Targeting protein-protein interactions for cancer therapy

Elizabeth Anscombe

LMB, Department of Biochemistry
and Magdalen College
University of Oxford

A thesis submitted in partial fulfilment for the requirements of the degree of Doctor of Philosophy at the University of Oxford

Trinity Term 2012

Word Count 48000

Abstract

Protein-protein interactions (PPIs) are key drug targets and recent breakthroughs in this area are providing insight into the types of molecules needed to selectively and potently inhibit a target traditionally seen as untractable. The rules that have been used to design classic substrate-competitive drugs (for example Lipinski's rule of five) may not apply in this new field in the same way. Here I present work performed in three systems that are well-validated drug targets for oncogenesis: the CDK2/cyclin A complex, the PLK1 Polobox domain and MDM2. In each case the site of the protein-protein interaction is defined and understood and the rationale for pharmaceutical intervention is clear. I use these as a model system to evaluate the characteristics of drugs that target protein-protein interaction sites and present work on the development of inhibitors as potential leads for subsequent drug development. In Chapter 1 I introduce the problems, challenges and rewards of PPI drug development; in Chapter 2 I present co-crystal structures of MDM2 with isoindolinone inhibitors; in Chapter 3 I detail attempts to co-crystallise the Plk1 Polobox with inhibitors and screen potential inhibitors; in Chapter 4 I present the results of screening to identify inhibitors of Cyclin A recruitment; and in Chapter 5 I discuss other strategies for inhibition of the CDK2/cyclin A complex, including results with a covalent inhibitor. Through these projects I have been able to demonstrate the wide applicability of the PPI inhibition approach, identify key features of drugs able to inhibit PPIs and contribute to drug design in each system.

Acknowledgement

No DPhil study can be undertaken without the support of family, friends and colleagues, and particularly in science we rely on a large network of collaborators and research contacts. I have many people to thank, without whom this thesis could never have been written.

Firstly, enormous thanks are due to Jane Endicott and Martin Noble, my supervisors, whose guidance and help at all times was priceless, especially when experiments did not work or gave strange results. Thanks for always being so supportive and keeping me on track! I must also thank Matt Higgins for taking over supervision of this 'orphan' student for the last year of my DPhil once Jane and Martin left Oxford and for always chatting and smiling and never getting annoyed if I've been uncommunicative whilst writing my thesis.

I have had the good fortune to work with some great people on the individual projects within this thesis, so next I would like to thank Christiane Riedinger and Burcu Anil, with whom I worked on the MDM2 project; Aude Echalier Glazer, with whom I worked on the Cyclin A recruitment project; and Elisa Meschini, with whom I worked on the covalent inhibitor project (NU6300). You have all moved on to other places or greater things by now but I have always remembered your friendship and input on the projects thankfully.

Next I would like to thank those people in our group and within the Biochemistry department who helped me use instruments and patiently explained technicalities to me: Ed Lowe, for his enormous help with all things crystallographic; Nick Brown, for help with all things protein purification; and David Staunton , for help with mass spec andall things biophysics. I would have been lost without you.

I addition I must thank all in the LMB for their friendship, laughter and support in tough times. I won't mention everyone by name but you know who you are! In particular I would also like to thank Elspeth Garman and Jan Gruber, with whom I did undergraduate projects, and who initially fired up my enthusiasm for lab work. I miss working with you both.

I would like to thank everyone in the collaborating group of chemists at Newcastle University, especially Ian Hardcastle, Roger Griffin and Herbie Newell; and our collaborator on the PLK1 project, Shudong Wang, from Nottingham University.

Of course some of the people who are vital to the success of this thesis were not scientists at all, so to my friends and family, thank you for your love and support from the bottom of my heart - especially to my mum, for always being there, and for putting up with infrequent phone calls.

Lastly, but of course not least, I would like to thank my husband and dedicate this thesis to him. Will, we met on the first day of university as undergraduates and since then I could never imagine my life without you. Thank you so much for your unconditional love and supporting me through good times and bad.

Table of Contents

Section	Page number
1 Introduction	***17***
1.1 Evolving drug design strategy	***17***
1.1.1 Strategies for drug design: target selection	*17*
1.2 A crisis in the pharmaceutical industry?	***18***
1.2.1 Observation of a decrease in research efficiency	*18*
1.2.2 Possible causes	*20*
1.2.3 Funding, patents and the relationship between industry and academia	*22*
1.3 Drug development strategies and timelines	***24***
1.31 Strategies for drug development	*24*
1.3.2 Stages in drug development and timeline	*27*
1.4 Targeting protein-protein interactions	***29***
1.4.1 Potential drug targets	*31*
1.4.2 Features of PPIs as druggable interfaces	*31*
1.5 Recent successes with PPI inhibitors	***33***
1.5.1 PPI inhibitors currently in clinical trials	*33*
1.5.2 Case study: JQ1 and BRD4/histones	*35*
1.5.3 Case study: Navitoclax and Bcl2/substrate	*37*
1.5.4 Case study: Tramiprosate (Alzhemed)	*40*
1.6 A strategy for finding new PPI inhibitors	***44***
1.6.1 Target identification: interactomics	*44*
1.6.2 High-throughput screening	*46*
1.6.3 Natural product libraries	*48*
1.6.4 Substrate-focused design	*52*
1.7 Aims of this thesis	***54***
2 Disrupting the MDM2/p53 interaction with isoindolinones	***55***
2.1 Background and introduction	***55***
2.1.1 p53: function and structure	*55*
2.1.1.1 The 'guardian of the genome': function as a tumour	*55*

suppressor	
2.1.1.2 Binding partners and interactions	*58*
2.1.1.3 p53 and cancer development	*58*
2.1.2 MDM2: inhibitor of p53 function	*58*
2.1.2.1 Effect of MDM2 binding	*58*
2.1.2.2 MDM2 structure	*59*
2.1.3 MDM2: drug target	*60*
2.1.3.1 Effect of disrupting the MDM2/p53 interaction	*60*
2.1.3.2 Existing compounds targeting the p53 binding pocket	*60*
2.1.4 Development of the isoindolinone series of compounds	*64*
2.2 Expression and purification of MDM2 constructs	***66***
2.2.1 Constructs produced to aid crystallisation	*66*
2.2.2 Expression of MDM2 constructs	*68*
2.2.3 Purification of MDM2 constructs	*69*
2.3 Co-crystal structures of MDM2 with compounds	***71***
2.3.1 Crystallisation	*71*
2.3.1.1 Broad screens	*71*
2.3.1.2 Follow-up screening	*71*
2.3.2 Data collection	*71*
2.3.3 Structure determination	*71*
2.3.4 Structure of MDM2 bound to Nutlin 3a	*75*
2.3.5 Structure of MDM2 bound to NCL00018934	*77*
2.3.6 Structure of MDM2 bound to NCL00018710	*79*
2.3.7 Structure of MDM2 bound to NU8406A	*81*
2.3.8 Review of isoindolinone binding modes	*82*
2.3.8.1 Comparison with nutlins binding mode	*82*
2.4 MDMX expression and purification	***85***
2.4.1 Cloning	*85*
2.4.2 Expression	*86*
2.4.3 Purification	*87*
2.5 MDMX crystallisation attempts	***90***
2.5.1 Broad screening	*90*

2.5.2 Narrow screening	*91*
2.5.3 Identification of possible crystallisation conditions	*92*
2.6 Conclusions and further experiments	***93***
3 The PLK1 Polobox: the search for a co-crystal structure	***95***
3.1 Background and introduction	***95***
3.1.1 PLK1: regulating mitosis	*95*
3.1.1.1 PLK1: sequence, structure and function	*95*
3.1.1.2 Regulation, binding partners and substrates	*97*
3.1.1.3 Plk1 and carcinogenesis	*102*
3.1.2 The Polobox domain: substrate binding	*103*
3.1.2.1 Function of the PBD	*104*
3.1.2.2 Structures of the PBD	*104*
3.1.3 Plk1 as a drug target	*110*
3.1.3.1 Effects of inhibiting Plk1 for cancer therapy	*110*
3.1.3.2 Strategies for inhibiting Plk1	*112*
3.1.3.3 Known Plk1 inhibitors targeting the PBD	*116*
3.1.4 Aims of this project	*118*
3.2 Expression and purification of the Plk1 PBD	***119***
3.2.1 Full-length Plk1 PBD	*119*
3.2.2 Truncated Plk1 PBD	*122*
3.3 Fluorescence anisotropy: SAR of compounds	***124***
3.3.1 Thymoquinone and poloxin derivatives	*124*
3.3.1.1 Confirmation of IC_{50} value for thymoquinone	*124*
3.3.1.2 Determination of IC_{50} values for thymoquinone derivatives	*127*
3.3.2 Screening of other known Plk1 inhibitors for PBD binding	*131*
3.4 Crystallisation attempts: soaking the crystals	***132***
3.4.1 Soaking protocols and strategies	*132*
3.5 Crystallisation attempts: co-crystallisation	***135***
3.5.1 Co-crystallisation: full-length PBD	*135*
3.5.2 Co-crystallisation: truncated PBD	*136*
3.6 Conclusions and further experiments	***136***

4 ***CDK2/cyclin A: targeting the recruitment site***	***138***
4.1 Background and introduction	***138***
4.1.1 CDKs, the role of CDK2 in the cell cycle and control of CDK2 activity	*138*
4.1.2 CDK2 structural biology	*140*
4.1.3 CDK2/cyclin A as a drug target	*144*
4.1.4 Aims of the project	*146*
4.2 Preparation for library screening	***147***
4.2.1 Protein expression and purification	*147*
4.2.2 Mutagenesis of p107: producing an optimal substrate for CDK2/cyclin A assays	*148*
4.2.3 Mutagenesis of cyclin A: investigation of Trp217 function in substrate recognition	*151*
4.2.4 Protein production and expression: cyclin A W217I/F/K	*152*
4.2.5 Characterisation of cyclin A W217 mutants	*155*
4.3 Selection of compound libraries	***157***
4.4 Medium-throughput screening of libraries: thermal denaturation assays	***158***
4.5 Screening of ChemDiv and natural product inhibitors by bandshift experiments	***161***
4.6 Radioactive kinase assays: screening compound sets for binding to CDK2/cyclin A	***167***
4.7 Summary of screening results	***171***
4.8 Validation of hits from ChemDiv screening	***172***
4.9 Fluorescence polarisation assay: checking for displacement of a recruitment peptide	***175***
4.10 Co-crystallisation of CDK2/cyclin A with hit compounds	***178***
4.11 Conclusion and further work	***179***
5 Other drug design strategies targeting CDK2/cyclin A	***181***
5.1 Background and introduction	***181***
5.1.1 Functions of CDK2	*181*
5.1.2 Structural basis for CDK2 activity and functions	*181*

5.1.3 Differences and similarities between CDK2 and other CDKs	*182*
5.1.4 ATP-competitive inhibitors	*183*
5.1.5 Covalent inhibitors: recent successes, issues and development	*186*
5.1.5 Aims of the project	*187*
5.2 NU6300: irreversible inhibition	***188***
5.2.1 Preparation of CDK2 mutants for testing NU6300	*188*
5.2.1.1 Mutagenesis	*188*
5.2.1.2 Expression	*188*
5.2.1.3 Purification	*188*
5.2.2 Does NU6300 covalently modify CDK2?	*190*
5.2.2.1 Mass spectrometry: checking for covalent modification	*190*
5.2.2.2 Methodology: pull-down bandshift assay	*191*
5.2.2.3 Timescales of modification	*192*
5.2.2.4 Degradation of compound in solution	*192*
5.2.3 Determination of covalently modified residue	*192*
5.2.3.1 Testing mutants for covalent modification	*193*
5.2.4 Determination of CDK2/cyclin A structure modified by NU6300	*197*
5.2.4.1 Crystallisation	*197*
5.2.4.2 Data collection	*197*
5.2.4.3 Data analysis and refinement of structure	*198*
5.2.4.4 Crystal structure of NU6300-modified CDK2/cyclin A	*198*
5.3 Co-crystal structures of CDK2/cyclin A with traditional ATP-competitive drugs	***200***
5.3.1 Background and introduction	*200*
5.3.1.1 Compound series development	*200*
5.3.1.2 Other ATP-competitive kinase inhibitors	*201*
5.3.1.3 Aims of this project	*202*
5.3.2 Structure determination	*202*
5.3.2.1 Crystallisation	*202*

5.3.2.2 Data collection	*202*
5.3.2.3 Data analysis and refinement of structure	*203*
5.3.3 Comparison of binding mode with other compounds in the series	*205*
5.3.3.1 Co-crystal structure of NCL14550 with CDK2/cyclin A	*205*
5.3.3.2 Co-crystal structures of NCL-00018365 and NCL-00018046 with CDK2/cyclin A	*206*
5.4 Conclusions and further experiments	***208***
6 Concluding remarks	***209***
6.1 Advantages of alternative strategies	***209***
6.2 Future direction for the projects	***210***
References	*213*
Appendix A: Materials and methods	*225*
A0 Materials	*225*
A1 Molecular biology: cloning	*225*
A1.1 Cloning templates	*225*
A1.2 Purification of recombinant plasmid	*225*
A1.3 Polymerase chain reaction	*226*
A1.4 Ligation	*227*
A1.5 Restriction digest	*227*
A1.6 Alkaline phosphatase	*228*
A1.7 Transformation	*228*
A1.8 Preparation of competent cells	*228*
A1.9 Site-directed mutagenesis	*229*
A1.10 Agarose gel electrophoresis	*229*
A2 Protein expression and purification	*229*
A2.1 Bacterial culture	*230*
A2.2 Protein purification	*231*
A2.2.1 Preparation of cell lysate	*231*
A2.2.2 Affinity chromatography	*231*
A2.2.3 Gel filtration chromatography	*232*
A2.2.4 Dialysis	*232*

A3 Molecular biology: protein characterisation	*233*
A3.1 Determination of protein concentration	*233*
A3.2 SDS-PAGE	*233*
A3.3 Mass spectrometry	*234*
A4 SDS-PAGE bandshift assay	*234*
A4.1 Preparation of reagents	*234*
A4.2 Timecourse of reaction	*235*
A5 Radioactive kinase assay	*235*
A5.1 Preparation of reagents and equipment	*235*
A5.2 Timecourse of reaction	*236*
A5.3 Washing and mounting P81 papers	*236*
A5.4 Imaging and analysing P81 papers	*236*
A6 Fluorescence anisotropy assay	*238*
A6.1 Preparation of reagents	*238*
A6.2 Preparation of completed plate	*238*
A6.3 Measurement of fluorescent anisotropy	*239*
A6.4 Analysis of results	*239*
A7 Thermal denaturation assay	*239*
A7.1 Preparation of reagents	*240*
A7.2 Preparation of completed plate	*240*
A7.3 Measurement of thermal denaturation curves	*241*
A7.4 Analysis of results	*241*
Appendix B: X-ray crystallography	*242*
B1 Protein crystallisation	*242*
B1.1 Conditions for crystal growth	*242*
B2 Unit cell, symmetry and space group	*243*
B3 Theory of diffraction	*246*
B3.1 The electron as an X-ray scatterer	*246*
B3.2 Representation of a wave as a Fourier series	*246*
B3.3 Bragg's law	*247*
B3.4 The reciprocal lattice and the Ewald sphere	*248*
B3.5 Atomic scattering factor	*249*

B3.6 The Laue condition for diffraction	*250*
B3.7 Fourier transform and the structure factor equation	*250*
B3.8 Friedel's law	*251*
B4 Solving the phase problem	*252*
B4.1 Molecular replacement	*253*
B4.2 Patterson function	*253*
B4.3 Rotation and translation search	*254*
B5 Macromolecular structure refinement with REFMAC5	*255*
B6 Manual rebuilding using Coot	*256*
B7 Assessment of quality of the structure with MolProbity	*257*
B8 Crystallography jargon	*257*
B8.1 Matthews coefficient	*257*
B8.2 Temperature factor	*258*
B8.3 R factors	*259*
Appendix C: Structures of compounds in libraries	*261*
C1 ChemDiv library from virtual screening	*261*

List of Figures

Figure	Page number
Figure 1.1 Observation of a decline in research and development efficiency in the pharmaceutical industry	17
Figure 1.2 Stages in the development of a new drug candidate	26
Figure 1.3 'Hot spots' within proteins: a small number of residues contribute the majority of the free energy of the interaction	31
Figure 1.4 The structure of BRD4 in complex with JQ1	35
Figure 1.5 Structure of navitoclax (ABT-263) ***Figure 1.6*** Abnormally cleaved amyloid precursor protein gives the fragment amyloid beta, which forms aggregate	37 40
Figure 1.7 Scaffold-based drug design libraries allow iterative drug design to build an effective inhibitor	53
Figure 2.1 Cellular effects of p53 activation	56
Figure 2.2 p53 protein architecture	57
Figure 2.3 Control of p53 activity in an unstressed cell	59
Figure 2.4 1YCR structure	60
Figure 2.5 Binding modes of known inhibitors of the MDM2/p53 interaction	62
Figure 2.6 Structure of the isoindolinone scaffold	64
Figure 2.7 NMR solution structure of MDM2 residues 17-125	67
Figure 2.8 Purification of MDM2 E69K70A residues 17-109	70
Figure 2.9 Co-crystals of MDM2	71
Figure 2.10 Structure of Nutlin 3a bound to MDM2	75
Figure 2.11 Structure of NCL00018934 bound to MDM2	75
Figure 2.12 Structure of NCL00018710 bound to MDM2	79
Figure 2.13 Structure of NU8406A bound to MDM2	81
Figure 2.14 Effect of addition of *t*Butyl group on the position of the side chain of Y67	83
Figure 2.15 Summary of isoindolinone binding modes when	84

compared to Nutlin 3a	
Figure 2.16 Purification of MDMX	88
Figure 2.17 New possible crystal conditions identified for MDMX	92
Figure 2.18 Development of NCL-00019583	93
Figure 3.1 Primary structure of Plk1	96
Figure 3.2 Cellular functions of Plk1	99
Figure 3.3 Models for Plk1 action, proposed by Lowery *et al*, 2005	101
Figure 3.4 Structure of the Plk1 Polobox domain	107
Figure 3.5 Interactions between the phosphopeptide MQSpTPL and residues of Plk1 PBD	108
Figure 3.6 BI2536, a specific inhibitor of Plks	114
Figure 3.7 Poloxin and thymoquinone bind the PBD	117
Figure 3.8 Purification of Plk1 PBD full-length	120
Figure 3.9 Secondary structure of the Polobox domain	122
Figure 3.10 Basis of fluorescence anisotropy competition assay	124
Figure 3.11 The response of Plk1 PBD to a freeze/thaw cycle	126
Figure 3.12 Fluorescence anisotropy competition assay	127
Figure 3.13 Derivatives of thymoquinone	128
Figure 3.14 OC series compounds bind to the Plk PBD with varying affinity	129
Figure 3.15 ON01910.Na and related compounds do not bind to the Plk1 PBD	131
Figure 3.16 Crystals of apo PBD (subtilisin-cleaved) used for compound soaks	134
Figure 3.17 Plk1 PBD co-crystals with the MQSpTPL optimal phosphopeptide	135
Figure 4.1 CDKs and cyclins control the cell cycle	138
Figure 4.2 Control of CDK2 activity	140
Figure 4.3 pCDK2/cyclin A complex structure	141

