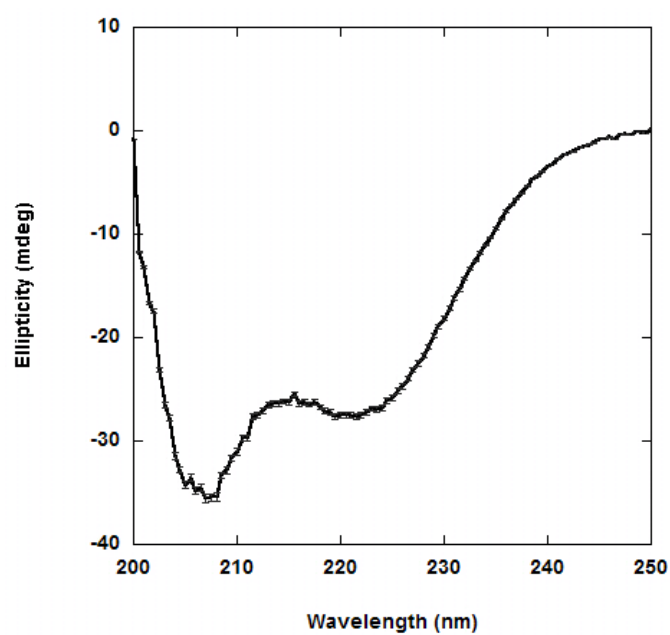


Figure 2.2 Far-UV CD spectrum of the ankyrin domain showing two minima at 208 nm and 220 nm.

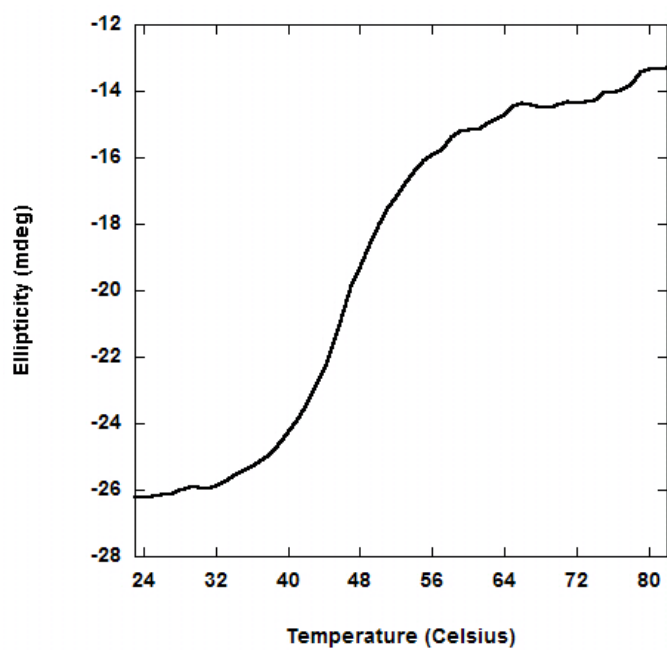


Figure 2.3 Denaturation curve showing the ellipticity as a function of temperature to determine the unfolding temperature at the midpoint of the curve.

2.3 Analysis of the ankyrin interactions in Notch transcription complex

PICCOLO is a database created by Dr Richard Bickerton to characterise interactions in multi-protein systems. PyMOL was integrated in PICCOLO as a visualisation tool to assist in analysis. Interactions and different interface properties could be visualised in different colours.

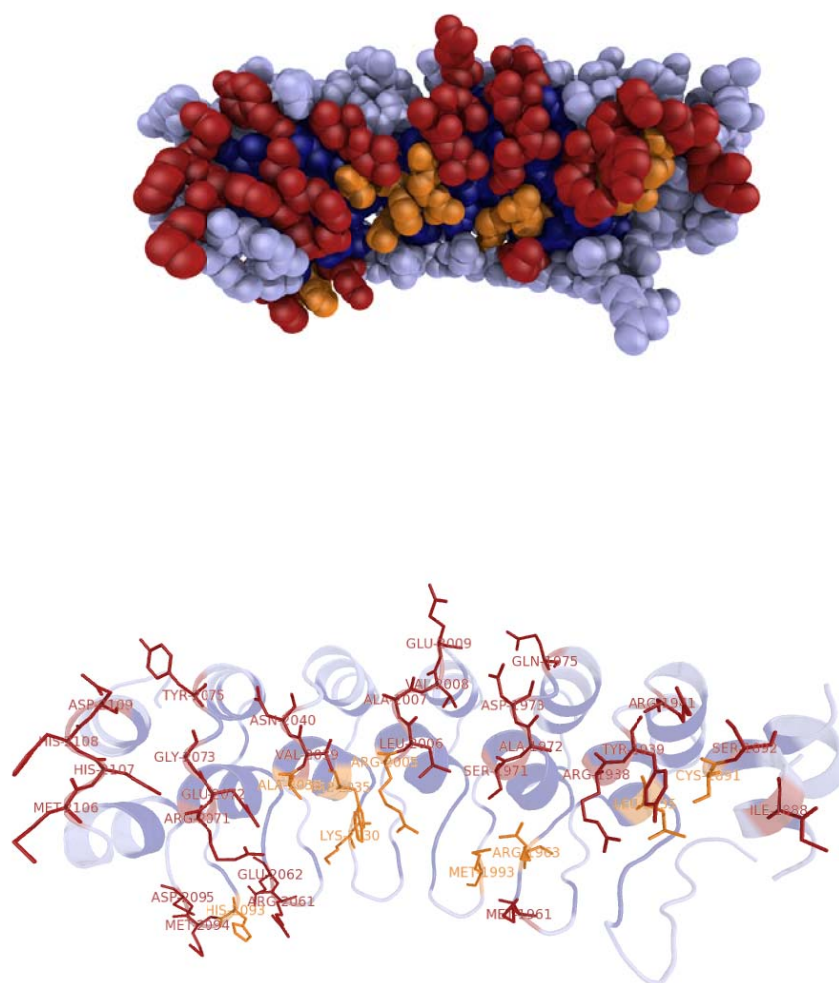


Figure 2.4 A schematic representation of the solvent accessibility of the interacting molecules of the ankyrin domain (PDB 2F8Y) at the interface between ank-CSL and ank-MAML as detected by PICCOLO; red represents the residues at the interface periphery and orange the residues at the interface core. The upper panel shows a sphere representation and in the lower panel a cartoon representation of the residues.

The interactions between the ankyrin domain and MAML predicted by PICCOLO were predominantly ionic interactions (Figure 2.5). This included charged interactions between acidic residues at loops between repeats three and four of the ankyrin domain and turns three and four of MAML. As shown by the crystal structure, the helical MAML polypeptide bends at Pro46 (Nam *et al.*, 2006). The ankyrin domain interacts with the first region of MAML before it bends. A second subset of hydrophobic interactions are observed between turns seven to nine of MAML and repeats six and seven of the ankyrin domain (Nam *et al.*, 2006). The residues of the ankyrin domain involved in the interaction with MAML are mainly at the interface periphery as predicted by PICCOLO and are mostly charged residues.

Charge reversal mutations of selected residues of MAML were carried out to investigate their importance in the formation of the transcription complex (Del Bianco *et al.*, 2007). Single mutation of R25 to glutamate prevented the formation of the complex whereas single mutations of residues R22, R26, R31, E38, R53 and T56 did not have any inhibitory effect. Double mutation of R22 and R25 had a dominant negative effect.

Table 2.1 A list of the properties of the interactions between the ankyrin domain and MAML

Ankyrin residues	Ankyrin domain interface solvent accessibility	MAML residues	Type of interaction
Tyr2075	Interface periphery	Arg40	Pi-cation
His2107	Interface periphery	Arg40	Pi-cation
Asp1973	Interface periphery	Arg22	hydrogen-bond
Gly2073	Interface periphery	Arg40	hydrogen-bond
Asp1973	Interface periphery	Arg40	hydrogen-bond
Asp2109	Interface periphery	Arg26	ionic
Glu2009	Interface periphery	Arg22	ionic
Glu2009	Interface periphery	Arg25	ionic
Asp2109	Interface periphery	Arg40	ionic
His2108	Interface periphery	Glu47	ionic
Ala2007	Interface periphery	Glu47	hydrophobic
Met2106	Interface periphery	Val44	hydrophobic
Met2106	Interface periphery	Leu29	hydrophobic

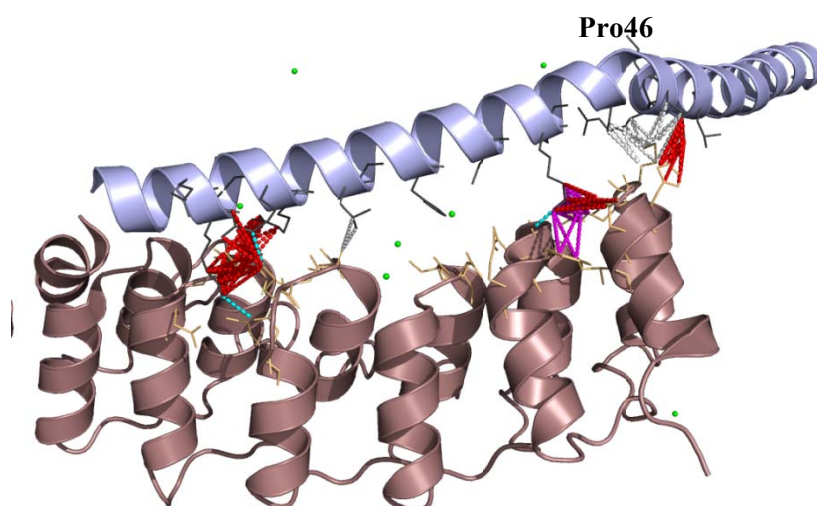


Figure 2.5 Cartoon representation of the interactions detected by PICCOLO between the ankyrin domain (maroon) with MAML (light blue) from the complex crystal structure (PDB 2F8Y). Ionic interactions are depicted in dashed red lines, hydrogen-bond interactions in cyan, pi-cation in magenta and hydrophobic interactions in grey. Water molecules are shown as green dots.

The crystal structure of the transcriptional complex shows that CSL interacts with the ankyrin domain in an unexpected manner. The CSL protein is formed of three domains; an N-terminal domain (NTD), a central β -trefoil domain (BTD) and a C-terminal domain (CTD). It was expected that the CSL domains were arranged in the same order upon complexation, however the crystal structure revealed otherwise. A long β strand connects BTD to the CTD in a manner that gives the impression that CTD is sitting on top of BTD (Gordon *et al.*, 2008). The interaction between CSL and the ankyrin domain occurs through the BTD and CTD of CSL in a discontinuous arrangement. The NTD interacts with the positive region of the ankyrin domain through the loop residues that connect repeats five to six and six to seven of the ankyrin domain (Nam *et al.*, 2006). The β sheet of CTD interacts with the concave surface of the ankyrin domain creating a groove parallel to the long axis of the ankyrin domain. This groove accommodates the N-terminal region of MAML before it kinks by 40° as highlighted previously. The C-terminal region of MAML after the curvature is nestled in a hydrophobic pocket formed predominantly by a four-stranded β sheet of NTD of CSL (Nam *et al.*, 2006). PICCOLO analysis showed that most of the ankyrin domain residues interacting with CSL are located in the interface core (Figure 2.6). The residues are polar and engaged predominantly

in ionic and hydrogen-bond interactions. The basic residues are more abundant than acidic residues.

Similarly, mutational studies were conducted to investigate the importance and significance of the interacting residues on stabilising the transcription complex (Deepti Gupta, unpublished; Del Bianco *et al.*, 2007). A limited alanine scan of was carried out of the ankyrin domain residues at the ankyrin-CSL interface and ankyrin-MAML interface. This included residues; R2005A, W2035A, E2072A, and E2076A at the ankyrin-CSL interface and residues D1973A, E2009A, N2040A and D2109A at the ankyrin-MAML. None of these mutations affected the formation of the ternary complex in size-exclusion chromatography. However, charge reversal mutations of two particular residues prevented the formation of the ternary complex. Single site mutations of D1973R and E2072R disrupted the formation of the complex in chromatographic assays. These mutations also inhibit the induction of transcription in cell-based reporter assays. This suggested that the electrostatic complementarity is essential for the assembly of the transcription complex. In addition, multisite substitution of residues (E2072K/ D2095V; R1963E/ R2005E/ E2072K; V2039D/ E2072K/ D2095V and R1963E/ R2005E/ E2072K/ D2095V) appeared to have a more pronounced effect in transcription inhibition suggesting an additive effect (Deepti Gupta, unpublished).

There is no conformational rearrangement observed in either the ankyrin domain or CSL structures after complex formation. Differences in the backbone chain of CSL when compared with the *Caenorhabditis elegans* CSL:DNA complex (Kovall & Hendrickson, 2004) alone can be attributed to crystal packing.

Table 2.2 A list of the properties of the interactions between the ankyrin domain and CSL

Ankyrin residues	Ankyrin domain interface solvent accessibility	CSL residues	Type of interaction
Cys 1891	Interface core	Pro434	H-bond
Arg2005	Interface core	Arg382	H-bond
Arg1963	Interface core	Gln347	H-bond
Arg2005	Interface core	Lys123	H-bond
Glu2072	Interface periphery	Tyr381, Gly384	H-bond
Arg2071	Interface periphery	Glu385	H-bond
Arg2061	Interface periphery	Gln362	H-bond
His2093	Interface core	His124	Aromatic
His2093	Interface core	His124, Arg 146	Pi-cation
Arg1941	Interface periphery	GLN358	ionic
Arg1963	Interface core	Asp353	ionic
Arg2005	Interface core	GLN358	ionic
Lys 2030	Interface core	Glu385	ionic
Asp2095	Interface periphery	Glu385	ionic
Glu2091	Interface core	His124	Aromatic
Arg2071	Interface periphery	Glu385	ionic

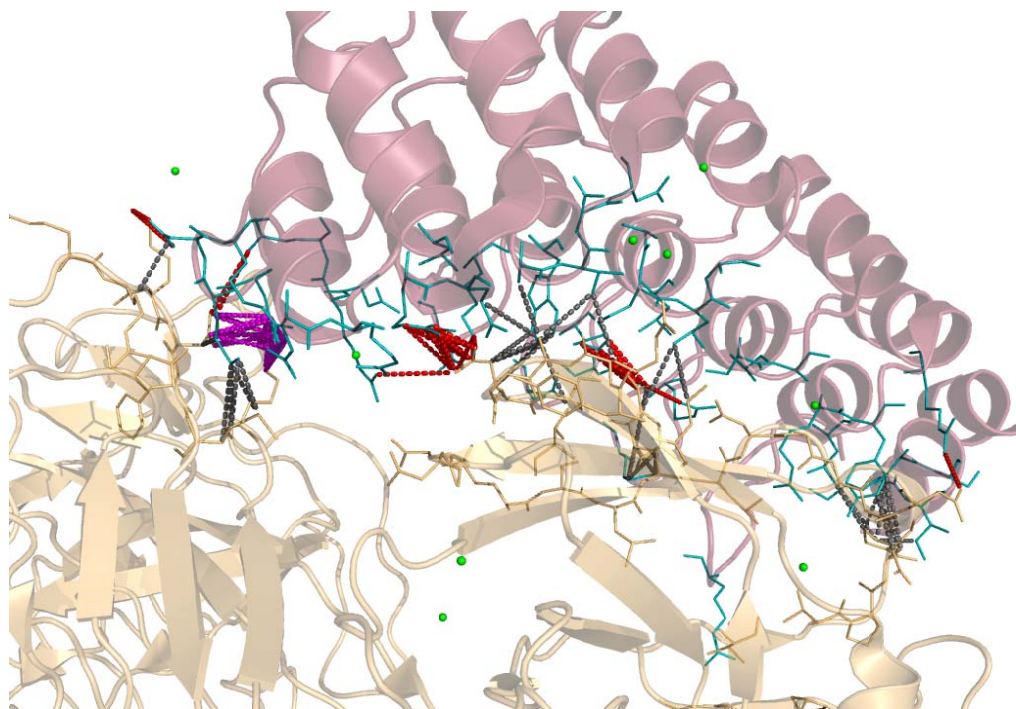


Figure 2.6 Cartoon representation of the interactions detected by PICCOLO between the ankyrin domain (maroon) with CSL (light orange) from the complex crystal structure (PDB 2F8Y). Ionic interactions are depicted in dashed red lines, hydrogen-bond interactions in cyan, pi-cation in magenta and hydrophobic interactions in grey. Water molecules are shown as green dots.

The Notch transcription complex appears to be stabilised by numerous additive electrostatic interactions. The lack of evident hotspot residues makes it difficult to modulate the protein-protein interactions. Inhibiting several interactions simultaneously may be required in order to inhibit complex formation. Identifying small molecules such as fragments that could interfere with these numerous interactions seems to be a sensible approach.

The ankyrin residues interacting with MAML lie at the interface periphery of the ankyrin interface which is solvent accessible. This can suggest that polar molecules are more likely to bind to this region. The interface periphery of protein-protein interfaces is usually rich in arginines and histidines (Richard Bickerton, unpublished). However, only two histidine residues were detected at the interface periphery of the ankyrin domain with MAML. Acidic residues are more abundant forming ionic and hydrogen bond interactions.

On the other hand, the interface of the ankyrin domain with CSL constitutes mainly of residues at the interface core. This region is usually enriched with tryptophan, tyrosine and methionine residues (Richard Bickerton, unpublished). However, none of those residues were observed at the interface core of the ankyrin domain. Instead, arginine residues were favoured engaging in hydrogen bonds and ionic interactions.

The nature of the interactions of the ankyrin residues interacting with CSL and MAML is broad and diverse. Finding small molecules that could interfere with these interactions may require screening a more diverse library. Fragment libraries can provide a larger chemical space for targeting the ankyrin domain interface.

Chapter 3

Fragment Library Screening

3.1 Introduction to Fragment Screening

3.1.1 Fragment library design

The design of a screening library is of paramount importance, as it will influence the direction of drug discovery and development of the initial hits. The factors that should be considered when establishing a fragment library (Leach *et al.*, 2006) depend on the screening technique and on screening concentrations, the higher the concentration the smaller the library. The novelty of the target and knowledge of natural or known ligands also determine how focused the library should be.

Physical properties of the fragments are key to setting up a library. The solubilities of the fragments are important, although difficult to predict and need to be measured experimentally. The numbers of hydrogen-bond donors and acceptors, together with the cLogP value can give an indication of the solubility. Since the fragments are usually screened at higher concentrations (up to 10 mM) in aqueous buffers, the fragments should have considerable water solubility. The solubility of fragments can be improved by using organic solvents as co-solvents. DMSO is usually used at concentrations of 1% to 10%, depending on the screening technique, in order to minimise aggregation and false positive hits. The purity of the fragments is critical in screening, as impurities can interfere by encouraging aggregation or interacting with the target irreversibly, leading to false positive hits. However, this will depend on the sensitivity of the screening technique as well.

The chemical properties of the fragments define the diversity of the library. It is preferable that the fragment library contains similar molecules to confirm positive hits as well as to define SAR even with small molecules. The fragments should be devoid of any toxicophore or reactive groups; several filters allow

excluding such molecules (Verheij *et al.*, 2006; Roche *et al.*, 2002). The feasibility of synthesis and chemical elaboration should be considered as well: the molecules should contain a “chemical handle”, which allows for further elaboration through growing or linking. Several functional groups are used as chemical handles, including carboxylic acid and nitrile groups. Although this provides an attractive way to develop the fragment from hit to lead, it limits the search for novel hits. There has always been a search for new scaffolds representing a much broader chemical space. Interestingly, very few new synthetic compounds are published every year even though it has been demonstrated that around 3000 synthetically tractable new compounds need to be explored (Pitt *et al.*, 2009). Many of these will be challenging to synthesise, or incompatible with the biological systems as the authors suggested (Pitt *et al.*, 2009).

The physical and chemical requirements for constructing a representative fragment library led to the suggestion of a “rule of three” (Congreve *et al.*, 2003). This rule is derived by analogy with the “Lipinski rule of five” and indicates that molecular weight should preferably be ≤ 300 , the number of rotatable bonds ≤ 3 , the number of hydrogen bond acceptors and donors both ≤ 3 , $cLogP \leq 3$ and the polar surface area $\leq 60 \text{ \AA}^2$.

The fragment library used in this study is a commercial library supplied from Maybridge and complies with the rule of three (Congreve *et al.*, 2003). . This library that I used included 100 extra compounds prepared in the Chemistry Department, University of Cambridge in the laboratory of Professor Chris Abell, in addition to the Maybridge set. The library is maintained in Chemistry for screening a range of targets; it is not target-tailored but it is reasonably diverse.

3.1.2 Screening Techniques

The screening methods should be sensitive enough to identify weakly binding fragments. Ideally, biological *in vivo* assays would be the best way for screening and identifying potential hits. However, these molecules are weak binders and interact in low μM or mM concentration ranges, which are usually not detected

in assays. Biophysical *in vitro* screens are normally carried out instead. They have the advantage that they can detect the fragments that bind and interact directly with the target without interference from other cellular proteins that are often found in biological assays. The biophysical techniques vary in their sensitiveness, their hit rate, their advantages and disadvantages, and their requirements for both target and fragment library. The choice of screening method depends primarily on the protein target, i.e. whether it can meet the requirements of the technique, its availability and the ease of interpreting the screening data.

3.2 Fluorescent-based thermal shift assay screening

3.2.1 Background of fluorescent-based thermal shift assays

The fluorescent-based thermal shift assay can be considered a high-throughput screening technique in drug discovery. It has emerged as a rapid technique for hit identification, based on energetic coupling events between the ligand and protein unfolding. These energy transformation processes cause ligand-induced changes in protein unfolding curves (Pantoliano *et al.*, 2001). The unfolding is thermally induced and measured in the presence and absence of ligands. Protein unfolding can be monitored by environmentally sensitive dyes, which have different quantum yields depending on the dielectric constant value of the solvent or surrounding environment (Pantoliano *et al.*, 2001). As the protein unfolds, it exposes hydrophobic regions that resemble a low dielectric constant solvent resulting in an increased fluorescence. In this study, SYPRO® orange (Invitrogen) was used as the fluorescent dye.

3.2.2 Advantages of fluorescent-based thermal shift assays

This method has become a popular method in the past few years as a preliminary screening method. It is regarded as a high-throughput screening technique owing to its rapidness, cheapness and simplicity. It is suitable for screening protein targets of unknown functions or unknown ligands. There is no

special treatment or labelling required for either the protein or the compounds. Minimal amounts of protein are required for screening large libraries; 10 mgs of protein can be used to screen about a 1200 compound library. Highly purified protein samples are not necessary, 75% pure samples can be used for screening as indicated by Pontoliano *et al.* (2001).

3.2.3 Limitations of fluorescent-based thermal assay screening

This assay is a simple preliminary screening method that gives a quick yes/no answer. However, as for any screening method, several limitations should be considered when screening and analysing data. Compounds can bind to both the folded and unfolded state of the protein, decreasing the ΔT_m value and causing a false negative. In addition, highly fluorescent compounds produce a high fluorescent background, giving a lower ΔT_m value and leading to false negative hits as well. The method is limited to proteins that unfold in a two-state manner, where the T_m value can be easily determined. Binding affinities can be difficult to calculate, as the binding enthalpy can be overestimated. A technical limitation can arise from the instrumentation itself, because of the optical edge effects of using a CCD (closed-circuit device) camera for fluorescence detection. This leads to variation of excitation light and fluorescence intensities between different samples.

In this chapter, I describe the optimisation of the screening conditions by buffer screening and the screening of a small fragment library against the ankyrin domain.

3.2.4 Buffer Screening

3.2.4.1 Materials and methods

The ankyrin domain was expressed as described previously (See Chapter 2). Stock protein solution was frozen at a concentration of about 315 μM (10 mg/ml). It was diluted with original buffer, which constituted of 50 mM Tris pH 8, 50 mM NaCl to reach a concentration of 10 μM (about 1.65 mg/ml). The buffer screening was carried out in thin-walled plates (Bio-Rad) and sealed with caps (Bio-Rad). The solutions were added to form a final volume of 100 μl as shown in table 3.1. The buffers were screened at 10 mM and some were screened at 20 mM. Water was added instead of the buffer in the control samples. The plates were then heated in an iCycler iQ Real Time PCR Detection System (Bio-Rad) from 25 $^{\circ}\text{C}$ to 60 $^{\circ}\text{C}$ with 0.5 $^{\circ}\text{C}$ increments. Fluorescence was detected by a CCD camera. SYPRO® orange dye has fluorescence excitation and emission wavelengths 490 and 525 nm, respectively. The buffers used for screening are shown in table 3.2.

Table 3.1 Reagents used in buffer screening

Reagent	Concentration
Protein	8 μl (final concentration 10 μM)
SYPRO orange	25 μl (1:250)
DTT	5 μl (final concentration 5 mM)
Buffer	10 μl or 20 μl (final concentration 10 mM or 20 mM)
NaCl	50 mM and at 400 mM

Table 3.2 Buffer list used for screening, the pH values and the concentration used for screening.

Number	Buffer	pH	concentration
--------	--------	----	---------------

1	CHES	8.5	10 mM, 20 mM
2	Sodium Phosphate	7	10 mM, 20 mM
3	Bicine	9	10 mM, 20 mM
4	Sodium Phosphate	5	10 mM, 20 mM
5	Sodium Acetate	4.6	10 mM, 20 mM
6	Glycine	10	10 mM, 20 mM
7	Borate	8.5	10 mM, 20 mM
8	Tris	8	10 mM
9	bis Tris	6.7	10 mM
10	PIPES	6	10 mM
11	Sodium Citrate	5.6	10 mM
12	MES	6.5	10 mM
13	Imidazole	8	10 mM
14	CAPS	10.5	10 mM
15	Sodium cacodylate	6.5	10 mM

3.2.4.2 Results

The ankyrin domain was screened in 15 buffers at pH values varying from pH 4.6 to pH 10.5 at usually at 10 mM but sometimes at 20 mM buffer concentration. None of the buffers screened at 50 mM NaCl salt concentration showed any stabilisation of the protein (Figure 3.1). The Ankyrin domain did not unfold in a two-state manner as demonstrated before (Zweifel & Barrick, 2001), but showed some melting profiles indicating complex unfolding patterns and instability. When the concentration of the salt was increased to 400 mM, one buffer only, buffer CHES at pH 8.5 was able to stabilise the ankyrin domain (Figure 3.2). Then the protein unfolded in a two-state transition and showed reproducible unfolding curves. The average T_m value was 43 °C. That is approximately 3 °C higher than previously reported in *Drosophila* (Zweifel & Barrick, 2001), and this could be attributed to the higher salt concentration.

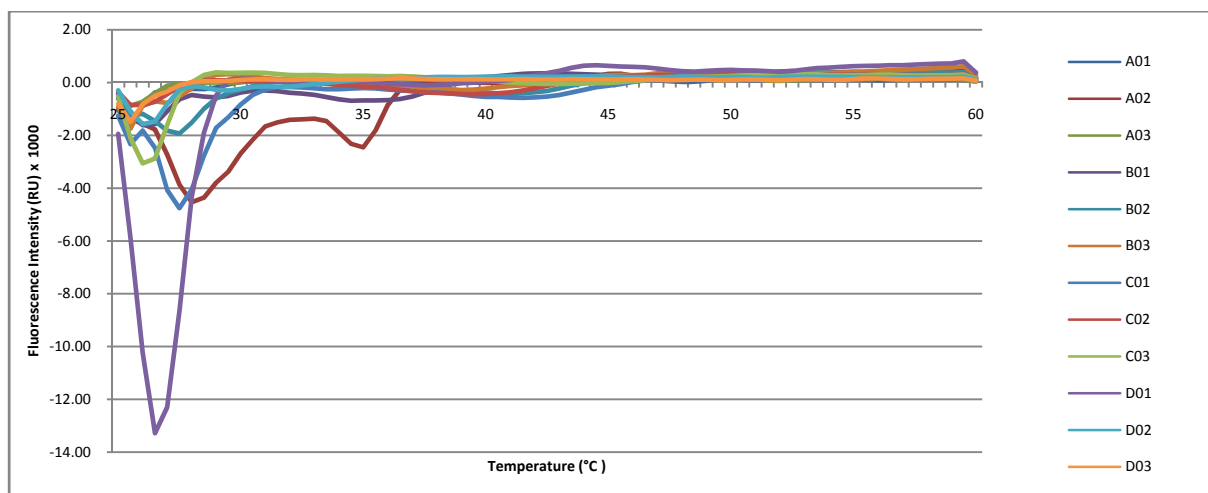


Figure 3.1 The derivative unfolding curves of the protein in different buffers at low salt concentration. Many peaks appeared indicating some instability followed by what seems to be aggregation.

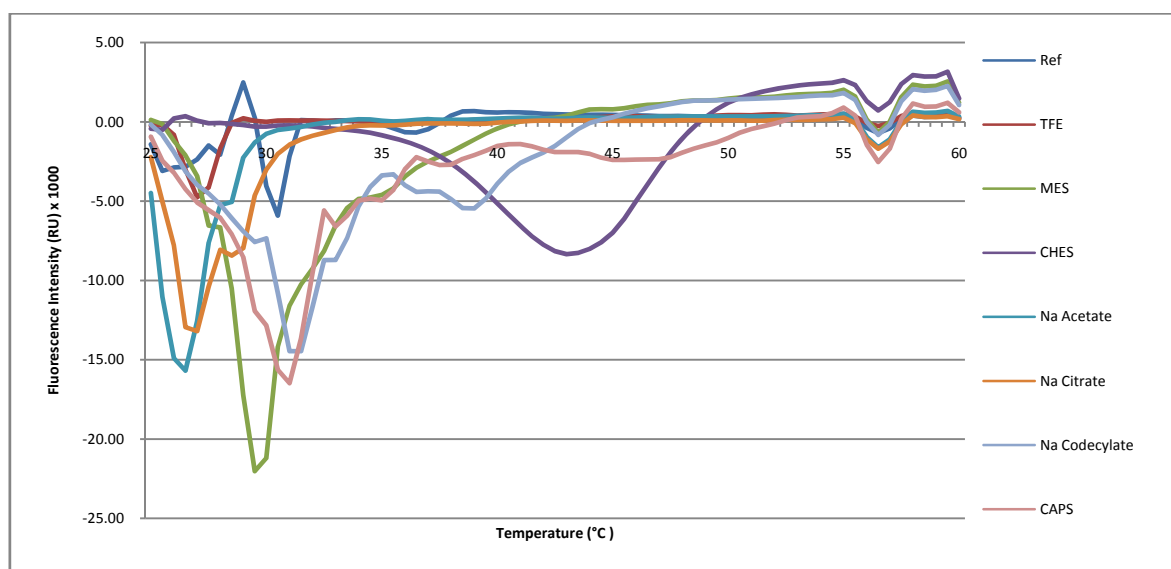


Figure 3.2 Derivatives of the melting curves for the buffer screens in high salt concentration. The only buffer that showed a reproducible two-state unfolding transition was CHES pH 8.5 as shown which appears as a chevron.

3.2.4.3 Discussion

Buffer screening is a useful way to optimise buffer conditions for various purposes. Optimising biophysical properties of proteins, such as solubility, homogeneity and stability is important for protein crystallisation. Thermofluor-based buffer optimisation has been widely used to improve crystallisation-hit rates (Ericsson *et al.*, 2006).

The ankyrin domain is a repeat protein that comprises seven ankyrin repeats. The characteristic architecture of the ankyrin repeats would suggest that each repeat would behave as an independent module, and unfold independently of the adjacent repeats. This is due to the fact that the repeats are stacked against each other. These repeats are stabilised through local and short-range interactions with respect to the sequence positions, unlike globular proteins that are stabilised by long-range contacts of distant residues (Mosavi *et al.*, 2004). It would be expected that the unfolding events will involve a population of intermediates. However, thermodynamic data and folding studies on ankyrin domain in *Drosophila* Notch2 receptor (Zweifel & Barrick, 2001) and in rat myotrophin (Lowe & Itzhaki, 2007) suggest otherwise. The studies reveal that the ankyrin domains adopt a cooperative unfolding behaviour and unfold in a two-state fold manner.