Figure 4.4 The cyclin A substrate recruitment site	143
Figure 4.5 p107 purification	149
Figure 4.6 Modelling cyclin A mutations	150
Figure 4.7 Purification of mutated (W217I/F/K) bovine cyclin A	152
Figure 4.8 Validation of mutated (W217X) cyclin A: pulldowns and bandshifts	154
Figure 4.9 Mutation of W217F in the cyclin A recruitment site ablates ability to bind recruitment peptide	155
Figure 4.10 Thermal denaturation assay	159
Figure 4.11 Screening of natural product inhibitors by bandshift assay	162-166
Figure 4.12 Bandshift assay: hits from ChemDiv library screening	167
Figure 4.13 Radioactive kinase assay calibration	168
Figure 4.14 Results of the radioactive kinase assay	170-171
Figure 4.15 Results of screening ChemDiv inhibitor set with different techniques	172
Figure 4.16 The radioactive kinase assay shows that recruitment site inhibitors affect phosphorylation of p107 but not histone H1	173
Figure 4.17 Validation of hits	174
Figure 4.18 E6 and structurally related compounds	175
Figure 4.19 Fluorescence polarisation: displacement of a labelled recruitment peptide	176
Figure 4.20 Co-crystals of CDK2/cyclin A with hit compounds	178
Figure 5.1 Features of the ATP binding site	183
Figure 5.2 Structures of key ATP-competitive CDK2 inhibitors	184
Figure 5.3 NU6300 can covalently modify CDK2/cyclin A	189
Figure 5.4 NU6300 covalently inhibits wild-type CDK2	190
Figure 5.5 Hydrolysis of NU6300 in solution	191

Figure 5.6 NU6300 modifies either K88 or K89	192
Figure 5.7 NU6300 can be removed by dialysis from CDK2 K88EK89V, showing that the normal residue of modification is one of these.	193
Figure 5.8 Identification of the CDK2 residue covalently modified by NU6300	193
Figure 5.9 K88 is a minor reaction site of NU6300	194
Figure 5.10 Mutation of K89 abolishes NU6300 covalent modification	195
Figure 5.11 The crystal structure of NU6300 bound to CDK2/cyclin A	197
Figure 5.12 Suggested possible mechanism of modification of pCDK2 by NU6300	198
Figure 5.13 Binding of NU6102 to CDK2 (A) or PfPk5 (B)	199
Figure 5.14 Binding pose of NCL14550 within the ATP binding site	204
Figure 5.15 Binding poses of NCL18365 and NCL18046 within the ATP binding site	205
Figure 5.16 Illustration of the binding mode of NCL18365 (A) and NCL18046 (B) to CDK2	206
Figure A.1 Format of results for radioactive kinase assay	237
Figure B.1 Phase diagram showing the progress of a vapour diffusion experiment	243
Figure B.2 The unit cell	243
Figure B.3 Bragg's law: scattering from two lattice planes	248
Figure B.4 The Ewald sphere	249

List of Tables

Table	Page number
Table 1.1 Summary of PPI inhibitors in development	32
Table 1.2 Anti-cancer drugs from natural product sources in clinical or pre-clinical development	49
Table 2.1 Structures of MDM2 bound to inhibitory peptides or small-molecule inhibitors	63
Table 2.2 Progress of purification of MDM2 constructs for crystallisation	69
Table 2.3 Data collection and refinement statistics for structures of isoindolinones bound to MDM2	73
Table 2.4 MDMX mutants produced	85
Table 2.5 Expression and purification of MDMX constructs	89
Table 3.1 Previously identified Plk1 crystal forms	134
Table 4.1 Results of fluorescence anisotropy: effect of changing chemical group at B	176
Table 5.1 X ray data collection and refinement statistics for CDK2/cyclin A/inhibitor complexes	203
Table A.1 Recipes for bacterial culture media	230
Table A2 Recipes for SDS-PAGE: gels and running buffer	234
Table B.1 The seven crystal systems and the fourteen Bravais lattices	245

List of abbreviations

Abl	Abelson tyrosine kinase
AML	Acute myeloid leukemia
APC/C	Anaphase-promoting complex/cyclosome
Bcl-2	B-cell lymphoma 2
BH3	Bcl-2 homology 3
BRD4	Bromodomain-containing protein 4
BSA	Bovine serum albumin
cAPK	cAMP dependent protein kinase
Cdc	Cell division cycle
CDK	Cyclin-dependent kinase
Cip/Kip	CDK interacting protein/kinase inhibitory protein
CLL	Chronic lymphocytic leukemia
CSF Aβ	Cerebrospinal fluid amyloid beta
CTC	Circulating tumour cell concentration
CYFRA 21-1	Cytokeratin 19 fragment antigen 21-1
DMSO	Dimethylsulfoxide
DTT	Dithiothreitol
E2F	E2 promoter binding factor
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ESRF	European synchrotron radiation facility
FDA	Food and Drug Administration
GFP	Green fluorescent protein
GPCR	G-protein coupled receptor
GST	Glutathione S transferase
H1	Histone 1
HBS	Hepes buffered saline
HIF-1	Hypoxia-inducing factor 1
HIV	Human immunodeficiency virus
HTS	High-throughput screening
IC_{50}	Inhibitory concentration 50
IgG	Immunoglobulin G
IMS-RP	Intermembrane space reporter protein
IP	Intellectual property
IPTG	Isopropyl-β-D-thiogalactopyranoside
JNK	Jun N-terminal kinase
KD	Kinase domain

Km	Michaelis constant
MDM2	Murine double minute 2
MITF	Microphthalmia-associated transcription factor
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MTG	Monothioglycerol
Ni-NTA	Nickel-nitrilotriacetic acid
NMC	NUT midline carcinoma
NMR	Nuclear magnetic resonance
NP	Natural product
NSCLC	Non small-cell lung cancer
NSE	Neuro-specific enolase
NUT	Nuclear protein in testis
OD	Optical density
pRb	Retinoblastoma protein
PB1/PB2	Polo-box 1/2
PBD	Polo-box domain **or** Phosph-Ser/phospho-Thr binding domain
PEG	Polyethylene glycol
PET	Positron emission tomography
PI3K	Phosphatidylinositol-3-kinase
Plk	Polo-like kinase
PKC	Protein kinase C
PPI	Protein-protein interaction
pro-GRP	pro-gastrin releasing protein
P-TEFb	Positive transcription elongation factor
QSAR	Quantitative structure-activity relationship
R&D	Research and development
RMSD	Root mean square deviation
RNAi	RNA interference
SAR	Structure-activity relationship
SCLC	Small-cell lung cancer
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
siRNA	small interfering RNA **or** silencing RNA **or** short interfering RNA
SPR	Surface plasmon resonance
VEGF	Vascular endothelial growth factor
V_{max}	Maximum velocity
WHO	World Health Organisation

1 Introduction

1.1 Evolving drug design strategy

1.1.1 Strategies for drug design: target selection

The choice of a drug target has historically been complicated by the fact that basic knowledge regarding the structure and function of the target on a molecular level was absent. Typically, drug design has focused on the modification and synthesis of structural analogues of a lead compound identified by serendipity or in traditional binding assays (eg penicillins, cisplatin, aspirin) to optimise pharmacological properties (such as absorption, metabolism, distribution and elimination: ADME properties). In reality, design was often very much focused on the structure of the drug molecule without much focus on how this would work within the protein's binding site.

When genes for proteins involved in disease processes had to be identified one by one, the fundamental science behind the target selection often took a long time, and an incomplete view of the network of proteins involved in creating the disease state could cause failure in later stages of the drug design process due to unexpected interactions. However, genomics has had a strong influence on drug design, since the knowledge of the genetic basis for disease states allows a more informed decision to be made with regard to the most useful protein target for a putative drug (Emilien *et al,* 2000). The promise of personalised medicine, in which drugs might be prescribed for a patient based on their genetic profile, is likely to be realised in the next few decades, and this provides a particular boost for the drug design process for cancer, due to the genetic heterogeneity of morphologically similar tumours.

The impact of the availability of high-resolution structures of proteins involved in disease processes has also been immense, since a potential target selection can be understood on a molecular level. Inhibitors can then be designed based on the geometry of the binding site, and the subtle differences between the

structure of the binding site in homologues or related proteins can be employed to derive specificity.

1.2 A crisis in the pharmaceutical industry?

1.2.1 Observation of a decrease in research efficiency

Over the last fifty years, FDA figures for drugs approved per billion US$ spent, show that the efficiency of research and development programs has decreased, with an overall trend of efficiency decreasing by three times every twenty years (Figure 1.1).

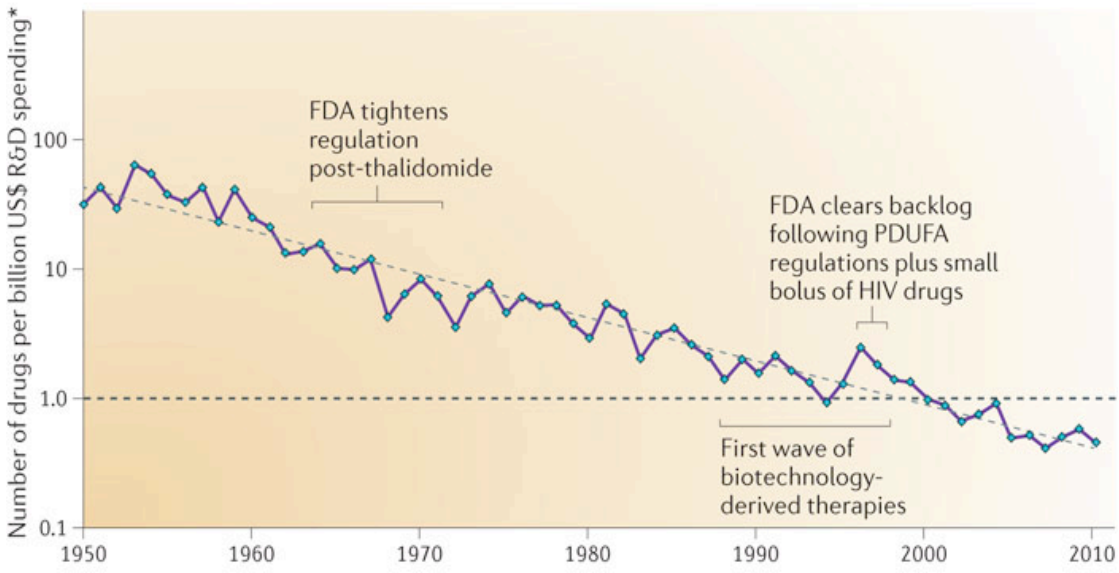


Figure 1.1 Observation of a decline in research and development efficiency in the pharmaceutical industry Figure reproduced from Scannell *et al*, 2012. Adjusted for inflation, the number of new drugs approved by the FDA per $bn spent on research and development in the US is decreasing year on year.

The discussion about the reasons behind this decline in research efficiency is ongoing (reviewed in Scannell *et al*, 2012) but it is clearly a cause for concern. There is a constant need for new drugs, not only to target diseases for which no therapy currently exists, but also to guard against resistance to current drugs, for example as a result of the evolution of antibiotic resistant 'superbugs'. New

drugs will also improve current therapies by increasing their efficacy, and decreasing off-target side effects.

1.2.2 Possible causes

The "low-hanging fruit" hypothesis suggests that some of the easiest targets for drug development might already have been exhausted, so that those that remain unexploited may be challenging pharmacologically. Although there is little direct evidence for this view, it appears to be the case in some specific areas, such as antibiotic development. The rate of discovery of new antibiotics remained high from the 1940s to the 1990s as new classes were discovered. However there have been few new discoveries since 2000, despite the pressing need as a result of antibiotic resistance (Shlaes, 2003; Hogberg *et al*, 2010). It may be that the main pathways by which bacterial growth and replication can be inhibited have already been targeted, and therefore the 'low-hanging fruit' has been plucked. A related problem is that the standards required of a new drug have increased, since each new drug candidate must represent a therapeutic advance on the current treatment to make it through clinical trials (Scannell *et al*, 2012). This increases the attrition rate during clinical development of new drugs.

Increasingly stringent regulations and testing may also play a part in making drug discovery more difficult. The attrition rate of new drugs in Phase III trials has been particularly high in recent years with failures mainly due to insufficient efficacy compared to placebo or commercially available drugs (Pammolli *et al*, 2011). It can often be difficult to anticipate this early in clinical trials, thereby wasting time and money developing drugs that fail at this late stage in the development process (David *et al*, 2009).

A change in the types of disease state targeted may also have made the problems more difficult, with an increased focus on chronic or fatal diseases, as well as rare conditions. In particular, a high proportion of current drug development projects focus on diseases with no external causative agent, such as Alzheimer's disease or cancer. Analysis of the Pharmaceutical Industry Database (a collection of data from public sources and companies, regarding classification, timescales and organisation of projects) shows that development of antineoplastic agents in

particular have the largest potential market but the lowest probability of success (Pammolli *et al*, 2011).

The use of more sophisticated approaches to drug design, with the potential for fewer side effects, may present more difficulties for development in the short term. For example, the development of antibody- or stem cell-based therapies for cancers presents significant new challenges since it represents a fundamentally different approach to traditional drug development. For new types of therapies such as this, the target mechanism of action needs to be understood before the drug development process can begin, so naturally this is likely to take much longer and potentially have a higher failure rate. In addition, the change towards stratified and personalised medicine – which aims to rationalise the use of treatments for cancer based on an understanding of how the drug has performed in large cohorts of patients, in order to understand the response for differing tumour genetic profiles – allows opportunities to refine treatment but may also be a more difficult challenge for drug development.

1.2.3 Funding, patents and the relationship between industry and academia

The patent system aims to encourage research and development (R&D) by protecting the rights of pharmaceutical companies to profit from the drugs they have developed, whilst allowing the sale of cheaper generic versions after a reasonable time period has elapsed, in order to benefit patients. At the moment the time period covered by intellectual property (IP) laws is 20 years, although this can be extended by various means. These means include reaching a private settlement with generics manufacturers, or obtaining one or more agreed 30-month 'stays' on generic product sales, designed to resolve potential disputes over patent infringements (Kesselheim, 2007). The cost of bringing new drugs to market is increasingly prohibitively expensive and IP agreements are normally signed quite early in the process of drug development, despite the fact that there may still be as much as a decade of pre-clinical and clinical testing to complete in order to commercialise the drug. There is some argument therefore that a longer period of IP protection would incentivise innovation.

In some circumstances, additional incentives are required in order to promote R&D by pharmaceutical companies. Orphan diseases, neglected tropical diseases and rare conditions are examples for which the potential profits would not themselves provide a sufficient inducement to spend large amounts of money and time on R&D programs. Such incentives include tax breaks, advanced commitments by governments to buy stocks of a product (typically vaccines or anti-virals), prize funds (for example in difficult therapeutic challenges such as Alzheimer's disease) and partnership between public and private sector efforts (Moran *et al*, 2009; Villa *et al*, 2009; Maurer, 2006). The World Health Organisation (WHO) makes efforts to advise on public health, identify areas of therapeutic need and promote policies that maintain such incentives and encourage research into new drugs for the most critical targets. The WHO Secretariat on Public Health, Innovation and Intellectual Property, established in 2006, is vital for coordinating public and private drug development efforts,

informing IP policy and promoting global access to medical innovations (Mackey and Liang, 2012).

The use of better decision-making processes within the pharmaceutical industry, and better cooperation between academia and industry in drug development, is vital for the ongoing viability of the R&D process. Academic partners are ideally set up to investigate the basic science which backs up drug development efforts, and to perform the early stages of drug design and pre-clinical development. Many academic centres do not have the resources required to perform clinical trials so a close collaboration with industrial partners from this point is important in order to provide the impetus to quickly bring promising drug candidates from the pre-clinical stages of development to a commercially viable medicine (Vallance *et al*, 2010). The introduction of new strategies and techniques for target validation, compound screening and drug design are also important ways in which the process can be made more efficient (discussed in Section 1.2.1).

1.3 Drug development strategies and timelines

1.3.1 Strategies for drug development

The initial identification of a compound with some activity towards a particular drug target sometimes occurs by serendipity, allowing further development of a family of compounds with improved pharmacological properties. For example, some early antibiotics were identified by chance, allowing for further development of the penicillins, and cisplatin and other platinum-containing anti-cancer drugs were identified by a chance observation in an unrelated experiment, allowing for the development of more sophisticated platinum-based analogues (reviewed in Fischer, 2007 and Monneret, 2011). Obviously this is not a planned approach to drug discovery but serendipitous discovery can often play an important role in the early stages of the drug design process.

Alternatively, high-throughput screening of compound libraries or databases can generate the initial lead compound for drug design. This can be particularly useful in the case of a new molecular target, for which little information is known. In combination with structure-based drug design, this approach has been successfully employed to identify kinase inhibitors from libraries of existing compounds, resulting in potent lead compounds (IC_{50}<5nM) consisting of a phenyl urea moiety on a pyrimido-diazapinone scaffold (Akritopoulou-Zanze and Hajduk, 2008). It can be particularly important for lead generation in projects with no known inhibitors as a starting point, exemplified by the discovery of the inhibitors of the neurokinin-1 receptor (a GPCR) using a privileged scaffold (the spiropyrrolopyrrole motif) as a starting point (Bleicher *et al,* 2002; Bleicher *et al,* 2003).