However, the unfolding of the ankyrin domain using the thermal-fluorescent method did not show a two-state transition. The melting profile indicated more than one unfolding event. This was irreproducible and varied with different runs in same buffer and salt conditions. This could be explained in several ways. The first repeat is partially unfolded and disordered, and only when it binds to CSL, the transcription factor, does it adopt the unique ankyrin fold. This would interfere with the measurement. Another possibility is that the fluorescent dye can bind to individual repeats at the same time and each repeat will act as an individual domain, appearing as multi-state folding.

The aim of this screen was to find an optimum condition that stabilises the structural conformation of the protein, particularly the first unfolded repeat.

Screening with buffers at low salt concentration did not stabilise the protein and gave the same melting profiles as the original buffer. Increasing the NaCl concentration to 400 mM did not show much improvement, except when using CHES pH 8.5 as a buffer, where the protein was unfolded in a reproducible two-state manner. High NaCl concentrations appear to have stabilised the protein as indicated by a higher T_m value. The salt effect in stabilising α -helical conformation has been studied recently (Xiong *et al.*, 2009). It was found that various salts have the ability to induce the formation of α -helical structures in polyalanine peptides through different effects. NaCl was found to have both ion-screening effects and specific ion-binding interactions. Ion screening reduces the electrostatic interactions between the protein charges depending on the ionic strength (Xiong *et al.*, 2009). Specific ion binding can stabilise the α -helix due to preferential ion pairing between oppositely charged ions of similar charge density (Xiong *et al.*, 2009). NaCl could have the same affect on the ankyrin domain, inducing α -helical conformation in the first repeat when it is unbound in solution.

Another explanation for the stabilisation of the ankyrin domain in CHES buffer and high salt concentration might be the optimisation of the dielectric constant of the solution so that the SYPRO® orange dye can bind only to one ankyrin repeat.

3.2.5 Screening of Fragment Library

3.2.5.1 Materials and methods

The fragment library is composed of about 1200 compounds. The fragments were dispensed in 96-well thin-walled plates (Bio-Rad) and sealed with caps (Bio-Rad). Solutions of 5 mM fragment and 5% DMSO as control were added to make a 100 μ l final volume. The solutions were prepared as shown in table 3.3. The plates were heated in an iCycler iQ Real Time PCR Detection System (Bio-Rad) in the same manner as described in section 3.2.4.1.

Table 3.3 The composition of the fragment and control wells showing the final concentration value in brackets

	Fragment well	Control well
Fragment	5 ul (5mM)	-
DMSO	-	5 ul (5%)
Protein	8 ul (10 uM)	8 ul (10 uM)
SYPRO orange	25 ul (1:250)	25 ul (1:250)
DDT	5 ul (5 mM)	5 ul (5 mM)
CHES pH 8.5	10 ul (10 mM)	10 ul (10 mM)
NaCl	47 ul (400 mM)	47 ul (400 mM)

3.2.5.2 Results and discussion

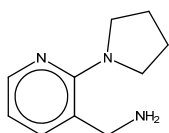
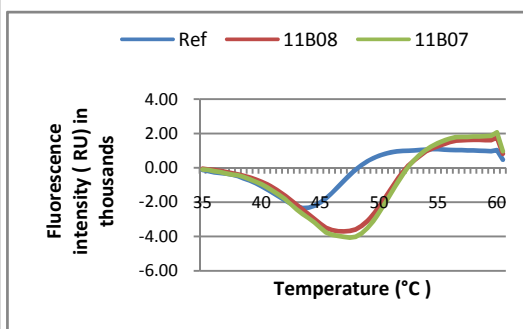
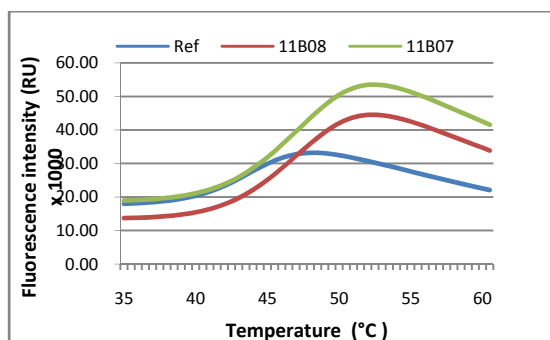
The melting profiles of the ankyrin domain were monitored in presence and absence of the fragment compounds. In the presence of 5% DMSO, the protein unfolded in a two-state manner and showed an average T_m value of 43°C indicating that the DMSO at this concentration did not have any detrimental effects. Each melting curve was investigated individually to confirm its two-state unfolding behaviour. The melting curves showed some irregularities in the temperature range (25 -35 °C); this could be due to the introduction of bubbles on sampling. This region was discarded on measuring T_m . Although each well contained solutions of identical composition apart from the fragment compound, there were variations in the fluorescence intensity. This may have resulted from variations in liquid handling, quenching effects of some fragments, bubbles again introduced in the solutions, or the non-uniformity across the CCD camera. Yet, these variations did not affect the results, as each T_m was measured independently in every unfolding curve. Calculating the derivatives of the melting curves helped in estimating and comparing T_m values, especially with small ΔT_m . The Excel macro sheet for analysing data was prepared by Dr. Duncan Scott (Department of Chemistry, University of Cambridge).

For simplicity, each fragment was designated a code number; the first part indicated the plate number and the second part the fragment position in plate. Each plate contained 72 fragments and controls were added in each plate and for every run.

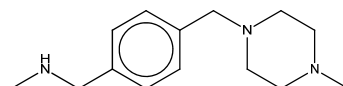
Positive hits were identified as fragment compounds that caused a shift in T_m by 0.5 °C or more. This was taken to indicate that the fragment is binding and stabilising the protein. The final number of hits was 36 out of a library of 1201 compounds giving a hit rate $2.99 \approx 3$ %. This figure is comparable to the hit rate generated by screening a traditional druggable enzyme target using the same fragment library (hit rate 3.25 %) (Leonardo Silvestre, personal communication). The hit rate can give an indication of the druggability of the target. Although protein-protein interfaces have been emerging as targets for drug discovery, their druggability is expected to be lower than other traditional targets using HTS libraries for screening. That can be due to fact that large molecules are difficult to accommodate small cavities found at protein-protein interfaces. Whereas smaller fragments can occupy these small pockets and hence enhance the hit rate.

However, the hit rate could be affected by the sensitivity of the screening technique itself. There could be a number of false-negative hits using thermofluor-based assays due to fluorescence quenching by the fragment molecules. Screening the fragments at high concentrations in the mM range would increase the probability of the excited molecules interacting with each other and would reduce the intensity of fluorescence. The presence of impurities and increasing temperature can contribute to fluorescence quenching as well.

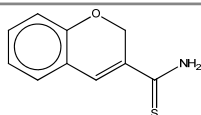
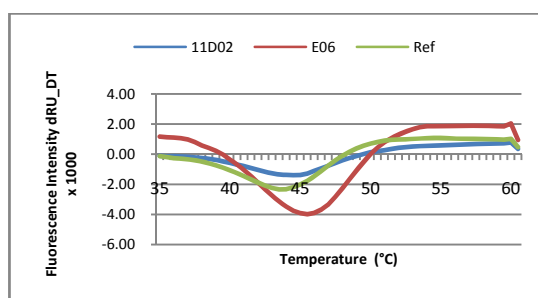
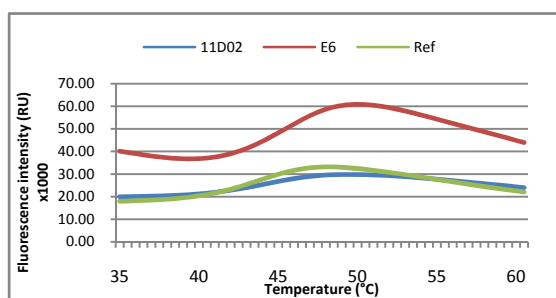
Different fragments from the library had different effects on the protein. One group stabilised the protein and caused a positive ΔT_m , a second group destabilised the protein decreasing its T_m , a third group precipitated the protein, and a fourth group had no effect. The first group, which was considered as a positive hit group, was analysed and further classified according to similar structural features. There are five subgroups that could be identified through their common scaffolds (Figure 3.9).



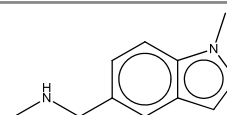
11B7 $\Delta T_m = 3.75^\circ\text{C}$



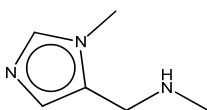
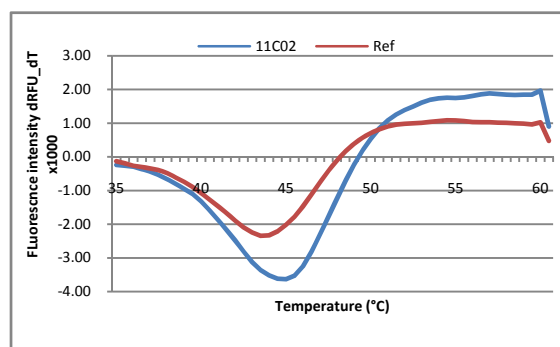
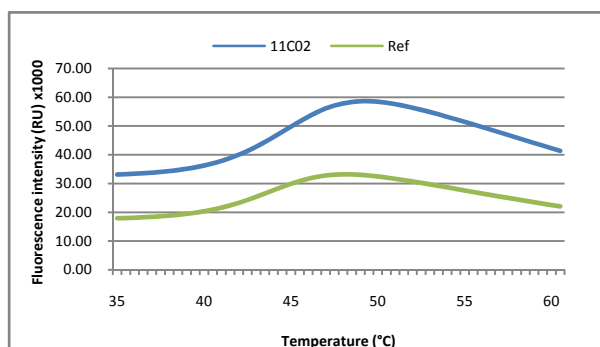
11B8 $\Delta T_m = 3.25^\circ\text{C}$



11D02 $\Delta T_m = 0.75^\circ\text{C}$

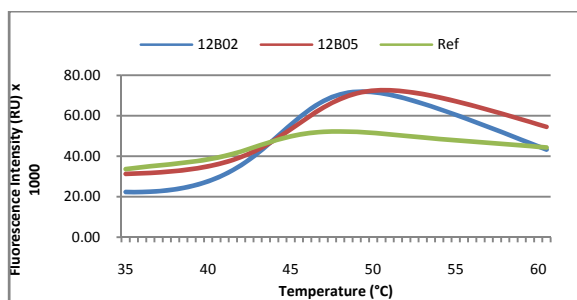


11E06 $\Delta T_m = 1.75^\circ\text{C}$

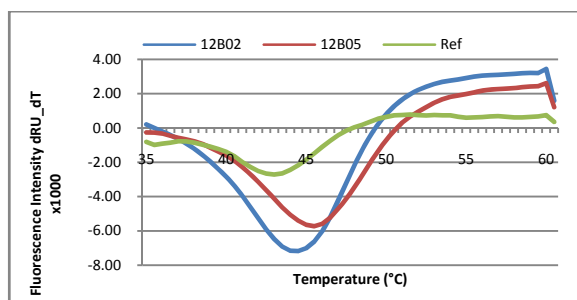
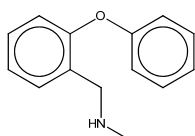


11C02 $\Delta T_m = 1^\circ\text{C}, 2^\circ\text{C}$

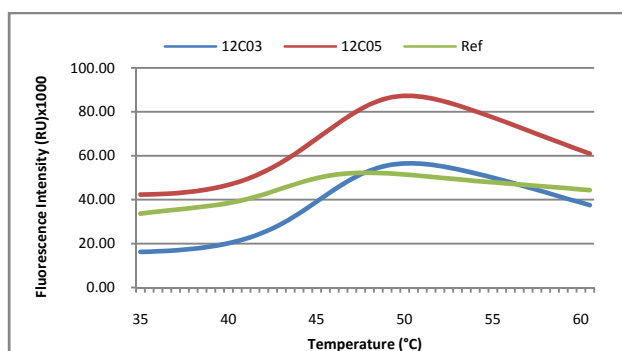
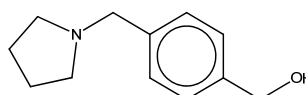
Figure 3.3 The thermal unfolding curves of the fragments 11B07, 11B08, 11D02, 11E06 and 11C02. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.



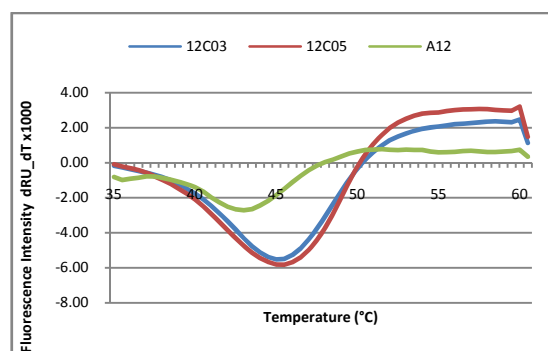
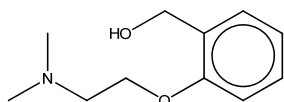
12B2 $\Delta T_m = 1.25^\circ\text{C}$



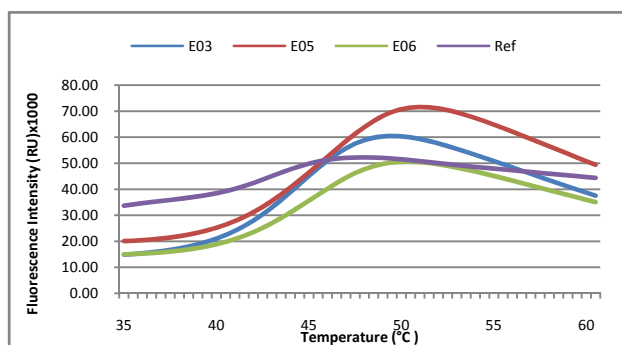
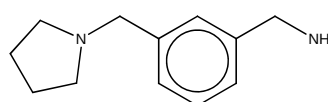
12B5 $\Delta T_m = 2.5^\circ\text{C}, 1^\circ\text{C}$



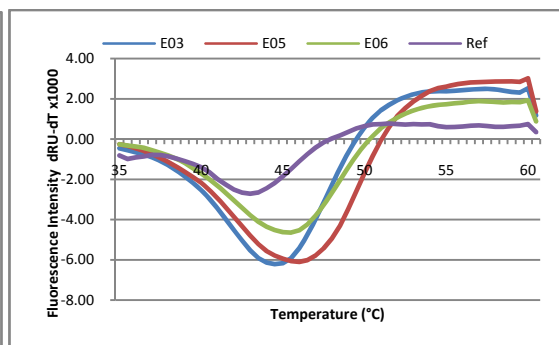
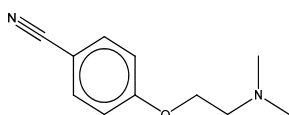
12C3 $\Delta T_m = 2.25^\circ\text{C}, 2^\circ\text{C}$



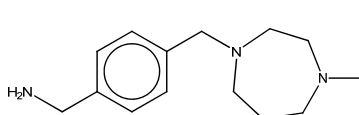
12C5 $\Delta T_m = 2.25^\circ\text{C}, 2.5^\circ\text{C}$



12E3 $\Delta T_m = 1.5^\circ\text{C}$



12E5 $\Delta T_m = 3^\circ\text{C}, 3.5^\circ\text{C}$



12E6 $\Delta T_m = 2.25^\circ\text{C}, 1^\circ\text{C}$

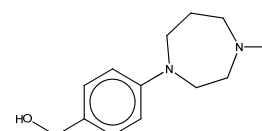
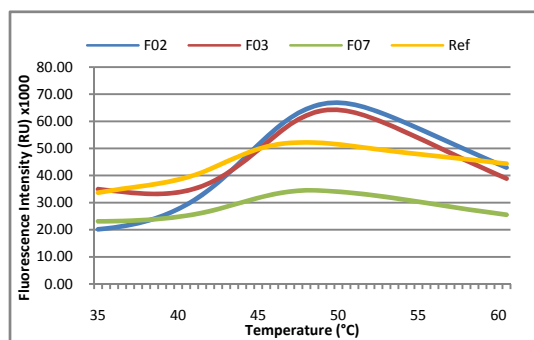
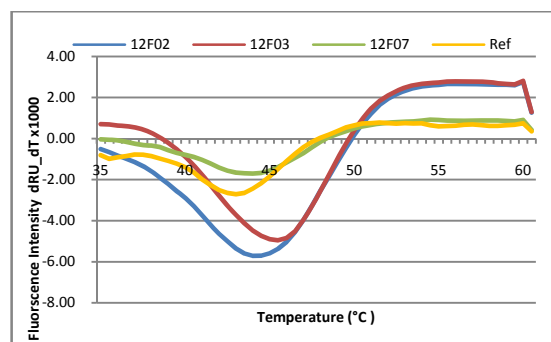
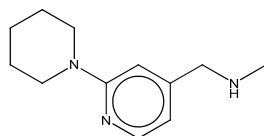


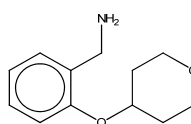
Figure 3.4 The thermal unfolding curves of 12B2, 12B5, 12C3, 12C5, 12E3, 12E5, 12E6. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.



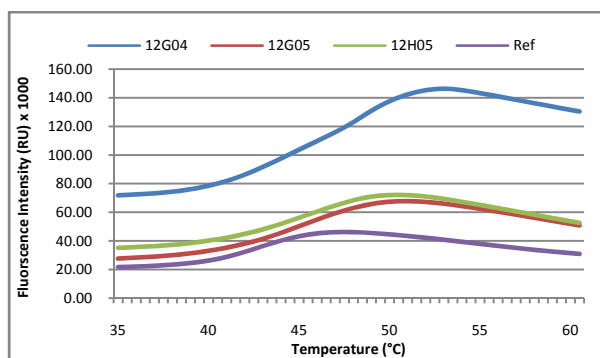
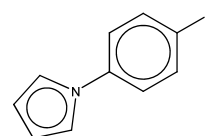
12F2 $\Delta T_m = 1.25^\circ\text{C}$



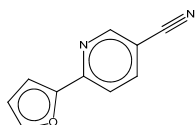
12F3 $\Delta T_m = 2.5^\circ\text{C}, 1^\circ\text{C}$



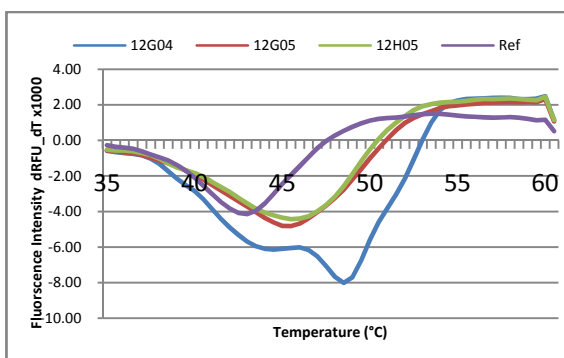
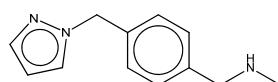
12F7 $\Delta T_m = 1^\circ\text{C}$



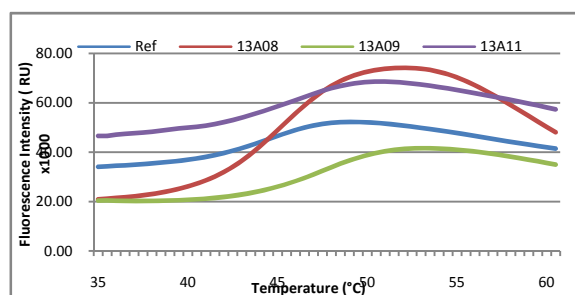
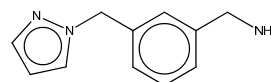
12G04 $\Delta T_m = 1.25^\circ\text{C}$



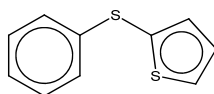
12G05 $\Delta T_m = 2.5^\circ\text{C}, 2^\circ\text{C}$



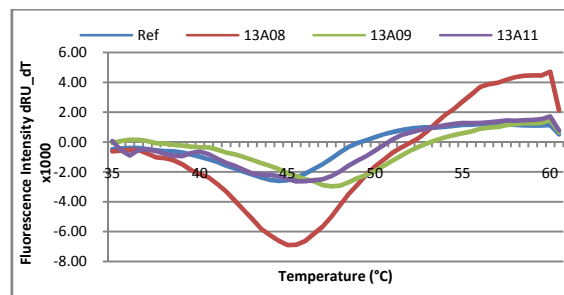
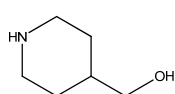
12H05 $\Delta T_m = 2.5^\circ\text{C}, 2.5^\circ\text{C}$



13A8 $\Delta T_m = 0.75^\circ\text{C}$



13A9 $\Delta T_m = 3^\circ\text{C}$



13A11 $\Delta T_m = 1.25^\circ\text{C}$

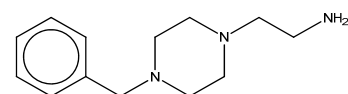
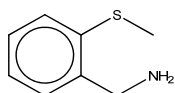
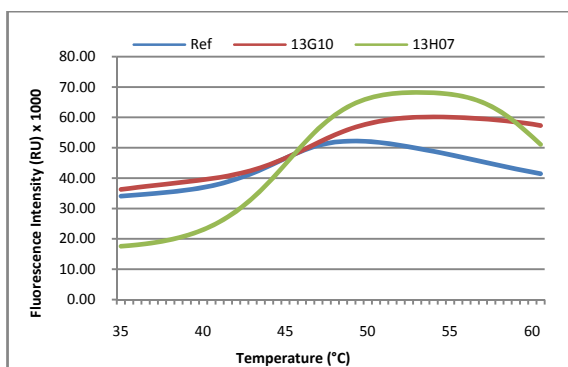
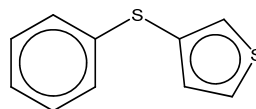
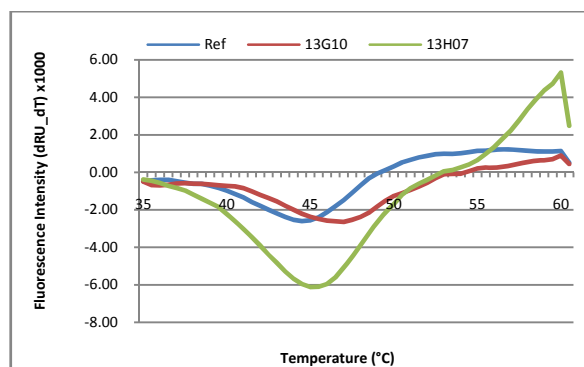


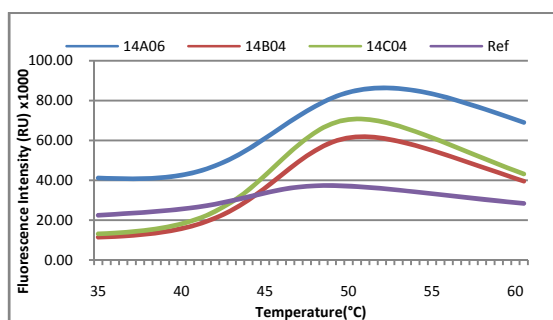
Figure 3.5 The thermal unfolding curves of 12F2, 12F3, 12F7, 12G4, 12G5, 12H5, 13A8, 13A9 and 13A11. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.



13G10 $\Delta T_m = 2.5^\circ\text{C}, 1.5^\circ\text{C}$



13H07 $\Delta T_m = 0.75^\circ\text{C} \text{ } .0.5^\circ\text{C}$



14A06 $\Delta T_m = 1.5^\circ\text{C}$

14B04 $\Delta T_m = 1.75^\circ\text{C}$

14C04 $\Delta T_m = 1.75^\circ\text{C}$

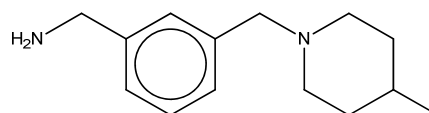
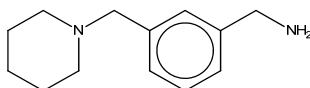
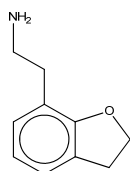
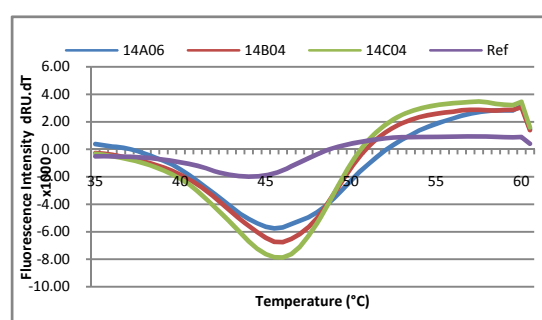
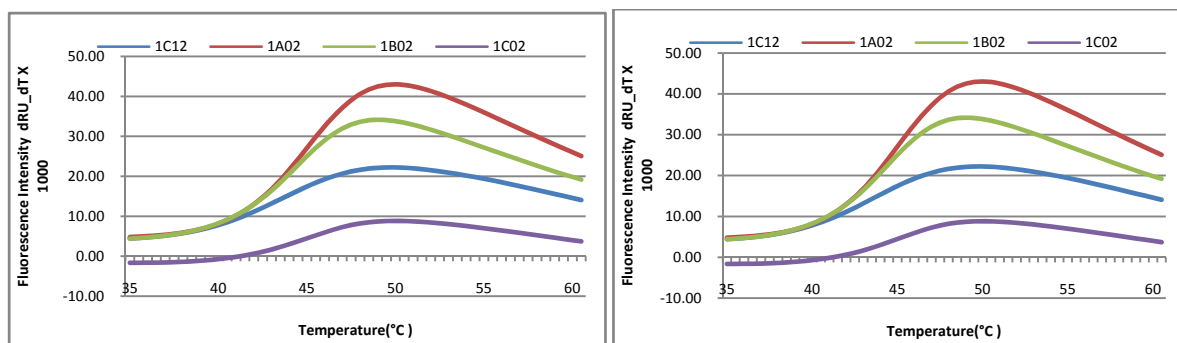


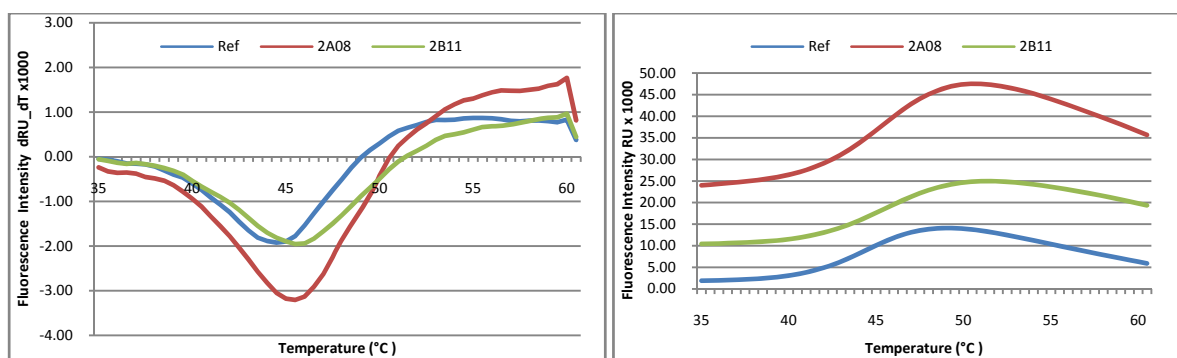
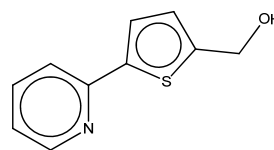
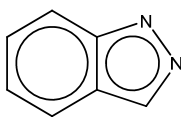
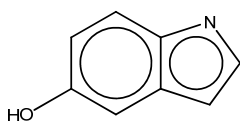
Figure 3.6 The thermal unfolding curves of 13G10, 13H07, 1406, 14B04, 14C04. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.



1A02, $\Delta T_m = 1.5^\circ\text{C}$

1B02, $\Delta T_m = 1^\circ\text{C}$

1C02, $\Delta T_m = 1^\circ\text{C}$



2A08 $\Delta T_m = 1^\circ\text{C}$

2B11 $\Delta T_m = 1.5^\circ\text{C}$

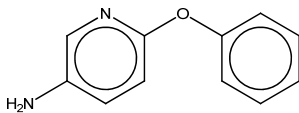
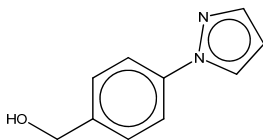
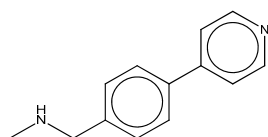
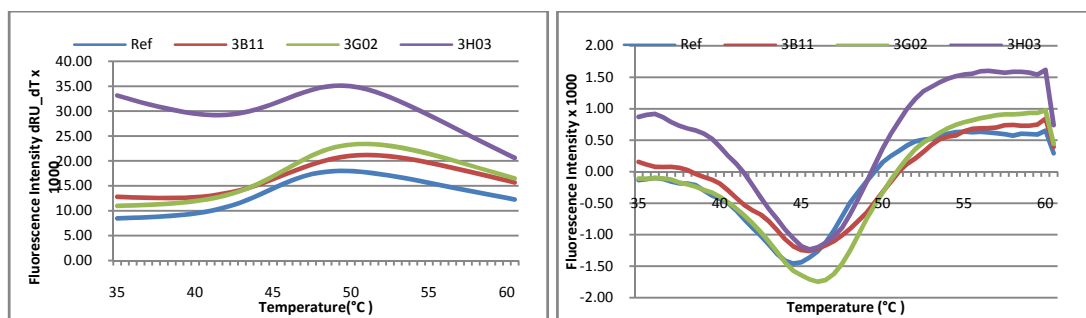
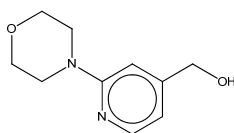


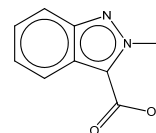
Figure 3.7 The thermal unfolding curves of 1A02, 1B02, 1C02, 2A08, 2B11. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curves.



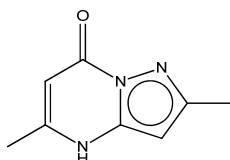
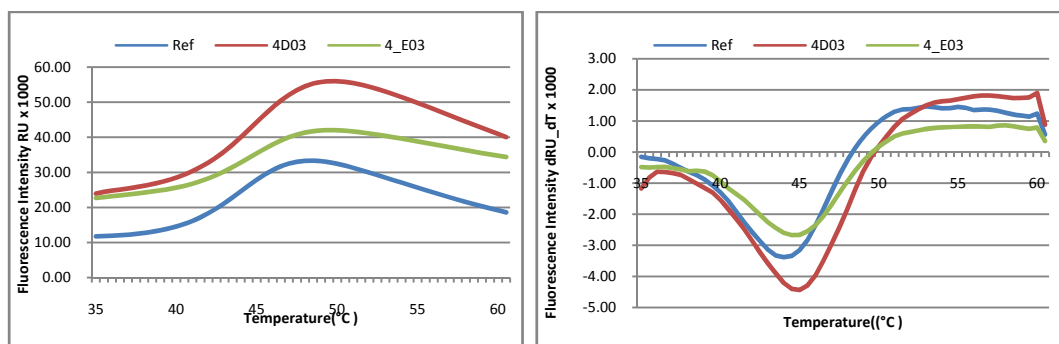
03B11 $\Delta T_m = 0.75^\circ\text{C}$



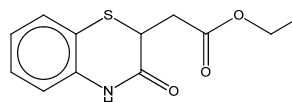
03G02 $\Delta T_m = 1.5^\circ\text{C}$



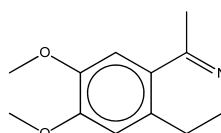
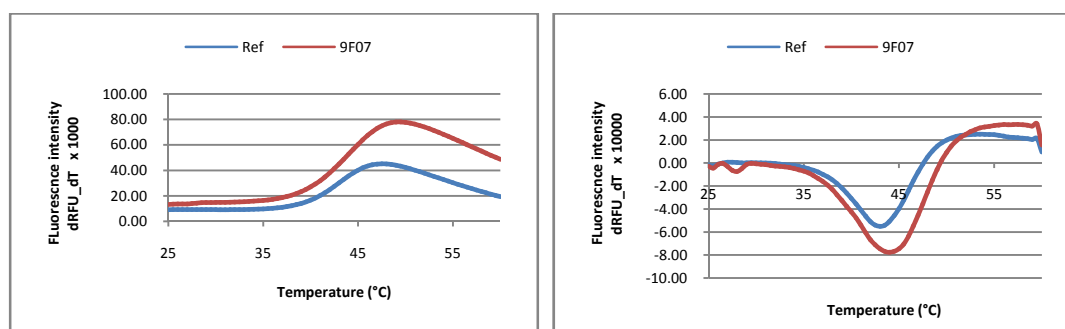
03H03, $\Delta T_m = 1^\circ\text{C}$



04D03 $\Delta T_m = 0.75^\circ\text{C}$



04E03 $\Delta T_m = 0.75^\circ\text{C}$



9F07 $\Delta T_m = 1^\circ\text{C}$

Figure 3.8 The thermal unfolding curves of 3B11, 3G02, 3H03, 4D03, 4E03 and 9F07. Figures on the left hand represent the native curves. Figures on the right represent the corresponding derivatives of the melting curve

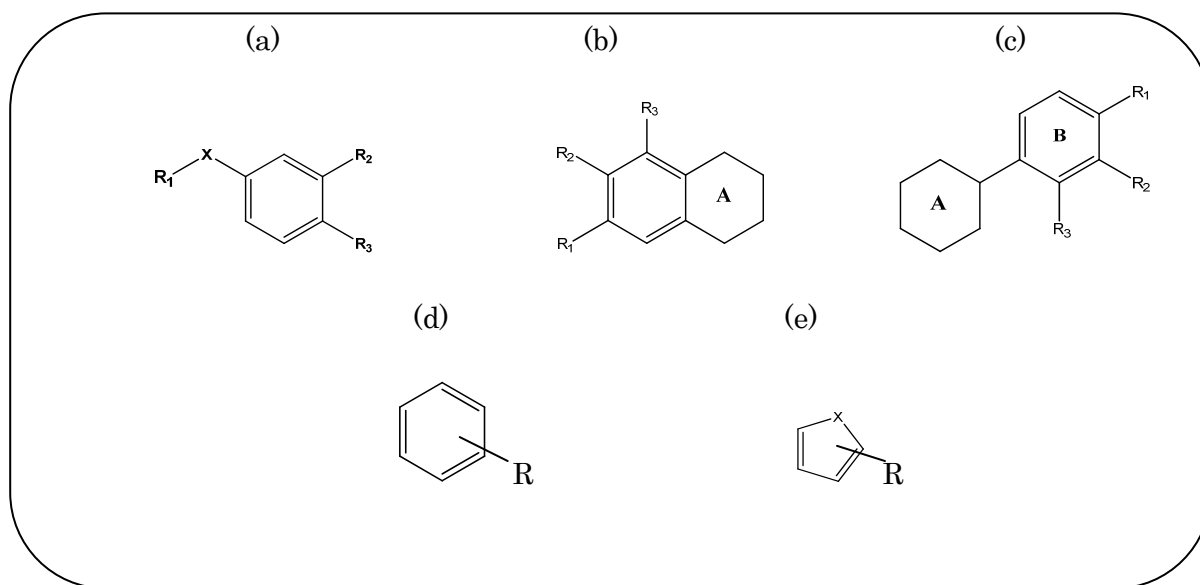


Figure 3.9 Classification of fragment hits according to distinctive chemical structure. Five main classes were found to bind by fluorescent-based thermal shift screening. These include: (a) benzyl derivatives, (b) fused bicyclic rings, (c) biaryl compounds, (d) phenyl derivatives, and (e) 5-membered heterocyclic rings.

The first subgroup consists of 13 fragments with a common phenyl ring linked through one atom to a heterocyclic 5- or 6- or even 7-membered ring, which could be either aliphatic or aromatic. The preference for binding to this scaffold could be attributed to a certain degree of flexibility inferred by the one-atom bridge. Although the ΔT_m values correlate with the binding affinities, ranking and comparing should not be solely based on ΔT_m values, as affinity is dependant on binding enthalpy as described in the following equation (Lo *et al.*, 2004):

$$K_L^{T_m} = \frac{\exp\left\{-\Delta H_u^{T_0} / R\left[\frac{1}{T_m} - \frac{1}{T_0}\right] + \Delta C_{pu}^{T_0} / R\left[\ln\left(\frac{T_m}{T_0} + \frac{T_0}{T_m} - 1\right)\right]\right\}}{[L_{T_m}]}$$

$K_L^{T_m}$ = ligand association constant at T_m

T_m = midpoint of protein unfolding transition in presence of ligand

T_0 = midpoint of protein unfolding transition in absence of ligand

$\Delta H_u^{T_0}$ = enthalpy of protein unfolding in the absence of ligand at T_0

$\Delta C_{pu}^{T_0}$ = change in heat capacity on protein unfolding in the absence of ligand

$[L_{T_m}]$ = concentration of free ligand at T_m

R = gas constant

However, as the binding enthalpy data was unavailable for any of the fragments, ranking was conducted among fragments of the same chemical structural classes assuming they have the same binding enthalpy. The compound that contained a 7-membered 1,4-diazepan ring was found to have the highest ΔT_m within this group, suggesting that the 1,4-diazepan ring could be a favourable ring system.

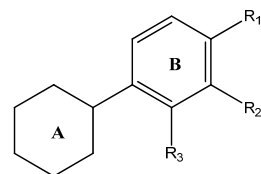
The second subgroup consists of 6 bicyclic fused rings where the phenyl ring is the common feature. According to the ΔT_m values, they have comparable ΔT_m and it was difficult to rank them. It appeared there is no preference for a particular bicyclic pair of rings.

The third subgroup consists of 14 biaryl compounds, where at least one ring is a phenyl or pyridyl ring. These compounds are similar to biphenyl systems, which were found to be the most preferable moiety to bind to proteins as shown in a previous study (Hajduk *et al.*, 2000). The biaryl ring systems may behave like biphenyls, which can be involved in many interactions with proteins due to their flexibility that allows them to accommodate to protein surfaces (Hajduk *et al.*, 2000). This could explain the higher hit rate than other chemical scaffolds.

The phenyl rings was found to be the least preferred moiety that binds to proteins as shown previously (Hajduk *et al.*, 2000). However, the fourth subgroup includes 10 phenyl derivatives and a piperidine ring. The size and shape of phenyl and 6-membered rings could be more suitable for binding to shallow protein surfaces and targeting protein-protein interfaces that tend to be more lipophilic, even with less interactions than larger and more complex ring systems (Figure 3.10).

The final subgroup is the smallest and least complex, consisting of 5-membered rings. It includes only 5 compounds, the imidazole ring being the most favourable. These small rings might be mimicking the side chain of histidine amino acid and capable of binding to shallow and flat pockets across the interface (Figure 3.11).

Table 3.4 A table of the third subgroup, ring (B) is a phenyl ring unless mentioned otherwise.



A	R ₁	R ₂	R ₃	dTm
	F	H	H	1/1
	-CN	H	H	0.75/1
	-CH ₂ OH	H	H	0.75/1
				0.75/0.5
				0.85 /0.5
	H	NH ₂	H	1/1
	H	H	NH ₂	1.5
	H	H	-OCH ₃	1.25/2

A	R ₁	R ₂	R ₃	B	dTm
	-CH ₂ OH	H	H		2.25/ 1
	F	H	H		1/ 0.5
	H	H	H		0.75
	-CF ₃	H	H		0.75/ 1.5
					2.25
					3.75/1.5

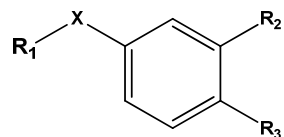


Table 3.5 First subgroup of benzyl derivatives

R_1	X	R_2	R_3	ΔT_m	R_1	X	R_2	R_3	ΔT_m
	C	-CH ₂ NH ₂	H	2.5 °C		C	H	H	1.75(1.5) °C
	C	NH ₂	H	2.75(2.5) °C		C	H	H	2.5(2) °C
	C	H	-CH ₂ NH ₂	2.5(1)°C		C	H	H	1.25°C
	C	-CH ₂ NH ₂	H	2.5(2.5) °C		C	H	H	2 °C
	C	H	-CH ₂ NHCH ₃	2.5(2) °C		S	H	H	0.75(0.5) °C
	C	H	-CH ₂ NH ₂	1.75(1.5) °C		O	H	H	0.75(0.5) °C
	C	-CH ₂ NH ₂	H	3(3.5) °C		C	H	H	1.75(1.5) °C
						C	H	H	2.5(2) °C

Table 3.6 Second group of bicyclic derivatives

A	R ₁	R ₂	R ₃	dTm
	-OCH ₃	-OCH ₃	H	1.5/1.5
	-OCH ₃	-OCH ₃	H	1 / 1
	-CH ₂ NHCH ₃	H	H	1 / 1.5
	-CH ₂ NH ₂	H	H	1.75
	H	H	OH	1.75/1
	H	H	-CH ₂ CH ₂ NH ₂	1/2

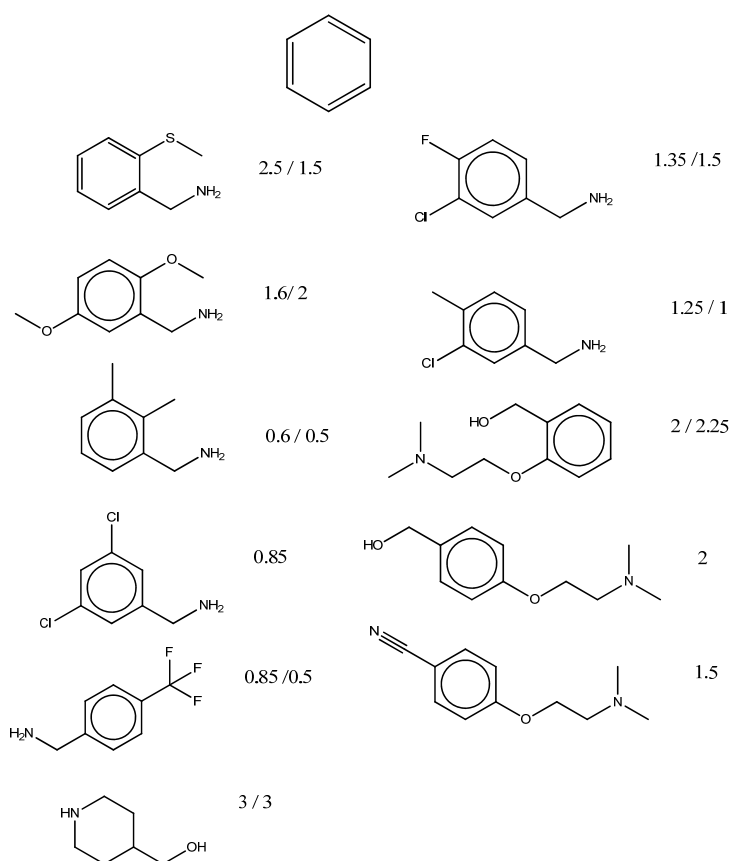


Figure 3.10 The fourth subgroup of phenyl derivatives and the corresponding ΔT_m values.

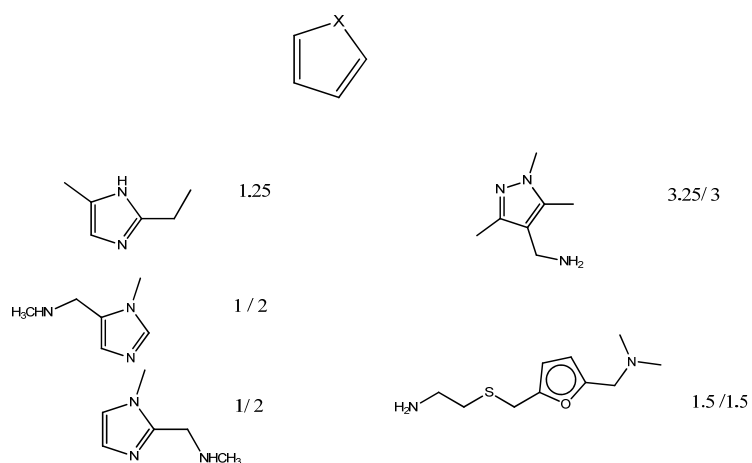


Figure 3.11 The fifth subgroup of five-membered ring derivatives and the corresponding ΔT_m values.

The second group, which included those fragments that destabilised the protein, can be further sub-grouped according to common functional groups rather than common scaffolds, which were not observed amongst the positive hits. The first subgroup was found to contain a carboxylic acid functional group (-COOH), attached to aliphatic and aromatic heterocyclic rings, whereas it constituted 43.8 % of this group. None of the positive hits had a -COOH group, suggesting that it could have a destabilising effect. The -COOH group has previously been found to bind to various protein targets, particularly to DNA- or RNA- binding proteins (Hajduk *et al.*, 2000). In addition, the -COOH group is a common “chemical handle” that is designed for further chemical elaboration of small fragments. However, the carboxylic acid containing compounds will likely be ionised at pH 8.5, the pH at which the screening is carried out and the electrostatic potential of the ankyrin domain interface is negative. The negatively charged -COOH could destabilise the protein.

The second subgroup, numbering seven compounds, had a trifluoromethyl group, which often leads to substituted compounds being strong acids. It is incorporated in many drugs and pharmaceuticals, as it is capable of replacing the methyl and chlorine groups (Yale, 1959). It is a highly electronegative group that lies between chlorine and fluorine and causes destabilisation of the protein (True *et al.*, 2003). Although some of positive hits had the less electronegative chlorine group.

In addition to those groups that were found to destabilise the protein, there appear to be particular ring systems that are not favourable for binding. Some of the molecules that caused negative ΔT_m contained morpholine, thiazole and oxygen containing heterocyclic rings, whereas these ring systems were mostly absent from the positive hits.

3.3 Conclusions

The quality of positive and negative hits depends on the design of the fragment library, the nature of the target and the sensitivity of the technique used in screening. Although having compounds that allow for further optimisation and elaboration by introducing “chemical handles” enhances the ease of moving from hit to lead, it has some limitations. As has been found here, the –COOH or nitrile were unfavourable groups and precipitated or destabilised the protein.

Targeting protein-protein interactions may require a more tailored fragment library of scaffolds similar to amino acid structures that could mimic peptides or proteins. The hit rate is also influenced by the nature of the screening method; the ability to identify false-positives and false-negatives would minimise such variations. The fluorescent-based thermal shift assay used here, is greatly affected by the fluorescent quenching properties of the fragments.

Although this screen could be considered a yes/no answer screening method, especially in the absence of binding enthalpy values, it can still give an indication of the type of molecular scaffolds that the protein target prefers to bind. These positive hits could be of no significance, as they might be binding in locations other than the binding sites. However, they can still help in identifying functional groups that should be avoided in hit-to-lead optimisation.

Chapter 4

Structural Characterisation of Ankyrin-Fragment Complexes

4.1 Introduction

Structure-based drug design (SBDD) relies to an increasing extent on information obtained from crystallography. It comes as no surprise that the increase in the number of structures in the Protein Data Bank is associated with an increase in the number of protein-ligand structures. Crystallography has become an integral part of many drug discovery projects, and that does not only include later stages of lead optimisation, but also protein structures solved by X-ray crystallography used for target assessment and validation. In addition, fragment screening has employed crystallography for picking up hits in preliminary screening, an approach that has been pioneered by Abbott and Astex therapeutics.

Protein-ligand interactions derived from crystal structures assist in developing hits to leads and drug-like molecules through an iterative process. The early analysis of protein-ligand structures facilitates the design of easily synthesised scaffolds and the elimination of toxicophores. Early apo-protein crystal structures permit optimisation of the crystals in order to provide a system amenable to SBDD.

4.2 Optimising a crystallographic system for SBDD

Selection of a successful SBDD system starts with identification of the protein form that truly represents the biological target. Ideally, full-length human proteins are the targets of choice. However, large amounts of protein are required for iterative SBDD steps and in poorly expressed proteins a

miniaturisation of the protein can be the solution, for example by utilising catalytic or receptor domains for flexible multi-domain proteins. Homologues that show high expression levels and high sequence similarity at the site of interest provide an alternative approach. In some cases, mutations in homologous proteins can be designed to mimic a human binding site creating a humanised form, or proteins can be engineered to enhance solubility and stability. The production of a reproducible form of the protein using such approaches can avoid batch-to-batch variation that is unfavourable in long repetitive SBDD processes.

Although obtaining crystals is often the bottleneck for elucidating protein structures, this does not seem to have been the case in drug discovery, possibly because most companies have worked on well characterised families of protein kinases, phosphatases, aspartic proteinases and so on. The quality of the structure is critical for moving from a hit to lead and the evaluation of crystal structures for drug discovery has been widely discussed (Anderson, 2003). The resolution should be better than 2.5 Å, to allow correct placement of residues and atoms of proteins and ligands in electron density maps. This is important for characterising the main interactions based on accurate distances. Crystal structures of poorer resolutions can be detecting the presence of ligands or other molecules that are in the crystallisation conditions. However, these structures should be assessed carefully if they are to be used for drug design.

The R factor and R_{free} value should not exceed 28% and preferably 25%. The conformation of the ligand molecules should be evaluated as well; the bond length deviations should not exceed 0.015 Å, and bond angles deviations not more than 3° (Anderson, 2003). The thermal motions/disorder of the atoms is indicated by the B-factor or temperature factor. The value of the B-factor should not exceed the average for the molecule; otherwise it could indicate an error in the atomic positions (Anderson, 2003). There is often an assumption that the protein model is correct without realising that it is the personal interpretation of the data by crystallographers. Following the recent retraction of crystal structures (Chang *et al.*, 2006) there has been a requirement by the PDB (Protein Data Bank) to

deposit structure factors together with coordinates, so allowing validation of structures in the PDB by non-authors. Even with high-resolution structures, some regions of poor electron density can be interpreted in different ways. This is particularly important if these regions are interesting binding or catalytic sites.

Another limitation that could be found even with structures of reasonable resolutions is the assignment of nitrogen and oxygen in asparagine and glutamine residues. The placement of these atoms in these residues could be difficult as it is not possible to distinguish by the electron density. In addition, assigning the correct tautomeric state of histidine and other ligands is problematic. However, judgments can be made according to the hydrogen-bond environment after adding solvent and water molecules. The same problem occurs with ligands. Placement of nitrogen in asymmetric pyridine or pyrrole rings is often problematic and can affect the interpretation of possible interactions and the way to proceed with elaboration and optimisation.

The crystallisation conditions are often different from those of other biophysical screening methods and biological assays. For instance, the pH can differ and affect the conformations of both protein and ligand. Even small differences in pH can alter the ionisation states of ligands and protonation of the protein, which in turn affects the binding mode or even inhibits binding altogether. It is important to consider the compatibility of the methods used, though it is often difficult to achieve since every technique may require optimisation of screening conditions independently.

In this chapter, I describe the crystallisation of the ankyrin domain, my attempts to produce protein-ligand structures, and the results and implications of these experiments.

4.3 Materials and methods

4.3.1 Crystallisation and optimisation

The protein was concentrated to 10 mg/ml in a buffer constituting of 50mM Tris pH 8 and 50mM NaCl. The ankyrin domain was crystallised in 0.9 M (NH₄)₂HPO₄, 0.2 M NaCl and 0.1 M imidazole (pH ~8.5) as precipitant as reported previously (Ehebauer *et al.*, 2005) and crystallisation was optimised around these conditions, which proved to be reproducible.

- a) The ankyrin domain was screened using different crystallisation screens: Classics, SM1, PEG, PH clear I, PH clear II and PEG I. Crystals were formed in various conditions and screening grids around initial conditions were designed for optimising these conditions.
- b) The ankyrin domain was prepared in another buffer 10 mM CHES pH 8.5, 400 mM NaCl and screened again using PEG1, PH clear I, PH clear II.

For microseed matrix-screening, the seed stock was prepared by crushing crystals formed in conditions 0.7 M (NH₄)₂HPO₄, 0.2 M NaCl, and imidazole pH 8. This was reconstituted in 50 µl of the crystallisation buffer, mechanically homogenised by vortexing, and the seed stock was frozen at -20 °C. Seeding was carried out by adding 0.2 µl of screening solution, 0.1 µl of seed stock and 0.3 µl of protein at 10 mg/ml.

4.3.2 Soaking and Co-crystallisation

i) Crystal Soaking:

Different approaches for soaking were used:

- a) Solutions of the fragments in the crystallisation buffer were prepared at different concentration ranges; 10 mM, 20 mM, 50 mM, 100 mM, 200 mM and 500 mM.
- b) Soaking times varied between 30 seconds, 10 minutes, 30 minutes, 1 hour, 2 hours, 4 hours, 24 hours, one week and up to 3 weeks.

c) Stepwise soaking was carried out by gradually increasing the fragment concentration in order to minimise the effect of DMSO.

d) Fragment solutions were prepared in solutions of the primary screen, the crystals were transferred by gradually changing the conditions of the original crystallisation solution to the final condition 10mM CHES pH 8.5 and 400 mM NaCl.

e) Cross-linking of the crystals with 25% glutaraldehyde:

Crystals were transferred to 1 ul drop containing the crystallisation buffer on a cover slip. A microbridge containing 25% glutaraldehyde was placed in a well. The cover slip was then placed over the well to expose the crystal to the glutaraldehyde via vapour diffusion for 1 minute. The crystals were then transferred to fragment solutions of a stabilising solution containing higher precipitant and lower salt content. The fragments were at a concentration of 100mM and soaked overnight.

f) The solutions were prepared with and without cryoprotectant. Glycerol (25%) was used as a cryoprotectant.

g) Stabilising solutions were prepared by increasing salt concentrations or adding xylitol, which was added at a concentration 4% and the fragment concentration was 50 mM.

ii) Co-crystallisation:

The protein was incubated with the fragment at different concentrations: 50mM, 100 mM and 250 mM overnight at 4 °C and crystallised using the original conditions.

4.3.3 Data collection and processing

Several data sets were collected at various synchrotrons at Diamond, ESRF and ALS. Datasets of 120 images were collected to obtain completeness. The images were indexed and integrated by MOSFLM, then scaled using SCALA program, a part of the CCP4 package. The structures were solved by molecular replacement using the program PHASER and the coordinates of the Notch ankyrin domain (PDB 1YYH) as a probe. The σ A-weighted $2F_o - F_c$ and $F_o - F_c$ electron-density maps were visualised to allow rebuilding and refitting using Coot.

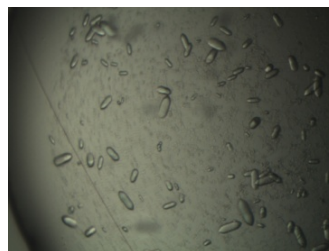
4.4 Results and discussion

4.4.1 Crystallisation

Structures of protein-ligand complexes can be obtained by soaking or cocrystallisation. Protein crystals comprise 30% to 80% solvent (Matthews, 1968), which is found in channels determined by the crystal lattice network (Vilenchik *et al.*, 1998). The ability to soak ligands depends on the diffusion of compounds in the solvent channels. The availability of crystal forms with different space groups is helpful in overcoming problems arising from occlusion of binding sites by intermolecular interactions in the crystal lattice. The ankyrin domain was originally crystallised in the space group P6₅ in two different conditions (Ehebauer *et al.*, 2005; Nam *et al.*, 2006). In this crystal form, the interface that is involved in the interaction with CSL and MAML is exposed to the solvent channels and is not blocked by the crystallographic contact residues (Figure 4.2). This crystal form is likely compatible with diffusion of fragments to the protein interaction sites. However, the lack of distinctive binding pockets to target and knowledge of any known or natural ligands make it hard to predict where small fragments bind. The ankyrin domain was screened again for different crystal conditions in search of other space groups that could improve soakability of the crystals.



Ank in 10mMCHES 400mMNaCl
25 %(w/v) PEG2000 MME(PEG1)



Ank in 50mM Tris pH 8 50 mM NaCl
15% (w/v) PEG 20,000 (PEG 1)



Ank in 50mM Tris pH 8 50 mM NaCl
1M LiCl 0.1 M Bicine pH 9 (PH Clear II)



Ank in 50mM Tris pH 8.5 50 mM NaCl
0.05 MgCl₂, 0.01 M MES pH 6.5, 10% isopropanol,
5% PEG 4000 (Matrix seeding)

Figure 4.1 Crystal hits of the ankyrin domain found in different screens.

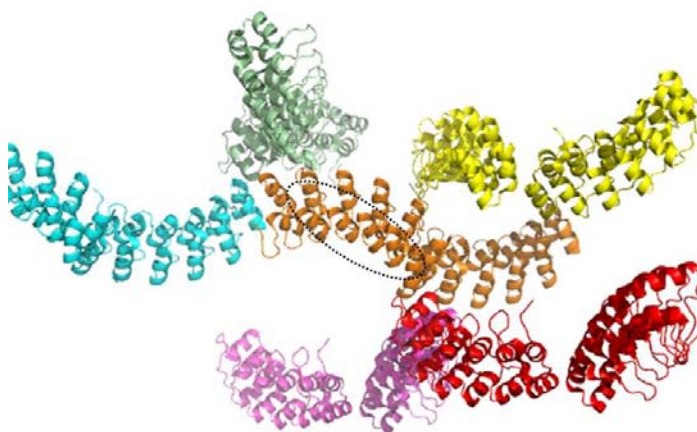


Figure 4.2 A cartoon representation of the ankyrin domain showing symmetry related molecules of the crystal structure (PDB 1YYH). The dotted circle highlights the interface which the ankyrin participates in protein-protein interactions in the ternary transcription complex with CSL and MAML. The crystal contacts in this space group with other symmetry related molecules do not interfere with the binding interface. (The symmetry molecules that are only in contact with ankyrin molecule (in orange) are shown for simplicity).

Three approaches were used for screening:

- a) The ankyrin domain was screened in the original buffer of 50 mM Tris pH 8 and 50 mM NaCl and also in
- b) 10 mM CHES pH 8 and 400 mM NaCl, which was found to stabilise the protein as it has been demonstrated in the buffer screening assay using the fluorescent-based thermal shift screening method. A more stable form of the protein would improve the quality of crystals.

c) Microseed matrix screening is an established method that has been used to increase the crystallisation hit-rate (D'Arcy *et al.*, 2007). Seed stocks were prepared by mechanical homogenisation through vortexing. Introducing seeds in the crystallisation conditions can induce a nucleation event. For the ankyrin domain, seeds were prepared from crystals that were formed in condition 0.7 M $(\text{NH}_4)_2\text{HPO}_4$, 0.2 M NaCl, and imidazole pH 8.

Crystals were formed in various conditions (Figure 4.1); some of which appeared to be of different space groups judging by their crystal morphology. However, it was difficult to reproduce these conditions and hence determine their space groups. And in an iterative SBDD process, it is essential to be able to easily reproduce crystals.

Crystals formed in conditions 0.7 M $(\text{NH}_4)_2\text{HPO}_4$, 0.2 M NaCl, and imidazole pH 8 and in 16 % (w/v) PEG 8000, 0.2 M NaCl and 0.1 M CAPS pH10.5 were used for soaking.

4.4.2 Soaking and Co-crystallisation

The electron density of the ligand or fragment is visible if the ligand occupies at least 30% of the binding sites as suggested in an earlier study (Wu *et al.*, 2001). It is important to understand how the protein binds with the ligand to reach

equilibrium. The fraction Y of protein (P) that binds with the ligand (L) to form a protein-ligand complex (PL) with a dissociation constant, K_d , can be derived from the following (Danley, 2006):

$$K_d = \frac{[P][L]}{[PL]} \quad (1)$$

$$Y = \frac{[PL]}{[P] + [PL]} \quad (2)$$

From the previous equations (1) and (2), Y can be expressed as a function of K_d as following:

$$Y = \frac{[L]}{K_d + [L]}$$

It can be concluded that the dissociation constant and the ligand concentration are the factors that determine the binding equilibrium. This should be considered when designing the soaking experiments regarding soaking times and ligand concentrations in the soaking solutions. This would be primarily affected by the ligand solubility. A major problem for these hydrophobic small fragments is the solubility in aqueous buffers where most of the screenings are carried out. In crystal soaking, it is the fraction of the compounds that is soluble that actually can diffuse and move freely in the solvent channels.

The fragments are solubilised in 100% DMSO at 1M concentration as a stock solution. Solutions of fragments were prepared by dilution to various concentrations with various DMSO amounts. The concentration of DMSO should not exceed 20%, and typically not more than 10%. The DMSO can improve the solubility of some of the fragments; however DMSO can have deleterious effects on the crystal packing of the crystal lattice leading to reducing the quality of crystals or can even cause the crystals to crack and dissolve. The soaking was monitored over a time-course and step-wise soaking was used by gradually increasing the fragment concentration when higher concentrations of fragment were used.

Fragments that were soluble in the crystallisation buffer were directly added to the soaking solutions without solubilising in DMSO. The crystals used for

soaking were formed in two conditions; one contained PEG 8000 which can improve the fragment solubility.

The crystals were soaked with fragments that were hits in the screening by the fluorescent-based thermal shift assay. This screening was carried out in a buffer containing 10mM CHES pH 8.5 and 400 mM NaCl, which differs from the crystallisation conditions. The binding of fragments is affected by their protonation and conformation states, and by the constituents found in solution. The fragments were dissolved in the screening buffer to maintain the same conditions that seem to be favourable for binding. The crystals were transferred from their original crystallisation buffer to the fragment solution by gradual buffer exchange in order to keep the crystal intact. This could maintain the same state of the fragments, but the assumption is that the protein is maintained in the conformation determined by the crystal contacts.

Different “stabilising” solutions are used to stabilise the crystal and protect it from damage. These solutions often contain the same composition as the crystallisation buffer but with increased concentration of one or more of the reagents. Cross-linking with glutaraldehyde is also used to stabilise crystals against mechanical stress during freezing (Lusty, 1999). It was used here to stabilise the crystal against the damaging effect that can be caused by introducing an organic solvent as DMSO or by the fragment itself.

Crystals can also be stabilised by the addition of xylitol to the fragment solution in a concentration 2-5%; xylitol can improve the solubility of the fragment as well.

Co-crystallisation is another method used to obtain protein-ligand structures where the ligand is incubated with the protein prior to crystallisation. This may be suitable for ligands with poor solubility or proteins that tend to aggregate. However, the ligand can alter the solubility of the protein, requiring screening for conditions other than the original. Sometimes each individual inhibitor requires screening making it rather tedious, but it has the advantage of allowing for conformational changes that might occur during ligand binding. The

conformational changes are restricted when soaking, particularly when another crystal form cannot be found. If possible, it is preferable to obtain structures by both soaking and co-crystallisation. The binding mode in solution obtained from co-crystallisation may vary than that in the soaked form.

As described above, different soaking and co-crystallisation methods have been utilised in an attempt to elucidate the binding mode of the fragment hits previously identified in the initial screen (Chapter 3). One problem that was encountered during the soaking was the cracking and dissolving of the crystals, which greatly affected the crystal quality. Some crystals did not diffract at all and others diffracted at very low resolution. The cracking of the apo-crystals could be due to a deleterious effect of DMSO or an indication of conformational change induced by ligand binding that is incompatible with the crystal packing

Several datasets were collected. Data from crystals that diffracted at a resolution worse than 2.7 Å was not collected. The resolution of the collected datasets ranged between 1.9 Å and 2.7 Å and the structures were solved as described previously. The $2F_o - F_c$ maps were visualised using Coot software to search for any unmodelled electron density. The binding modes of only two fragments were elucidated and there was no evident electron density for the other fragment hits. This could be due to the low occupancy of the weakly binding fragments

Table 4.1 Data collection and refinement statistics ankyrin-fragment structures. Values in parenthesis are for the highest resolution shell

	Fragment 12C05	Fragment 9F07
Space group	P6 ₅	P6 ₅
Unit cell	a=b=97.94 Å c=109.44 Å α = β = 90°, γ = 120°	a=b= 97.43 Å c= 110.23 Å α = β = 90°, γ = 120°
Resolution	21.20 2.53 Å (2.40- 2.40 Å)	33.69 -2.42 Å (2.30 -2.30 Å)
Completeness	99.18 %	99.98 %
R _{sym}	5% (27.5%)	3.5 % (22.20 %)
Average I/σ(I)	6.6	7.3
Number of unique reflections	21935	22060
Wilson plot B factor	34.53 %	50.93 %
Refinement		
R _{cryst}	22.0 %	23.90 %
R _{free}	30.3 %	30.95 %
Molecule per asymmetric unit		
Number of ligand molecules	2	2
Estimated co-ordinate error	1	1
RMSD bonds	0.30 Å	0.31 Å
RMSD angles	0.024Å 2.17°	0.026 Å 2.42°

$R_{sym} = \sum_h |I_h - \langle I \rangle| / \sum_h \sum I_h$, where I_h is the intensity of reflection h and $\langle I \rangle$ is the mean intensity of all symmetry related reflections. $R_{cryst} = \sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$, where F_{obs} and F_{calc} are observed and calculated structure factor amplitudes. R_{free} as for R_{cryst} using a random subset of data excluded from the refinement, 5% of the total dataset was used. Estimated co-ordinate error based on R value was calculated using Refmac

4.4.3 The first crystal structure

The first structure was solved at a resolution 2.5 Å with two molecules in the asymmetric unit. The electron density for the first ankyrin repeat in both chains was very poor; the first 52 residues in chain A and first 51 residues in chain B were unmodelled. The fragment was found located in two different sites (Figure 4.2).