Once the initial lead is identified, the drug design process can be split into two possible approaches: either target-orientated or ligand-orientated. In the former approach, the molecular target, usually a protein, is used as the basis for drug design. This is also known as *de novo* drug design or structure-based drug design. High-resolution structures of the target from NMR or X-ray

crystallography are combined with computational approaches, allowing potential drug compounds to be screened. Computer-aided molecular modelling can help to design a compound that fits the binding site, is likely to make interactions with particular residues and has drug-like properties (reviewed in Schneider, 2012). This approach has been used to identify androgen receptor antagonists (Guo *et al*, 2012), gankyrin ligands (Thakur and Hassan, 2011) and BRAF kinase inhibitors (Park *et al*, 2012).

Ligand-based design focuses on improving the properties of the drug by identifying the chemical groups that contribute most to the interaction. Rational drug design uses structure-activity relationship (SAR) or quantitative structure activity relationship (QSAR) data, relating the identity of each chemical group in a compound series to potency. Computational and biophysical techniques assess binding and also allow pre-selection of drug-like properties, to improve the compound function. This approach has been used to design non-nucleoside inhibitors of human immunodeficiency virus (HIV) reverse transcriptase (Jain *et al*, 2012) and to identify analogues of 2-phenoxy-N-substituted acetamide, which inhibit Hypoxia-inducing factor -1 (HIF-1) (Noolvi *et al*, 2012).

Another influential method of ligand–based drug design is scaffold or fragment-based design, in which libraries of small chemical moieties or central scaffold groups are assessed for binding and the hit compound is iteratively improved to provide a potential lead molecule. Because this technique uses combinations of small fragments, a lot of chemical space can be sampled within a small library. However, it can be difficult to identify binding from fragments because the binding affinity may be weak, due to the compound's size. Biophysical techniques are usually employed to assess binding. The optimised compound is built by iteratively adding groups to a scaffold, or by identifying two or more fragments that bind close together in a binding site so that they can be linked to form a larger, more potent inhibitor, which exploits all of the interactions identified (Feyfant *et al*, 2011). Fragment-based screening has produced several successful drug candidates, including ABT-518 (a matrix metalloprotease

inhibitor) and LY-517717 (a Factor Xa serine protease inhibitor) (reviewed in Fattori *et al*, 2008).

These and other strategies for drug design and development are discussed more thoroughly in Section 1.5.

1.3.2 Stages in drug development and timeline

The main stages in the pre-clinical and clinical development of a new drug are illustrated in Figure 1.2.

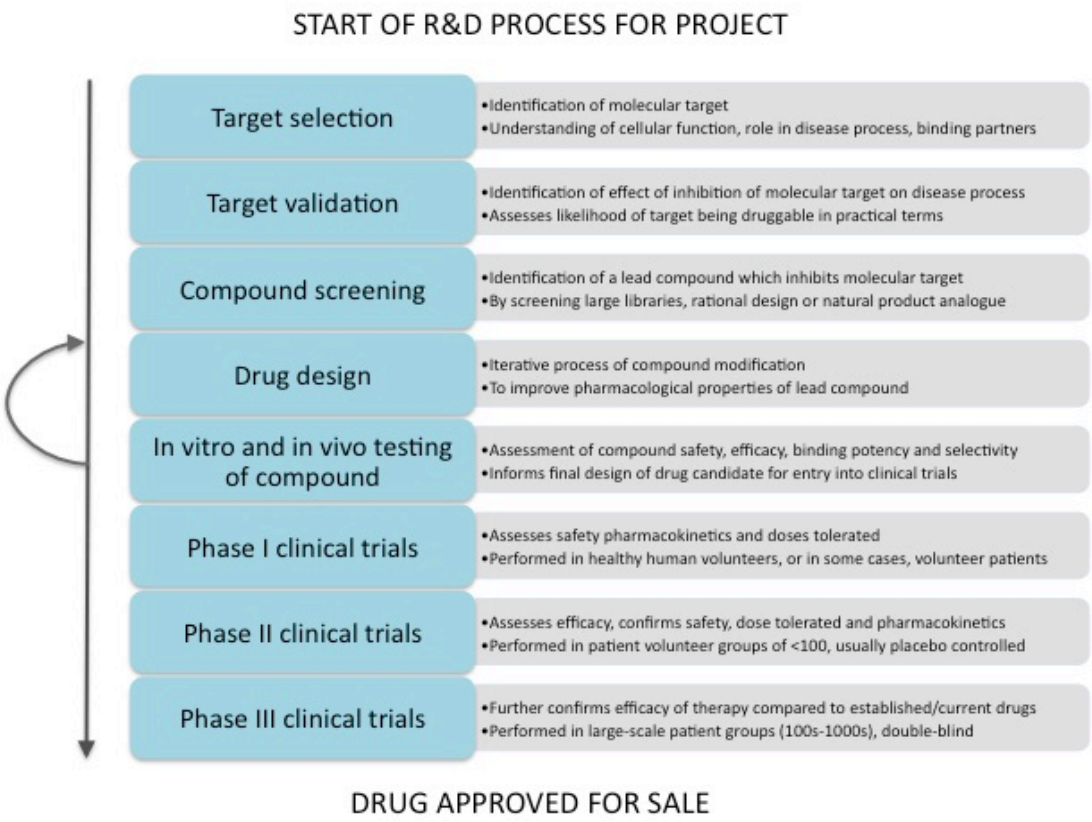


Figure 1.2 Stages in the development of a new drug candidate. The figure shows the preclinical and clinical stages. At any stage of the process a project may become unviable, due to the difficulty of drug design or problems with the selected candidate including side effects, lack of efficacy or delivery. For the purposes of clarity, some stages have been simplified in this diagram (see text for details).

The process of drug development usually takes at least 10 years, although it may be much longer in difficult cases. Although the process is represented in Figure 1.2 as largely linear, it can be extremely iterative and may involve multiple false leads and restarts. The vast majority of potential drugs do not attain clinical approval. The compound may fail early in the process, perhaps because a target is too difficult to drug effectively, or because of functional redundancy. Otherwise, the compounds may have undesirable pharmacological properties, or insufficient potency or selectivity. In addition, a large proportion of drug candidates fail prior to entering clinical trials. This may be because they show

side effects, off-target effects or poor efficacy compared to placebo or current therapies.

The pre-clinical drug development process starts with the identification of a molecular target, normally a protein. This process has been improved by our increasing understanding of signalling pathways within the cell, and by the human genome project, which identified new potential targets (reviewed in Emilien *et al*, 2000). The use of databases such as the Therapeutic Target Database, DrugBank and STITCH ('search tool for interactions of chemicals') may aid the process of target selection, as well as early compound screening carried out *in silico*. The process of target validation identifies interactions of the protein target and investigates its role in the disease process. Knowledge of this kind is usually vital to back up a drug design effort. At this point, the search for compounds that modify the function of the molecular target can begin. The methods for compound screening and drug design are discussed further in Section 1.5, but usually include high-throughput screening, or substrate-based or ligand-based rational design. At this point, *in vitro* or *in vivo* testing of the lead compound is conducted, usually involving assays such as ELISA or SPR, and cell-based and whole animal studies, to confirm basic safety and establish dosing regimes.

The clinical stage of drug development begins with Phase I clinical trials, conducted in healthy human volunteers, testing safety, tolerated dose, methods of administration and pharmacokinetics (the effect of human metabolism on the drug). Phase II and III studies can each be split into stages a and b (early and late respectively). Phase IIa trials involve a small number of patients and assess dose regimes, confirm safety and identify whether any therapeutic effect exists. Phase IIb trials are larger, double-blind and placebo controlled, measuring efficacy and confirming dose regimes and side effects. Phase IIIa trials involve administering the selected dose regime to a larger group of patients using double-blinding and placebo and existing drug controls.

At this point the drug may be registered but must undergo additional testing in Phase IIIb trials to establish what additional benefits the potential therapy offers in comparison to existing therapies. Sometimes these studies are carried out over several years in order to assess long-term effects of the therapy. At this point, if clinical trials have been successful, the drug may be submitted for approval (by the FDA in the US or other regulatory bodies). However, Phase IV clinical studies are often carried out at this point before the drug is used in a clinical setting, to detect any potential rare side effects, drug-drug interactions and long-term effects. This study might identify which patient populations or disease sub-types might be most effectively treated with the therapy.

1.4 Targeting protein-protein interactions

1.4.1 Potential drug targets

Protein-protein interactions (PPIs) are an untapped reserve of potential drug targets that have been previously ignored. Traditionally they have been viewed as large, flat, featureless surfaces with no distinct binding site for an inhibitor to exploit and therefore difficult to target. However, PPIs are ubiquitous in signalling cascades, gene expression, and cell surface interactions and aberrant interactions can lead to disease. In particular, carcinogenesis is highly dependent on the interactions and regulation of membrane receptor-dependent signalling cascades and cell cycle regulatory proteins. It has become clear in recent years that there is no need to ignore this large class of potential drug targets, since their reputation as 'undruggable' is not always deserved (reviewed in Mullard, 2012).

PPIs may have an advantage as drug targets since compared to traditional protein-ligand interactions, they are likely to be more selective and present a less general target, for example compared to an ATP-binding site (highly conserved between all kinases). A PPI inhibitor may therefore have a more specific binding profile and fewer side effects or off-target effects. In addition, PPIs are often of higher affinity and more dynamic than the interactions of proteins with small molecule ligands. As a result, lower concentrations of inhibitor are required to disrupt the interaction. Despite these advantages there are also some difficulties: the process of drug design is likely to be more difficult and take longer, since there is no cellular small-molecule ligand to be used as a starting point for design. The peptide representing the protein ligand can be large and therefore not represent a good starting point for design (Wells and McClendon, 2007). Furthermore libraries often do not contain compounds that make good PPI inhibitors, so high-throughput screening is problematic (see Section 1.5.2).

1.3.2 Features of PPIs as druggable interfaces

The surfaces involved in a protein-protein interface are typically large (~1500-3000 Å^2) and flat compared to those surfaces involved in protein-small molecule interactions (the normal drug target). These surfaces by contrast are usually (~300-1000 Å^2) and involve a deep groove or binding pocket in which the ligand is buried (Jones and Thornton, 1996). However, this characterisation of PPIs is not universal, since some do involve a more traditional deep binding pocket, in which large side chains from residues of one binding partner project into a cleft on the other. A good example is the p53-binding site on MDM2, where a Trp, Phe, and Ile/Leu side chain each fit into three deeper pockets within the overall hydrophobic binding site (discussed in detail in Chapter 3).

Since the binding site mediating a protein-protein interaction is much larger than that of a protein/small-molecule ligand, much larger molecules may be required to disrupt the interaction. There is already some indication that this is the case as many of the molecules so far identified as PPI inhibitors are larger and more hydrophobic, with more H-bond acceptors and donors, compared to drugs currently in the clinic (reviewed in Morelli *et al*, 2011).

It is also becoming apparent that PPIs are largely governed by 'hot spots': a small number of residues that contribute the majority of the binding energy to the interaction. These residues mediate H-bonds, or form salt bridges or hydrophobic interactions (see Figure 1.3). Disruption of the interactions made by these key residues may be sufficient to displace the protein ligand. Thus designing a PPI inhibitor in these cases may be more straightforward than expected (Arkin and Wells, 2004; Buckingham, 2004).

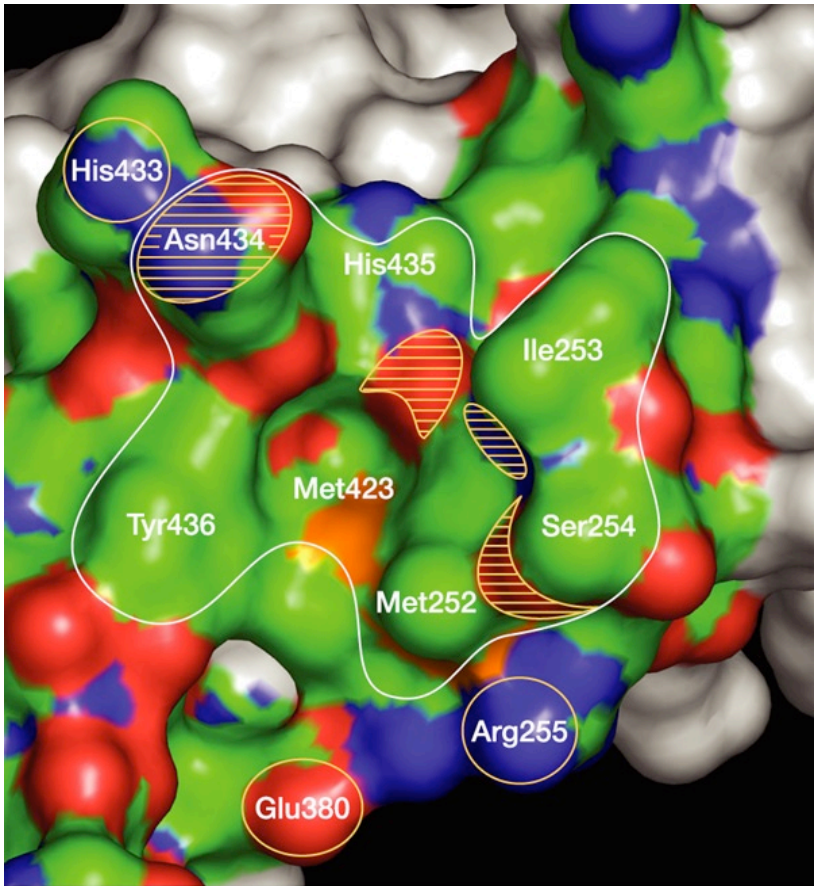


Figure 1.3 'Hot spots' within proteins: a small number of residues contribute the majority of the free energy of the interaction The Fc domain of an immunoglobulin (IgG) is drawn to illustrate this point. The immunoglobulin binding site for its epitope is coloured by atom type and the consensus binding site outlined. Residues contributing to the central hydrophobic core of the site are circled in white and residues forming salt bridges are circled in yellow. Residues that make H-bonds are marked by dashed yellow. Determined from co-crystal structures with ligand bound. Reproduced from Arkin and Wells, 2004.

The hotspot theory may also explain how one protein binding site is able to bind several structurally related proteins in slightly different orientations. Each partner will have a different set of surface residues, and as a result slightly different hotspots. In each case, binding will be maintained, despite the orientation and contacts being slightly altered.

1.5 Recent successes with PPI inhibitors

1.5.1 Some PPI inhibitors currently in development

The field of PPI inhibitor development is now very competitive. The main targets are receptors and other cell surface proteins, selected because cellular uptake is not required. However, PPI inhibitors have also been identified that target proteins within the cell, for example the MDM2/p53 and BRD4/histone interactions. Some of the PPI inhibitors currently in development are listed in Table 1.1.

Name	Company	Disease process	Molecular target	Stage of development
SAR1118	SARcode (Sunesis Pharmaceuticals)	Dry eye	Cell adhesion molecule 1/lymphocyte function-associated antigen 1	Phase III
Navitoclax (ABT-263) and ABT-737	Abbott/Genentech	Cancer	BCL2/substrates	Phase II
Obatoclax	Teva	Cancer	BCL2/substrates	Phase II
JQ1	n/a	Cancer	BRD4/histones	Pre-clinical
Tramiprosate	Neurochem	Alzheimer's disease	Amyloid beta peptide	Phase III (discontinued)
RG7112 and RO5503781	Roche	Cancer	p53/MDM2	Phase I
SP4206	Sunesis Pharmaceuticals	Organ and tissue graft rejection	IL-2/receptor	Pre-clinical
Compound 23 (an indandione)	Boehringer Ingerheim	Cervical cancer	HPV E1/E2 proteins	Pre-clinical
Compound 1 (a pyridylpyrimidine)	Wyeth	Bacterial diseases	ZipA/FtsZ	Pre-clinical
SP304	n/a	Inflammatory diseases	TNF/TNFR	Pre-clinical
Eltrombopag	GSK/ Ligand Pharmaceuticals	Thrombocytopenia (low platelet count)	Thrombopoietin receptor	FDA approved 2008

Table 1.1 Summary of PPI inhibitors in development These drug candidates are all currently in pre-clinical or clinical trials with the exception of tramiprosate, where the Phase III trials were discontinued. Information: clinicaltrials.gov

1.5.2 Case study: JQ1 and targeting BRD4-histone interactions

BRD4 is a bromodomain-containing protein that can bind acetylated chromatin to modulate gene activation and repression (reviewed in Wu and Chiang, 2007). BRD4 associates with different transcription complexes including Mediator and P-TEFb, which phosphorylate RNA polymerase II on the C-terminal domain (Devaiah *et al*, 2012). Although the exact mechanism by which this process regulates gene transcription is still unclear, the effects of BRD4 on transcriptional activity explain why it is implicated in a wide range of cancers, including breast cancer and acute myeloid leukemia (AML).

BRD4 is the major driver of NUT midline carcinoma (NMC), a disease in which as a result of a genetic translocation BRD4 is expressed as a chimera with NUT (nuclear protein in testis). The BRD4-NUT fusion causes uncontrolled proliferation and is associated with a very poor disease prognosis (French *et al,* 2003). For this reason BRD4 is a major drug target. Disruption of the BRD4 interaction with transcriptional elongation complexes should decrease the proliferative ability of cancer cells. Another relevant function of BRD4 is as a potentiator of expression of Myc an oncogenic transcription factor, which when overexpressed, leads to unregulated gene expression. (Zuber et al, 2011).