The first binding site for fragment 12C05

The first site of the fragment 12C05 was positioned between the upper helices of repeats six and seven. Although the fragment 12C05 is sandwiched between the two protein molecules in the asymmetric unit, it is held by a number of interactions (Table 4.2). It appears to be stabilised by weak hydrogen bonds, a salt bridge, and a π - π interaction. The weak hydrogen bonds C-H...O are formed between aspartate 2109 and glutamate 2076 side chains. The pK_a values of aspartate and glutamate are 3.9 and 3.3, respectively. That makes them negatively charged at the pH of the crystallisation conditions (pH 8.5). However, at a resolution 2.5 Å it is impossible to assign the negatively charged oxygen atom, so it is assumed that it is the atom that is involved in a hydrogen bond predicted on basis of distances. The weak hydrogen bonds have variable geometries, as the distances are shorter when involved with main chain than with side chain donors and acceptors. The hydrogen-bond distance ranged from ≈ 2.8 Å to ≥ 3 Å (Figure 4.3).

Salt bridges are important interactions that have been observed in many protein-ligand structures. The pK_a for the nitrogen atom in the pyrrolidine ring was not provided by the supplier (Maybridge), so it was calculated using an online website calculator <http://sparc.chem.uga.edu/sparc/>. It was predicted to be 9.36, which means it is positively charged at the crystallisation solution pH. It could be involved in a salt bridge interaction with the negatively charged oxygen of aspartate 2109. It is worth mentioning that these interactions were observed in high salt concentrations, suggesting that these interactions are specific.

Aromatic interactions play an important role in the stabilisation of proteins (Chakrabarti & Bhattacharyya, 2007). The phenyl ring is involved in a face-to-face π - π interaction with the phenyl ring of tyrosine residue 2075 where the distance between the two planes is 3.5Å.

Table 4.2 Summary of the interactions the fragment 12C05 forms at the first site

Type of interaction	Residue involved	Distance
Weak hydrogen bond	D2109	2.76 Å
Weak hydrogen bond	E2076	3.06 Å
Salt Bridge	D2109	3.4 Å
π - π aromatic interaction	Y2075	3.5 Å

The electron density of chain B of the ankyrin domain was weaker and poorer than that of chain A. There was very little electron density of the fragment at the corresponding site in chain B and it was difficult to place the fragment. It is quite common to find only one ligand in monomers of multimeric proteins (Sevcik *et al.*, 1991; Marcio Dias, unpublished). This could be the binding site being blocked by crystal contacts in one monomer, or the affinity being weaker due to different conformational restraints. The relative occupancies would also vary with the technique used, whether it was soaking or co-crystallisation of the ligands.

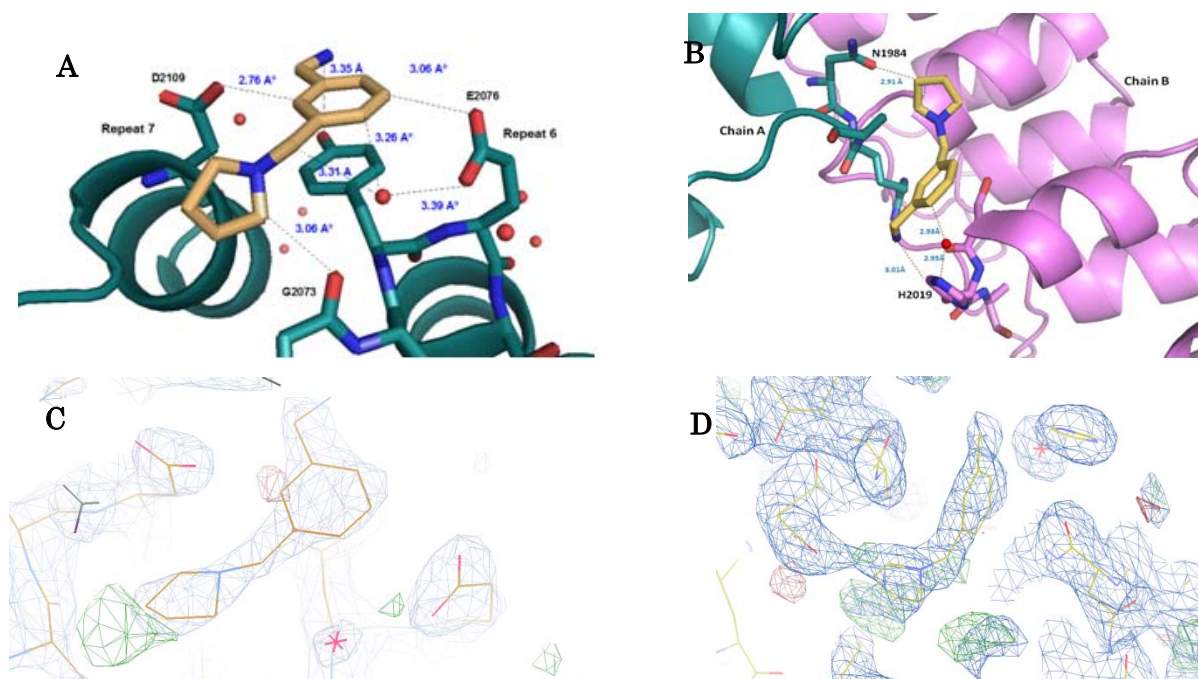


Figure 4.3 Figures showing the two sites the binding fragment 12C05. The interactions are shown in A) with residues D2109, E2076, D2109 and Y2075. B) Interactions of the fragment at the dimer interface N1984, H2019 and H2019. In C) and D) electron density maps $2|F_o - F_c|$ at a contour level of 1σ showing the fragments at the two sites in A) and B) respectively

The second binding site for fragment 12C05:

The fragment 12C05 was found binding in another site between the two protein chains in the asymmetric unit (Figure 4.3B). The interactions are mainly weak hydrogen bonds with residues N1984 in chain A and H2109 in chain B, and one interaction mediated through a water molecule. The carbonyl group of N1984 side chain forms a weak hydrogen bond with a $-C-H$ of the pyrrolidnyl group of the fragment. Another weak hydrogen bond through the alkyl-amino side chain of the fragment formed with $-C-H$ of the imidazole ring of H2104. A water molecule mediates the interaction between $-C-H$ of the phenyl ring of the fragment and the nitrogen atom of the H2019 imidazole side chain. Again, these interactions are suggested by the distances and assumptions made by assigning the nitrogen atoms in both the asparagine and histidine residues (Figure 4.3).

Table 4.3 Summary of the interactions that fragment 12C05 forms at the second site

Type of interaction	Residue involved	Distance
Weak hydrogen-bond	N1984	2.91 Å
Weak hydrogen-bond	H2019	3.01 Å
Water-mediated bond	H2019	2.93 Å - 2.95 Å

4.4.3.1 Implications of complex formation

The crystal structure was superposed with the structure of ankyrin domain in the complex with MAML and CSL (PDB 2F8X). In the first site where 12C05 is binding to chain A, the fragment was found at the interface between ankyrin and MAML (Mastermind). In fact, the pyrrolidine ring was found in the position where the arginine R40 of MAML lies. It seems that the nitrogen atom of the pyrrolidine ring mimicks the guanidine group of the R40. R40 of MAML lies at the interface with ankyrin where it interacts with residues G2073 and D2109 of ankyrin through its guanidine group (Figure 4.4). The pKa of the guanidinium group is 12.48, making it positively charged at pH 8.5 of the crystallisation condition. The positive charge is delocalised as a result of conjugation between the double bond and the nitrogen's lone pair. This allows it to form many interactions, mainly hydrogen bonds. It forms a hydrogen bond with the carbonyl group of G2073, and is involved in an ionic interaction with the side chain of D2109, as indicated by PICCOLO (Chapter 2).

Mutational studies have been carried out to analyse the protein-protein interactions that are involved in the complex-formation (Del Bianco *et al.*, 2008; (Deepti Gupta, unpublished). Single mutation of D2109 to alanine did not disrupt the complex-formation. The authors reached the conclusion that the stabilisation of the complex depends on numerous interactions rather than individual “hotspot” residues. The fragment 12C05 may seem to disrupt a single interaction across the interface, but that may not be sufficient to prevent the formation of the ternary complex. In addition, these fragments are weak binders with low affinities. However, further optimisation of the fragment could improve the affinity.

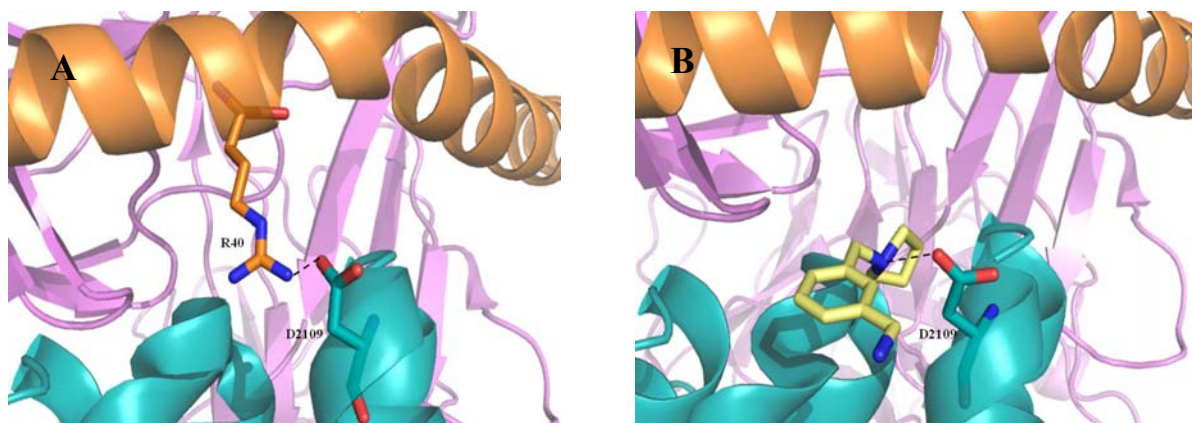


Figure 4.4 Implications of fragment 12C05 for complex formation. A) Crystal structure of the ternary complex shows a salt bridge interaction between R40 of MAML and D2109 of the ankyrin domain (PDB 2F8X) B) Superposed structure of the ankyrin domain with the bound fragment 12C05 shows a salt bridge interaction between the fragment and the same D2109 of ankyrin domain. The ankyrin domain is in teal, MAML in orange and CSL in violet.

4.4.3.2 Implication of dimerisation

The ankyrin structure is a dimer with few crystallographic contacts. There have been conflicting views on the significance of dimerisation of the ankyrin domain. It was assumed that a dimeric ankyrin domain of *Drosophila* was merely a crystallographic dimer; as it was found as a monomer in solutions of different pH values and in different ionic strengths (Zweifel *et al.*, 2003). However, yeast-interaction-trap assays demonstrated that the ankyrin domain in *Drosophila* could be involved in homotypic interactions (Matsuno *et al.*, (1997)). This confirmed earlier studies on the Glp-1 homologue of *Caenorhabditis elegans* that showed a homotypic interaction using constructs encoding ankyrin domain through yeast-two-hybrid assays (Roehl *et al.*, 1996). Dimerisation of human Notch1 ankyrin in solution was also observed from small angle X-ray scattering (SAXS) data (Matthias Ehebauer, unpublished data). Analysis of SAXS data for the Notch RAMANK (molecule consisting of the RAM region and the ankyrin domain) indicated a monomeric form (Matthias Ehebauer, unpublished data). This suggested that dimerisation is characteristic of the ankyrin domain only when expressed on its own and that it was unlikely to be of any biological relevance in complex formation and gene transcription.

However, recent studies have investigated the role of dimerisation in transcription of Notch target genes (Nam *et al.*, 2007). These studies were based on evolutionary conserved regions in CSL-binding sites and on crystal contacts observed between ankyrin molecules. Dual CSL-binding sites in target genes of mammalian homologues were found to be conserved and oriented head-to-head. These sites, known as “sequence-paired” binding sites (SPSs), are separated by nucleotide spacers that vary in length according to species. The crystal-contact residues between ankyrin-ankyrin molecules in the complex structure (Nam *et al.*, 2007) (Figures 4.5C and 4.5D) and in isolation (Ehebauer *et al.*, 2005) (Figures 4.5A and 4.5B) were found to be evolutionarily conserved. These contacts are not involved in interactions with the other molecules of the ternary complex. These observations suggested a cooperative dimerisation assembly model for the Notch transcription complex on the target genes. This model was supported by mutational studies of the ankyrin-ankyrin contact residues and the impact of these mutations on transcription. The model proposes that after assembly of the Notch transcription complex, the complex dimerises on the target Notch genes on “sequence-paired” sites and switches on transcription in a dose-dependant manner.

The contact residues between the two chains are engaged mainly through electrostatic interactions. Residues R1985, K1946 and E1950 were found to be essential for dimerisation and single mutations of these residues led to inhibition of transcription (Nam *et al.*, 2007). The Fragment 12C05 was found at the dimerisation interface binding to residue N1984 adjacent to R1985. The fragment did not interfere with any of the electrostatic interactions that stabilise the two ankyrin-domain chains. However, it is possible to grow the fragment or optimise it to engage in interactions with the guanidino group of R1985. This could potentially inhibit the dimerisation of the assembled complex; and inhibit the gene transcription according to the proposed model.

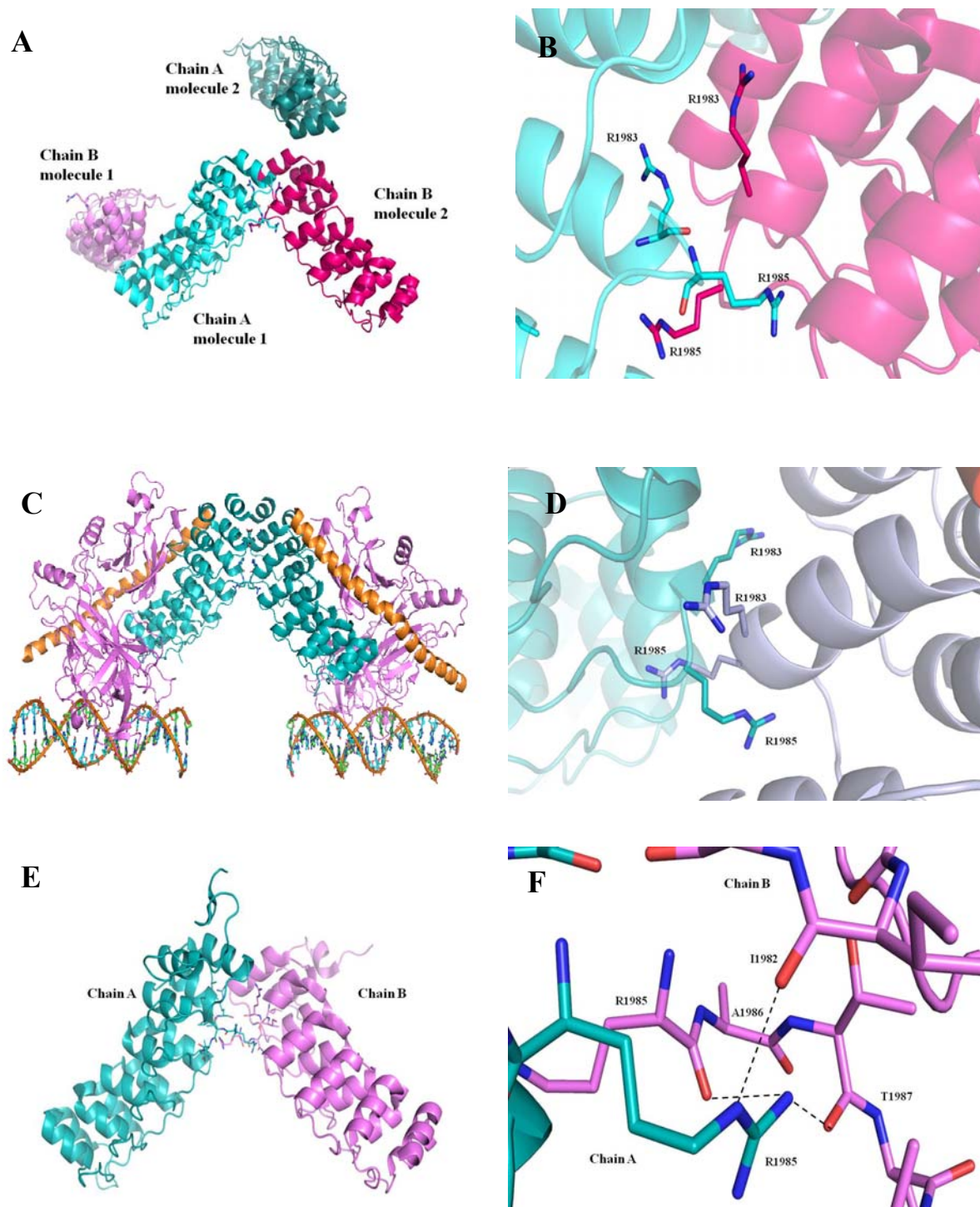


Figure 4.5 Crystal structures of the ankyrin domain both in the isolated structures (PDB 1YYH and PDB 2F8Y) and in complex (PDB 2F8X) show conserved head-to-head interactions through conserved residues of the ankyrin domain. In A) Crystal contact residues between molecules in the asymmetric unit show head-to-head interaction (PDB 1YYH) B) Residue R1985 in one chain of one molecule interacts with the other chain of another molecule in the asymmetric unit (PDB 1YYH). C) Crystal contacts of the ternary complex are involved through the ankyrin domain pseudo-twofold symmetry axis of the crystal (PDB 2F8X). D) The head-to-head interaction of the ankyrin domain in the ternary complex is involved again through residue R1985. E) Another crystal structure of higher resolution of the isolated ankyrin domain (PDB 2F8Y) shows contact residues between the two chains of the same molecule. F) Residue R1985 is involved in hydrogen-bond interactions with backbone carbonyl groups (PDB 2F8Y).

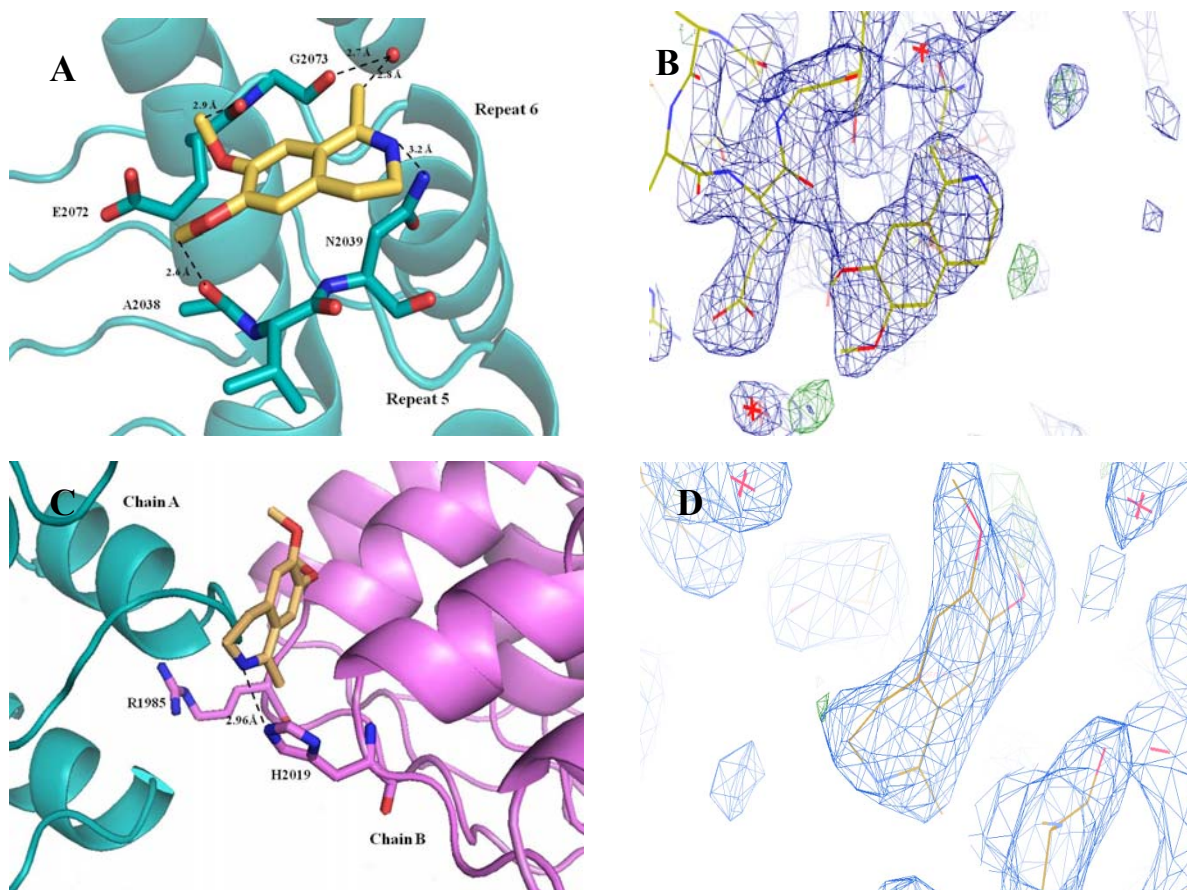


Figure 4.6 A) The crystal structure of the ankyrin domain with fragment 9F07 showing weak hydrogen-bond interactions with residues E2072, G2073, A2038 and N2039 (ankyrin domain in teal and 9F07 in orange) C) The crystal structure of 9F07 at the second site between the two ankyrin chains (Chain A in teal and chain B in violet). B) and D) electron density maps $2|F_o - F_c|$ at a contour level at 1σ showing the fragments at the two sites in A) and C) respectively.

4.4.4 The second crystal structure

The second structure was solved at a resolution $2.5 \text{ \AA}$ and the first repeat was again missing due to its unstructured nature. The electron density of the first 50 residues in chain A and chain B was very poor, making it impossible to model any residues. Similarly, the fragment was found binding in two different locations (Figure 4.6).

The first binding site for fragment 9F07:

The first site was located between the upper helices of repeats five and six (Figure 4.6A). The fragment is bound mainly through weak hydrogen bonds with the carbonyl groups of the main chain of the protein (Table 4.4). One methyl group of the methoxy side chain forms a weak hydrogen bond with the carbonyl

group of main chain amide bond of E2072, and the other forms a hydrogen bond with the carbonyl group of main-chain amide bond of A2038. The nitrogen atom of the dihydroisoquinoline ring of the fragment forms a hydrogen bond with the amino side-chain group of N2040. A water molecule mediates the interaction between the methyl side chain of the fragment and carbonyl group of main-chain of G2073 through weak hydrogen-bonds. The fragment binds at an unusual solvent-exposed site that is too shallow and flat to accommodate small molecules or even smaller fragments. However, it was possible to fit the fragment to the electron density and to derive interactions with the ankyrin domain. The electron density corresponding to the fragment at chain B of the ankyrin domain was very weak and it was again difficult to place the fragment

The second binding site for fragment 9F07:

The second site was found between the two chains. The fragment is involved in only one hydrogen bond with H2019 between the NH of the fragment and the nitrogen of the imidazole ring of H2019. The fragment appears to be stacked between the two chains through hydrophobic interactions.

Table 4.4 Summary of the interactions the fragment 9F07 forms

Type of interaction	Residue involved	Distance
Weak hydrogen-bond	E2072	2.9 Å
Weak hydrogen-bond	A2038	2.6 Å
Weak hydrogen-bond	N2039	3.2 Å
Water-mediated interaction	G2073	2.7 – 2.8 Å
Hydrogen-bond*	H2019	2.96 Å

*This interaction is formed at the dimerisation interface.

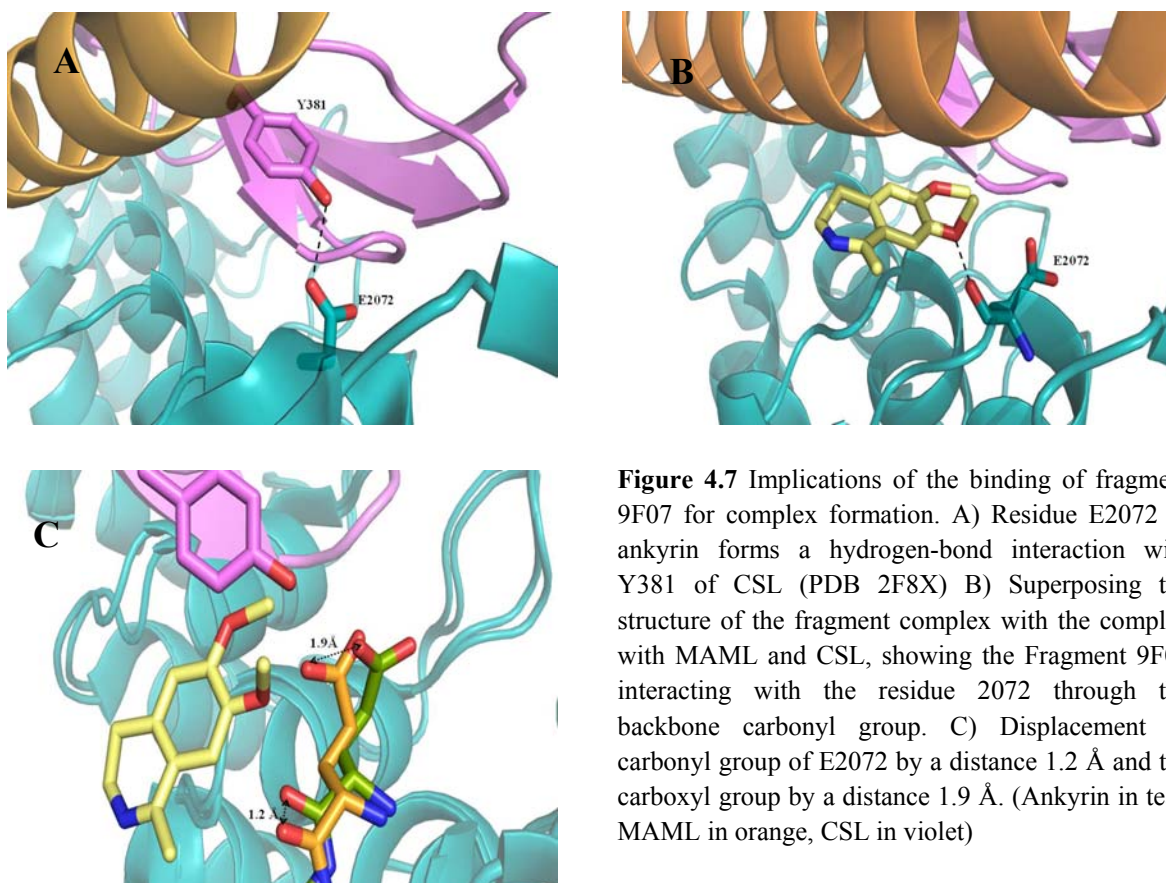
4.4.4.1 Implication of the fragment 9F07 on complex formation

The structure of the ankyrin domain was again superposed with the structure of ankyrin domain in the complex structure (2F8X). The fragment 9F07 lies at the

interface of the ankyrin domain with MAML and CSL. The fragment interacts with residue E2072 through its main-chain carbonyl group. This residue is involved through its carboxylic side chain with the hydroxyl group of Y381 of CSL. The fragment does not interfere directly with this interaction. There is no great conformational change induced by the fragment when compared to both the ankyrin structure in isolation (core rmsd = 0.564 Å) and in complex (core rmsd = 0.636 Å). The ankyrin domain itself does not undergo any change when it forms the ternary complex (Nam *et al.*, 2006). However, the flexible carboxylic side chain of E2072 residue is displaced by a distance 1.9 Å distance which makes it about 3.3 Å further from the –OH of Y381 which would normally form a hydrogen bond with and the carbonyl group of E2072 is displaced by a distance 1.2 Å (Figure 4.7C). These are small conformational changes that may not be significant, given that the ankyrin domain does not undergo large conformational changes upon complexation. However, further optimisation or fragment-growing could provide a starting point to a molecule that potentially diminishes this specific interaction.

4.4.4.2 Implication of the fragment 9F07 on dimerisation

The fragment was also found at the dimerisation site of the two ankyrin domain chains. As previously highlighted, the complex dimer assembly through the ankyrin domain could have an important role in the Notch transcriptional activity. The fragment 9F07 binds to the same H2019 as fragment 12C05 through the same nitrogen of the imidazole ring. This suggests that this site is a possible target to explore in order to interfere with the transcription process.



4.4.5 The relative position of the two fragments to each other

The two structures of the ankyrin domain with the two fragments were superposed. The positions of the two fragments were found to be at a relatively close distance of about 6 Å. One fragment is positioned at the interface with the CSL and the other is located at the interface with MAML (Figure 4.8). The presence of two fragments at these sites suggests the possibility of linking the two fragments or developing a molecule that could target the two interfaces. However, it is difficult to link the two fragments and maintain the same conformation and geometry. This will be discussed in more detail in chapter 6.

The two fragments are located at the dimerisation interface and directed in the same orientation. These two molecules have different scaffolds and do not share any common functional groups though they both form a common hydrogen-bond interaction with residue H2019 of the ankyrin domain. The binding of the two fragments in this site could be a crystallographic artefact by stacking between

the two chains or a true binding site that could be targeted. This must be validated by mutating the residues involved and testing for binding, but this was not possible due to time limitation.

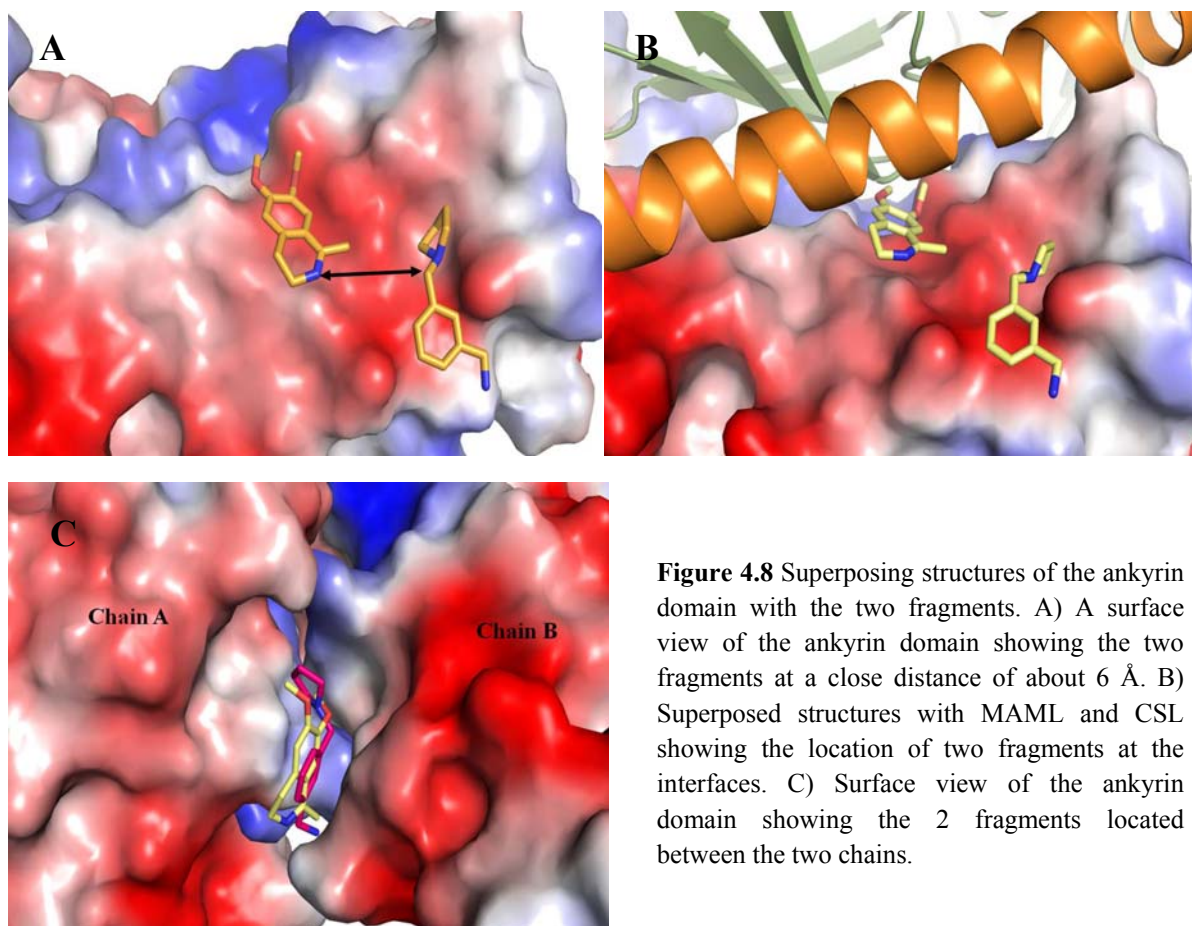


Figure 4.8 Superposing structures of the ankyrin domain with the two fragments. A) A surface view of the ankyrin domain showing the two fragments at a close distance of about 6 Å. B) Superposed structures with MAML and CSL showing the location of two fragments at the interfaces. C) Surface view of the ankyrin domain showing the 2 fragments located between the two chains.

4.5 Conclusions

Structure-based drug design is a powerful tool in drug discovery and drug development. The availability of structures solved by NMR or X-ray crystallography facilitates structure-guided drug design. In fact, X-ray crystallography has been used as a screening method in fragment-based approaches in an automated manner. It is a valid screening method that determines the actual binding mode of ligands with very few false positives.

False positives may arise from artefactual structures that are usually easy to distinguish. Although the ankyrin domain was easily crystallised in various conditions and in a reproducible fashion, structures of only two different fragments were obtained. Different methods and approaches were used for either soaking or co-crystallisation that could overcome problems such as solubility or low diffusion rates of the small hydrophobic fragments. The failure to obtain structures could be attributed to many reasons: incompatibility of the conditions of the initial screening step with the crystallisation conditions, fragment binding at positions of crystal contacts, weak binding affinities of these fragments or the failure to reach the binding equilibrium owing to their poor solubility.

However, the structures obtained of the two fragments here are the first reported examples so far of small molecules that can bind to the ankyrin domain of the Notch receptor. There have been speculations about targeting protein-protein interfaces with small molecules. The two fragments were found to be binding at the interfaces of the ternary complex and the dimerisation interface. This could provide a starting point to develop inhibitors, though this would seem challenging owing to the flat surfaces of these interfaces and lack of adjacent significant pockets to grow to. The two fragments lie close to each other, which suggests the possibility of linking them together. This poses another challenge to maintain the positions and conformations of both fragments.

Chapter 5

Kinetic Studies Using Surface Plasmon Resonance

5.1. Introduction

Surface plasmon resonance (SPR) is an important biophysical method for studying bio-molecular interactions, which has the advantage of measuring affinities in addition to interesting thermodynamic and kinetic parameters. Developing highly sensitive biosensors has attracted many drug discovery groups to implement SPR as a high throughput method for screening large compound libraries. Now that fragment-based drug discovery is being integrated in many drug discovery programs, the availability of SPR as a label-free detection system has made it a primary fragment-screening method. SPR is now an established method in companies such as the Genetech division of Roche, the Heidelberg-based Graffinity and Vernalis

5.1.1 Surface Plasmon Resonance

Surface plasmons are coherent oscillations of electrons that can exist at a metal-dielectric interface as electromagnetic waves. Surface Plasmon Resonance (SPR) is the phenomenon of exciting the surface plasmons by light in thin conducting films that separate media of different refractive indices. Biosensors have implemented this phenomenon to monitor and study the interaction between molecules in real time. In Biacore systems, the optical setup is designed so that a glass of a sensor chip is covered by a thin layer of gold as a conducting film. When the incident light on the reflecting surface exceeds the critical angle a condition of total internal reflection takes place. This forms an electric field intensity known as *evanescent wave field* whose amplitude decreases

exponentially with the distance from the surface. Varying the angle of incidence or the wavelength of the incident light will excite plasmons in the gold film. In resonance conditions, absorption of energy through the evanescent wave field creates a decrease in intensity of the reflected light known as a SPR minimum or SPR angle. The angle and wavelength of SPR minimum are sensitive to changes in the refractive index of the medium adjacent to the metal surface. Changes in the refractive index can be caused by variations in the solute concentration at the surface of the sensor chip. One interacting partner is attached to the surface of the sensor chip and the other is passed over the surface. In case of interaction, the change of the local refractive index caused by adsorption of interacting molecules is monitored as a change in the SPR minimum reflecting the presence of molecules as complexes.

5.1.2 Advantages of Surface Plasmon Resonance

SPR offers many advantages that make it a primary screening method for some targets. An ideal SPR binding experiment consists of two stages. In the first stage, k_{on} , the association rate constant of the ligand and receptor, can be measured. In the second dissociating stage, the sample is rinsed by the running buffer and k_{off} , the dissociation rate constant of the receptor-ligand complex to free ligand and receptor molecules, is calculated. The binding constant K_d can then be computed ($K_d = k_{off} / k_{on}$).

The fact that SPR consumes only a few micrograms makes it suitable for proteins that have very low expression yields or are expensive to produce. Another advantage is that it does not require labelling of the compounds, so facilitating its use for high-throughput screening. Furthermore, it is possible to use organic solvents such as DMSO for poorly soluble compounds. Monitoring the stability of the ligand protein by observing the baseline behaviour is possible using biosensors. This is very important in the drug optimisation process, as it gives information about the drug residence time, which can reflect the pharmacological effect and target selectivity (Copeland *et al.*, 2006). Target selectivity can be estimated by comparing the half-life times for both targets and non-targets, which are often referred to as off-targets (Copeland *et al.*, 2006).

This highly sensitive technique can measure K_d values ranging from 1mM and 1pM and in real-time, so allowing characterisation of kinetic events.

5.1.3 Limitations of Surface Plasmon Resonance

Surface Plasmon resonance is still restricted in its use for high-throughput screening by several factors that need to be addressed. The coupling of the target protein can occur at the binding site, so interfering with recognition of potential ligands. This can be avoided by using different coupling and immobilisation methods; thiol coupling for instance might help to immobilise the proteins in a defined orientation with less coupling sites than amine coupling, thus allowing for more selectivity. The conditions of the assay can affect the behaviour of the binding of ligands. The choice of buffer, pH, and temperature may either improve or worsen the binding sensorgram profiles. These will require initial optimisation before carrying out the binding studies.

Regeneration will depend on the nature of interaction of the ligands with the protein. Interactions with small molecules such as fragments, which likely involve ionic or hydrogen-bond interactions, can be reversed using NaCl as a regenerating solution. On the other hand, some ligands may interact with the protein irreversibly through covalent bonding, which will require more drastic conditions for regeneration. These conditions may destroy the protein activity and impair the performance of the assay.

In highly sensitive biosensors, other components of the buffer could interfere with the data interpretation. This can lead to masking of the actual binding kinetic data or providing false measurements. Such molecules could be the buffer itself, for example HEPES, or molecules such as DMSO in the ligand sample solutions and running buffer, which are both small molecules that bind to many proteins and may compete at the binding site. These problems have been overcome either by using a different buffer such as phosphate or by applying the solvent correction method to exclude the effect of DMSO.

The distinction between valid and false-positive hits is an important part of the screening process. False positives must be identified and discarded from any initial medicinal chemistry programs. They can be detected by inspecting their SPR behaviour over a range of concentrations. A classification system has been developed to help in identifying these promiscuous inhibitors (Giannetti *et al.*, 2008). The first group consists of compounds that bind with a stoichiometry five times higher than expected; these are known as superstoichiometric compounds and are eliminated from further follow up (Giannetti *et al.*, 2008). The second group encompasses compounds that have a stoichiometry lying between 1:1 and 5:1 (protein: compound); these are labelled as non-stoichiometric compounds and should be tested in other assays (Giannetti *et al.*, 2008). The third group, which includes concentration-dependant aggregators, consists of compounds that tend to aggregate at higher concentrations and bind non-specifically to the target protein (Giannetti *et al.*, 2008). These compounds are eliminated from the screen.

In this chapter I describe the characterisation of a selected set of fragments identified in the primary screen (Chapter 3), the optimisation of the fragment 12C05 and the study of indirubin-3'-monoxime binding to the ankyrin domain.

5.2 Characterisation of fragments and small molecules

5.2.1 Characterisation of some of the fragment hits

A fluorescent-based thermal shift assay has been used to identify initial hits in a primary screen (Chapter 3). The binding constants of the fragment hits detected were measured using the Biacore T100 system which implements the surface plasmon resonance phenomenon.

5.2.1.1 Materials and methods

***i)* Immobilisation**

a) Buffer pH and protein pre-concentration

An important step prior to protein immobilisation is electrostatic pre-concentration in a dextran matrix. A sensor chip CM5 was used where the carboxymethylated dextran on the sensor chip surface acquires a negative charge at a pH above 3.5. The optimum conditions required for the ankyrin protein to become positively-charge were scanned using pH-scouting experiment. The ankyrin protein solutions were prepared in different coupling buffers to be tested at different pH values; acetate pH 4, acetate pH 4.5, acetate pH 5 and pH 5.5 at a final concentration of 1 μ M. The running buffer used was HBS which is commercially available from Biacore (10 mM HEPES pH 7.4, 150 mM NaCl, 0.05 % P20) at a temperature 25 °C.

b) Immobilisation procedures

After choosing the optimum buffer solution for the protein, the surface of the sensor chip was activated using a mixture of 0.4 M EDC (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) and 0.1 M NHS (N-hydroxysuccinimide). The ankyrin domain was prepared in acetate buffer pH 4 to reach a final concentration 1 μ M. It was immobilised by injecting at a flow rate 10 μ l/min until it reached the target RU. This was followed by deactivating the remaining active groups on the surface by injecting 1M ethanol-amine-HCL pH 8.5.

The amount of protein to immobilise was calculated using the following formula:

$$\frac{\text{Immobilised protein(RU)}}{\text{ligand binding capacity(RU)}} = \frac{\text{Protein MW}}{\text{ligand MW}}$$

ii) Kinetic analysis of fragments

Serial dilutions of the fragment solutions were prepared either in buffer HBS (10 mM HEPES pH 7.4, 150 mM NaCl and 0.05 % P20) or in PBS buffer (10 mM phosphate buffer pH, 150 mM NaCl and 0.05 % tween 20) depending on the

solubility of the fragments. Some fragments were dissolved in 5% DMSO and the running buffers were also prepared in 5% DMSO. In these cases, solvent correction working solutions were designed by preparing a series of aliquots using PBS or HBS containing 4.5% and 5.8% DMSO according to the following table:

Table 5.1 Preparing the aliquots for solvent correction solutions

Buffer\Vial	1	2	3	4	5	6	7	8
4.5%		200	400	600	800	1000	1200	1400
5.8 %	1400	1200	1000	800	600	400	200	
Total volume	1400	1400	1400	1400	1400	1400	1400	1400

The fragment solutions were injected onto the surface of the chip at a flow rate 30 µl/min, contact time was 60 seconds and dissociation time was 60 seconds.

The sensorgrams were analysed using the Biacore T100 evaluation software. Solvent correction was applied in cases where DMSO was used.

5.2.1.2 Results and discussion

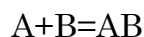
Immobilisation was carried out using the amine coupling method which is the most common approach used. In this method the carboxylic group of the activated surface of the sensor chip is covalently linked with free amino groups. Although these should not be at or near the active or binding site, in the case of the ankyrin domain there is no structural information about where the fragments might bind. The ankyrin domain protein was prepared in acetate buffer pH 4 to pre-concentrate the protein to the surface of the chip. In this approach the immobilisation target depends on the purpose of the analysis; lower levels of immobilisation are usually recommended for kinetic measurements in order to reduce the effects of mass transport of the ligand to the surface. However, kinetic analysis of low molecular weight ligands requires higher level of immobilisation as the SPR response depends on the mass. The immobilisation

target according to the formula (see Materials and Methods) was calculated to be about 7647 RU. The immobilisation level only reached about 4400 RU which was still high enough to carry out kinetic studies.

Biacore has been used to characterise the binding of some fragments that have been picked from the preliminary screen. Some fragments were not readily available for subsequent kinetic analysis. The solubility of the fragments varied, some being soluble in the running buffer whilst others required 5% DMSO to improve the solubility. The common buffer which was used as a running buffer in kinetic and binding screening was HEPES buffer which may interfere with binding (see comments in Introduction). Other buffers were investigated; phosphate buffer was considered the buffer of choice in fragment analysis. However, the HEPES buffer, which is a sulfonic acid derivative, was shown not to bind in the initial screening of the fragment library and in the buffer screen. This allowed the use of HEPES in the running buffer in cases where it improved the solubility of the fragment without concern about possible non-specific or interfering binding. In addition, the buffer was always run as a control in the kinetic analysis experiments and it was subtracted from the fragment run indicating that any response observed would be solely a result of binding of the fragment.

Two methods are used for analysis of the sensorgrams to measure the binding constant. The first fits the experimental data to a mathematical model. This often requires knowledge of the interactions, for example about multiple interaction sites or conformational changes. The fitting procedure includes an iterative process to find the best fit for an equation that can describe the interaction. In kinetic assays, at least five different concentrations are necessary to extract kinetic parameters. Parameters in the fitting equation are assigned either locally or globally. Local parameters are usually concentration and bulk refractive index contribution which are assigned independently for each curve. Global parameters are assigned for the whole dataset; this provides more robust values for the rate constants. Local parameters can be used for determining rate constants in cases where the protein activity is variable in different cycles such

as a change of capture level in capture assays. The simplest binding model is a 1:1 binding interaction:



The second method is based on affinity determination where the K_d is measured as the ligand concentration that gives a response corresponding to half R_{max} . Fragments are weakly binding molecules that usually have very fast on- and off-rates. The typical sensorgram for fragments is characterised by a square-shaped plot. Although the various kinetic parameters cannot be derived from these sensorgrams, the dissociation constant can still be measured by the steady state method where steady state binding levels (R_{eq}) values are plotted against sample concentration (C). The K_d can be computed from the following equation:

$$R_{eq} = \frac{CR_{max}}{K_d + C} + RI$$

RI is the refractive index contribution

It should be pointed out that these weakly binding fragments might bind in the millimolar range. The concentration range of the samples prepared for analysis should include concentrations higher and lower than the expected K_d . The fragments are quite hydrophobic molecules with limited solubility and they tend to aggregate at higher concentrations. To avoid the aggregation of the compounds at the surface of the chip, the samples are usually centrifuged down to eliminate the effect of binding of aggregates. However, the concentration of samples is then less than it estimated and the computed K_d value would be underestimated. It is therefore advisable to measure the new concentrations of the samples after centrifugation for a more accurate calculation. However, this seemed difficult with the small-volume samples and the lack of an adequate method to measure concentrations of small volumes. The calculated K_d then would be a rough estimate of the binding affinities but could still be used to establish SAR tables in some cases.

Specificity has often been a concern when the binding site is uncertain. As a consequence, on screening using biosensors, a counter-protein is usually used that is either the same target but the active site is blocked, or another protein that is known for not binding to these molecules. However, in the early stages of drug discovery, specificity and selectivity of fragments is generally of less importance than in more advanced stages. The aim of the first hit-identification phase is to identify the main chemical moieties that bind to the protein and to measure their potency. This was the purpose of the experiments described here.

Some fragments did not show any binding despite being identified as positive hits in screening using the thermal shift method (Chapter 3). This may be a consequence of one or more factors. First these fragment hits could have been false-positives although some of them were analogues of confirmed hits and it would have been expected that they would give a binding response. Secondly, the immobilisation of the ankyrin domain may have been unfavourable for the fragment binding. Thirdly, the solubility of the fragments may be too low even after solubilising in 5%DMSO. Fourthly, the compatibility of the initial screening conditions with the running buffer conditions could also introduce a source of variation between the results from screening and kinetic analysis. Possible variations include different buffers or pH values that could alter the ionisation or protonation states of the fragments. Very small differences in pH can be very critical in determining the ionisation state that is favourable for binding; and this depends on the ionisation constants of the fragments as well. In the experiments described here the screening of the fragment library was carried out at a pH 8.5 while the kinetic analysis was carried out pH 7.4.

Ideally, the fragments should all be tested in the same conditions. However, the conditions for each fragment in these experiments were individually optimised to improve solubility; this included changing the buffer, the pH and sometimes the temperature. This could introduce errors in comparing K_d values or other kinetic parameters, but on balance was considered worth doing. Regeneration of the surface using 1M NaCl was sufficient to restore the baseline, especially for those fragments with fast off-rates.

Fragment 12C05 was fitted to a 1:1 kinetic model. Although it was shown to bind at two different sites in the crystal structure, the sensorgram seemed to fit well in a 1:1 model. In addition, the bivalent kinetic model describes the binding of two molecules to two equivalent binding sites on the protein. It has been demonstrated crystallographically that the fragment 12C05 binds at the upper helices between repeats six and seven in one position. The second site observed in the crystals is between the two ankyrin chains. The fragment is likely binding to individual ankyrin molecules only very weakly and is not observed in solution.

The affinities of three other compounds analogous to 12C05 were measured. From these experiments it was clear that differences in side chains of the phenyl ring caused variations in affinity. Replacing the methyl amino group with an amino group improved the affinity from 7.80 mM to 1.58 mM while replacing it with a hydroxyl methyl group decreased the affinity to 10.45 mM. The highest affinity was observed when the pyrrolidine ring was replaced with a piperdinyll ring showing an affinity of 0.968 mM.

The interactions of fragment12C05 with the ankyrin domain were elucidated by the protein-fragment crystal structure. The fragment is stabilised by many interactions including a π - π aromatic interaction with residue Y2075 and a salt bridge with D2109 with the nitrogen atom of the pyrrollidine ring. The pKa values of the nitrogen atom of the pyrrollidine and piperdine ring in case of 14B04 were calculated using an online calculator <http://sparc.chem.uga.edu/sparc/>. These variations did not cause any dramatic changes in the pKa values. Under the running buffer and sample buffer conditions, which are at a pH 7.4, the nitrogen atoms are positively charged. This means that the salt bridge with D2109 is unaffected by changes in the phenyl side chain. Instead, the difference in affinities is likely caused by a change in the strength of the π - π aromatic interaction by the direct effect of these substituents on the phenyl ring. In case of 14B04, the replacement of the pyrrollidine ring with a piperdine ring could favour the positioning of the nitrogen atom to form a better salt bridge. The five-

membered pyrrolidine ring adopts a puckered conformation whereas the six-membered piperidine ring is found in a chair conformation.

Fragment 9F07 was identified in the first screen and detected in the crystal structure as well. It showed a binding response in the Biacore experiments but failed to reach a saturation level with increasing concentrations. This could indicate that it is binding very weakly and to reach saturation would require higher concentrations at which the fragment would start to aggregate.

Table 5.2 A table list of the fragments tested and their corresponding kinetic parameters

Fragment	Binding	K _d (mM)	K _{on} (M ⁻¹ s ⁻¹)	K _{off} (s ⁻¹)
12C05	Binding	7.876 ± 1.05	2.356±0.22	0.018±0.0007
12D05	Binding	1.586 ± 0.15	12.43± 0.90	0.019±0.003
12B05	Binding	10.45 ±1.668	-	-
14B04	Binding	0.968 ± 0.237	-	-
14H04	Binding	1.90 ± 0.78	-	-
1605	Binding	0.3908	-	-
12E03	No binding	-	-	-
12H05	No binding	-	-	-
12E06	No binding	-	-	-
9A03	No binding	-	-	-
10H10	No binding	-	-	-
13A09	No binding	-	-	-

A

K_d=7.87 mM

B

K_d= 1.58 mM

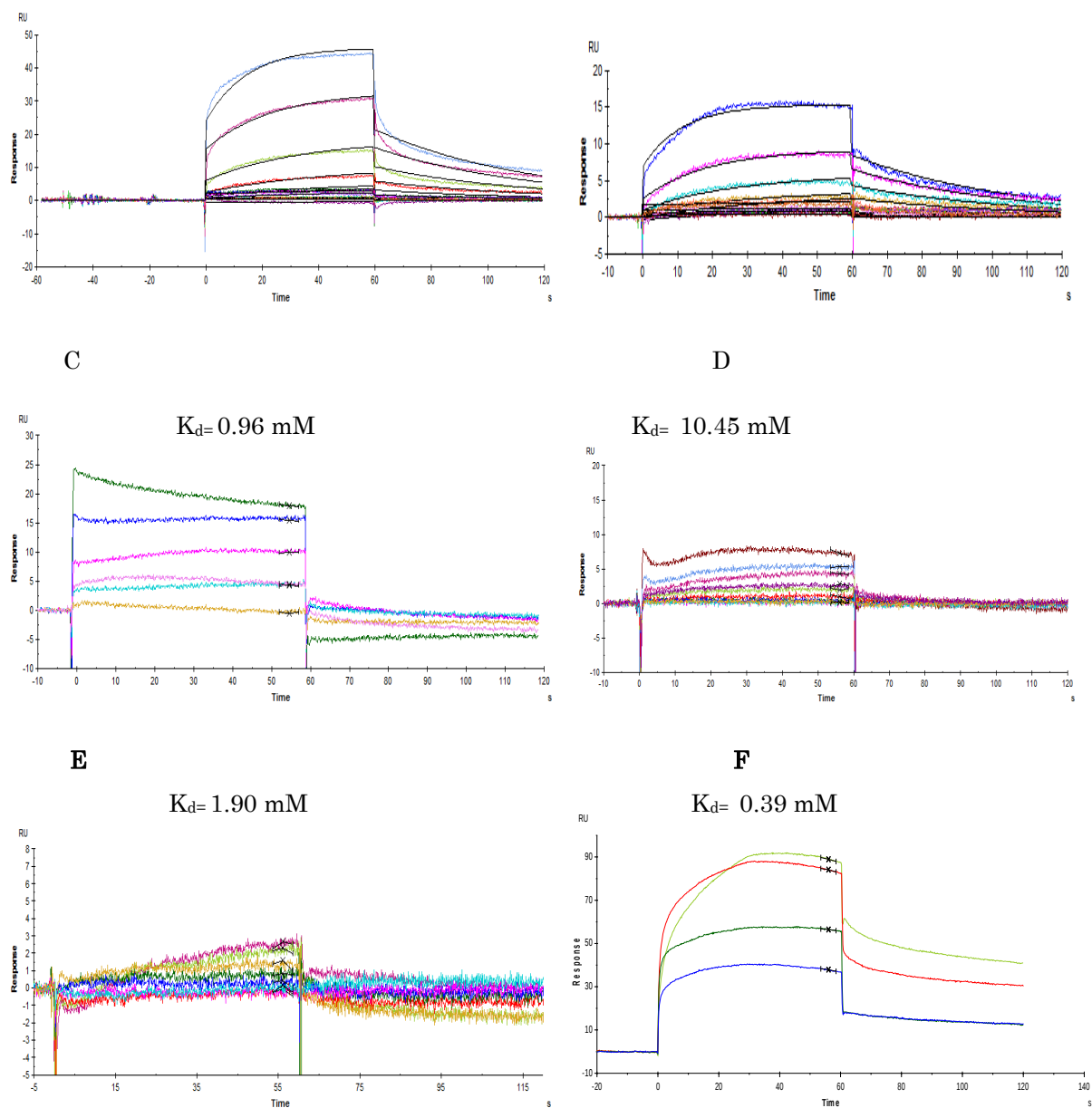
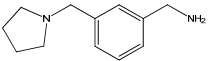
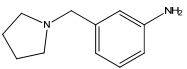
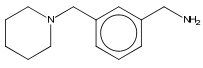
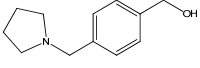


Figure 5.1 Sensorgrams of fragments shown to bind using thermal shift and estimated K_d values. In A) the fragment 12C05 and in B) fragment 12D05, the sensorgrams were fitted to a 1:1 model and kinetic parameters were derived. In C) fragment 14B04 , D)fragment 12B05, E)fragment 14H04 and F)fragment 1605, K_d values were only measured by steady state affinity.

Table 5.3 List K_d and pKa values of fragments 12C5 and its analogous

Fragment	K_d	pKa
12C05 	7.876 ± 1.05 mM	9.36
12D5 	1.586 ± 0.1 mM	9.43
14B4 	0.968 ± 0.237 mM	9.36
12B5 	10.45 ± 1.668 mM	9.43

5.2.2. Characterisation of optimised fragment

Fragment 12C05 has been validated as a positive hit using three different biophysical techniques: fluorescent based-thermal shift assay, SPR and X-ray crystallography. The fragment has been optimised on the basis of the information gathered from these methods. The crystal structure suggested that the fragment was involved in a number of interactions. These interactions include a face-to-face π - π aromatic interaction, weak hydrogen bonds and a salt bridge. Optimisation of the fragment aimed at finding the heterocyclic ring that would bind with highest affinity and enhance the interaction (Figure 5.2).

5.2.2.1 Materials and methods

The Biacore run was set up as described earlier (see section 5.2.1.1). Compounds A, B and C were synthesised by Dr Hamid Nasiri (Department of Chemistry, University of Cambridge). The concentration series for the compounds were prepared as following:

For compound A: 37.5, 18.75, 9.38, 4.69, 2.34, 1.17, 0.58 mM.

For compound B: 18.75, 9.38, 4.69, 2.34, 1.17, 0.58, 0.29 mM.

For compound C: 37.5, 18.75, 9.38, 4.69, 2.34, 1.17, 0.58 mM.

The sensorgrams were analysed using the Biacore T100 evaluation software. Solvent correction was applied in cases where DMSO was used.

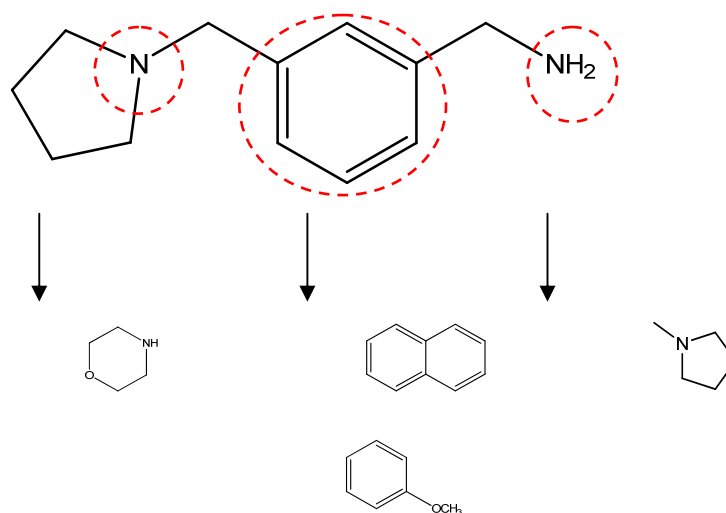


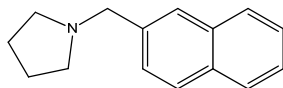
Figure 5.2 A summary of the optimisation approaches designed to optimise the fragment 12C05. The chemical moieties in dashed red circles were identified as pharmacophoric points involved in important interactions with ankyrin amino acid residues. These moieties were replaced by other rings or substitutions that were thought likely to enhance these interactions

5.2.2.2 Results and discussion

The π - π aromatic interaction should be improved by enhancing the stacking; this might be achieved by increasing the electron density at the phenyl ring. Replacing the phenyl ring by a naphthalene ring would increase the delocalisation of the electrons through the conjugated system.

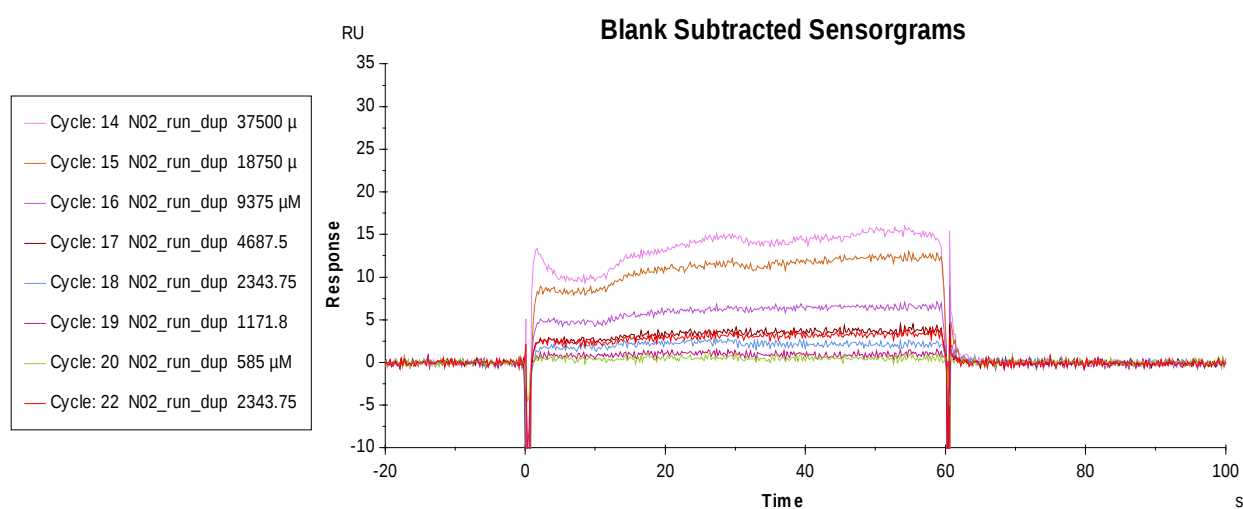
The affinity of the compound A was measured using SPR by the steady state method. It bound to the ankyrin domain with a K_d 21.95 mM, an approximately three-fold decrease in affinity compared to the original fragment. This could indicate that the naphthalene ring is not favourable for binding as expected. The resonance structure of the naphthalene ring could have affected the basicity of the nitrogen atom of the pyrrolidine ring. The nitrogen atom was involved in a salt

bridge interaction with the acidic carbonyl group of the aspartate residue Asp2109 of the ankyrin domain.