JQ1 is a small-molecule inhibitor of BRD4. It was identified from molecular modelling and docking studies using the apo BRD4 structure as a template (Filippakopoulos *et al,* 2010). JQ1 binds in the acetylated lysine binding pocket and competitively inhibits the interaction of BRD4 with chromatin. The compound shows activity *in vitro* using NMC cell lines, including significant growth arrest, and slows down tumour growth in mouse xenograft models. The structure of BRD4 (and of BRD2) in complex with JQ1 has been determined (Filippakopoulos *et al,* 2010: shown in Figure 1.4)

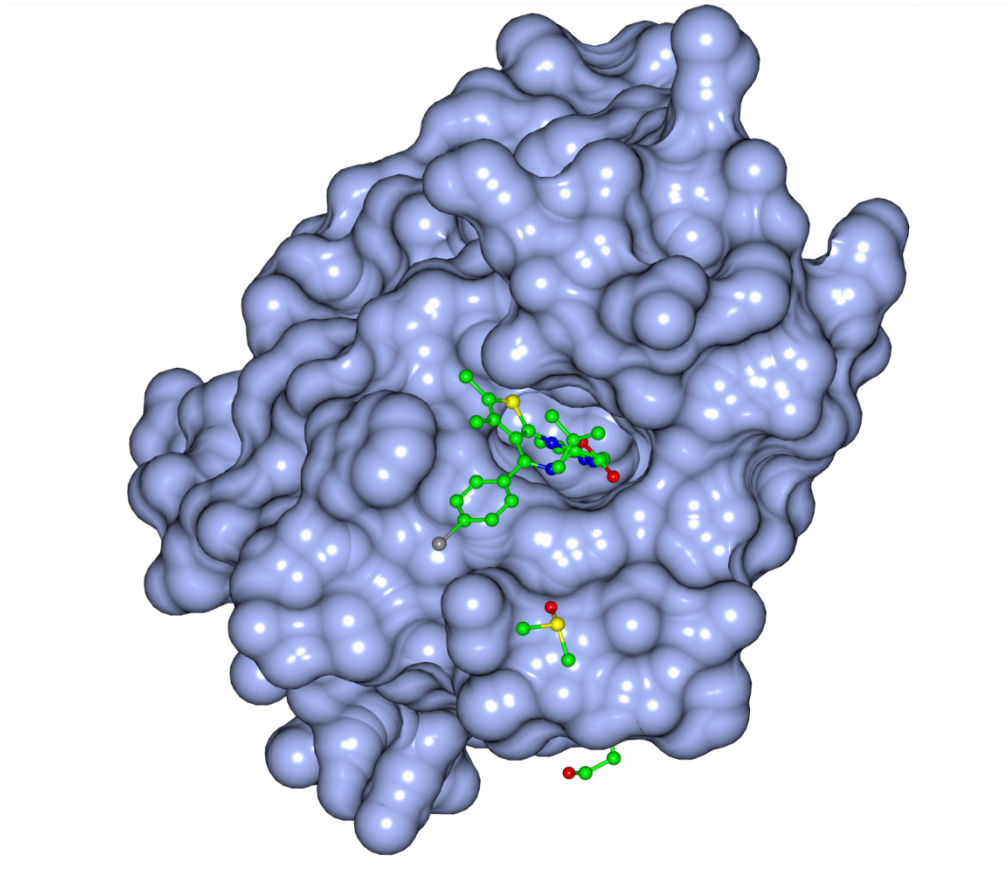


Figure 1.4 The structure of BRD4 in complex with JQ1 The protein is displayed as a blue surface, with the inhibitor as ball and stick (with carbon atoms in green). JQ1 inserts deep into the chromatin binding pocket and as shown here has very good shape complementarity with the binding site. Figure prepared in CCP4MG, from molecule 3MXF (Filippakopoulos *et al,* 2010)

Aside from NMC, JQ1 has also been tested on AML cell lines (Zuber *et al*, 2011). BRD4 activity was suppressed in these cell lines, either by JQ1 administration or by RNAi, and this led to growth inhibition (IC_{50} of <500 nM for JQ1) and apoptosis. In addition, JQ1 treatment elicited secondary anti-tumour effects that included increased likelihood of terminal myeloid differentiation, reduced or eliminated leukemic stem cell populations required to sustain tumour growth, and down-regulated Myc expression.

Although JQ1 development is still in a very early stage, there are causes for optimism: the molecule has good pharmaceutical properties and appears to be specifically and potently targeting BRD4 with few off-target effects. In addition, it has robust cellular outcomes on tumour cell growth, proliferation and differentiation.

1.5.3 Case study: Navitoclax and targeting Bcl2-substrate interactions

Navitoclax (Figure 1.5), also known as ABT-263, is a small-molecule PPI inhibitor of the anti-apoptotic factors Bcl-2 and Bcl-x(L).

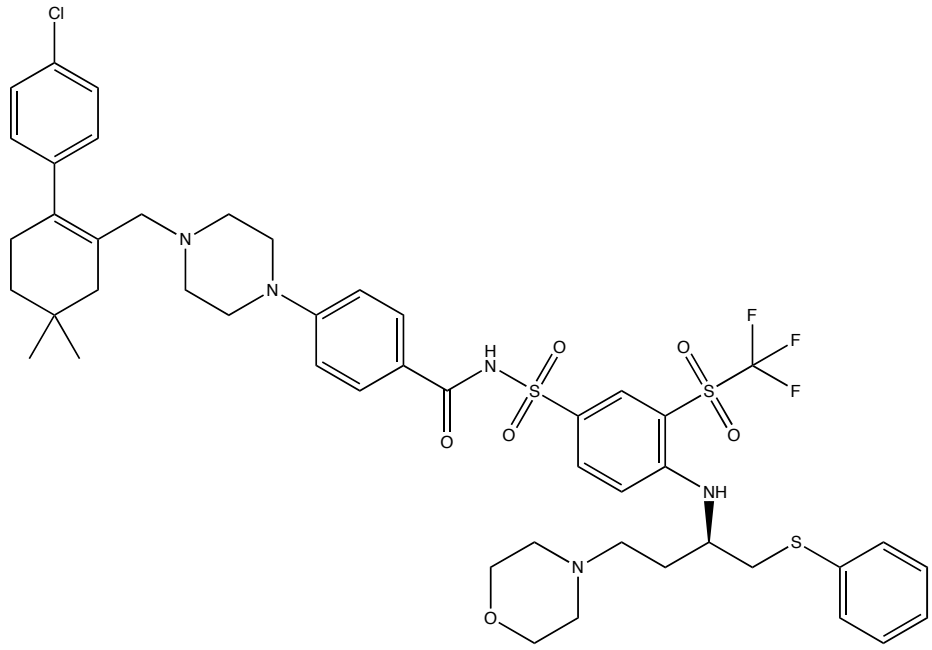


Figure 1.5 Structure of navitoclax (ABT-263)

Abbott and Genentex have developed this inhibitor following identification of a BH3 mimetic named ABT-737 from high-throughput NMR and parallel synthesis (Petros *et al*, 2006; Bruncko *et al*, 2007). ABT-737 was very potent against Bcl-2 and Bcl-x(L) but suffered from low aqueous solubility. A molecule was therefore sought which retained the efficacy of ABT-737 but had improved pharmaceutical properties, allowing oral bioavailability. This search led to the development of ABT-263 (Navitoclax). The drug shows Bcl-2/Bcl-x(L) –dependent cytotoxicity *in vitro* using small cell lung cancer (SCLC) cell lines, and leads to tumour regression in mouse xenograft models (Tse *et al*, 2008).

Navitoclax inhibits the interaction of Bcl-2 and Bcl-x(L) with the pro-apoptotic BH3-like proteins Bid, Bim, Bad, Bik, Bmf, Hrk, Noxa and Puma. The release of Bcl-2 and Bcl-x(L) allows restoration of apoptosis, and overcomes the block to

apoptosis represented by the overexpression of anti-apoptotic factors. A Phase I dose escalation study with Navitoclax (Wilson *et al*, 2010) showed a clinically relevant outcome with relatively low off-target toxicity in patients with relapsed or refractory lymphoid malignancies, at the selected dose. Some tumour reduction was seen in around half of patients.

The effects of navitoclax within cells have been visualised using high-resolution *in vivo* imaging of fluorescently-labelled inter-membrane space reporter protein (IMS-RP). This protein is normally located in the mitochondrial inter-membrane space but is released on apoptosis as the mitochondrial membrane loses structural integrity and so can be used as a genetic reporter of apoptosis. Using both live mouse xenografts and breast and pancreatic cancer cell lines *in vitro*, the induction of mitochondrial outer membrane permeabilisation was observed when navitoclax was administered (Earley *et al*, 2012). This result directly confirms the mode of action of navitoclax and shows that it can induce apoptosis *in vivo* as well as *in vitro*.

A small-scale Phase II study of navitoclax with 39 small cell lung cancer (SCLC) patients observed an effect on previously reported SCLC biomarkers: circulating tumour cell concentration (CTC), cytokeratin 19 fragment antigen 21-1 (CYFRA 21-1) and neuro-specific enolase (NSE) (Rudin *et al,* 2012). The patient responses in terms of biomarker improvement were also correlated with pro-gastrin releasing peptide (pro-GRP) expression. The pro-GRP gene is close to that coding for Bcl-2, so pro-GRP levels could also be used as a biomarker or an estimate of prognosis for success with this drug. This study also confirmed safety and the dosing regimen.

Navitoclax has been suggested to be especially useful in combination with other drugs, and has recently been tested in xenograft models in combination with bendamustine (an alkylating agent) and rituximab (an anti-CD20 antibody) (Ackler et al, 2012). These combinations had been postulated as an alternative therapy for non-Hodgkin's lymphoma to the standard combination therapy R-CHOP (rituximab, in combination with the cytotoxic drugs cyclophosphamide,

doxorubicin, vincristine and the immunosuppressant prednisolone). Navitoclax potentiation of both bendamustine and rituximab *in vivo* is myc- and p53-dependent and causes tumour growth inhibition and improves the durability of the response.

One of the interesting lessons that can be learned from the success of navitoclax in early trials with lymphoma and SCLC is the difference between the type of molecule that might be expected to be a PPI inhibitor and a "conventional" drug. Navitoclax violates three out of four of Lipinski's rules, being too large (at 975 Da), too hydrophobic (with a cLogP of 12), and having 11 hydrogen-bond acceptors. The lack of traditional 'drug-like' properties suggests that some of our previous attempts to search for new PPI inhibitors have had too narrow a focus, particularly by restricting the search to molecules that satisfy Lipinski's rule of five.

1.5.4 Case study: Tramiprosate (Alzhemed) and targeting amyloid β aggregation

Alzheimer's disease is a type of dementia that is driven by biochemical changes in the brain and in particular, by the build up of plaques, made of fibrils containing amyloid and tau proteins. This process is illustrated in Figure 1.6.

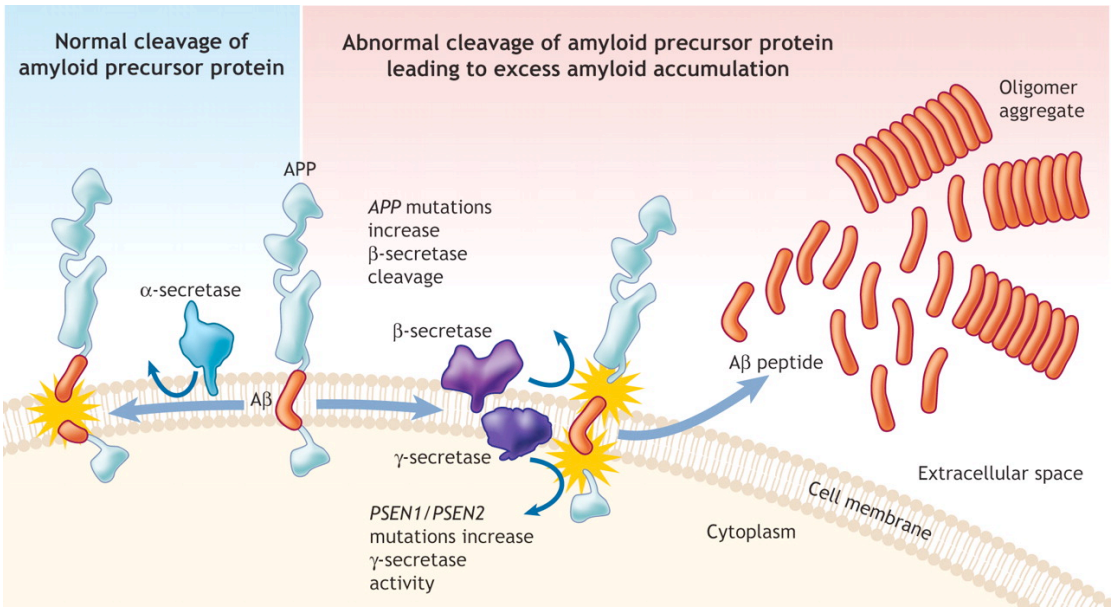


Figure 1.6 Abnormally cleaved amyloid precursor protein gives the fragment amyloid beta, which forms aggregate The amyloid precursor protein (APP) is normally cleaved in one position by α secretase. However, under certain circumstances, such as mutations in the APP; cleavage by β or γ secretase gives a shorter amyloid fragment called the Amyloid beta (Aβ), a prion that is prone to aggregation, initiating the process of plaque formation. Figure reproduced from Patterson *et al,* 2008.

Alzheimer's disease is frequently treated with neurotransmitter mimetics that only improve symptoms, rather than interrupting the fundamental processes involved in the development of the disease. Recently the identification of novel molecules that can interfere with the amyloid and tau cascades has provided new hope of improved treatment.

One of the problems with assessing the efficacy of Alzheimer's treatments is that the decline in cognitive function, which is the main symptom of the disease, can occur over many years making trial design difficult. In order to circumvent this

issue, novel Alzheimer's treatments are often assessed using biomarkers as reporters of efficacy. In addition, magnetic resonance imaging (MRI) and positron emission tomography (PET) can monitor changes in brain volume and function (reviewed in Thal *et al,* 2006).

The most useful biomarker for early clinical trials is the measurement of amyloid beta levels in cerebro-spinal fluid (CSF Aβ). This level reflects the amyloid beta concentrations available to diffuse into brain tissue and adhere to a growing plaque. In this way, changes in CSF Aβ can be used as an indirect measure of plaque growth and therefore disease progression. This approach has been validated using animal models: rats and transgenic mice dosed with an inhibitor of gamma secretase (the protease which causes the cleavage of the amyloid precursor protein, to form Aβ) show a decrease in Aβ in brain and CSF simultaneously (Best *et al,* 2006; Barten *et al,* 2004).

Tramiprosate (Alzhemed) was identified by Neurochem, (initially called NC-531) as a sulphated glycosaminoglycan mimic that showed activity *in vivo* when tested for inhibition of amyloid plaque growth (Gervais *et al,* 2001). The molecule appears to bind preferentially to soluble Aβ peptide and stabilise it in a non-fibrillar state, inhibiting the interaction with other molecules of Aβ peptide and interrupting the process of amyloid deposition in plaque. A phase II clinical trial lasting 36 months showed a reduction of CSF Aβ levels in patients with mild-to-moderate Alzheimer's disease (Aisen *et al,* 2007). In addition, this trial showed that the doses used were safe and well-tolerated, with few side effects, and that the effects were dose-dependent. A measurable improvement in clinical and cognitive tests such as the Clinical Dementia Rating Sum of Boxes (CDR-SB) was seen for some patients.

However, the promising results from this trial did not follow through to similarly impressive clinical outcomes from Phase III trials. The result of the North American Alphase trials carried out on 790 patients with mild-to-moderate Alzheimer's disease, across a timescale of 78 weeks, was reported recently (Aisen *et al*, 2011). In this trial some improvements were noted using

hippocampal volume loss as a proxy reporter for disease progress, but there was no significant slowing of cognitive decline using other tests. Due to the problems encountered within the Alphase trial, another large-scale Phase III trial being carried out in Europe with tramiprosate is now being reported as discontinued (clinicaltrials.gov).

The failure of tramiprosate as a viable drug at a relatively late stage in clinical trials raises important questions for the development of PPI inhibitors as drugs. In this case an initially promising drug candidate, which had shown good activity *in vitro* and had some effect on biomarkers and other proxy measures of disease progression, nevertheless failed to demonstrate the same significant effect on a symptomatic basis. Before yielding to pessimism over the fate of PPI inhibitors, however, it is worth considering possible reasons for the lack of activity of this particular drug within the Alphase trials.

Firstly, mild Alzheimer's disease is a particularly difficult disease to diagnose and assess, since the cognitive decline may be slow and gradual and can reach natural plateaus (Bozoki *et al*, 2009). In addition this phase of the disease can produce quite a heterogeneous set of symptoms and outcomes for patients. One issue that comes out of this trial is a lack of statistical power, that is the ability to identify a positive outcome to a statistically significant degree. In this case the patients showed more variance in their symptoms than expected, and the cognitive decline of the placebo group was much more slight than expected, making it harder to identify a positive result. However, this should not lead us to despair for the design of clinical trials for other PPI inhibitors. Diseases such as cancer, which is likely to represent the target class of the vast majority of PPI inhibitors, is unlikely to show the same problems regarding slow disease progression, although problems surrounding disease heterogeneity may still occur in some instances.

In terms of Alzheimer's research, the late-stage failures of clinical trials has led to some pharmaceutical companies pulling out of Alzheimers drug development: for example in August 2012, Pfizer and Johnson & Johnson withdrew from

development of intravenous bapineuzumab, after a failure in Phase III trials to produce better cognitive outcomes compared to placebo; Phase III results with Pfizer's Dimebon were negative against a placebo; Eli Lilly's Phase III trial of semagacestat failed catastrophically, as the drug actually appeared to worsen symptoms, and early results with solenazumab were unconvincing, although this drug is still in large-scale trials as a preventive and for patients with mild symptoms (clinicaltrials.gov).

A lack of confidence in the robustness of clinical outcomes and in the use of Aβ as a biomarker, in the face of the massive cost of running the long-term studies required to identify an effect on dementia, may be hampering efforts in this area of research (Abbott, 2008; Karran *et al.,* 2011). AstraZeneca, Pfizer, Merck, Sanofi, Novartis and GlaxoSmithKline all recently reduced their efforts in development of Aβ-targeting therapies and downsized their European neuroscience R&D teams.

In the case of Alzheimer's disease, other inhibitors of amyloid and tau cascades are in development, which might use a similar mechanism (reviewed in Blazer and Neubig, 2009). These include curcumin, benzoquinone derivatives and N744. The work on tramiprosate validates this approach to some extent since a decrease in hippocampal volume loss and mild possible effects in cognitive tests were seen. Maybe other PPI inhibitors that use a similar mechanism will have more pronounced effects measured against cognitive and other clinical outcomes.

1.6 A strategy for finding new PPI inhibitors

1.6.1 Target identification: interactomics

The first step in the search for a new PPI inhibitor is to identify a molecular target. In this respect, many of the key issues remain the same as those guiding target identification within traditional drug discovery areas. Firstly, the protein should be of a clearly defined function within the disease process of interest. The effects of inhibiting the function of the protein should be known, so that an understanding of the physiological effects of administering the putative drug can be assessed. Ideally, detailed information about the structure and function of the protein should be available in order to make the process of drug discovery easier. However, the process of target identification for PPI inhibitors will necessarily be somewhat different to that for traditional drugs which compete with small-molecule ligands, since the target in this case is not really one but two or many proteins.