Compound A

A



B

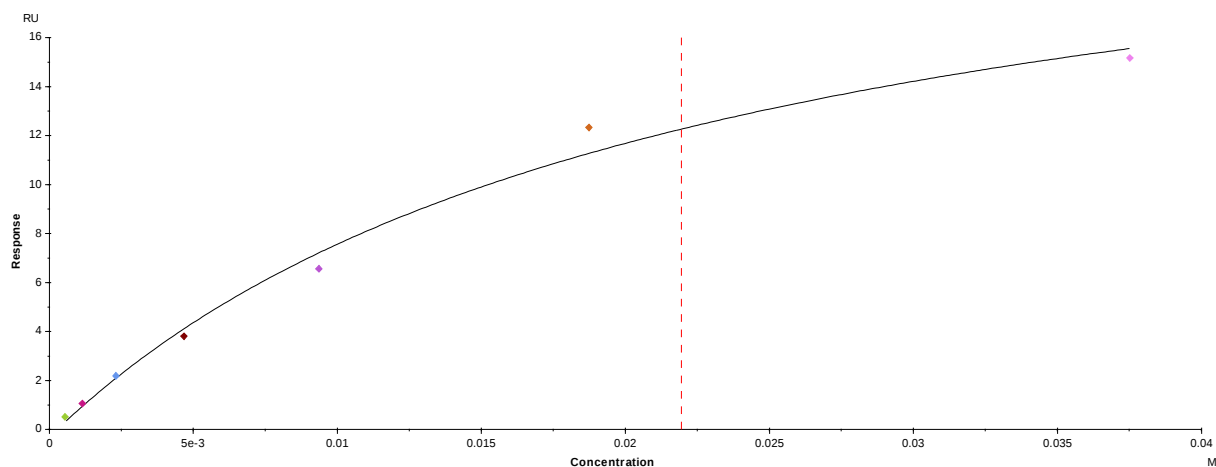
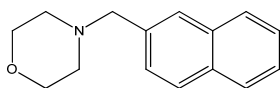


Figure 5.3 The effect of replacing a phenyl ring with a naphthalene ring and removing the methyl amino side chain on the affinity. In A) the sensorgram of the compound showing fast on- and off-rates, and in B) a steady state affinity measurement of a K_d value of about 21 mM



Compound B

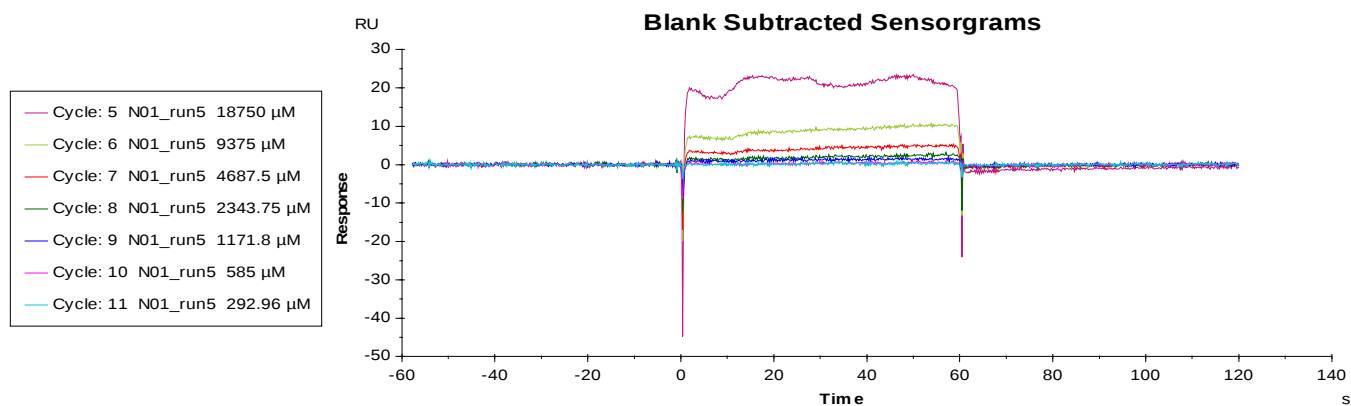
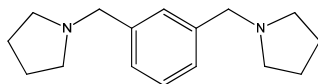


Figure 5.4 Sensorgram showing the effect of replacing the pyrroline ring with a morpholine ring. Aggregation is observed at higher concentration (18.75 mM)



Compound C

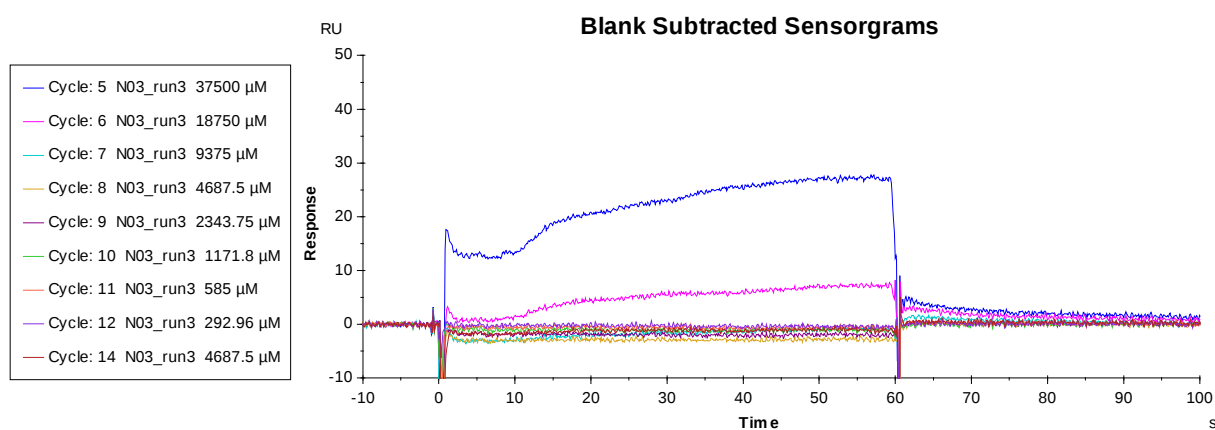


Figure 5.5 Replacing the methyl amino side chain by a methyl pyrroline inhibited the binding and aggregation was observed in sensorgram of two high concentration values (37.5 and 18.75 mM)

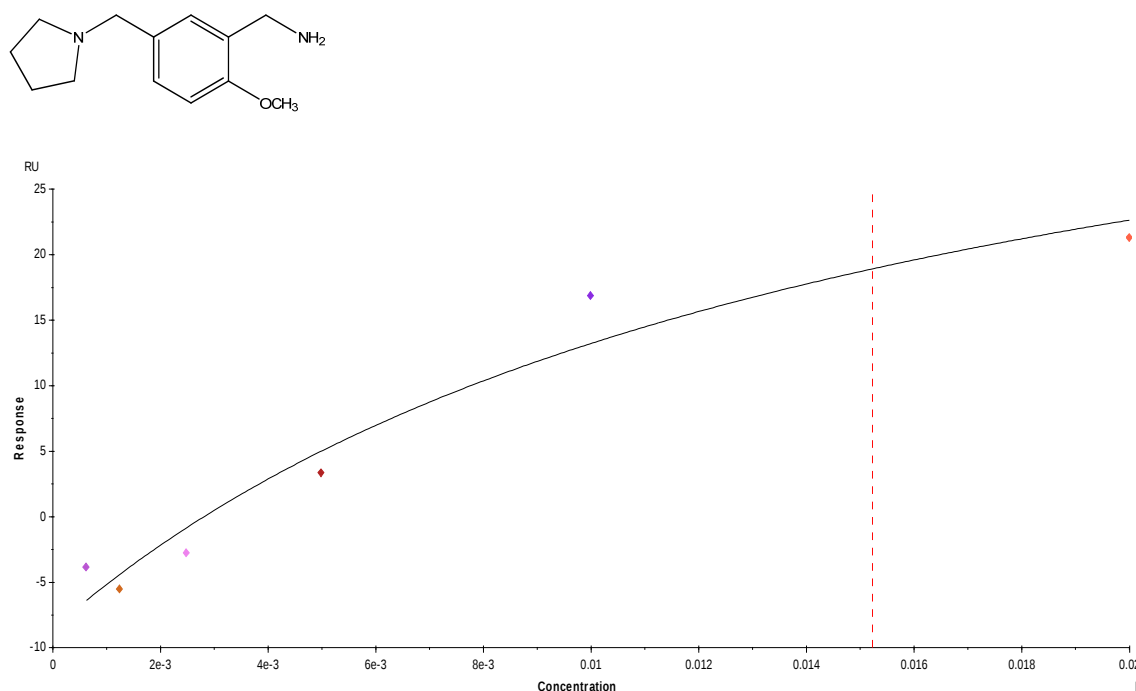


Figure 5.6 Adding a methoxy group did not have a significant affect on the affinity which was measured by steady-state which showed that the new molecule was binding at a K_d 15.23.

The pyrrolidine ring was replaced by a morpholine group and the phenyl ring by a naphthalene ring in compound B. These changes were made in order to explore different ring systems in order to identify the ring with highest affinity. The sensorgram showed binding as the RU values increased with increasing concentration, but this could be a false positive as a result of aggregation. The K_d values were not calculated as compound B starts to aggregate and precipitate at higher concentration as seen at 18.75 mM. The K_d values are likely to be between 9 and 30 mM. This result suggested that the morpholine ring is not favourable for binding as it has been confirmed by fluorescent-based thermal shift assay (Chapter 3).

The methyl amino side chain was replaced with a methyl pyrrolidinyl group in compound C. The larger group was targeted to bind to the small adjacent pocket

in an attempt to grow the fragment 12C05. It did not show any binding response at lower concentrations but showed a jump in the signal at higher concentrations indicating aggregation. This compound falls in the category of concentration-dependant aggregators that are normally eliminated in the screen. However, the compound might be improved by replacing the side chain with an ethyl-pyrrolyl group instead.

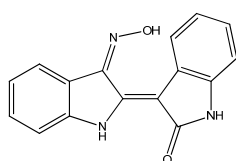
These results showed that the fragment 12C05 is the best scaffold binding to the ankyrin protein despite its weak affinity. Optimising this fragment has proven to be difficult, probably due to the many interactions through which it is involved with ankyrin domain residues. These interactions seem to be finely balanced so that optimising one interaction may adversely affect the rest of the interactions.

5.2.3 Characterisation of Indirubin-3'-monoxime

Indirubin is an active component that can be found in the Chinese herb Danggui Longhui Wan. It has been traditionally used for its therapeutic benefits in treating chronic diseases (Eisenbrand *et al.*, 2004; Xiao *et al.*, 2002). Indirubin and its derivatives have been identified as inhibitors for a number of targets (Adachi *et al.*, 2001; Bian *et al.*, 2003; Bertrand *et al.*, 2003; Nam *et al.*, 2005). They have recently been found to inhibit the Notch1 signalling pathway as well. The main mechanism of inhibition was through inhibiting the function of glycogen synthase-3 β (GSK-3 β). GSK-3 β regulates the Notch1 signalling pathway by controlling the proteosomal degradation of Notch intracellular domain through its direct phosphorylation. Indirubin derivatives caused different inhibition activities on the Notch signalling (Foltz *et al.*, 2002; Epinosa *et al.*, 2003). In addition, indirubin-3'-monoxime was found to inhibit the interaction between Notch intracellular domain and RBP-Jk (Lee *et al.*, 2008). This suggested that the indirubin-3'-monoxime could bind with either of the two proteins so I tested whether it binds to the ankyrin domain using the SPR technique.

5.2.3.1 Results and discussion

Indirubin-3'-monoxime is the only small compound that has been reported so far to have any inhibitory effect on the physical interaction of the Notch1 intracellular domain and RBP-Jk. This was demonstrated by co-immunoprecipitation assays, showing that indirubin-3'-monoxime could interact with either protein and prevent the interaction. Indirubin-3'-monoxime is a small molecule of molecular weight of 277.28 Da. Fragmenting indirubin-3'-monoxime through the double bond would result in indole derivatives, which are similar to the chemical moieties that were observed in primary fragment screening. It was tested against the ankyrin domain but showed no binding at any concentration that experimentally possible to test. This suggested that the indirubin-3'-monoxime does not bind to the ankyrin domain. However, it could bind to other regions of the Notch intracellular domain or exhibit its inhibitory action by binding to the RBP-Jk instead. This could be because indirubin-3'-monoxime is a very planar molecule with limited flexibility. There is a tendency for protein-protein inhibitors to be large, rigid and hydrophobic (Higueruelo *et al.*, 2009). They also possess a higher ring count than drugs and ligands found in the PDB. Indirubin-3'-monoxime consists of four aromatic rings, if each indole ring is considered as two rings. This affects the solubility of the molecule as well as other physical properties such as logP.



Indirubin-3'-monoxime

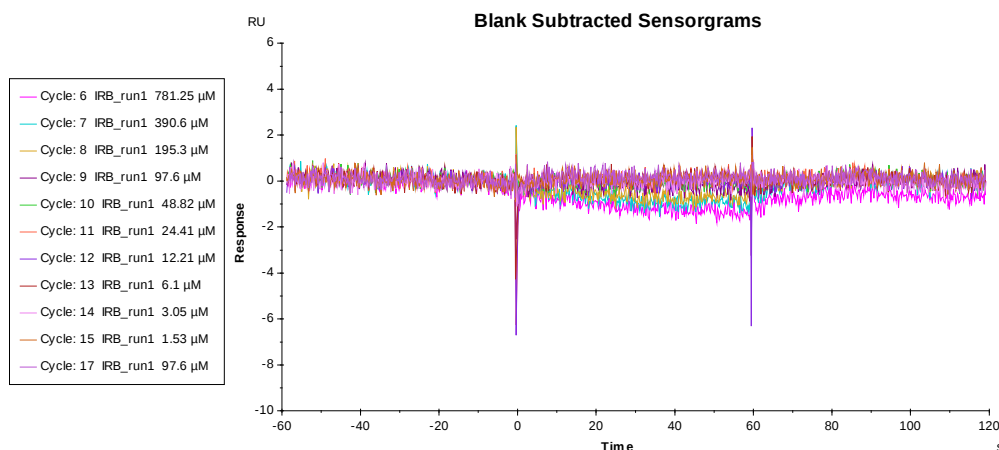


Figure 5.7 Sensorgrams of indirubin-3'-monoxime at various concentrations showing no binding responses to the ankyrin domain.

The concentration of indirubin-3'-monoxime that caused the inhibition of the protein-protein interaction was not stated (Lee *et al.*, 2008). The ankyrin domain could be an off-target that binds with less specificity at a higher concentration than used here, but indirubin-3'-monoxime was insoluble at higher concentrations in the conditions used in this kinetic analysis.

5.3 Conclusions

Measuring binding affinities for molecules is an important step in the drug discovery process. In early screening stages, its importance lies in prioritising small molecule hits for proceeding from hit-to-lead. The challenge at this stage is to find a sensitive technique that can help in measuring weakly-binding affinities. The development of sensitive biosensors has made it possible to quantify and analyse these molecules. In fact, it has been used as a primary high-throughput screening for fragments. However, interpretation of the data retrieved may not be straightforward in some cases and requires additional information to explain the binding behaviour observed in different sensorgrams. Another hurdle to overcome in analysing small fragments is their poor water solubility; the use of organic solvents as DMSO can help to improve the solubility in some cases, but there is always a risk of aggregation that could lead to false-positive results.

The binding affinities measured combined with other structural information can help in optimising and developing fragments in a structure-based approach. However, fragment 12C05 was the only fragment that was characterised both structurally and kinetically. Nevertheless, the information helped in designing some compounds I thought might improve the binding affinity. Unfortunately, none of the proposed compounds has significantly changed affinities, although small changes were observed that help in understanding the interactions.

Chapter 6

Computational Methods for Fragment Elaboration

6.1 Introduction

Computational approaches have proved useful tools to complement experimental approaches in the drug discovery process. Computational methods have been integrated at different stages starting with hit identification using virtual screening, on to hit-to-lead optimisation and through to improving pharmaceutical and physical properties. Virtual screening selects molecules from larger libraries by docking. *De novo* design on the other hand generates new molecules within the boundaries of a binding pocket. A wide variety of programs have been developed for structure-based drug design (Table 6.1), most of which implement techniques to explore the translational, rotational and conformational spaces of small molecules in the site of interest. They also employ scoring functions to estimate the free energy of binding for a pose, to predict the binding mode and to rank molecules in large libraries.

Other important computational analysis tools in the drug design process include pharmacophoric screening to identify the three-dimensional geometric arrangement of the essential features in the binding site. This suggests applying pharmacophoric descriptors as constraints for virtual screening and as descriptors for library design.

Table 6.1 A list of selected programs for docking and *de novo* design

Docking		
GOLD	Genetic Algorithm, empirical scoring function	(Jones <i>et al.</i> , 1995)
Autodock	Simulated annealing, Genetic Algorithm, AMBER scoring function	(Morris <i>et al.</i> , 1998)
DOCK	Combinatorial docking, multiconformer rigid-body docking, AMBER scoring	(Ewing <i>et al.</i> , (2001)
<i>De novo</i> design		
SPROUT	Fragment-based, sequential growth and combinatorial search	(Gillet <i>et al.</i> , 1993)
SPLICE	Recombination of ligands from database search	(Ho & Marshall, 1993)
LEGEND	Atom-based and stochastic search	(Itai & Nishibata, 1991)

6.2 *De novo* design of molecules using SPROUT

Creating new compounds in *de novo* drug design programs with no chemical restrictions can generate too many suggestions to deal with in limited computer time. Nevertheless, *de novo* design remains an important idea-generating tool in early stages of a drug discovery program. Synthetic tractability of the generated molecules could be solved by restricting the chemical search space to compounds that can be easily synthesised through known chemical reactions. *De novo* programs that employ various techniques have been developed (Table 6.1). SPROUT produces molecules through fragment-based sequential growth and combinatorial search (Gillett *et al.*, 2004). One approach is to grow the ligand by sequential adding of fragments to a specified point in the binding site. Another involves fragment placement and linking by placing fragments and docking them to the energetically favourable pose and then connecting them with spacer linkers.

6.2.1 Fragment growing and fragment linking

SPROUT, a useful *de novo* structure building software, generates structures according to set specifications, including ring sizes, numbers of rings, rotatable bonds and other parameters selected by the user from a built library template or from a library designed beforehand. SPROUT represents a modularised system that facilitates structure-based drug design at various stages. This can be achieved through five modules that run consecutively, starting from binding-pocket identification and followed by target site identification. This detects hydrogen bonding and hydrophobic regions in the binding pocket that can be useful for positioning potential ligand atoms. Functional groups are then docked to selected positions and linked together by generating skeletons that comply to preset steric constraints. The results generated can then be clustered and scored for evaluation. SPROUT can also be used in hit optimisation in order to design molecules by linking two fragments at adjacent sites.

In the following sections, I describe the design of compounds using SPROUT by linking the two fragments 9F07 and 12C05 that were characterised crystallographically. SPROUT was also used to design larger compounds of 12C05 by growing the fragment to adjacent pockets.

6.2.2 Materials and methods

SPROUT is part of a three-package product line that has been developed at the University of Leeds. SPROUT version 6.2 was employed here for *de novo* drug design. A license for using this software to implement it in the fragment elaboration process was kindly provided by Professor Peter Johnson, University of Leeds.

i) Protein Template

The protein-ligand structure of the ankyrin domain bound with fragment 12C05 was used here as a protein template. The protein was prepared by adding hydrogen atoms and assigning the protonation states according to the

surrounding environment. The amide groups of asparagines and glutamines were flipped according to the most likely hydrogen-bond network. Histidine residues were visually inspected and tautomeric states were assigned to allow for the most favourable interactions. This was performed using the software SYBYL® 8.0.

ii) SPROUT modules

The SPROUT software is an interactive system in which the modules are set as following:

- a) CANGAROO module: The acronym stands for *Cleft ANalysis by Geometry based Algorithm Regardless Of the Orientation*. In this module the protein-ligand template PDB file was used as an input file. The protein-fragment structure is divided into three sets: protein, water molecules and fragment.

In fragment growing: the fragment 12C05 is selected and defined as a cavity (ligand) file. The receptor file was defined by cutting a section around the cavity rather than using the whole protein structure. A section of a radius 15 Å around the fragment was selected and identified as the receptor file.

In fragment linking: The same protein-fragment structure was used and the fragment 9F07 was imported. Both fragments 12C05 and 9F07 were selected and defined as the cavity (ligand) file. The receptor file was selected by defining a cleft 10 Å around the selected cavity.

- b) HIPPO module: The acronym stands for *Hydrogen Bonding Interaction Site Prediction as Positions with Orientations*. In this module, the possible binding targets within the selected receptor were explored and defined. Hydrogen bond donors and acceptors were identified and hydrophobic regions and surfaces are investigated.

In fragment growing: a spherical target site with a radius of 0.5 Å was created at both site1 and site2 (Figure 6.2).

In fragment linking: The two fragments; 12C05 and 9F07 were imported and the points of attachment for linking the two fragments were selected as vertices.

- c) ELEFANT module: The acronym stands for ***EL**Ection of **F**unctional groups and **AN**choring them to **T**arget sites*. Functional groups from a template library were selected as docking start templates. The selected groups are then docked and positioned to satisfy the selected site.

In fragment growing: The docking start templates for both site1 and site2 included five-membered rings, six-membered rings, methyl group and derivatives, amino groups and derivatives, guanidinium group and sulfonyl group.

In fragment linking: No flexibility was inferred on either fragment so as to maintain their conformations.

SPIDER module: This acronym stands for ***S**tructure **P**roduction with **I**nteractive **DE**sign of **R**esults*. The docking start templates are connected with selected spacers in this module. Spacer templates are defined from a spacer template library. A large number of spacers was selected, including flexible saturated and unsaturated alkyl chains and derivatives, three-membered, four-membered rings, five-membered rings and derivatives.

In fragment growing: This module was run separately for site1 and site2.

In fragment linking: The same spacer templates were selected.

- d) ALLIGATOR module: This acronym stands for ***AL**gorithms for **LIG**And **T**esting and **O**rdering of **R**esults*. This is the final module where the results are visualised, scored according to a built-in scoring scheme and ranked. The first one hundred highest-ranking were selected as a set and imported as a SDF and PDB file for further analysis.

6.2.3 Results and Discussion

6.2.3.1 Fragment linking

Linking two fragments that bind at proximal sites is not as straightforward as it may appear. There are many factors to consider while attempting to link two molecules. First, the protein conformation that binds one fragment may differ from the one that binds the other. This could be useful in cases of positive allosteric cooperativity, where the binding of one fragment causes a favourable conformation change allowing the binding of another fragment in a distant site. It could even create a new binding pocket that may be a new binding site. It is also preferable in cases of configurational cooperativity, where the number of unproductive configurations of the protein is reduced (Whitty, 2008). Fragment linking could actually be more significant if the two fragments or molecules were observed to bind in the same protein structure so that variations in conformation are avoided.

The next decision concerns the choice of the size, orientation and chemistry of the linker, which should not be too short to cover the distance and not too long; and it should maintain the orientation and conformation of both fragments to allow proper binding. The flexibility of the linker is also important to consider. Rigid linkers have the advantage of minimal conformational entropy, although they restrict the conformation of the linked molecule. This could increase the chances of unfavourable interactions with the protein or cause improper binding due to incorrect orientation. On the other hand, flexible linkers can adopt different conformations without causing steric strain, allowing for sampling of conformational space.

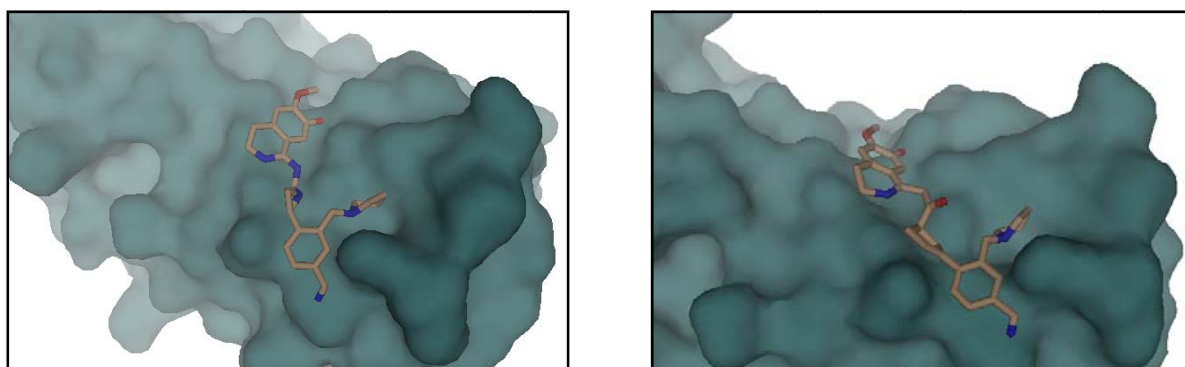
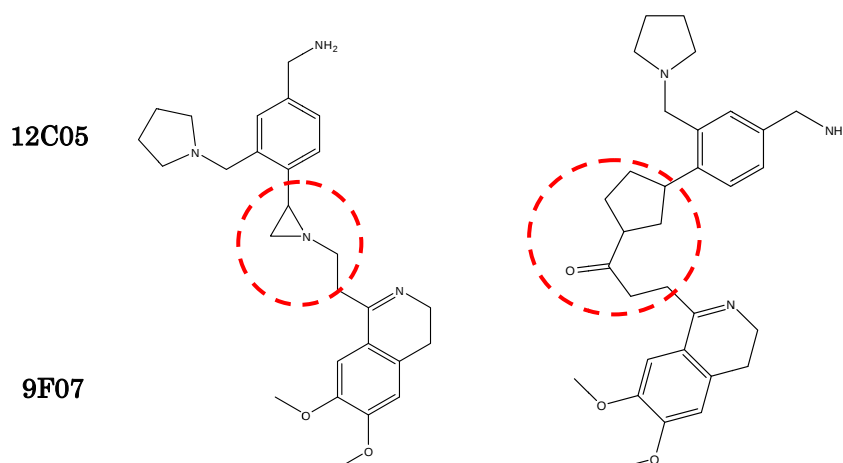


Figure 6.1 A surface view of the ankyrin domain with the two structures generated by SPROUT after attempting to link the two fragments together. The conformations of both fragments were maintained as in the original crystal structure and skeleton of linkers (within red circles) were built to connect them together without causing any clashes with the protein residues.

The fragments 12C05 and 9F07 were found to be binding at positions in two different crystal structures that are about 6 Å apart. This appeared to be an attractive opportunity for attempting to link the two fragments in order to provide a ligand of higher affinity. The linkers were selected from the library provided by SPROUT. The positions for linking were selected and SPROUT built skeletons in an attempt to connect the two selected points. This generated two structures that maintained the same conformation of the original fragments in the crystal structure (Figure 6.1). These compounds were commercially unavailable and were hard to synthesise as they had many chiral centres.

Nevertheless, they gave an idea of the linker length and the appropriate chemistry that could be considered.

6.2.3.2 Fragment growing using SPROUT

The fragment 12C05 is located at a very shallow surface that is devoid of any distinct pockets. This makes it difficult to elaborate and grow the fragment to give molecules of higher affinity and good ligand efficiency. There are two potential sites that the fragment could grow to (Figure 6.2). The first is a small pocket lined with polar and hydrophobic residues: D2109, I2110, R2112 and L2113. The fragment 12C05, identified in the preliminary screening of the fragment library, was used as the starting point for adding new groups and docking the new compounds to the binding site. The new groups were added from a template library provided in the software. The distance from the methyl amino side chain of the fragment to site 1 is about 8 Å. The new groups selected for docking in this site were small groups such as methyl or amino groups or 5 membered-rings. This appeared to be more suitable for the small pocket rather than larger groups. Simple alkyl chains were selected as linkers. These alkyl chains could be involved in lipophilic interactions with the hydrophobic residues. These interactions are essential in many protein-ligand stabilising interactions which arise mainly by the replacement and release of ordered water molecules. The fragment molecule was maintained in the same conformation by applying constraints to keep the same pharmacophoric points that preserve the primary interactions. The skeletons formed were then scored according to the interactions they formed using a SPROUT scoring scheme, the higher the score the more favourable the structure. The compounds that scored higher were compounds of aliphatic branched side chains while compounds containing cyclic groups scored less (Figure 6.3).

There is a second potential site that could be targeted for fragment growing. It is a shallow and flat surface where it would be difficult to introduce groups that form contacts with the protein residues. This site (site 2) is lined with residues

E2072, G2073, V2039, N2040, and E2076. This site lies proximal to the interface of the ankyrin domain with MAML, although there are no residues that directly interact. This could be useful for an enhanced inhibitory role rather than improving the binding affinity. Again, the fragment 12C05 was used as the starting point and groups were added to position 4 of the benzene ring of the fragment 12C05. The groups that were selected from the template library were more flexible chains containing functional groups such as sulfonamides and guanidinium groups. The choice of these chains should be more favourable in the case of flat protein surfaces rather than cyclic side chains. The compounds that were ranked top contained branched chains (Figure 6.4).

Certain limitations and caveats should be considered on assessing the generated molecules. The protein residues remain constrained and rigid throughout the docking, and only the ligands are allowed limited flexibility. Even here the fragment was remained constrained to maintain the original interactions and the new groups only were allowed flexibility. The suggested molecules are selected with respect to their availability or ease of synthesis. However, the approach remains complementary to other computational tools in drug design.

The fragment 9F07 was also characterised crystallographically suggesting that it could be optimised using computational methods. However, this appeared to be challenging for two reasons. First, it binds very weakly with an affinity greater than 40 mM which would not make it the best candidate hit to start with. Secondly, the fragment was found to bind at a very shallow surface with no adjacent pockets or contacts that could provide an anchor to grow the fragment to.

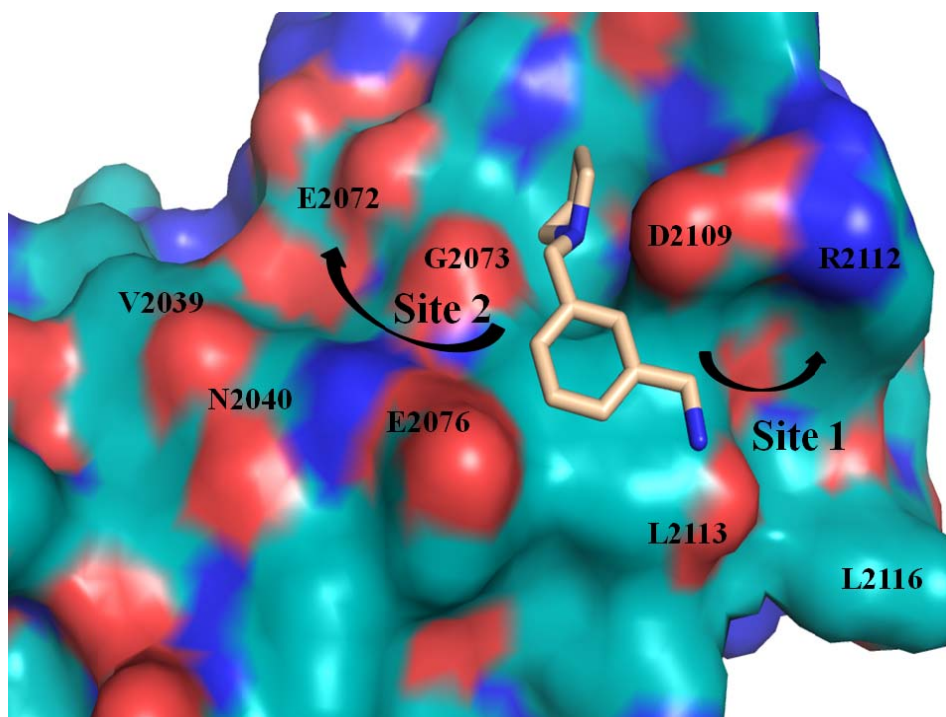


Figure 6.2 A surface view of the ankyrin domain binding to the fragment 12C05 showing two potential sites that could be targeted for fragment optimisation. Site 1 provides a deep small pocket that could enhance the affinity. On the other hand, Site 2 lies close to the protein-protein interface with MAML and could provide an enhanced inhibitory effect. The residues that line both sites can aid in selecting the potential groups that can be added to introduce a new interaction .