Firstly, the competing molecule is a large protein, so the structure of the ligand cannot be used as a starting point for drug design in the same way. This constraint places more importance on characterising the interaction site, so that potential hotspots in the interface that might prove druggable can be identified. In addition, because many proteins will form multiple interactions with different partners at the same site all the binding partners should be identified in order to understand the downstream effects of PPI disruption. Some inhibitors may disrupt selective interactions depending on the overlap of different binding sites on a protein. Due to the unique nature of the target for a PPI, one of the techniques widely anticipated to make a significant impact in the process of target selection is interactomics.

A growing appreciation that structural and biochemical characterisation of individual proteins in isolation gives an incomplete picture of their functions within cells, has led to the development of interactomics, the study of the details of protein-protein, protein-ligand and protein-nucleic acid interactions. Using a

combination of experimental techniques such as yeast-2-hybrid, co-immunoprecipitation, mass spectrometry (MS), phage display, protein-fragment complementation and tandem affinity purification assays, a large volume of information about the interaction networks found within the cell has been generated (reviewed in Santonico *et al,* 2005). In order to organise and employ this data usefully, it has been assembled into numerous databases of interactions, associated with predictive and docking tools, including, for PPIs: DOCKGROUND, Protein3D, PRISM, SCOP, PsiBase and PICCOLO. Many of these databases use information from the structures deposited in the PDB along with published interactome data, to identify and characterise PPI interfaces (reviewed in Lee *et al*, 2009).

The main use of interactomics in drug development is for target identification, for example as used recently by Ramen *et al* (Ramen *et al*, 2012) to identify targets for tuberculosis drugs. In addition to the requirement for a well-characterised role in a disease process, an ideal drug target would be at a 'hub' or 'bottleneck' in an interaction network. If this is the case, then inhibiting the function of the protein will disrupt the whole network and lead to a significant physiological outcome, rather than inhibition being compensated for by other routes through the network.

Unfortunately, the reported interactomes can be quite different, suggesting that our toolkit for determining this information completely and accurately is still insufficient. While this remains the case interactomics cannot be regarded as a truly viable single tool to identify PPI drug targets. For example, two reports of the *S. cerevisiae* protein interactome (Ito *et al*, 2001; Uetz *et al*, 2000) produced only about 10-20% overlap. However, methods are continually improving (Koegl and Uetz, 2007) and interactomics may provide a useful first tool for narrowing down potential PPI drug targets within a network of interest.

1.6.2 High-throughput screening

High-throughput screening (HTS) allows hundreds or thousands of drug candidates to be screened per day, using experimental or computational methods. This throughput means that far more chemical space can be sampled during screening to increase the likelihood of identifying a suitable lead compound. Experimental techniques that can be used for HTS include fluorescence and light-based assays, ELISA, RNAi and cell proliferation assays (reviewed in Mayr and Bojanic, 2009). For example, a luciferase-based assay carried out in cells measures the stabilisation of the protein ATAD5 due to DNA damage, and can therefore identify genotoxic compounds (Fox and Myung, 2012). Typically, HTS assays are carried out in 96-, 384- or 1536-well plates, require specialist equipment such as robots and sensitive detection equipment, and as a result historically have been limited to pharmaceutical companies.

However, HTS can also be performed *in silico* using virtual screening, which uses docking software to evaluate the calculated binding capacity of pharmacophores. Much of the software required is open-source and relatively simple to use. It has recently been used for lead compound selection in a couple of difficult cases: against multi-drug resistant "superbugs" (Skariyachan *et al,* 2012) as well as for PPI inhibitor development, in order to identify inhibitors of the p53/mortalin interaction (Utomo *et al,* 2012). This latter study provides proof-of-concept that HTS and virtual screening in particular can be useful for PPI drug development (Voet and Zhang, 2012).

HTS techniques and approaches may need significant optimisation before use, since screening such large numbers of compounds makes false positives more likely. Hits should be validated using alternative techniques after HTS, especially for virtual screening (Coma *et al*, 2009). However, HTS (virtual screening) provides a useful approach for first-pass screening, especially with a new or poorly-characterised target where ligand-based drug development proves difficult. For PPI inhibitors, HTS has sometimes not produced positive results because the types of molecules present in the libraries under evaluation were

unsuitable to act as PPI inhibitors, having retained typical “drug-like” features and properties (Spencer, 1998; Cochran, 2000). Often larger, more hydrophobic molecules will not be included in the libraries since they would not be expected to behave as typical competitive inhibitors.

1.6.3 Natural product libraries

Natural products identified from plant, marine, fungal or microbial sources have always been a major source of new pharmaceuticals. The most famous examples are the analgesic and anti-inflammatory drug aspirin, a derivative of salicylic acid found in willow bark, and the antibiotics penicillin and tetracycline, from microbial sources. Often plants and other organisms produce high concentrations of insecticide, antimicrobial, toxic or bitter-tasting compounds in order to protect themselves against predators. Some of these compounds also turn out to have therapeutic possibilities within human medicine. Several natural products or natural product derivatives are approved drugs and many are currently in clinical trials (reviewed in da Rocha *et al*, 2001).

The recent increase in natural product drug discovery success has been partly due to the use of high-throughput screening, the assembly of commercially available natural compound libraries and the increasing availability of compounds extracted from marine sources such as sponges and tunicates. The impact of genome sequencing has also been profound since it allows the identification of novel biosynthetic pathways and helps inform drug discovery. Marine natural products have proven to be a particularly rich source of very pharmacologically active compounds. These include the approved anti-cancer drugs trabectedin, cytarabine and eribulin mesylate (reviewed in Mayer *et al,* 2010, Gerwick and Moore, 2012).

Natural products have proven a particularly rich source of compounds with anti-cancer activity. One of the first successful cancer drugs from natural product sources was the anti-proliferative agent taxol (paclitaxel), originally identified from the yew tree *Taxus brevifolia* during natural product screening at the US National Cancer Institute. Paclitaxel binds tubulin and interferes with microtubule disassembly and detachment from centrosomes, leading to a cell cycle block at the mitotic checkpoint (Jordan *et al,* 1993). Paclitaxel is prescribed for ovarian, breast and lung cancers. The Vinca alkaloids, vincristine and vinblastine, isolated from the Madagascan periwinkle *Catharanthus roseus* also

act on microtubules, instead by inhibiting microtubule assembly (Jordan and Wilson, 2004) and are prescribed for non-Hodgkin's lymphoma and ALL.

Many other anti-cancer agents derived from natural product sources are currently in clinical trials. Although too numerous to list exhaustively, some examples are given in Table 1.2.

Name	Source Organism	Class of Agent	Molecular target	Stage of development	Cancer use
Plitidepsin (Aplidine)	Tunicate	Cyclic depsipeptide	Rac1, JNK, VEGF	Phase III	Advanced myeloma
Plinabulin (NPI 2358)	Fungus (NP derivative)	Diketopiperazine	Microtubules, JNK	Phase III	NSCLC
Elisidepsin	Mollusk (NP derivative)	Cyclic depsipeptide	Plasma membrane fluidity	Phase II	Wide range eg. GI, NSCLC
Zalypsis (PM00104)	Nudibranch	Alkaloid	DNA binding	Phase II	Myeloma
Irvalec (PM02734)	Marine derivative	Cyclic depsipeptide	Erb-B/HER	Phase II (development suspended)	Oesophagal (large cell undifferentiated carcinoma)
PX-866 (wortmannin derivative)	Fungus (NP derivative)	Viridin	PI3K, PLK, Akt	Preclinical	Unknown
Glembatumumab vedotin (CDX-011)	Mollusk (NP derivative)	Antibody-drug conjugate	Glycoprotein NMB, microtubules	Phase II (Fast Track)	Breast, advanced melanoma
Flavopiridol	Plant (NP derivative)	Alkaloid	CDKs	Phase II	Resistant ovarian, maybe others
Bryostatin 1	Bryozoan	Polyketide	PKC	Phase I	Leukemia, prostate?

Table 1.2 Anti-cancer drugs from natural product sources in clinical or pre-clinical development

One of the most promising candidates is flavopiridol, a pan-CDK inhibitor derived from the plant alkaloid rohitukine isolated from *Dysoxylum binectifarium*, (reviewed in Kelland, 2000). Flavopiridol causes cell cycle arrest in G1 or at the G2/M boundary and can induce apoptosis in some cell types

(Shapiro *et al,* 1999; Schrump *et al,* 1998; Konig *et al,* 1997). The compound also shows significant anti-tumour activity in mouse xenograft models and has therefore progressed to human clinical trials (Sedlacek *et al,* 1996; Arguello *et al,* 1998; Drees *et al,* 1997). Flavopiridol may not be of use as a single-agent therapeutic as numerous Phase II studies in CLL, NSCLC, non-Hodgkin's lymphoma, kidney, colorectal and gastric cancers have shown a limited response. In some cases serious vascular thrombotic side effects were seen, which limited the doses that could be given (Liu *et al,* 2004; Schwartz *et al,* 2001).

However, recent studies employing flavopiridol in combination with genotoxic drugs have shown much more promise. When administered with doxorubicin (which intercalates into DNA) flavopiridol potentiated its activity and side effects were more limited (Luke *et al,* 2012). Phase II clinical trials with flavopiridol in combination with cisplatin showed particular efficacy in the case of platin-resistant ovarian cancer (Bible *et al,* 2012). The rationale behind the use of combined treatment is that flavopiridol inhibits CDK9, which is vital for the function of Rad51, a key player in homologous recombination, since it suppresses p53 inhibition of Rad51 (Chao *et al,* 2000; Lam *et al,* 2001; Chen *et al,* 2005). As a result the normal DNA damage response is ablated and therefore the likelihood of apoptosis upon administration of genotoxics is increased.

It is becoming increasingly clear that natural product screening has wide applicability when searching for PPI inhibitors. The types of molecule likely to be identified as active natural products include large, complicated, rigid molecules, similar to validated PPI inhibitors such as navitoclax (as previously discussed). Recent examples of natural product derivatives identified as PPI inhibitors include chlorofusin, an inhibitor of the p53/MDM2 interaction, identified by ELISA and originally from the microfungus *Microdochium caespitosum* (Clark *et al*, 2009). Gambogic acid, a caged xanthone identified from the bark of the tree *Garcinia hanburyi* acts as an anti-proliferative, an anti-angiogenic and can induce apoptosis. Among other targets it inhibits the interaction of Hsp90 with substrates such as the heme-regulated eIF2α kinase (Davenport *et al,* 2011). It is likely that many further examples of PPI inhibitors derived from natural

products will emerge in the next few years due to the increased interest in this area.

1.6.4 Substrate-focused design

As discussed in Section 1.2.1, substrate-focused design involves altering the properties of the drug molecule by employing various techniques that assess binding of compounds and identify the contribution of each chemical group to binding. These properties are often measured as ligand efficiency, the free energy of binding divided by the number of non-hydrogen atoms within the compound. Optimising ligand efficiency by adding or removing groups to a compound improves binding whilst maintaining drug-like properties and ease of synthesis. Usually, molecular docking and computational analysis of binding is used in the early stages of the project, before the biophysical measurement of binding for members of a compound family to obtain SAR data. The rational drug design approach then uses this data to design new compounds with improved binding, solubility and selectivity profiles.

Structural information is extremely important for modern drug design, as it provides the exact pose of the inhibitor within a binding site and confirms whether interactions suggested from computational docking actually occur. Structure-based drug design also allows interactions with water molecules within the binding site to be assessed and identifies any conformational changes occurring upon compound binding, which might alter the shape of the binding pocket to which compounds should be designed. The selectivity of compounds can be assessed using biophysical techniques, whether an extremely specific target is required, or a broad panel of targets is desirable. This is usually controlled by a combination of the spatial and electrostatic 'fit' of the inhibitor within the binding site, and specific interactions with residues or bridging waters (Huggins *et al,* 2012).

Fragment or scaffold-based drug design utilises the same SAR information but relates activity more directly to the structure of the compound and particular chemical groups. This is particularly useful for kinase drug design, in which the highly conserved binding site and the large number of drugs already targeting this site make drug design problematic. A scaffold-based approach was used to

identify ‘privileged’ chemical scaffolds able to form interactions with the hinge region (Zhang and Bollag, 2010). The construction of a library of scaffolds for drug design is illustrated in Figure 1.7.

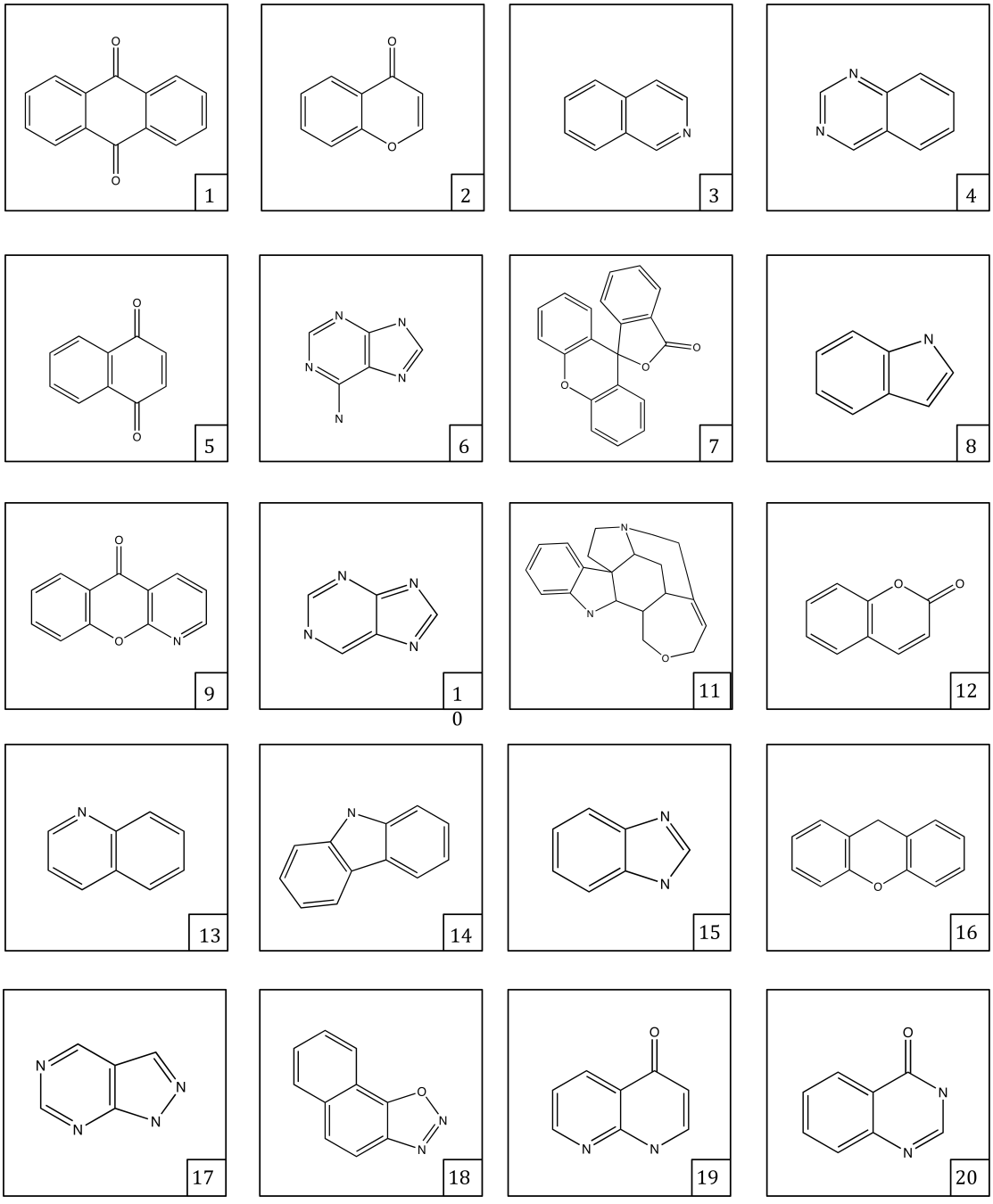


Figure 1.7: Scaffold-based drug design libraries allow iterative drug design to build an effective inhibitor Illustrated is a typical selection of scaffolds used for a drug design project. Chemical groups will be added to these scaffolds and the binding assessed by molecular docking. Figure adapted from Stahura *et al,* 1999.

1.7 Aims of this thesis

In this thesis I aim to show that inhibition of a variety of PPIs can be a drug development strategy where the molecular target is key to cancer development and is also a current drug target using conventional drugs. Taking up a project to develop a well-established compound series in collaboration with Newcastle University, I pursued structures of the most potent compounds in the series bound to MDM2, in order to help further inform drug development. The best compounds from this series are very close to being considered for clinical trials. Through identification of further lead compounds we hope to provide more evidence relating to the type of molecules that might be expected to act as protein-protein interaction inhibitors. Since this is still quite a new area, further compound characterisation is essential to allow the potential of this approach to be realised.

In Chapter 2 the results of screening polo-like kinase 1 (PLK1) inhibitors for binding to the Polobox domain are presented, as well as attempts to obtain a Polobox domain/inhibitor complex structure. This proved problematic, but docking was used to suggest a binding mode for these inhibitors. Chapter 3 contains work with an established series of MDM2/p53 PPI inhibitors, where I present three new structures of compounds bound to the p53 binding site of MDM2. I also present preliminary work with the MDM2 analogue MDMX. In Chapter 4, I present the results of medium-throughput screening of chemical libraries containing compounds that were expected to target the cyclin A recruitment site. After confirming that the compounds selected from this screen bind at the recruitment site, a hit compound was selected. Although this project is at a very early stage, the identification of this compound shows that this approach should allow an inroad into drug development. Finally, in Chapter 5 I present experiments using alternative approaches to inhibition of CDK2/cyclin A and compare the success of these approaches with that achieved from the work in Chapter 4.

2 Disrupting the MDM2/p53 interaction with isoindolinones

2.1 Background and introduction

2.1.1 p53: function and structure

2.1.1.1 The 'guardian of the genome': p53 as a tumour suppressor

p53 is a tumour suppressor involved in regulating the response to cellular stress such as DNA damage, hypoxia, oncogene activation or telomere shortening. It acts as a transcription factor and induces the expression of proteins coordinating the possible responses to stress of cell cycle arrest, apoptosis or cell senescence (reviewed in Ziffou and Lowe, 2009). These mechanisms are key to protecting the cell against transformation and tumorigenesis, as they prevent damaged cells progressing through the cell cycle, which could lead to uncontrolled or abnormal proliferation. Loss of functional p53, or dysregulation of the p53 pathway, therefore appears to be a necessary step in tumour formation (Vogelstein *et al*, 2000) and p53 gene is mutated in 50% of human cancers (Hollstein *et al*, 1991; Soussi *et al*, 2000).

Some of the pro-apoptotic p53 function is independent of its activity as a transcription factor (Essmann *et al*, 2005; Vaseva and Moll, 2009). p53 can act as a transcription factor for the caspases and death receptors (eg Fas), the pro-apoptotic Bcl2 family proteins (eg Bak, Bax, Puma and Noxa), various DNA repair components and p21 (an inhibitor of cell cycle progression). The exact cellular response induced depends on the balance of the resulting expression of these targets (Janicke *et al*, 2008; Qian and Chen, 2010). Figure 2.1 shows some of the main pathways involving p53.

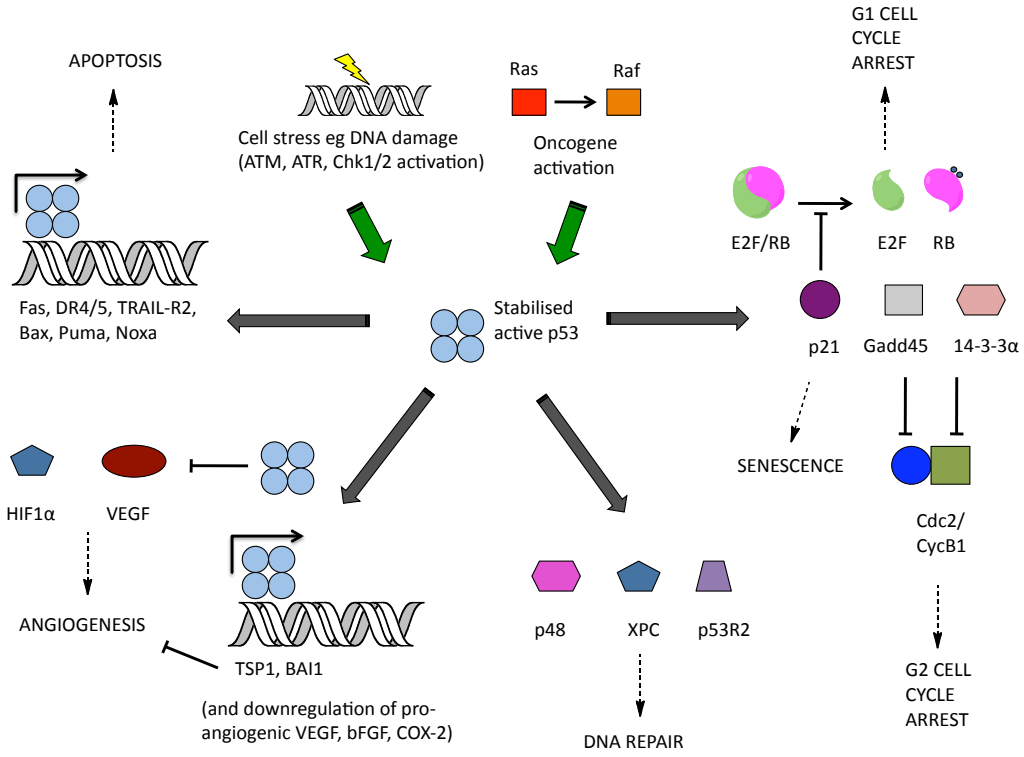


Figure 2.1 Cellular effects of p53 activation Upon cellular stress, caused by DNA damage, telomere shortening, oncogene activation, oxidative stress or DNA damage, p53 is activated, and produces a wide range of cellular effects. These include apoptosis, via either the extrinsic (death receptor) pathway or the intrinsic (mitochondrial) pathway, (mediated by p53-induced transcription of pro-apoptotic proteins); DNA repair, due to up-regulation of p48, XPC and p53R2 transcription by p53; senescence or cell cycle arrest in G1, mediated by p21; cell cycle arrest in G2 due to inhibition of the CDK1/cyclinB1 complex by p53; and suppression of angiogenesis, by p53 interference with VEGF and HIFα activity, and transcriptional control of pro-angiogenic and anti-angiogenic factors by p53. The stabilisation of p53 therefore has extremely wide-ranging consequences for cellular processes and active p53 effectively disrupts progression towards tumorigenesis.

Since high levels of p53 activity can lead to cell cycle arrest or apoptosis, and low levels of p53 activity can lead to loss of the DNA damage checkpoint, p53 activity must be tightly controlled within the cell cycle. In an unstressed cell, p53 is normally maintained at a low level due to constant turnover; its expression being balanced by proteasomal degradation (Hock and Vousden, 2010). However, upon sensing cellular stress such as DNA damage, p53 expression is maintained but degradation is inhibited, leading to an increase in p53 levels and, as a result, induction of p53-mediated cell cycle arrest, apoptosis or senescence (Harris and Levine, 2005). p53-null mice spontaneously develop tumours, which shrink following re-expression of endogenous p53 (Martins *et al*, 2006; Ventura *et al*,

2007; Xue *et al*, 2007). This tumour regression is mostly due to apoptosis, but clearance of tumour cells in p53-dependent senescence by the immune system is also observed.

Reflecting its many roles in the cell, p53 is composed of multiple domains (Figure 2.2).

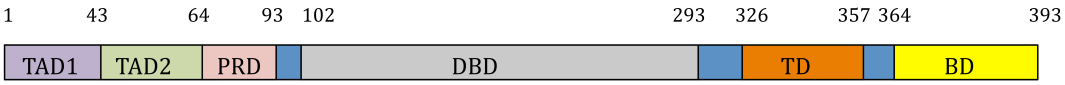


Figure 2.2 p53 protein architecture TAD1/2: transactivation domain 1/2; PRD: proline-rich domain; DBD: DNA-binding domain; TD: tetramerisation domain; BD: basic domain. Residue numbers are given above.

The DNA binding domain (DBD) is flanked by domains that assist in the regulation of p53. p53 is tetrameric when it is bound to DNA as a transcription factor and this oligomerisation is mediated by the tetramerisation domain (TD) located C-terminal to the DNA binding domain (Harms and Chen, 2006). The N-terminal transactivation domains (TAD1/2) are the binding site for components of the transcription machinery, and for MDM2 and MDMX (Zhu *et al,* 2000). Finally, the intracellular localization of p53 mediated by the basic domain is also an important factor in its regulation since export away from the nucleus prohibits p53 function and proteasomal degradation of p53 occurs in the cytoplasm. Notably, there are more than 10 different isoforms of p53 that arise due to alternative splicing and the use of alternative promoters, not all of which are functionally well characterised (Bourdon *et al,* 2005; Schwerk and Schulze-Osthoff, 2005).

2.1.1.2 Binding partners and interactions

2.1.1.3 p53 and cancer development

Most mutations of p53 in tumour cells are in the DNA-binding domain (Walker *et al,* 1999), and target a few 'hot spots' of residues directly involved in DNA binding. These mutations inhibit the binding of p53 to its target genes, or alter the sequence specificity of DNA binding, sometimes resulting in oncogenic function. Alternatively, mutations can cause destabilisation of the DNA binding domain structure, leading to loss of function, or dysregulation by interfering with interactions with binding partners.

2.1.2 MDM2: inhibitor of p53 function

2.1.2.1 Effect of MDM2 binding

MDM2 (murine double minute 2) was first identified as an oncogene able to cause transformation in mouse cell lines, giving the characteristic chromosome distribution known as double minutes (Cahilly-Snyder *et al,* 1987). MDM2 is the main negative regulator of p53 and controls p53 activity to avoid cell cycle arrest or apoptosis in undamaged cells (reviewed in Freedman *et al*, 1999). It is an E3 ubiquitin ligase and polyubiquitinates the C-terminal lysine residues of p53 (Toledo and Wahl, 2006). The transfer of the ubiquitin moiety from the E2 ubiquitin conjugating enzyme is dependent on the RING domain in MDM2 (Itahana *et al,* 2007). Ubiquitination of p53 promotes its nuclear export and degradation by the proteasome. MDM2 binding to p53 also inhibits its activity by blocking the transactivation domain.

The inhibition of p53 by MDM2 activity is counter-balanced by activating binding partners such as $p14^{Arf}$ (Harris and Levine, 2005). MDM2 is expressed at low levels in unstressed cells, as transcription of MDM2 is activated by p53, therefore MDM2 inhibits its own degradation. Upon p53 activation (due to cellular stress),

the MDM2/p53 interaction is disrupted and free p53 levels increase: in turn this causes an increase in MDM2 transcription, limiting further p53 activity in an autoregulatory feedback loop (illustrated in Figure 2.3). MDMX (MDM4) is an MDM2 homologue that does not ubiquitinate p53 directly but forms heteromers with MDM2 and enhances its inhibition of p53 (Linares *et al,* 2003; Hock and Vousden, 2010). A single knockout of either MDM2 or MDMX is embryonic lethal (Toledo and Wahl, 2006).

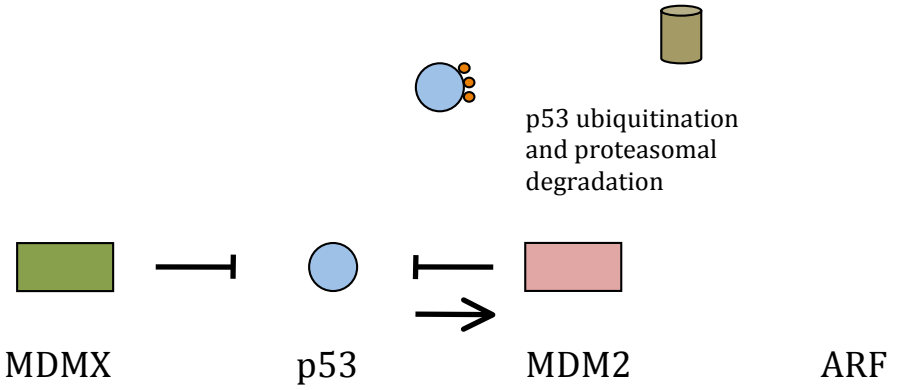


Figure 2.3 Control of p53 activity in an unstressed cell p53 and MDM2 are both kept at low levels in a normal unstressed cell by an autoregulatory feedback loop, in which MDM2 inhibits p53 activity by (i) causing its export into the cytoplasm, where it is ubiquitinated and subsequently degraded by the proteasome, and (ii) inhibiting its action as a transcription factor by blocking tetramerisation. MDMX also inhibits p53 activity and ARF inhibits MDM2 inhibition of p53.

2.1.2.2 MDM2 structure

The structure of MDM2 bound to a short peptide derived from p53 has been determined (Kussie *et al*, 1996), and is illustrated in Figure 4.4. The protein consists of two structurally similar portions of three β-sheets and two α–helices each, related by symmetry but with only 11% sequence identity. The two globular halves pack to form a large hydrophobic cleft at the interface, about 25Å long by 10Å wide, and 10Å deep at the deepest point. The p53 peptide inserts three amino acid side chains into three smaller hydrophobic pockets deep in the cleft, known (by the p53 residue number) as the W23, F19 and L26 pockets. This

crystal structure gave the first indication that the interaction may form a viable target for drug development, since the binding site is abnormally deep and well-defined for a PPI site.

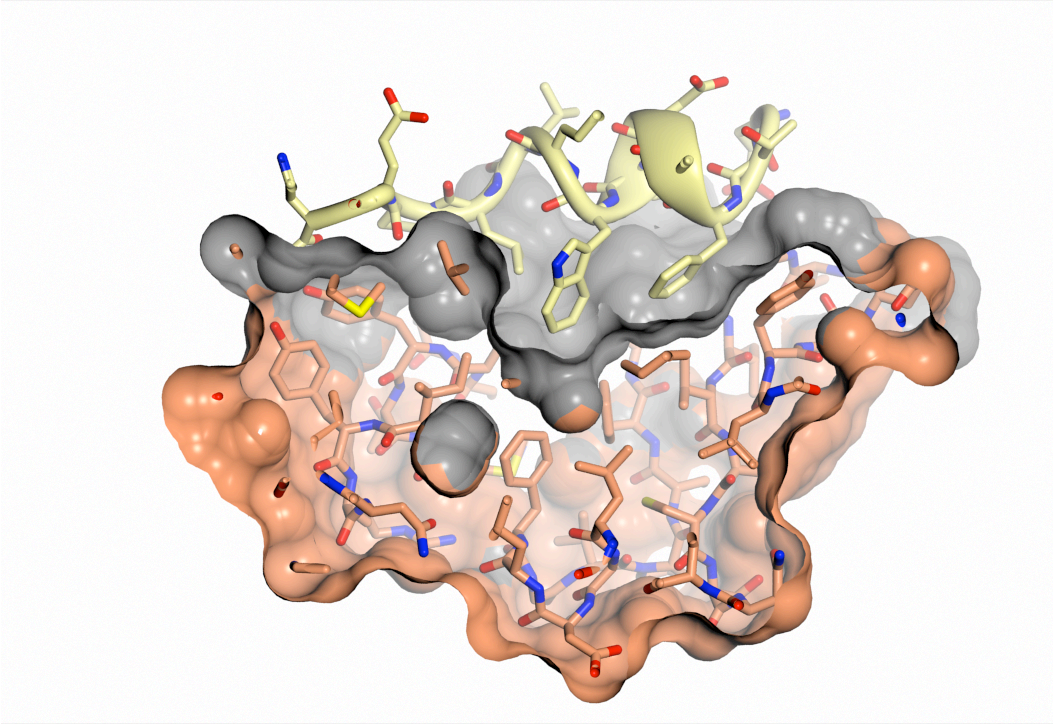


Figure 2.4: 1YCR structure MDM2 is shown in ball and stick representation, in coral. The surface of the protein is shown, with areas in direct contact with the p53 peptide coloured grey. The rest of the surface is coloured coral. The p53 peptide is represented as a light yellow ribbon, and the side view shows a leucine, tryptophan and phenylalanine residue each inserting their side chains into pockets on the MDM2 surface. Figure prepared in CCP4MG.

In this structure F19, L26 and W23 from the p53 transactivation domain peptide were inserted in a deep hydrophobic pocket in MDM2. The development of small-molecule inhibitors that could disrupt the p53-MDM2 interaction has focused on mimicking the shape of these amino acids.

2.1.3 MDM2: drug target

2.1.3.1 Effect of disrupting the MDM2/p53 interaction

Disruption of the MDM2/p53 interaction by small-molecule inhibitors or peptides, or removal of MDM2 using antisense technologies has been shown to be an effective strategy for activating p53-dependent pathways such as cell cycle arrest and apoptosis (Vassilev, 2007; Shangary and Wang, 2008). The development of small-molecule inhibitors targeting the MDM2/p53 interaction is therefore an important current area of investigation.

It is possible that MDM2/p53 PPI inhibitors may be especially effective in combination with genotoxic drugs. Such drug combination should limit toxicity and further encourage p53 activity in tumour cells once p53 is free from MDM2 inhibition (Secchiero *et al,* 2006; Coll-Mullet *et al,* 2006). However there are several factors which may limit the use of such a regime: firstly, the possible toxicities, which may be problematic in a combination therapy, although these do not appear too problematic in mouse models (Shangary *et al,* 2008); secondly, the difficulty of identifying tumours with wild-type p53 (and intact downstream pathways), without which this therapy would be ineffective; and finally, the problem of drug resistance, arising in tumour cells after treatment with MDM2/p53 disruptors, due to acquired defects in p53 itself. In addition, disruption of the MDM2/p53 interaction is compensated for in some tumour cell populations by upregulation of MDMX expression, so MDMX/p53 disrupting inhibitors (such as SAH-p53-8) may also need to be employed (Marine, 2010; Bernal *et al,* 2010).

2.1.3.2 Existing compounds targeting the p53-binding pocket

Small molecule non-peptidic inhibitors that target this interaction include the nutlins, cis-imidazolines identified during screening of large synthetic chemical libraries (Vassilev *et al,* 2004; Tovar *et al,* 2006), MI-63, a chromenotriazolopyrimidine inhibitor discovered by Amgen, the piperidinone

inhibitors rationally designed based on the binding modes of known MDM2/p53 PPI inhibitors (Rew *et al,* 2012), the spiro-oxindoles (Ding *et al,* 2005) and the benzodiazepinediones, (Grasberger, 2005). X-ray crystal structures show that each of these inhibitors disrupts the interaction of MDM2 and p53 by binding to MDM2 in the hydrophobic cleft and blocking p53 access (illustrated in Figure 2.5). Further series of small molecule inhibitors with high potency towards MDM2 have been reported recently, including the indolyl hydantoins (Graves *et al,* 2012) the imidazolyl indoles (Furet *et al,* 2012) and our series of isoindolinones (Hardcastle *et al,* 2006; Hardcastle *et al,* 2011; Hardcastle *et al,* 2011b; Watson *et al,* 2011).

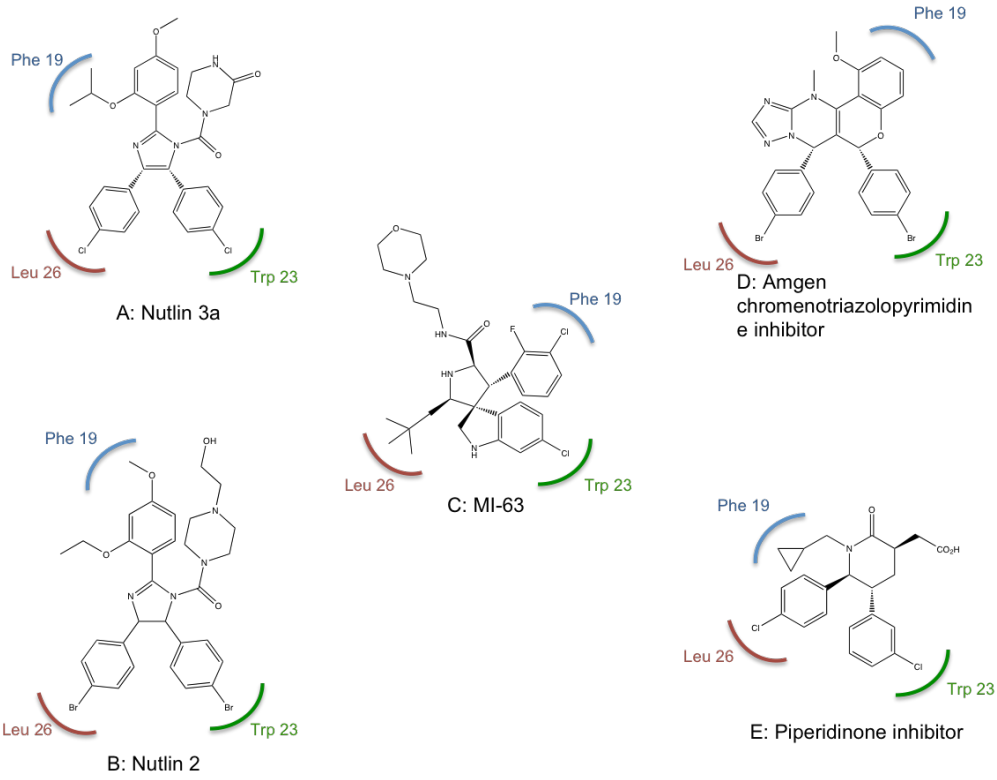


Figure 2.5 Binding modes of known inhibitors of the MDM2/p53 interaction The binding pockets on MDM2 are indicated by the p53 residue that normally occupies that pocket. Binding modes determined by X-ray crystallography, except for E, in which the binding mode was determined by molecular docking.