Site 1

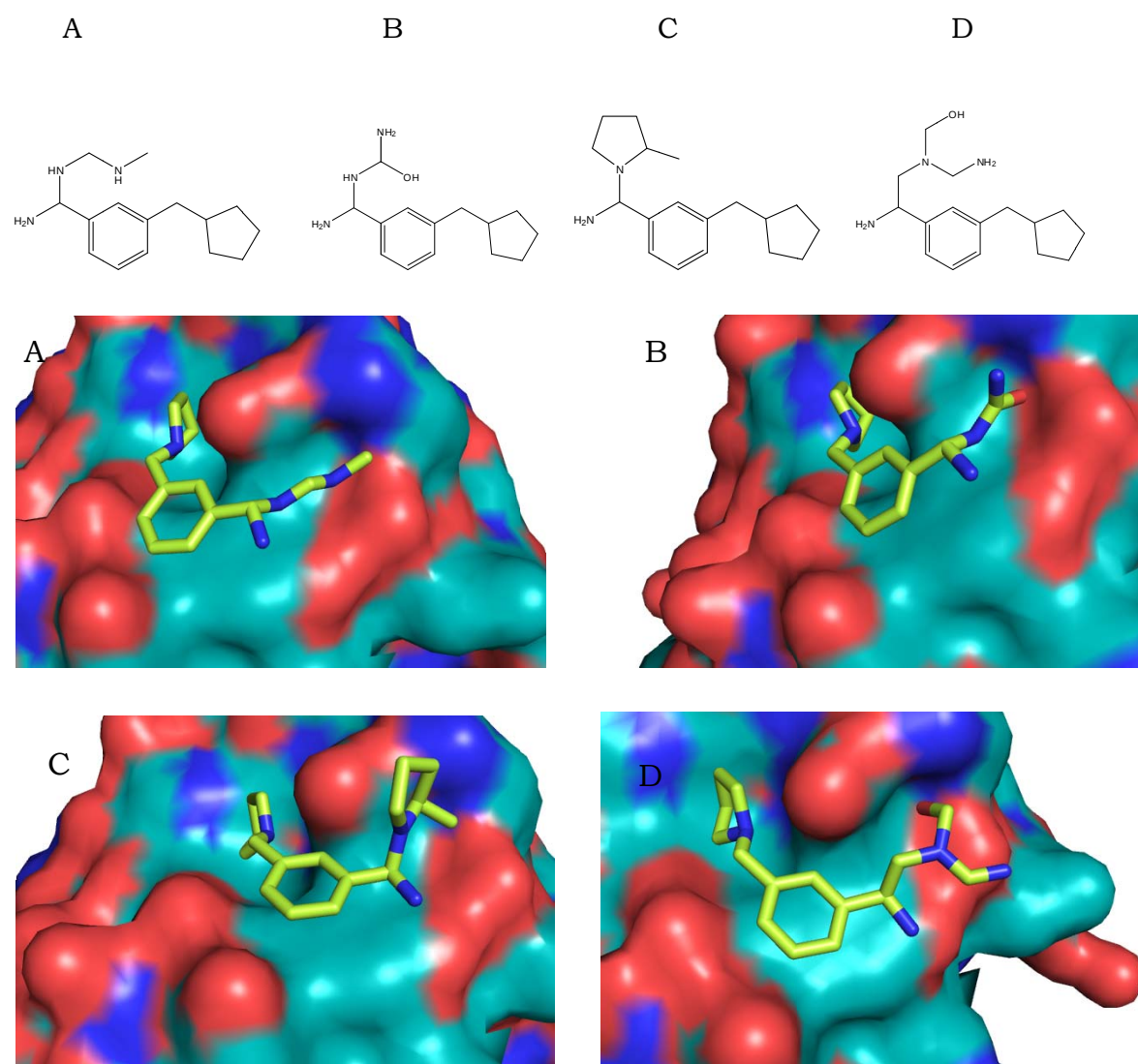


Figure 6.3 An example of the structures generated by docking selected functional groups to the adjacent pocket at (Site 1). Most of the output structures added an aliphatic branched side chain. The cyclic side chains were scored less as they might introduce some steric constraints.

Site 2

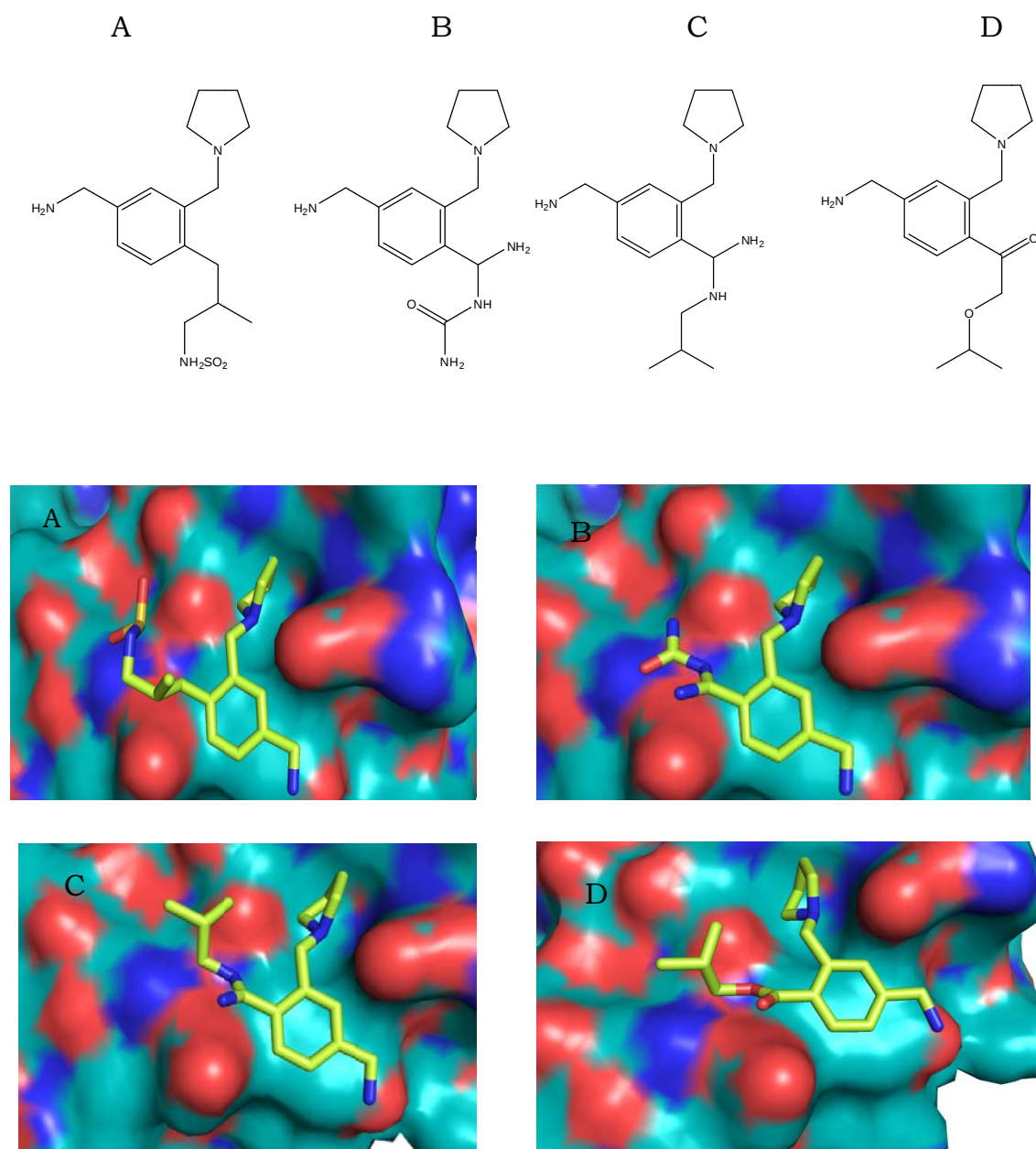


Figure 6.4 An example of the output structures that scored higher in optimising the fragment by growing to Site 2. The functional groups selected were restricted to aliphatic side chains that appeared more suitable for targeting a flatter surface. The sulphonamide group-containing side chain appeared at top of the list.

6.3 Virtual screening

The docking process consists of three main steps: posing, scoring, and ranking. The poses are possible docking solutions for the ligand on the surface of the receptor. The process of scoring estimates how well the ligand fits the docking site and ranking uses the scoring functions to order the poses according to their likelihood of being the correct solution. Three main types of scoring function are used: force-field-based, empirical and knowledge-based.

- (i) Force-field-based scoring functions depend on estimating the binding energies using molecular mechanics force fields. The internal energy of the protein is not calculated as only one protein conformation is used, and only the internal energy of the ligand and the protein-ligand interaction energies are considered. However, the major limitation of this approach is that it does not include solvation and entropic effects. The approach has recently been extended to include protein-ligand, hydrogen-bond terms, for example in GOLD.
- (ii) Empirical scoring functions are derived from experimental data where the binding energy is estimated as the sum of localised interactions, including hydrogen bonds, ionic interactions, hydrophobic interactions and binding entropy. These functions are simpler than the force-field functions but they cannot be easily incorporated in new functions as they use different weighting factors for different terms. These functions are derived from molecular data by fitting and regression analysis.
- (iii) Knowledge-based scoring functions do not estimate binding energies, but rather attempt to reproduce experimental structures instead. Knowledge-based scoring functions can be used for screening large compound libraries due to their simplicity, but they are restricted by the fact that they are derived from information of a limited and possibly unrepresentative sample of protein-ligand data.

6.3.1 Materials and Methods

6.3.1 .1 Sample preparation

A small dataset was retrieved from the ZINC database containing commercially available compounds. The selection was based on the following criteria: a substructure search of the fragment 12C05, molecular weight ≤ 450 Da, rotatable bonds less than 5, hydrogen-bond acceptor ≤ 5 and hydrogen-bond donor ≤ 5 . About 1729 compounds were found to meet these criteria and were used for subsequent docking.

The molecules were prepared using the software **SYBYL® 8.0** provided by Tripos. Hydrogen atoms were added and charges assigned by the Gasteiger-Hückel method.

6.2.1.2 Protein preparation

The protein structures used for this screen were derived from the high resolution, crystal structure analysis of the ankyrin domain with the fragment 12C05 and a high-resolution structure for the domain defined by PDB (2F8Y). Hydrogen atoms were added and the protonation states were assigned according to the surrounding environment. The amide groups of asparagines and glutamines were flipped according to the most likely hydrogen-bond network. Histidine residues were visually inspected and tautomeric states assigned to allow the most favourable interactions.

6.2.1.3 The docking configuration

Virtual screening and docking was carried out using the programme GOLD. The docking configuration file specifies the parameters used in the run. The amide bonds of the ligands were allowed to flip, protonated carboxylic acids were allowed to flip, non-planar sp^3 nitrogens were allowed to invert, free corners of cyclic systems were allowed to flip above or below the plane of neighbouring corners, planar nitrogens bound to sp^2 carbons were allowed to invert between the cis and trans conformations during docking. The ligand conformational space

was restricted by torsion angle distributions extracted from the Cambridge Structural Database (CSD). The docking run was terminated when the three top solutions were within 1.5 Å. Flexibility of the protein was achieved by applying soft potentials to residues TYR2075, ASP2109, ASN2040, GLU2076 LYS2079 and LEU2113. The functional score was GoldScore.

6.2.2 Results and discussion

6.2.2.1 Substructure analysis

Substructure analysis has proved a powerful tool in high-throughput screening, particularly in the early stages where it can be used to analyse hit results and detect false-positives (Merlot *et al.* 2003). Substructure analysis is useful in toxicity prediction by avoiding known toxicophoric functional groups in the molecules or their metabolites (Merlot *et al.* 2003). Substructure search has been used here retrospectively after identifying compounds through the primary screen and identifying pharmacophoric groups crystallographically. This was used as another method to elaborate fragments and introduce new groups that could improve the binding. Compounds containing the substructure fragment 12C05 and meeting the Lipinski rule of five of hydrogen-bond donors and acceptors and molecular weight were retrieved from the ZINC database library. A dataset library of 1729 compounds was created and these molecules were docked on the protein.

6.2.2.2 Protein flexibility

The ankyrin domain does not undergo any major conformational changes upon binding to the fragments or the complex as observed by the crystal structures. However, in reality the protein in solution is in continuous motion and this has been addressed by several methods as described by (Carlson & McCammon, 2000). These methods vary in complexity and flexibility. One approach is to apply conformational sampling of side chains in the receptor, by either allowing

free rotational movements of hydrogen atoms at the receptor site or creating a rotameric library of side chain orientations. Another approach is by generating a sub-ensemble of states by Monte Carlo (MC) or Molecular dynamics (MD) calculations, which are reliable but slow methods. An ensemble of conformations can also be provided by sampling NMR structures or using multiple crystal structures (Carlson & McCammon, 2000). Soft potentials were used to accommodate small changes in conformations. In this method, soft functions are allowed for some clashes between the ligand and the receptor, which has the advantage of a shorter calculation time. There is no distinct binding site at the ankyrin interface to target. Soft potentials were specified for certain residues surrounding the fragment 12C05 in the original crystal structure.

6.2.2.3 Choice of the target template

Virtual screening has been a useful tool where the structure of the target protein has not been validated experimentally either by X-ray crystallography or NMR. A homology model can be used instead for screening small or large libraries (Kitchen *et al.*, 2004). In fact, homology models can sometimes give better docking results. The choice of the crystal structure of the target would influence the docking results. The ideal structure to be used as a template in screening would be the ligand-bound protein structure as even the smallest conformational change could influence the docking results. The ankyrin domain structure was solved using X-ray crystallography at different resolutions. Docking was carried out using two structures: a structure solved at high resolution of 1.5 Å (PDB 2F8Y) and a lower resolution structure of the ankyrin domain with the bound fragment (2.5 Å). High-resolution structures are preferred in docking for obtaining more reliable results. However, the high-resolution structure failed to reproduce the conformation of the fragment 12C05 that has been found experimentally. It was expected that the flexibility of the protein would have little effect since the ankyrin structure remains unchanged before and after complexation, as observed in the crystal structures. This suggested rigidity of the binding interface and an easier docking run. However, the crystal structures do

not always reflect the actual conformation in the solution state. Docking the fragment 12C05 is quite challenging for many reasons; one reason is the lack of restraints in fragment binding in small molecules as small as fragments to predict the correct binding mode, another reason is that most scoring functions have been derived for larger molecules and are not suitable for fragments. It is also difficult to predict the promiscuous binding mode of the fragments. The lower resolution structure of the ankyrin domain was used instead to screen the larger molecules.

6.2.2.4 Choice of scoring function

The docking method and criteria used should be validated by using the fragment as a control molecule for docking. The scoring function used was GoldScore, which is an empirical scoring function chosen by default in the GOLD program. The docking result was then compared with the crystal structure of the fragment. The docking did not result in the exact conformation of the fragment but rather had an RMSD of 4.9 Å with the experimental result. Docking was repeated using Chemscore, another empirical fitness function, and a 1.5 Å resolution structure (PDB 2F8Y). However, the experiment failed again to reproduce the experimental conformation, resulting in an even larger RMSD of 8.4 Å from the reference structure. The ChemScore function differs from GoldScore that it has been derived from measured binding affinity protein-ligand sets. The inability to predict the exact binding mode using the docking method could be attributed to lack of a significant binding pocket or the fact that the fragment is stacked between the two molecules in the asymmetric unit and the docking run included only one ankyrin molecule. The docking was repeated again after including a water molecule that is involved in a water-mediated interaction, but this did not improve the docking result and the fragment was reproduced at RMSD value of 6.44 Å. Some water molecules are highly conserved in the crystal structure and some could be important in mediating an essential protein-ligand interaction. There is a general rule in docking that the RMSD value of the re-docked structure and the reference original structure should not

exceed 2 Å. The docking attempts here failed to reproduce the fragment structure within the recommended RMSD limits. The RMSD limits are applied generally for larger compounds. In fragments, the case could be different; as it has been shown recently that successful docking results of fragments to the beta lactamase CTX-M was proven successful both crystallographically and in inhibition assays (Chen & Shoichet, 2009). The docking results were compared to the crystal structures producing RMSD 2.86 Å, which is larger than the recommended value (Chen & Shoichet, 2009). Docking of the ZINC molecules here was carried out using protein template and scoring function scheme that generated the smallest RMSD. The ligand-bound conformation and the GoldScore function were used in the docking run (Figure 6.5).

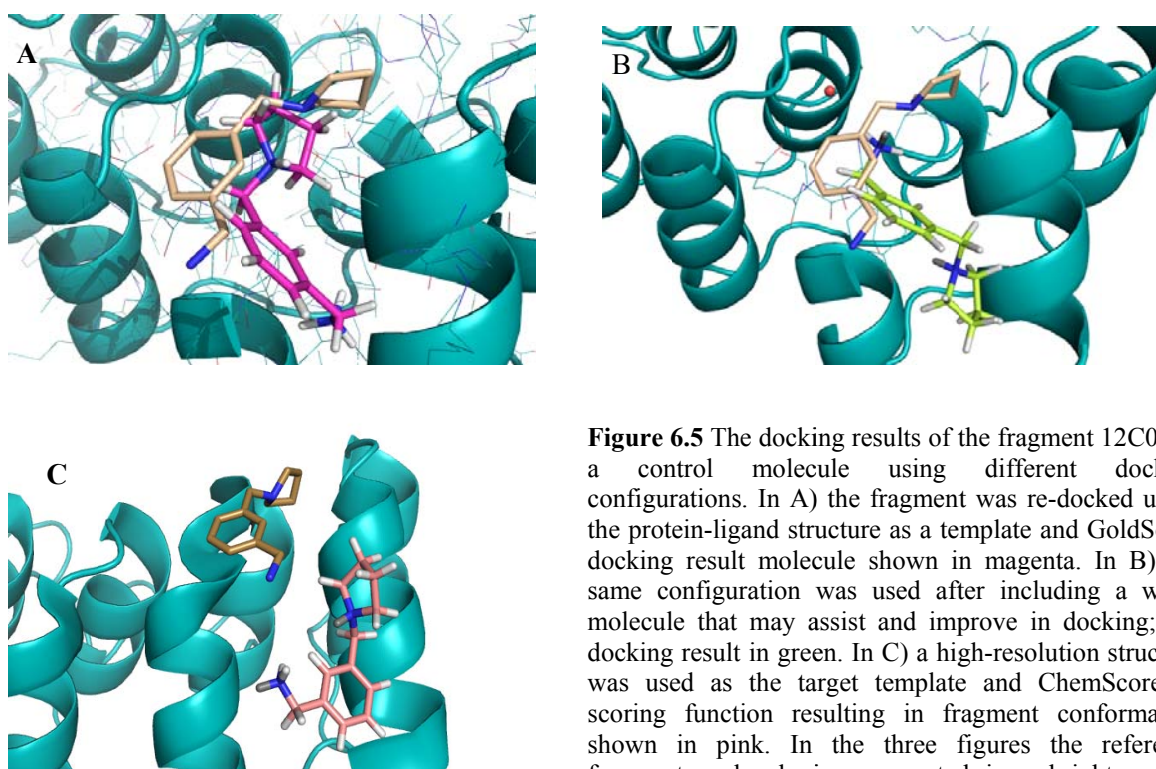


Figure 6.5 The docking results of the fragment 12C05 as a control molecule using different docking configurations. In A) the fragment was re-docked using the protein-ligand structure as a template and GoldScore docking result molecule shown in magenta. In B) the same configuration was used after including a water molecule that may assist and improve in docking; the docking result in green. In C) a high-resolution structure was used as the target template and ChemScore as scoring function resulting in fragment conformation shown in pink. In the three figures the reference fragment molecule is represented in a bright orange

Analysing the docking results requires some experience sometimes to identify the false-positives. Compounds with the highest ranking scores may not be the best candidates. Each compound should be inspected individually and visually.

There are different approaches to minimise the false positives caused by inaccurate scoring functions. One approach is docking and rescoring which is dependant on both the protein and the library to be screened. This would require using a combination of two scoring functions and finding the best combination. Another approach is to use consensus scoring, where top- ranked poses resulting from one scoring function are scored again with multiple scoring functions. The common compounds that are found in both are then considered for further experimental analysis. This method has proven useful in improving docking results. However, combinations of scoring functions need to be tested and chosen carefully. Choosing the correct scoring function or scoring method would require testing the biological or binding activity of these compounds experimentally. This was not available, so the scoring functions were selected on the basis of a comparison of the docking result of the fragment 12C05 and the actual binding mode in the crystal structure.

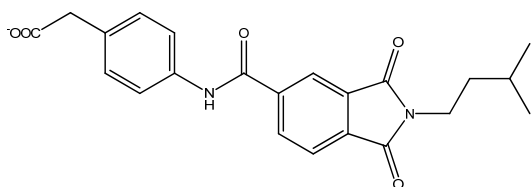
6.2.2.5 Selection of compounds

It is important to remember that the screening and docking process does not accurately measure the binding energies and that it is actually an enrichment process that facilitates the selecting step. It helps in short-listing a larger set of compounds to obtain a smaller dataset that can be easily inspected visually. The relative ranking of those compounds in the shortlist becomes of less significance; compounds ranked in the top five could be as effective as those in the top 50 or 100 lists as long as they are tested experimentally. The library of 1729 compounds extracted from the ZINC database was not large as compared to HTS, which may involve thousands or even millions of compounds. However, it is still subjected to filtering and screening for careful analysis.

The docking results were imported for analysis using the GoldMine software. Four descriptors, scored by the GoldScore function, were used to rank the docking results. The four descriptors were hydrogen bonds, van der Waals interactions, internal torsion angles and normalised fitness descriptors. Most scoring functions are additive resulting in higher scores for larger molecules. It is useful in this case to add a penalty value that is proportional to the molecular

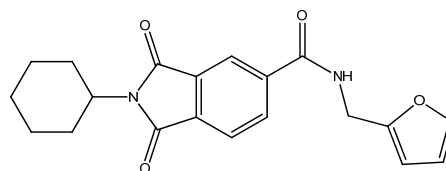
weight; a normalised fitness value was used here instead. The docking results were filtered by using these descriptors in combination with a scoring function. This resulted in 34 compounds with 99 different poses as the top-ranking compounds. The compounds were visualised in order to examine their binding modes and decide on the compounds that would be tested experimentally.

Many factors were considered in selecting the compounds: their overall score, their complementarity to the binding site and their commercial availability. Although the compounds were extracted from ZINC database which is a commercial source, some compounds were inaccessible and not easily available for testing. Two compounds were selected for further investigation. The first (compound 1) had the highest normalised fitness score with 2 different poses and ranked as first and second respectively (Table 6.2). It has 6 rotatable bonds, logP is 3.73, hydrogen bond donors are 2, hydrogen bond acceptors are 5 and its molecular weight was 394 Da. This compound obeys the “Lipinski rule of five” with only one violation of the rotatable bond rule. The isoindole ring of compound 1 is involved in an aromatic π - π interaction with the Y2075 phenyl ring. This is similar to the fragment 12C05, in which the phenyl ring of the fragment is involved in a similar interaction. The methylbutyl side chain was docked in the adjacent pocket and it could be involved in a hydrophobic interaction. The second compound (compound 2) had two poses and ranked in the 24th and 25th positions respectively (Table 6.3). It has a logP value of 3.77, 4 rotatable bonds, one hydrogen bond donor, 4 hydrogen bond acceptors and of molecular weight 352 Da. Both compounds were purchased from Chembridge. The docked structure showed an aromatic π - π interaction between the isoindole ring and Y2075 phenyl ring as well. The furylmethyl side chain was again docked in the adjacent pocket. The aromatic π - π interaction was the only interaction reproduced in docking the two substructure compounds (Figure 6.6).



[4-({2-(3-methylbutyl)-1,3-dioxo-2,3-dihydro
1,3-
-1H-isoindol-5-yl}carbonyl)amino)phenyl]acetic acid

Compound 1



2-cyclohexyl-N-(2-furylmethyl)-
dioxo-5-isoindolinecarboxamide

Compound 2

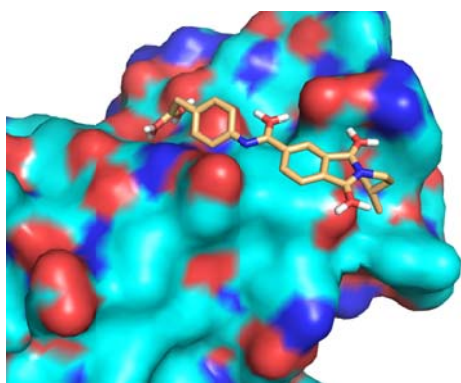
Table 6.2 Summary of scores of compound 1

	External HB	External Vdw	Internal torsion	Normalised fitness	Rank
Pose 1	6.3137	69.673	-0.4727	0.188188	1
Pose 2	6.9057	70.4112	-0.4668	0.188130	2

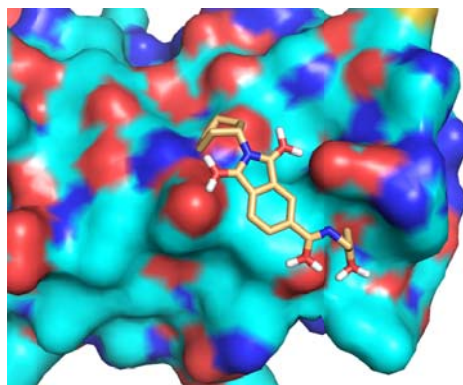
Table 6.3 Summary of scores of compound 2

	External HB	External Vdw	Internal torsion	Normalised fitness	Rank
Pose1	3.8179	56.4138	-0.2255	0.1706565	24
Pose 2	3.7001	56.5116	-0.2206	0.1705960	25

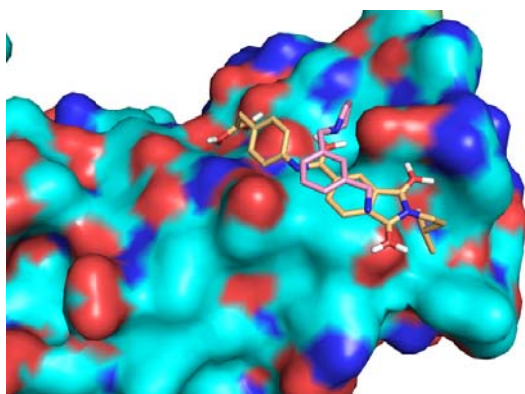
A



B



C



D

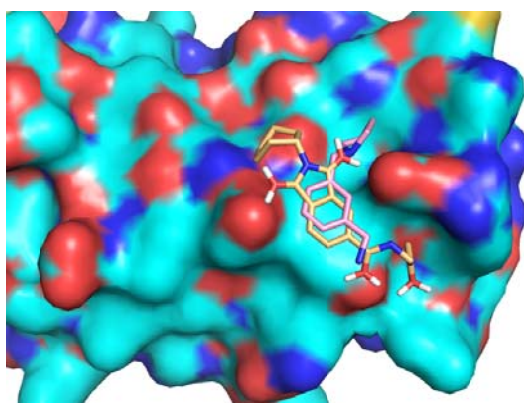


Figure 6.6 A surface view of the ankyrin domain with the docked structures of A) compound 1 and B) compound 2. The structures are then compared with the original fragment in C) and D) showing the common phenyl ring at an approximate same orientation compared to the fragment12C05.

6.2.3 Testing compounds experimentally

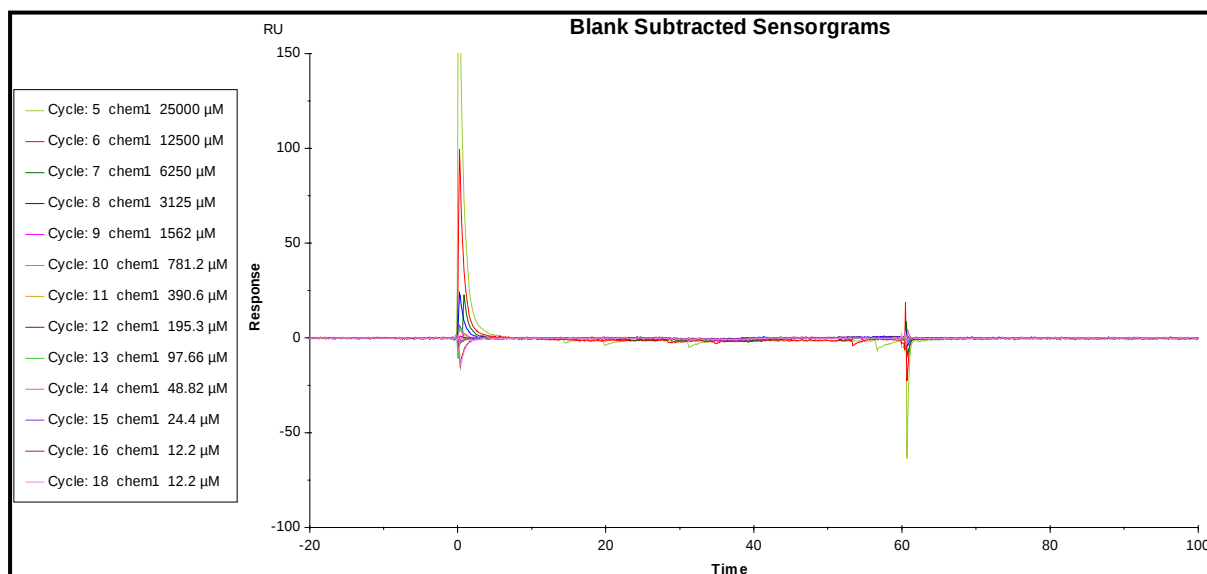
6.2.3.1 Detecting binding by Surface Plasmon resonance

The ankyrin domain was immobilised on a CM5 chip as it has been described in Chapter 5. A series of concentrations of the two compounds was prepared by serial dilution. The compounds were dissolved in a PBS buffer (phosphate buffer pH7.5, 150 mM NaCl, 0.05% Tween 20) in 5 % DMSO. The concentration series included the following concentrations: 25, 12.5, 6.25, 3.125, 1.562, 0.781, 0.391, 0.195, 0.977, 0.488, 0.244 and 0.122 mM. Solution of the compounds were injected onto the surface of the chip at a flow rate 30 μ l/min, contact time was 60 seconds and dissociation time was 60 seconds. A series of aliquots of PBS in DMSO was prepared for solvent correction as described in chapter 5. Regeneration was carried out by injecting 1M NaCl for 60 seconds. The run was carried out at a temperature 25°C. The sensorgrams were analysed using the Biacore T100 evaluation software and solvent correction was applied.

6.2.3.2 Results and discussion

Biacore has been used for screening the binding of the two compounds at various concentrations. The two compounds had poor solubility at higher concentrations but did not show any aggregation in the sensorgrams. The two compounds were dissolved in 5%DMSO and solvent correction was applied. Neither compounds showed any sign of binding and the sensorgrams returned rapidly to the baseline (Figure 6.7).

A



B

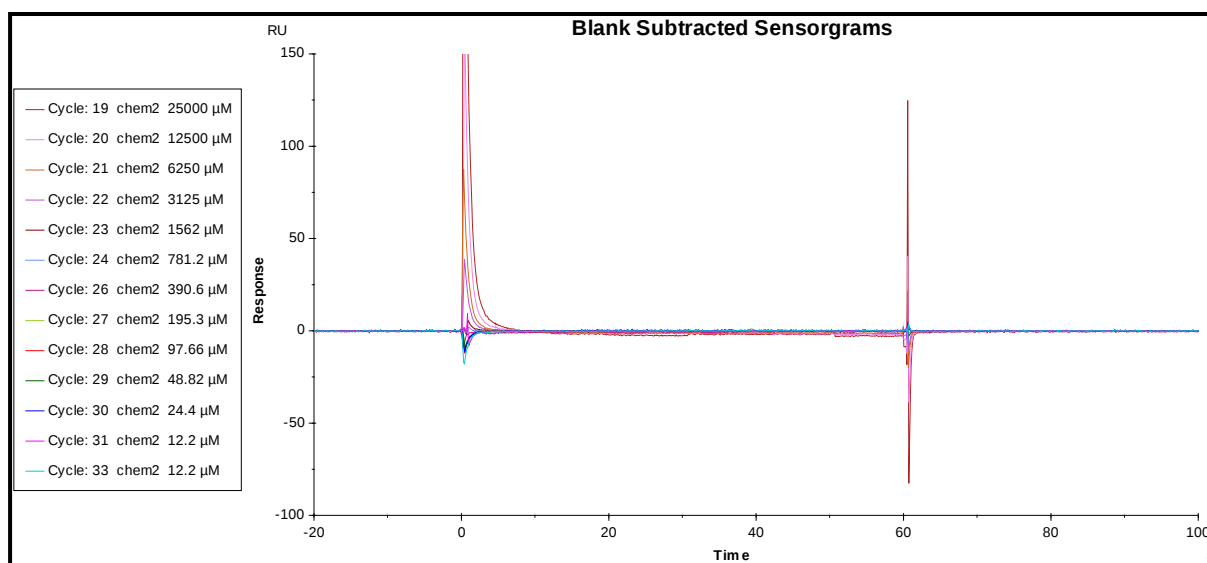


Figure 6.7 The sensorgrams of A) compound 1 and B) compound 2 showing no binding at a series of concentrations. The blank subtracted sensorgrams are solvent corrected to compensate for the effect of DMSO. The spikes observed mark the start and end of injection event.

Neither of the compounds showed any binding in SPR and no electron density was observed in the crystal structures. Although these compounds were highly-ranked in the docking screen, they failed to bind in experimental, solution

conditions. This could result from the inability of the virtual screening to predict the binding correctly or incorrect assignation of the protonation states of protein and ligands.

The docking protocol could have been improved by optimising the docking speed and accuracy by setting the docking efficiency and number of genetic algorithm attempts to different values. Exploration of different scoring functions was found to be the best way forward; in these experiments the docking protocol was optimised initially for small fragments but probably should be re-optimised for larger molecules. The limited solubility of the two compounds even in 100% DMSO also demonstrated that this can be an obstacle and it is not considered in the docking screens.