The nutlins and the spiro-oxindoles MI-63 and its derivative MI-219 are potent and specific inhibitors of MDM2. For example, Nutlin 3 (the most potent) has an IC_{50} of 90nM and is extremely selective for MDM2 over MDMX. Both Nutlin 3 and

MI-219 show significant anti-tumour activity in mouse xenograft models, causing increased p53 levels and activation of p53-dependent pathways, with downstream effects including upregulation of p21 expression. *In vitro*, cells are blocked in G1 or G2 phase and apoptosis is observed upon administration of low μM concentrations of Nutlin 3 (Secchiero *et al*, 2007). The activity of these inhibitors is dependent on wild-type p53 (Tovar *et al,* 2006; Sarek *et al*, 2007) and can result in apoptosis or senescence depending on the tumour type (Efeyan *et al*, 2007). They have also been tested successfully in a small group of B-CLL and AML patients (Kojima *et al*, 2006; Saddler *et al*, 2007).

Other MDM2/p53 inhibitors include the benzodiazepinediones (Koblish *et al*, 2006), HL198C, RITA (Reactivation of p53 and induction of tumour apoptosis), SJ-172550 (a compound identified from HTS that binds MDMX with a 25-fold preference over MDM2; Bernal *et al*, 2010), the quilinols and SAH-p53-8 (a so-called 'stapled' peptide, or stabilised α-helix). Due to the wealth of information now available about inhibitor binding to MDM2 from structural studies (summarised in Table 2.1) and the successes of some of these inhibitors through pre-clinical testing, the development of further PPI inhibitors targeting MDM2/p53 is an important current goal.

PDB ID	Paper	Content of structure
1YCR	Kussie and Pavletich, 1997	MDM2 with p53 peptide
1RV1	Lukacs *et al,* 2004	MDM2 with imidazoline inhibitor (Nutlin 2)
1T4E	Grasberger *et al,* 2005	MDM2 with benzodiazepine inhibitor
1TTV	Fry *et al,* 2004	MDM2 with imidazole piperazine inhibitor (NMR structure)
1T4F	Grasberger *et al,* 2005	MDM2 with optimised p53 peptide
3JZK	Huang, 2009	MDM2 with chromenotriazolopyrimidine 1 inhibitor
3LBJ, 3LBK	Popowicz *et al,* 2010	MDM2 with chloroindole inhibitor
3LBL	Popowicz *et al,* 2010	MDM2 with MI-63 analogue
3V3B	Baek *et al,* 2012	MDM2 with stapled p53 peptide
4ERE	Huang, 2012	MDM2 with compound 23
4ERF	Huang, 2012	MDM2 with compound 29 (AM-8553)
3TPX	Wu and Pazgier, 2012	MDM2 with trifluoromethylated D-peptide inhibitor

Table 2.1 Structures of MDM2 bound to inhibitory peptides or small-molecule inhibitors Unless otherwise stated, structures were determined by X-ray crystallography. Information extracted from the PDB

2.1.4 Development of the isoindolinone series of compounds

This project involved investigation of the isoindolinone inhibitors of the p53-MDM2 interaction, developed by our collaborators at the University of Newcastle (Hardcastle *et al*., 2005; Hardcastle *et al*., 2006; Riedinger *et al*., 2008; Hardcastle *et al,* 2011; Hardcastle *et al,* 2011b), using X-ray crystallography. The basic isoindolinone scaffold is illustrated in Figure 2.6.

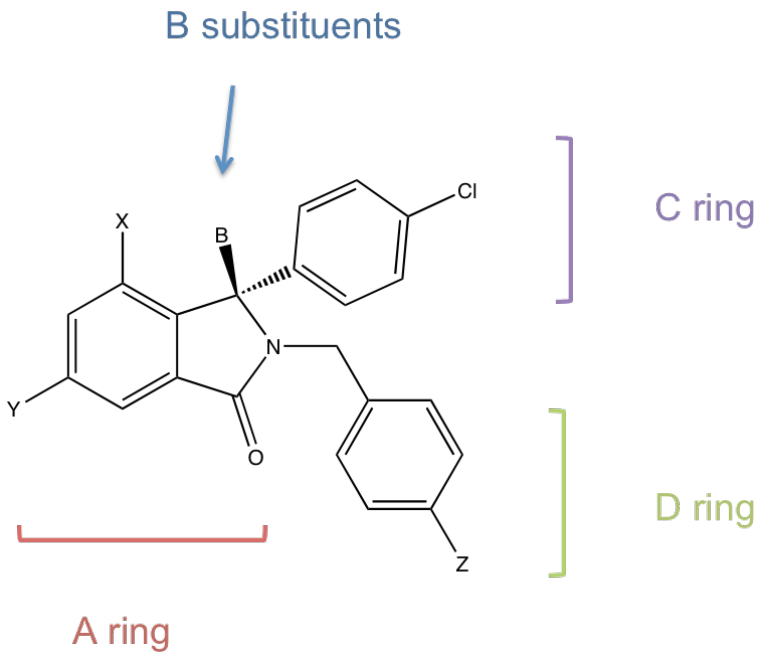


Figure 2.6 Structure of the isoindolinone scaffold The four parts of the molecule are henceforth known as the A, B, C and D rings. The substituent at B varies and different chemical groups have been tried at points X, Y and Z in order to increase potency.

This inhibitor series now contains several highly potent compounds that demonstrate significant anti-proliferative effects in tumour cells, dependent on p53 status.

A full structure determination of these inhibitors bound to MDM2 is an all-important step for their development and will provide detailed structural information about the interface, which cannot be determined from the previous NMR titration data (Riedinger *et al,* 2008). The titrations were informative in that they rapidly gave preliminary structural information about binding modes and likely affinity, as well as the relative affinity of enantiomers. MDM2 has

always been an extremely challenging crystallisation target, and the crystal structure of MDM2 has not been determined in its *apo* form, or bound to full-length p53, mainly due to its flexibility. The p53 binding site on MDM2 has been shown to be poorly structured and flexible in the absence of a ligand (Uhrinova *et al,* 2005). The available co-crystal structures of MDM2 and small molecule inhibitors demonstrate that ligands are able to increase the stability of MDM2, reduce flexibility within the p53-binding site and provide crystal contacts (Vassilev *et al,* 2004). The presence of a high-affinity ligand therefore appears to be a strict requirement for crystallization of MDM2 (Kussie *et al,* 1996; Grasberger *et al,* 2005; Allen *et al,* 2009; Sakurai *et al,* 2004).

The lower affinity of our compound series has therefore until recently limited the potential for their co-crystallisation with MDM2. For this reason we have previously concentrated on NMR-based approaches (particularly chemical shift analysis) to analyse the binding modes of the isoindolinone series with MDM2. However, recent progress in the development of the isoindolinones provided the impetus for the switch in focus to co-crystallisation of MDM2 with isoindolinone inhibitors. Since the crystallisation process is so dependent on the ligand provided, new crystallisation conditions need to be determined for each ligand-MDM2 complex. We therefore produced new MDM2 constructs optimally designed for crystallisation and utilised commercially broad screens to identify crystallisation conditions.

2.2 Expression and purification of MDM2 constructs

2.2.1 Constructs produced to aid crystallisation

The design of constructs for use in co-crystallography of protein-ligand complexes such that diffraction quality crystals can be obtained is becoming increasingly sophisticated (Dale *et al*, 2003; Bonnefond *et al*, 2011). One technique involves the surface engineering of proteins (Derewenda *et. al,* 2004). This strategy, known as 'surface-entropy reduction' (SER), attempts to reduce the entropic penalty contribution for the crystallisation process from large polar side-chains of amino acids (lysine, glutamate or glutamine), by the mutation of these residues to alanine. These residues are particularly suitable candidates for mutagenesis since they are likely to occur on the protein surface (Baud *et al,* 1999) and are not usually involved in PPIs (Conte *et al,* 1999). The strategy usually focuses on patches on the protein surface containing Lys/Glu/Gln clusters, to favour new hydrophobic interactions with other protein molecules and the formation of crystal contacts, and to reduce the entropy penalty of crystallisation. Exposure of carbonyl and amide groups when large side-chains are removed also allows the formation of ordered solvent networks, favouring the entropy component of crystallisation (Derewenda *et al,* 2011).

We successfully employed SER to engineer more crystallisable forms of MDM2, allowing us to sample a wide array of crystallisation conditions. Constructs were made by Burcu Anil Kirmizitas: the MDM2 constructs used employed two lengths (residues 17-108 or 17-125), building on previous work by Claire Chivers and Christiane Riedinger, in which the C-terminal tail (residues 109-125) was demonstrated to be extremely flexible, thus implying that its presence may hinder the crystallisation process (shown in Figure 2.7):

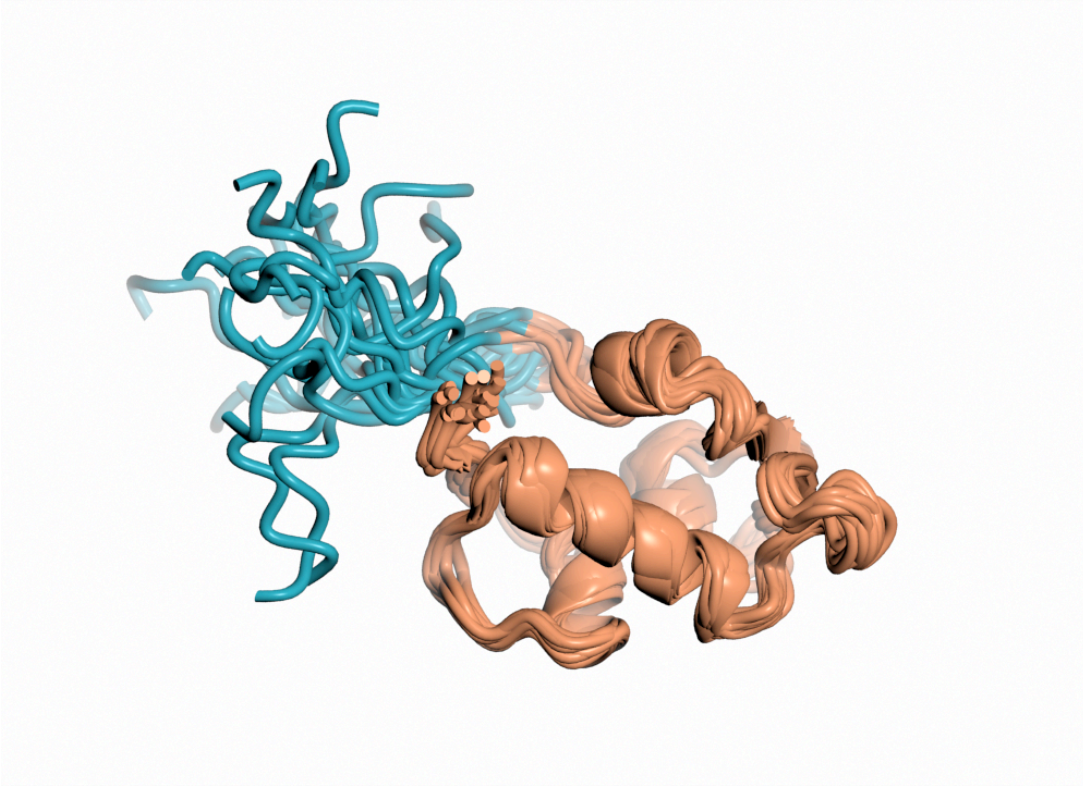


Figure 2.7 NMR solution structure of MDM2 residues 17-125: PDB ID code 1Z1M. The N-terminal globular domain is shown in coral, and the C-terminal flexible tail is represented in blue. The top 20 solutions are shown. The protein is shown in ribbon view. Figure prepared using CCP4MG.

The surface mutations remove highly polar residues that might inhibit crystallisation, and the removal of the flexible C-terminal tail aids crystallisation by removing another entropic barrier. Further MDM2 constructs were produced including the residues 18-109 or 18-125, with different surface mutations (to remove polar residues): the single mutants K51A, E52A, E69A, K70A, K94A, as well as the double mutants K51E52A, E69K70A and K94E95A. MDM2 was expressed as a GST-tagged fusion from pGEX6P-1 in BL21(DE3) cells. Protein purification was performed using a glutathione sepharose affinity column, elution with glutathione, cleavage of the GST tag by 3C protease and size exclusion chromatography. The purity of the protein was assessed by SDS-PAGE.

The crystal structures of the N-terminal domain of MDM2 (residues 17-109, E69K70A) in complex with Nutlin-3a (at 1.6 Å) and with the p53 peptide (residues 15-29) at 2.7 Å had previously been determined by Burcu Anil Kirmizitas (as detailed in Section 2.3.4), who also used a thermal denaturation assay (measuring protein stability) to validate the binding properties of the

double mutant MDM2 with respect to p53 and other ligands. We then built on this work and determined the crystal structures of three compounds from our isoindolinone series in complex with the N-terminal domain of MDM2 (in each case using a double mutant construct as described above).

2.2.2 Expression of MDM2 constructs

The different constructs of MDM2 were expressed as GST fusion proteins from pGEX6P-1 plasmids. The plasmids were transformed into *E coli* cells (strain Rosetta (DE3)/pLYS-S). Cultures were grown at 37°C in 1L Hyperbroth (supplemented with 100 μg/ml carbenicillin) in 5L flasks, until they reached an OD_{600} of 0.6. Protein expression was induced by addition of 0.2 mM IPTG, and maintained overnight at 18°C.

The yields from expression were variable but generally the levels of GST expressed were much higher than MDM2. Optimisation and repetition of the expression and purification process produced sufficient protein to test a large number of conditions in broad screens and to allow co-crystallisation trials with several NCL inhibitors.

2.2.3 Purification of MDM2 constructs

Purification of MDM2 was carried out from pellets frozen at -80°C. Cell pellets expressing MDM2 as a GST fusion were resuspended in HBS (50mM HEPES, pH 7.5, 250mM NaCl, 0.02% MTG, supplemented with Complete protease inhibitor cocktail, Roche), and after sonication and clarification (at 100,000g), the cell lysates were applied to a glutathione sepharose 4B gravity flow affinity chromatography column. GST-MDM2 bound to the column was eluted by addition of glutathione (10 mM in HBS) and then the GST tag was cleaved by 3C protease digestion overnight at 4°C (1:50 w/w). The mixture of MDM2 and GST was subsequently buffer-exchanged and further purified by size-exclusion chromatography (using an HR 26/60 S75 column, GE Healthcare). Fractions were pooled and the purity of the final product was checked using SDS-PAGE. The results of expression and purification trials are summarised in Table 2.2.

Construct name	Expression	Solubility	Purification status
MDM2_17-125	3 mg/l	≤ 15 mg/ml	Complete, no impurities
MDM2_17-125_K51A	3 mg/l	≤ 15 mg/ml	Complete, no impurities
MDM2_17-125_E52A	3mg/l	≤ 15 mg/ml	Complete, no impurities
MDM2_17-125_E69A	2 mg/l	≤ 15 mg/ml	Complete, no impurities
MDM2_17-125_K70A	2.5 mg/l	≤ 13 mg/ml	Complete, no impurities
MDM2_17-125_K94A	3 mg/l	≤ 15 mg/ml	Complete, no impurities
MDM2_17-125_K51E52A	2.5 mg/l	≤ 20 mg/ml	Complete, no impurities
MDM2_17-125_E69K70A	4 mg/l	≤ 20 mg/ml	Complete, no impurities
MDM2_17-125_K94E95A	2 mg/l	≤ 20 mg/ml	Complete, no impurities
MDM2_17-109	4 mg/l	≤ 17 mg/ml	Complete, no impurities
MDM2_17-109_K51E52A	4 mg/l	≤ 20 mg/ml	Complete, no impurities
MDM2_17-109_E69K70A	4 mg/l	≤ 20 mg/ml	Complete, no impurities
MDM2_17-109_K94E95A	3 mg/l	≤ 20 mg/ml	Complete, no impurities

Table 2.2 Progress of purification of MDM2 constructs for crystallisation

The progress of a typical purification is shown in Figure 2.8.