A major difficulty in virtual screening is its inability to identify potential binders. This is likely because they have been ranked poorly and disregarded incorrectly. Clashes with the protein receptor and other non-optimal ligand-receptor contacts, difficulties in predicting the ionisation state of groups on the protein or ligand and identifying changes in ionisation on ligand binding, and incorrect choice of ions or water molecules could all contribute to this (Klebe, 2006).

6.3 Summary and Conclusions

Computational methods have become essential tools and have been integrated fully into drug discovery, both in high-throughput screening and in lead optimisation. Two different structure-based methods have been employed here to design larger compounds with improved binding affinity. This study was focused only on one fragment (12C05) that was validated both crystallographically and in kinetic studies. The first approach used the software SPROUT to both link this fragment to another and to optimise it by building a small library of selected functional groups that were docked in selected sites. Some of these compounds

appeared unattractive for further follow-up either due to their challenging synthesis or to their relative lipophilicity or hydrophilicity. However, such compounds can still give some ideas about possible approaches for fragment optimisation. A major limitation of SPROUT is that it does not allow for protein flexibility. This could be overcome by re-docking the generated structures using docking programmes that allow flexibility of residues in binding sites.

The second approach used was virtual screening, which allowed docking a small commercially available library to the expected binding site. The program GOLD was used to shortlist the compounds that were likely to bind. Only two of the 34 compounds in the shortlist were selected for experimental validation. These did not show any binding when measured by SPR. The rest were not tested, so it is not known whether these two compounds were false-positives or whether the docking and screening protocol needed further optimisation. However, the virtual screening method is still vital in the drug design process. It does, nevertheless, require experimental validation to optimise the docking parameters, which may differ according to the protein target and the nature of the compounds in the library.

Chapter 7

Conclusions and future directions

7.1 Conclusions

The Notch signalling pathway is an important conserved pathway that has attracted much interest over the years. The Notch receptor and ligand are transmembrane proteins where ligand binding induces the release of an intracellular domain of the Notch receptor. The intracellular domain translocates to the nucleus where it is involved in the formation of an active transcription complex with CSL and MAML. Notch signalling is involved in different developmental processes; and abnormalities in the expression or function of different components of the Notch receptor are implicated in various diseases and cancers. There have been several approaches to development of therapeutic agents that modulate Notch signalling. Targeting the downstream events through interfering with the formation of the transcription complex is one direction that could be investigated. The ankyrin domain of the Notch intracellular domain plays an instrumental role in formation of the active transcription complex. In this work, I have explored the druggability of the ankyrin domain using a fragment-based approach

Several computational methods developed to predict druggability rely on geometrical and energy-based algorithms. However, many of these methods are incapable of predicting binding sites on protein-protein interfaces due to the lack of distinctive binding pockets.

In this study, I assessed the druggability of the ankyrin domain experimentally by screening against a commercial library. The hit rate and nature of hits depend on the design of the screening library, especially the nature of the

chemical moieties and the chemical space that it covers. In addition, it relies on the screening method, including its sensitivity and its ability to distinguish false positives and false negatives from true binders. In this dissertation, the screening library included fragments that complied with the “rule of three” (Congreve *et al.*, 2003). The assembly of the library was not target-specific as this project was part of a larger drug discovery campaign screening a wide range of targets. The fluorescent-based thermal shift screening assay required the protein to unfold in a two-state manner in order to compare unfolding in presence and absence of fragments. After optimisation of screening conditions, 36 fragment hits were identified from a library of 1201 compounds giving a hit rate approximately 3%. The identified hits were grouped in five main classes: benzyl derivatives, fused bicyclic rings, biaryl compounds, phenyl derivatives, and 5-membered heterocyclic rings. Although the frequency of these hits could be a function of the representation of these compounds in the screening library, these chemical scaffolds seemed favourable to bind with the ankyrin domain. These small molecules are the first identified to bind to the Notch1 ankyrin domain.

The affinities of the confirmed fragment hits were measured using Surface Plasmon Resonance (SPR). As expected, they were mostly in the millimolar range, not unusual for smaller molecules. The on- and off-rates of some fragments were measured after fitting a 1:1 model, but other fragments could not be analysed due to very fast on- and off-rates leading to a square-shaped sensorgram. The fragments were prioritised according to their affinity values, and then chemically optimised in order to improve their binding affinities. The lack of improvement led to the conclusion that the initial hits were the most favourable chemical scaffolds for binding the ankyrin domain.

The binding mode of the fragment hits was then investigated using X-crystallography. A small subset of two fragments, 12C05 and 9F07, was identified binding in the crystal structures. Probably owing to their very weak binding affinities, low occupancies and very fast off-rates, other fragments were not visible in the electron density maps. However, the two successful fragment-bound structures gave interesting insights into how small molecules can bind to

the ankyrin domain. Although it was expected that the fragments would bind in the concave region formed by the ankyrin-repeat architecture, they were found to bind residues in the upper helices involved in protein-protein interactions. This strongly suggests the possibility of mimicking protein-protein interactions by small molecules at the ankyrin interaction interface. However, elaborating the fragments and developing them to potential inhibitors is likely to be quite challenging given the flat surface and lack of anchoring small cavities adjacent to the two fragments.

Computational tools were employed to grow and link the two fragments together. A docking programme (GOLD) and *de novo* drug design software (SPROUT) assisted in finding and designing larger compounds based on the original scaffold of the fragment 12C05. Two compounds were selected and were tested experimentally but did not show any binding. However, these two compounds do not represent the complete docking output and other compounds need to be tested.

7.2 Lessons learnt for future directions

Choice of target

On embarking on a drug discovery program it is important to choose the target carefully. The first criterion to consider is the the biology and function of the target and how closely it is linked to the human disease. The availability and the ability to express and produce the target protein in a reproducible fashion becomes an essential requirement for a highly iterative process. In addition, the presence of well-defined pockets is usually assessed to identify the druggability of the target. The ankyrin domain is a good example of a protein whose biological function is well understood and its involvement in progression of T-ALL is characterised. However, the druggability of a protein fold comprising ankyrin repeats had not been investigated before. The results shown here indicate that it is possible to target the Notch1 ankyrin domain with small molecules, although optimisation of these molecules remains challenging. The ankyrin repeat is a

scaffold that mediates many protein-protein interactions and the discovery of small molecules that could bind and inhibit these interactions of wide interest for a number of different therapeutic interventions.

7.3 Future directions:

7.3.1 Choice of Target

Progress in targeting the ankyrin domain using a fragment-based approach has been encouraging and could be beneficial to inhibition of the formation of the active Notch transcription complex. However, in view of the fact that the ankyrin domain binds first with CSL and then MAML is recruited to stabilise the complex, the ankyrin-CSL binary complex might alternatively be used as a target to screen small molecules that either stabilise or destabilise the transcription complex.

7.3.2 Library design

Fragment libraries cover a larger chemical space and are easier to assemble and synthesise than HTS libraries. The relatively high hit rate that was shown by initial screening of the ankyrin domain suggests that the fragment-based approach could be a successful alternative for targeting protein-protein interactions. However, the hit rate could always be improved by designing a more tailored and target-specific library. Enriching the library with molecules similar to amino acids might enhance the success rates for difficult protein-protein interfaces. Although the physical properties are governed by the “rule of three” that restricts the choice of fragments, fragment solubility is particularly important for technical and practical reasons. As has been shown here with the ankyrin domain, solubility of fragments was a limiting factor for hit identification, kinetic characterisation of fragment hits and protein-ligand crystallisation attempts. It is arguable that fragments are hydrophobic small molecules and improving solubility could be achieved at later optimisation stages. However, hit-to-lead progression starts by identifying and prioritising the

most promising initial molecules that could be missed in early screening due to solubility problems.

A more focused library could be designed based on the confirmed hits that were found. Elimination of any functional groups that would destabilise the protein such as the carboxylic and trifluoromethyl groups and investigating a broader chemical space would help in enhancing the hit rates and exploring different scaffolds that could be more favourable and more potent for binding.

7.3.3 Fragment Screening

Orthogonal screening can confirm initial hits by testing the compounds using two different methods. The screening of the ankyrin domain was carried out using a thermal-based shift assay that has proved to be a successful screening approach to hit identification. The positive hits were then confirmed by SPR for kinetic and affinity measurements. However, different screening methods with different sensitivities could lead to different outputs, especially when different methods require particular screening conditions such as buffer composition and temperature. The mismatch of screening conditions could result in hit variations from one method to another, so careful assessment of the quality of hits by each method is necessary. Publications usually list the positive hits, discarding negative molecules that either do not bind or are regarded as promiscuous. Such information is useful when it comes to designing target-specific libraries by understanding which chemical moieties should be avoided. Screening the ankyrin domain using the thermal-shift assay has detected compounds that destabilise the protein by reducing its melting point. These compounds contained certain functional groups that could be avoided in future tailored libraries or during hit-to-lead optimisation. In addition, in case of difficult drug targets it is important to be able to distinguish if the hit-rate is a true representation of the druggability of target or due to technical and optimisation difficulties of the screening method itself.

Three biophysical methods were used here for screening and testing the binding of fragments to the ankyrin domain: fluorescent-based thermal shift assay, SPR and X-ray crystallography. Other complementary methods could be employed, such as heteronuclear NMR-based screening, which has very low false-positive and false-negative rates, and is able to identify new and different scaffolds.

References

Adachi J., Mori Y., Matsui S., Takigami H., Fujino J., Kitagawa H., Miller 3rd C.A., Kato T., Saeki K., Matsuda T. (2001). Indirubin and indigo are potent aryl hydrocarbon receptor ligands present in human urine, *J. Biol. Chem.* **276** :31475–31478.

Anderson A.C. (2003) The process of structure-based drug design. *Chem Biol.* Sep;**10**(9):787-97.

Arkin M.R., Wells J.A. (2004). Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov.* 2004 Apr;**3**(4):301-17.

Artavanis-Tsakonas S., Rand M. D., Lake R. J., (1999). "Notch signaling: cell fate control and signal integration in development." *Science*, **284**(5415): 770-6.

Aste-Amézaga M, Zhang N et al., (2010). Characterization of Notch1 antibodies that inhibit signaling of both normal and mutated Notch1 receptors. *PLoS One.* **5**(2):

Bain J., McLauchlan H., Elliott M., Cohen P. (2003). The specificities of protein kinase inhibitors: an update, *Biochem. J.***371** :199–204.

Bertrand J.A., Thieffine S., Vulpetti A., Cristiani C., Valsasina B., Knapp S., Kalisz H.M., Flocco M. (2003). Structural characterization of the GSK-3 β active site using selective and non-selective ATP-mimetic inhibitors, *J. Mol. Biol.* **333** :393–407.

Boehm, H.J. *et al.* (2000) Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.* **43**, 2664–2674

Bogan AA, Thorn KS. (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol.* **3**;280(1):1-9.

Bohacek, R.S., McMartin, C., Guida, W.C. (1996) The art and practice of structure based drug design: a molecular modelling perspective. *Med. Res. Rev.* **16**, 3–50.

Bradley CM, Barrick D. (2002). Limits of cooperativity in a structurally modular protein: response of the Notch ankyrin domain to analogous alanine substitutions in each repeat. *J Mol Biol.* **324**(2):373-86.

Bradley CM, Barrick D. (2005). Effect of multiple prolyl isomerization reactions on the stability and folding kinetics of the notch ankyrin domain: experiment and theory. *J Mol Biol.* **352**(2):253-65.

Bradley CM, Barrick D. (2006). The notch ankyrin domain folds via a discrete, centralized pathway. *Structure.* **14**(8):1303-12.

Bray S. (1998). Notch signalling in *Drosophila*: three ways to use a pathway. *Semin Cell Dev Biol.* 1998 Dec;9(6):591-7.

Brou, C., Logeat, F., Gupta, N., Bessia, C., LeBail, O., Doedens, J. R., Cumano, A., Roux, P., Black, R. A., Israel, A. (2000). "A novel proteolytic cleavage involved in Notch signaling: the role of the disintegrin-metalloprotease TACE." *Mol Cell* **5**(2): 207-16.

Bruckner, K., L. Perez, *et al.* (2000). "Glycosyltransferase activity of Fringe modulates Notch-Delta interactions." *Nature* **406**(6794): 411-5.

Carlson HA, McCammon JA. (2000). Accommodating protein flexibility in computational drug design. *Mol Pharmacol.* Feb;**57**(2):213-8.

Chakrabarti, P. & Bhattacharyya, R. (2007). Geometry of nonbonded interactions involving planar groups in proteins. *Progress in Biophysics and Molecular Biology* **95**:83-137.

Chang G, Roth CB, Reyes CL, Pornillos O, Chen YJ, Chen AP. (2006) Retraction. *Science.* Dec 22;**314**(5807):1875.

Chen Y., Shoichet B.K. (2009). Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat Chem Biol.* May;**5**(5):358-64.

Clackson T, Wells JA. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science.* **267**(5196):383-6.

Congreve M, Carr R, Murray C, Jhoti H. (2003). A 'rule of three' for fragment-based lead discovery? *Drug Discov Today.***8**(19):876-7.

Coleman ML., McDonough MA., Hewitson KS., Coles C., Mecinovic J., Edelmann M., Cook KM., Cockman ME., Lancaster DE., Kessler BM., Oldham NJ., Ratcliffe PJ., Schofield CJ. (2007). Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem.* **282**(33):24027-38.

Copeland R.A., Pompliano D.L., Meek T.D. (2006). Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov.* Sep;**5**(9):730-9.

Cordle J., Johnson S., et al. (2008). "A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition." *Nat Struct Mol Biol* **15**(8): 849-57.

D'Arcy A, Villard F, Marsh M. (2007). An automated microseed matrix-screening method for protein crystallization. *Acta Crystallography D Biol Crystallogr.* Apr;**63**(Pt 4):550-4.

Dalvit C, Pevarello P, Tatò M, Veronesi M, Vulpetti A, Sundström M. (2000). Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J Biomol NMR.* **18**(1):65-8.

Danley D.E. (2006) Crystallization to obtain protein-ligand complexes for structure-aided drug design. *Acta Crystallogr D Biol Crystallogr.* Jun;**62**(Pt 6):569-75.

de la Pompa JL, Wakeham A, Correia KM, Samper E, Brown S, Aguilera RJ, Nakano T, Honjo T, Mak TW, Rossant J, Conlon RA. (1997). Conservation of the Notch signalling pathway in mammalian neurogenesis. *Development.* **124**(6):1139-48.

Diederich, R. J., Matsuno, K., Hing, H. and Artavanis-Tsakonas, S. (1994) Cytosolic interaction between Deltex and Notch ankyrin repeats implicates Deltex in the Notch signaling pathway. *Development* **120**:473–481.

Del Bianco C., Aster J.C., Blacklow S.C. (2008). Mutational and energetic studies of Notch 1 transcription complexes. *J Mol Biol.* **376**(1):131-40.

DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. (2000). Convergent solutions to binding at a protein–protein interface. *Science* **287**, 1279–1283

Dexter, J. S. (1914). "The Analysis of a Case of Continuous Variation in *Drosophila* by a Study of Its Linkage Relations." *The American Naturalist* 48(576): 712-758.

Ehebauer MT, Chirgadze DY, Hayward P, Martinez Arias A, Blundell TL. (2005) High-resolution crystal structure of the human Notch 1 ankyrin domain. *Biochem J.* Nov 15;392(Pt 1):13-20.

Eisenbrand G., Hippe F., Jakobs S., Muehlbeyer S. (2004). Molecular mechanisms of indirubin and its derivatives: novel anticancer molecules with their origin in traditional Chinese phytomedicine, *J. Cancer Res. Clin. Oncol.* **130**:627–635.

Ellisen LW, Bird J, West DC, Soreng AL, Reynolds TC, Smith SD, Sklar J. (1991). TAN-1, the human homolog of the *Drosophila* notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell.* **66**(4):649-61.

Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem.* 357(2):289-98

Espinosa L., Ingles-Esteve J., Aguilera C., Bigas A. (2003) Phosphorylation by glycogen synthase kinase-3 beta down regulates Notch activity, a link for Notch and Wnt pathways, *J. Biol. Chem.* **278**: 32227–32235.

Ewing T.J., Makino S., Skillman A.G., Kuntz I.D. (2001) DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Design* **15**: 411–428.

Fejzo, J. *et al.* (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* **6**, 755–769

Foltz D.R., Santiago M.C., Berechid B.E., Nye J.S. (2002) Glycogen synthase kinase-3-beta modulates notch signaling and stability, *Curr. Biol.* **12** :1006–101.

Friedmann, D. R., J. J. Wilson, et al. (2008). "RAM-induced allostery facilitates assembly of a notch pathway active transcription complex." *J Biol Chem* **283**(21): 14781-91.

Fry DC. (2006). Protein-protein interactions as targets for small molecule drug discovery. *Biopolymers.* **84**(6):535-52.

Giannetti A.M., Koch B.D., Browner M.F. (2008). Surface plasmon resonance based assay for the detection and characterization of promiscuous inhibitors. *J Med Chem.* **51**(3):574-80.

Gillet V.J., Johnson A.P., Mata P., Sike S., Williams P.(1993). SPROUT: A program for structure generation. *J. Comput.-Aided Mol. Design* **7** :127-153.

Gillet V.J., Myatt G, Zsoldos Z, Johnson AP in *De Novo Design* (1995) K.Müller(Ed.) Escom ,Leidn. **3** :34-50.

Gillet V.J., Newell W., Mata P., Myatt G., Sike S., Zsoldos Z., Johnson A.P. (1994) SPROUT: recent developments in the de novo design of molecules. *J Chem Inf Comput Sci.* **34**(1):207-17.

Gillet VJ. (2004). Designing combinatorial libraries optimized on multiple objectives. *Methods Mol Biol.* **275**:335-54.

Gordon WR, Vardar-Ulu D, Histen G, Sanchez-Irizarry C, Aster JC, Blacklow SC. (2007). Structural basis for autoinhibition of Notch. *Nat Struct Mol Biol.* **14**(4):295-300.

Gordon, W. R., K. L. Arnett, et al. (2008). "The molecular logic of Notch signaling--a structural and biochemical perspective." *J Cell Sci* **121**(Pt 19): 3109-19.

Hajduk PJ, Bures M, Praestgaard J, Fesik SW. (2000). Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem.* **43**: 3443-3447

Hajduk PJ, Huth JR, Fesik SW. (2005a). Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem.* **48**(7):2518-25.

Hajduk PJ, Huth JR, Tse C. (2005b). Predicting protein druggability. *Drug Discov Today.* **10**(23-24):1675-82.

Hajduk PJ, Greer J. (2007). A decade of fragment-based drug design: strategic advances and lessons learned. *Nature Reviews Drug Discovery.* **6**, 211-219.

Hatakeyama J, Bessho Y, Katoh K, Ookawara S, Fujioka M, Guillemot F, Kageyama R. (2004). Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. *Development.* **131**(22):5539-50

Higueruelo A.P., Schreyer A., Bickerton G.R., Pitt W.R., Groom C.R., Blundell T.L. (2009). Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chem Biol Drug Des.* **74**(5):457-67.

Ho C. M.W., Marshall G.R. (1993). SPLICE: A program to assemble partial query solutions from three-dimensional database searches into novel ligands. *J. Comput.-Aided Mol. Design.* **7**(6): 623-647.

Hopkins AL, Groom CR, Alex A. (2004). Ligand efficiency: a useful metric for lead selection. *Drug Discov Today.* **9**(10):430-1.

Hunt TM (1917). "The theory of the gene".*American Naturalist* 51 (609): 513–54

Ishibashi M, Ang SL, Shiota K, Nakanishi S, Kageyama R, Guillemot F.(1995). Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects. *Genes Dev.* **9**(24):3136-48.

Itai.A, Nishibata.Y. (1991) Automatic creation of drug candidate structures based on receptor structure. Starting point for artificial lead generation. *Tetrahedron.* **47** :8985-8990.

Jones G., Willett P., Glen R.C. (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **245**: 43–53.

Joutel A, Andreux F, Gaulis S, Domenga V, Cecillon M, Battail N, Piga N, Chapon F, Godfrain C, Tournier-Lasserre E. (2000). The ectodomain of the Notch3 receptor accumulates within the cerebrovasculature of CADASIL patients. *J Clin Invest.* **105**(5):597-605.

Kidd S, Kelley MR, Young MW. (1986). Sequence of the notch locus of *Drosophila melanogaster*: relationship of the encoded protein to mammalian clotting and growth factors. *Mol Cell Biol* 6:3094-108.

Kim IC., Hamilton AD.(2006). Diphenylindane-Based Proteomimetics Reproduce the Projection of the i,i+3, i+4, and i+7 Residues on an α -Helix *Org.Lett.*, **8** (9), 1751–1754

Kim YW, Verdine GL. (2009). Stereochemical effects of all-hydrocarbon tethers in i,i+4 stapled peptides. *Bioorg Med Chem Lett.* **19**(9):2533-6.

Kitchen D.B., Decornez H., Furr J.R., Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov.* 2004 **3**(11):935-49.

Kurooka, H. and Honjo, T. (2000) Functional interaction between the mouse Notch1 intracellular region and histone acetyltransferases PCAF and GCN5. *J. Biol. Chem.* **275**:17211–17220

Ladbury JE, Klebe G, Freire E. (2010). Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat Rev Drug Discov.* **9**(1):23-7.

Lee M.J., Kim M.Y., Mo J.S., Ann E.J., Seo M.S., Hong J.A., Kim Y.C., Park H.S. (2008). Indirubin-3'-monoxime, a derivative of a Chinese anti-leukemia medicine, inhibits Notch1 signaling. *Cancer Lett.* **265**(2):215-25.

Li K, Li Y, Wu W, Gordon WR, Chang DW, Lu M et al.,(2008). Modulation of Notch signaling by antibodies specific for the extracellular negative regulatory region of NOTCH3. *J Biol Chem.* **283**(12):8046-54.

Liebeschuetz, J.W. et al. (2002) PRO_SELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery. The development of a series of highly potent and selective factor Xa inhibitors. *J. Med. Chem.* **45**: 1221–1232.

Lindner V, Booth C, Prudovsky I, Small D, Maciag T, Liaw L. (2001). Members of the Jagged/Notch gene families are expressed in injured arteries and regulate cell phenotype via alterations in cell matrix and cell-cell interaction. *Am J Pathol.* **159**(3):875-83.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ.(2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* **46**(1-3):3-26.

Lo MC, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, Ellestad G. (2004). Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem.* **332**(1):153-9.

Logeat, F., Bessia, C., Brou, C., LeBail, O., Jarriault, S., Seidah, N. G., Israel, A. (1998). "The Notch1 receptor is cleaved constitutively by a furin-like convertase." *Proc Natl Acad Sci U S A* **95**(14): 8108-12.

Lusty J.C. (1999). A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. *J. Appl. Cryst.* **32** :106±112.

Maillard I, Weng AP, Carpenter AC, Rodriguez CG, Sai H, Xu L, Allman D, Aster JC, Pear WS. (2004). Mastermind critically regulates Notch-mediated lymphoid cell fate decisions. *Blood*. **104**(6):1696-702.

Malecki MJ, Sanchez-Irizarry C, Mitchell JL, Histen G, Xu ML, Aster JC, Blacklow SC. (2006). Leukemia-associated mutations within the NOTCH1 heterodimerization domain fall into at least two distinct mechanistic classes. *Mol Cell Biol.* **26**(12):4642-51.

Marambaud P, Shioi J, Serban G, Georgakopoulos A, Sarner S, Nagy V, Baki L, Wen P, Efthimiopoulos S, Shao Z, Wisniewski T, Robakis NK. (2002). A presenilin-1/gamma-secretase cleavage releases the E-cadherin intracellular domain and regulates disassembly of adherens junctions. *EMBO J.* **21**(8):1948-56.

Matsuno, K., Diederich, R. J., Go, M. J., Blaumueller, C. M. and Artavanis-Tsakonas, S. (1995) Deltex acts as a positive regulator of Notch signalling through interactions with the Notch ankyrin repeats. *Development* 121, 2633–2644

Matsuno K., Go M.J., Sun X., Eastman D.S., Artavanis-Tsakonas S. (1997). Suppressor of Hairless-independent events in Notch signaling imply novel pathway elements. *Development*. 1997 Nov;124(21):4265-73.

Matthews, B. W. (1968). *J. Mol. Biol.* 33:491–497.

Matthews, B. W. (1974). *J. Mol. Biol.* 82:513–526.

McDaniell R, Warthen DM, Sanchez-Lara PA, Pai A, Krantz ID, Piccoli DA, Spinner NB. (2006). NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *Am J Hum Genet.* **79**(1):169-73.

Merlot C., Domine D., Cleve C. , Church D.J. (2003). Chemical substructures in drug discovery. *Drug Discovery Today* **8** (13):594-602.

Milano J, McKay J, Dagenais C, Foster-Brown L, Pognan F, Gadiant R, Jacobs RT, Zacco A, Greenberg B, Ciaccio PJ.(2004). Modulation of notch processing by gamma-secretase inhibitors causes intestinal goblet cell metaplasia and induction of genes known to specify gut secretory lineage differentiation. *Toxicol Sci.* **82**(1):341-58.

Moellering RE, Cornejo M, Davis TN, Del Bianco C, Aster JC, Blacklow SC, Kung AL, Gilliland DG, Verdine GL, Bradner JE. (2009). Direct inhibition of the NOTCH transcription factor complex. *Nature.* **462**(7270):182-8.

Mooij WT, Hartshorn MJ, Tickle IJ, Sharff AJ, Verdonk ML, Jhoti H. (2006). Automated protein-ligand crystallography for structure-based drug design. *ChemMedChem.* **1**(8):827-38.

Moloney, D. J., V. M. Panin, et al. (2000). "Fringe is a glycosyltransferase that modifies Notch." *Nature* **406**(6794): 369-75.

Morris G. M., D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**:1639–1662.

Mumm, J. S. and R. Kopan (2000). "Notch signaling: from the outside in." *Dev Biol* **228**(2): 151-65.

Nam Y, Sliz P, Song L, Aster JC, Blacklow SC. (2006) Structural basis for cooperativity in recruitment of MAML co-activators to Notch transcription complexes. *Cell*. Mar **10**; 124(5):973-83.

Nam Y., Sliz P., Pear W.S., Aster J.C., Blacklow S.C. (2007). Cooperative assembly of higher-order Notch complexes functions as a switch to induce transcription. *Proc Natl Acad Sci U S A*. **104**(7):2103-8.

Okajima T. and Irvine K. D. (2002). "Regulation of notch signaling by o-linked fucose." *Cell* 111(6): 893-904.

Oswald, F., T'auber, B., Dobner, T., Bourteele, S., Kostezka, U., Adler, G., Liptay, S., Schmid, R. M. (2001) p300 acts as a transcription coactivator for mammalian Notch-1. *Mol. Cell. Biol.* **21**:7761–7774.

Pantoliano Michael W., Petrella Eugene C., Kwasnoski Joseph D., Lobanov Victor S. et al. (2001) High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *J Biomol Screen* **6**: 429.

Panin V. M., V. Papayannopoulos, et al. (1997). "Fringe modulates Notch-ligand interactions." *Nature* **387**(6636): 908-12.

Pelletier L, Guillaumot P, Frêche B, Luquain C, Christiansen D, Brugière S, Garin J, Manié SN. (2006). Gamma-secretase-dependent proteolysis of CD44 promotes neoplastic transformation of rat fibroblastic cells. *Cancer Res.* **66**(7):3681-7.

Pitt WR, Parry DM, Perry BG, Groom CR. (2009). Heteroaromatic rings of the future. *J Med Chem.* **52**(9):2952-63.

Rebay, I., R. G. Fehon, et al. (1993). "Specific truncations of *Drosophila* Notch define dominant activated and dominant negative forms of the receptor." *Cell* **74**(2): 319-29.

Rizzo P, Osipo C, Foreman K, Golde T, Osborne B, Miele L. (2008). Rational targeting of Notch signaling in cancer. *Oncogene.* **27**(38):5124-31.

Roche P., Schneider O., Zuegge J., Guba W., Kansy M., Alanine A., et al. (2002). Development of a virtual screening method for identification of "Frequent Hitters" in compound libraries, *J. Med. Chem.* **45**:137–142

Roehl H., Bosenberg M., Blelloch R., Kimble J. (1996). Roles of the RAM and ANK domains in signaling by the *C. elegans* GLP-1 receptor. *EMBO J.* **15**(24):7002-12.

Schweisguth, F. (2004). "Regulation of Notch signaling activity." *Curr Biol* **14**(3): R129-38.

Sevcik J., Dodson E.J., Dodson G.G. (1991). Determination and restrained least-squares refinement of the structures of ribonuclease Sa and its complex with 3'-guanylic acid at 1.8 Å resolution. *Acta Crystallogr B.* **47** (2):240-53.

Schafmeister GE., Po J., and Verdine GL. An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *J. Am. Chem. Soc.*, **122** (24), pp 5891–5892

Shih, I.-M. & Wang, T.-L. (2007). Notch signaling, γ -secretase inhibitors, and cancer therapy. *Cancer Res.* **67**, 1879–1882.

Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science*. **274**(5292):1531-4.

Sperandio O, Reynès CH, Camproux AC, Villoutreix BO. (2010). Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov Today*. **15**(5-6):220-9.

Struhl, G., K. Fitzgerald, et al. (1993). "Intrinsic activity of the Lin-12 and Notch intracellular domains in vivo." *Cell* **74**(2): 331-45.

Verheij, H.J. (2006) Leadlikeness and structural diversity of synthetic screening libraries. *Mol. Divers.* 10, 377–388

Vidal GA, Naresh A, Marrero L, Jones FE. (2005). Presenilin-dependent gamma-secretase processing regulates multiple ERBB4/HER4 activities. *J Biol Chem*. **280**(20):19777-83. Epub 2005 Mar 3.

Vilenchik L. Z., Griffith, J. P., St Clair, N., Navia, M. A. & Margolin, A. L. (1998). Protein Crystals as Novel Microporous Materials *J. Am. Chem. Soc.* **120**:4290–4294.

Villa N, Walker L, Lindsell CE, Gasson J, Iruela-Arispe ML, Weinmaster G.(2001). Vascular expression of Notch pathway receptors and ligands is restricted to arterial vessels. *Mech Dev*. **108**(1-2):161-4

Wells JA, McClendon CL. (2007). Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*. **450**(7172):1001-9.

Weng AP, Ferrando AA, Lee W, Morris JP 4th, Silverman LB, Sanchez-Irizarry C, Blacklow SC, Look AT, Aster JC. (2004). Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*. **306**(5694):269-71.

Wenlock MC, Austin RP, Barton P, Davis AM, Leeson PD. (2003). A comparison of physiochemical property profiles of development and marketed oral drugs. *J Med Chem*. **46**(7):1250-6.

Wharton K. A., K. M. Johansen, et al. (1985). "Nucleotide sequence from the neurogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats." *Cell* **43**(3 Pt 2): 567-81.

Whitty A. (2008). Cooperativity and biological complexity. *Nat Chem Biol*. **4**(8):435-9.

Wu Y, Cain-Hom C, Choy L, Hagenbeek TJ, de Leon GP et al. (2010). Therapeutic antibody targeting of individual Notch receptors. *Nature*. **464** (7291):1052-7.

Xiao Z., Hao Y., Liu B., Qian L. (2002). Indirubin and meisoindigo in the treatment of chronic myelogenous leukemia in China, *Leuk. Lymphoma* **43**:1763–1768.

Xiong, K, Ascitutto EK. , Madura JD., Asher SA. (2009). Salt Dependence of an R-Helical Peptide Folding Energy Landscapes. *Biochemistry* **48**: 10818–10826.

Yale, Harry L. (1959). "The Trifluoromethyl Group in Medicinal Chemistry". *Journal of Medicinal and Pharmaceutical Chemistry*. **1**: 121–133.

Yin H, Lee GI, Sedey KA, Rodriguez JM, Wang HG, Sebt SM, Hamilton AD. (2005). Terephthalamide derivatives as mimetics of helical peptides: disruption of the Bcl-x(L)/Bak interaction. *J Am Chem Soc*. **127**(15):5463-8.

Zweifel ME and Barrick D. (2001). Studies of the Ankyrin Repeats of the *Drosophila melanogaster* Notch Receptor2. Solution Stability and Cooperativity of Unfolding Biochemistry **40**:14357-14367

Zweifel M.E., Leahy D.J., Hughson F.M., Barrick D. (2003). Structure and stability of the ankyrin domain of the Drosophila Notch receptor. Protein Sci. **12**(11):2622-32.