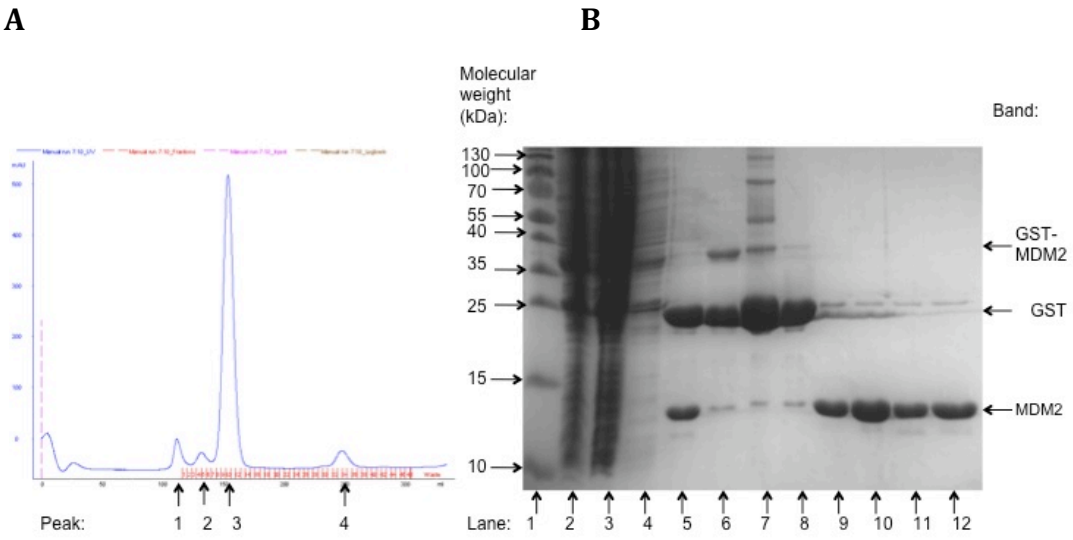


Figure 2.8 Purification of MDM2 E69K70A residues 17-109 A: chromatogram from gel filtration. Peak 1, 2: Aggregates; Peak 3: GST; Peak 4: MDM2. B: Photograph of SDS-PAGE gel. The positions of the GST and MDM2 bands are marked to the right, the sizes of bands in the ladder are marked to the left, and the number of the lane is marked below. Lane: 1 = protein standards; 2 = lysate; 3 = flow-through (affinity column); 4 = wash (affinity column); 5 = elution (affinity column) cleaved by 3C protease; 6-8 = fractions from peak 3 of gel filtration; 9-11 = fractions from peak 4 of gel filtration; 12 = final product

Generally, protein expression and purification was successful with these MDM2 constructs, although some had higher yields than others. We quickly switched to working mainly with MDM2 E69K70A residues 17-109 and MDM2 K51A residues 17-125, because these constructs grew crystals much more readily than other constructs.

2.3 Co-crystal structures of MDM2 with selected inhibitors

2.3.1 Crystallisation

2.3.1.1 Broad screens

Crystallisation trials were performed using MDM2 expressed and purified as detailed above, then incubated overnight at 4°C with a 2-fold molar excess of isoindolinone compounds. After incubation protein was concentrated to 12-15 mg/ml using Vivaspin concentrators (MWCO 5000 Da) with two washes in HBS.

Commercial broad screens were employed and trays were set up using the Tecan and Mosquito robotic pipetting systems (to dispense well solution and protein, respectively) in 96-well (3-subwells) SwissSci sitting drop plates. These plates allow for a wider range of crystallisation conditions to be covered for each construct since the proportion of well solution to protein can be varied. The broad screens used for crystallisation included Morpheus, PACT, JCSG+, MD1+2, PEG/ION, Wizard, Proplex and Midas. Crystals were grown in the following conditions (shown in Figure 2.9).

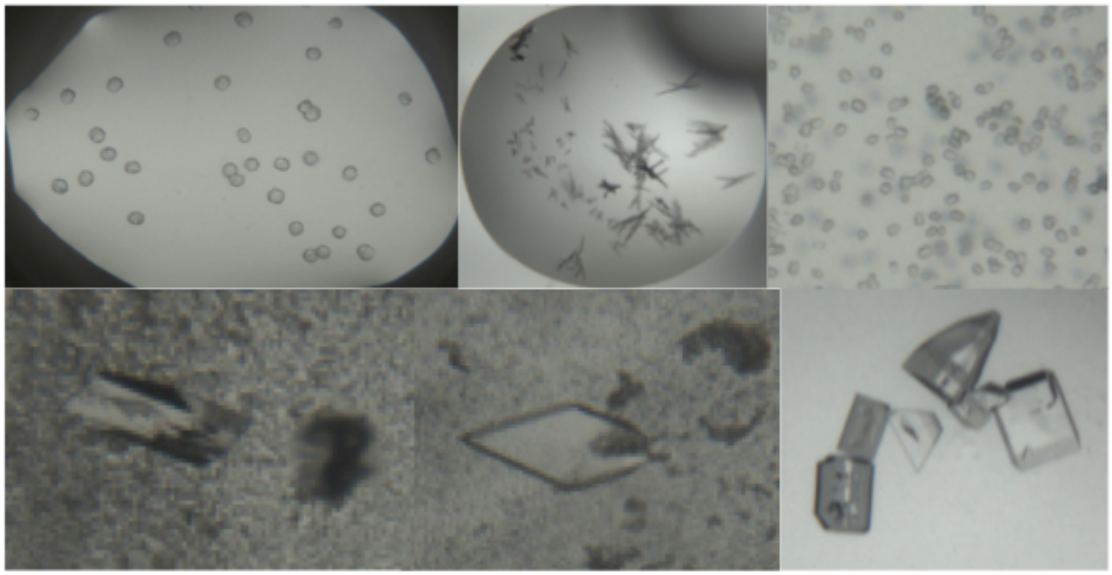


Figure 2.9 Co-crystals of MDM2 Clockwise from top left: MDM2 and NCL00018934 (14% PEG8k, 100mM imidazole pH 6.5); MDM2 and NCL18934 (0.1M Mes pH 6.5, 15% w/v PEG6k, 5% w/v MPD); MDM2 and

NCL00018046 (0.1M $(NH_4)_2SO_4$, 30% w/v PEG 5k MME); MDM2 and NCL00018710 (0.2M $(NH_4)_2SO_4$, 0.1M Na acetate, 20% w/v PEG 4k); MDM2 and NCL00018710 (0.2M $(NH_4)_2SO_4$, 0.1M Na acetate, 24% w/v PEG 4k); MDM2 and NU8406A (0.2M $(NH_4)_2SO_4$, 0.1M Na acetate, 20% w/v PEG 4k)

2.3.1.2 Follow-up screening

Following the identification of crystal conditions from broad screens, further work was undertaken to optimise these crystals and narrow screens were set up by hand in 24-well hanging and sitting-drop trays, using screens assembled by hand. The following optimisation screens were used:

- 6-28% PEG 8000; 100 mM imidazole pH 6.5-8
- 0-0.46M $(NH_4)_2SO_4$; 0.1M sodium acetate; 25% PEG 4000
- 0-0.46M $(NH_4)_2SO_4$; 25% PEG 4000
- 0.2M $(NH_4)_2SO_4$; 0.1M sodium acetate; 10-33% PEG 4000
- 0.2M $(NH_4)_2SO_4$; 10-33% PEG 8000
- 0.16-0.39M $(NH_4)_2SO_4$; 31% PEG 8000
- 0.2-0.35M $(NH_4)_2SO_4$; 28-33% PEG 8000
- 0-0.23M $(NH_4)_2SO_4$; 31% PEG 5000 MME
- 0.05-0.175M $(NH_4)_2SO_4$; 31% PEG 5000 MME
- 0.05-0.2M $(NH_4)_2SO_4$; 26-31% PEG 5000 MME

2.3.2 Data collection

Crystals were cryo-protected in 30% ethylene glycol (in 70% well solution) and flash frozen in liquid nitrogen. Data collection was carried out at Diamond Light Source (at the microfocus beamline I04). A data collection strategy was determined using Mosflm. The samples were loaded using a robotic sample changer.

2.3.3 Structure determination

Structure determination was carried out as detailed in Appendix B. All data were indexed and integrated using the Mosflm package (Leslie, 2006) and further processed using programs in the CCP4 suite (Win *et al,* 2011). The structures were solved by molecular replacement using the published MDM2-p53 complex structure (PDB ID: 1YCR, Kussie *et al,* 1996) with the peptide coordinates removed as a search model in Phaser (McCoy *et al,* 2007). Refinement was carried out using REFMAC5 (Murshudov *et al,* 1997) and manual rebuilding as well as identification of waters was performed using Coot (Emsley, 2004). The statistics for crystallographic data collection and refinement are shown in Table 2.3.

	MDM2 + NCL00018934	**MDM2 + NCL00018710**	**MDM2 + NU8406A**
Data collection parameters:			
Wavelength (Å)	0.96	0.96	0.96
Space group	C121	P212121	P3121
a	102.74	49.11	96.47
b	43.9	59.26	96.47
c	67.59	85.04	73.01
Resolution (Å)	50.59 – 1.55	48.62 – 2.24	83.54 – 2.20
Total reflections	31352	11420	18773
Unique reflections	8156	6423	6991
Completeness (%)	95.89	95.63	97.62
Refinement parameters:			
Resolution (Å)	50-1.55	45 – 2.24	75 – 2.20
R_{work}	20.4	19.9	21.2
R_{free}	22.1	24.7	22.9
Number of atoms:			
Protein	1414	1532	1578
Peptide/Ligand	73	70	34
Solvent	142	135	29
Average B-factor	30.1	11.3	32.7
Root mean square deviation			
Bond lengths (Å)	0.004	0.009	0.007
Bond angles (°)	1.765	1.976	1.511

Table 2.3 Data collection and refinement statistics for structures of isoindolinones bound to MDM2

The structures were assessed using MolProbity (Chen *et al,* 2010) This program assigns a quality value to the structure, by comparing it against other structures deposited in the PDB. It compares the statistics from data refinement and assesses amongst other information the number of clashes, how ideal are the bond lengths and angles and the likelihood that residues will take up the conformations assigned. The refined structure was accepted when the program gave a value for the structure that was in the top 30% of the structures deposited, or when further refinement and rebuilding made no further difference to the quality.

2.3.4 Structure of MDM2 bound to Nutlin 3a

The structure of the MDM2-Nutlin-3a complex, determined by Burcu Anil Kirmizitas is illustrated for comparative purposes in Figure 2.10.

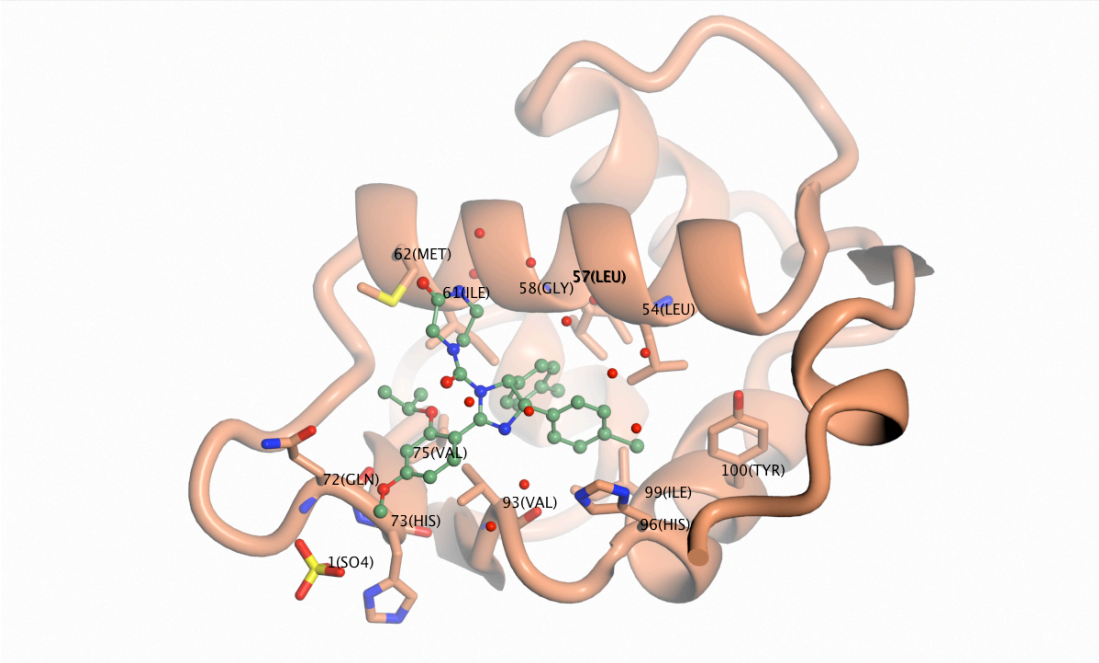


Figure 2.10 Structure of Nutlin 3a bound to MDM2 The fold of MDM2 is unchanged by the presence of nutlins. The Nutlin occupies the p53 binding pocket and inserts its side groups into the tryptophan, leucine and phenylalanine pockets, contacting polar and non-polar residues. MDM2 is shown in coral ribbon form, with Nutlin 3a in green ball-and-stick representation. Figure prepared in CCP4MG.

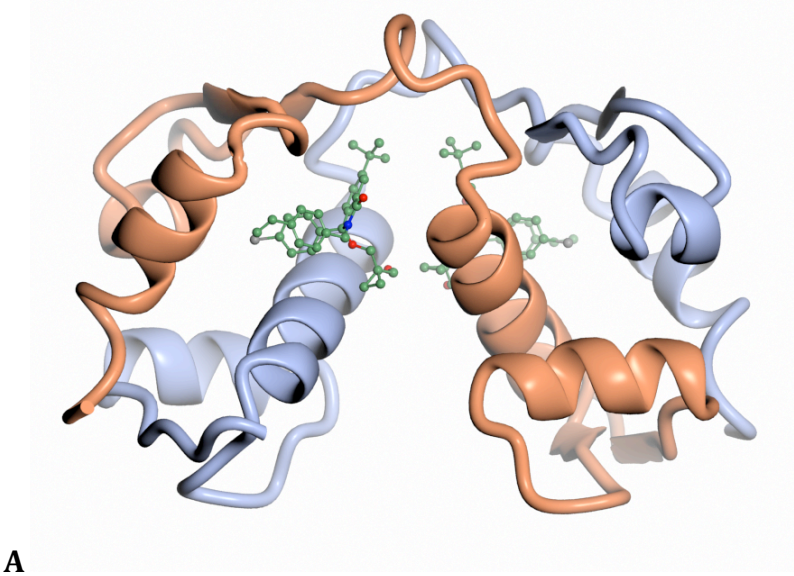
Mutation of E69 and K70 to alanine does not substantially alter the binding mode of the p53 peptide, although L26 does not insert as deeply into its pocket as in the structure with wild-type MDM2 (PBD ID: 1YCR: Kussie *et al,* 1996). As predicted, mutation of E69 and K70 to alanine alters the crystal contacts, with these residues now packing against Q18, I19 and P20 and the amide group of A69 hydrogen bonding with a water molecule (in the p53 peptide-bound structure). In the structure determined with our construct, the p53 ligand is only peripherally involved in making crystal contacts (mediated by a salt bridge between the side chains of K24 and D84 from the symmetry mate). By comparison, in the Kussie *et al* structure, the peptides from each MDM2 copy pack directly against one another. These observations using the 17-109

E69K70A construct suggested that it would be extremely useful for co-crystallisation with our isoindolinones.

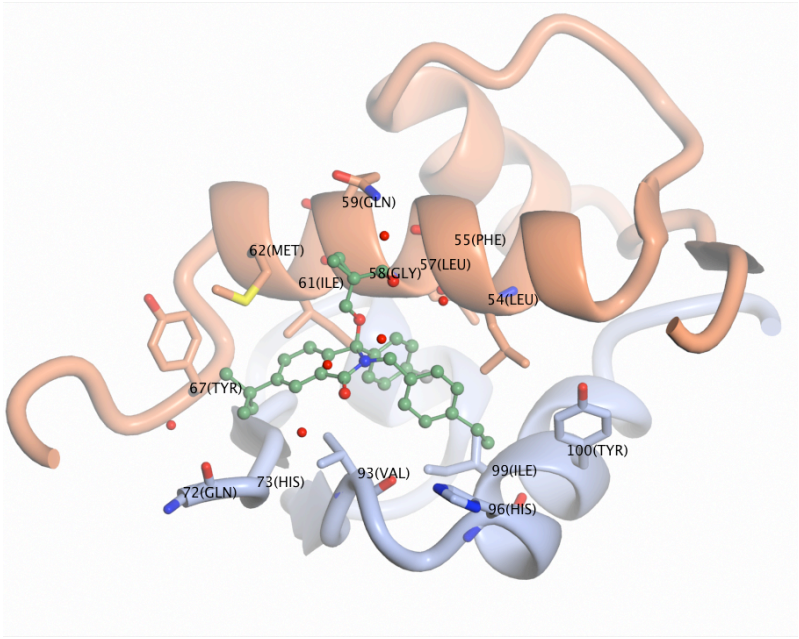
In the Nutlin-3a bound structure, the ligand is also less involved in crystal contacts than was seen in the Nutlin-2 bound MDM2 structure (PBD ID: 1RV1: Vassilev *et al,* 2004). The Nutlin contacts the first α-helix in the MDM2 symmetry mate and the carbonyl groups of A69 and A70 form hydrogen bonds with proximate waters. L33 and D84 from the symmetry-related MDM2 provide additional crystal contacts to A69 and A70. A key feature of these structures that differs between the p53-peptide-bound and Nutlin-3a-bound versions is the location of the loop formed by residues 17-25. In the p53-bound structure this points outwards into the solvent. However, in the Nutlin-3a-bound structure this segment is able to pack onto the protein in order to cover an exposed hydrophobic patch. This loop may mediate binding interactions and increase MDM2 stability.

2.3.5 Structure of MDM2 bound to NCL00018934

The structure of the NCL00018934-MDM2 complex (residues 17-109, E69K70A) is illustrated in Figure 2.11.



A



B

Figure 2.11 Structure of NCL00018934 bound to MDM2 (A) MDM2- NCL00018934 complex. The fold of MDM2 is unchanged by the presence of NCL00018934. The unit cell of this crystal form contains two copies of MDM2 and two copies of NCL00018934. The two MDM2 molecules formed a dimer with a strand exchange, whereby each MDM2 fold is composed of part of each copy of MDM2. **(B)** MDM2-NCL00018934 complex. NCL00018934 occupies the p53 binding pocket and inserts its side groups into the tryptophan, leucine and phenylalanine pockets, contacting polar and non-polar residues. The MDM2 copies are shown in ribbon form in light blue and coral, with NCL00018934 in green ball-and-stick representation. Figure prepared in CCP4MG.

In this structure two copies of MDM2 and two copies of NCL00018934 were in the unit cell. The two MDM2 molecules formed a dimer by strand exchange, whereby each MDM2 fold was made up of part of each copy of MDM2. NCL00018934 occupies the p53 binding pocket and inserts its side groups into the tryptophan, leucine and phenylalanine pockets as expected. The isoindolinones pack back-to-back with one another and also contact residues from each copy of MDM2, but are not directly involved in crystal contacts. The binding mode of NCL00018934 is illustrated in Figure 2.15.

2.3.6 Structure of MDM2 bound to NCL00018710

The structure of the NCL00018710-MDM2 complex (residues 17-125, K51A) is illustrated in Figure 2.12.

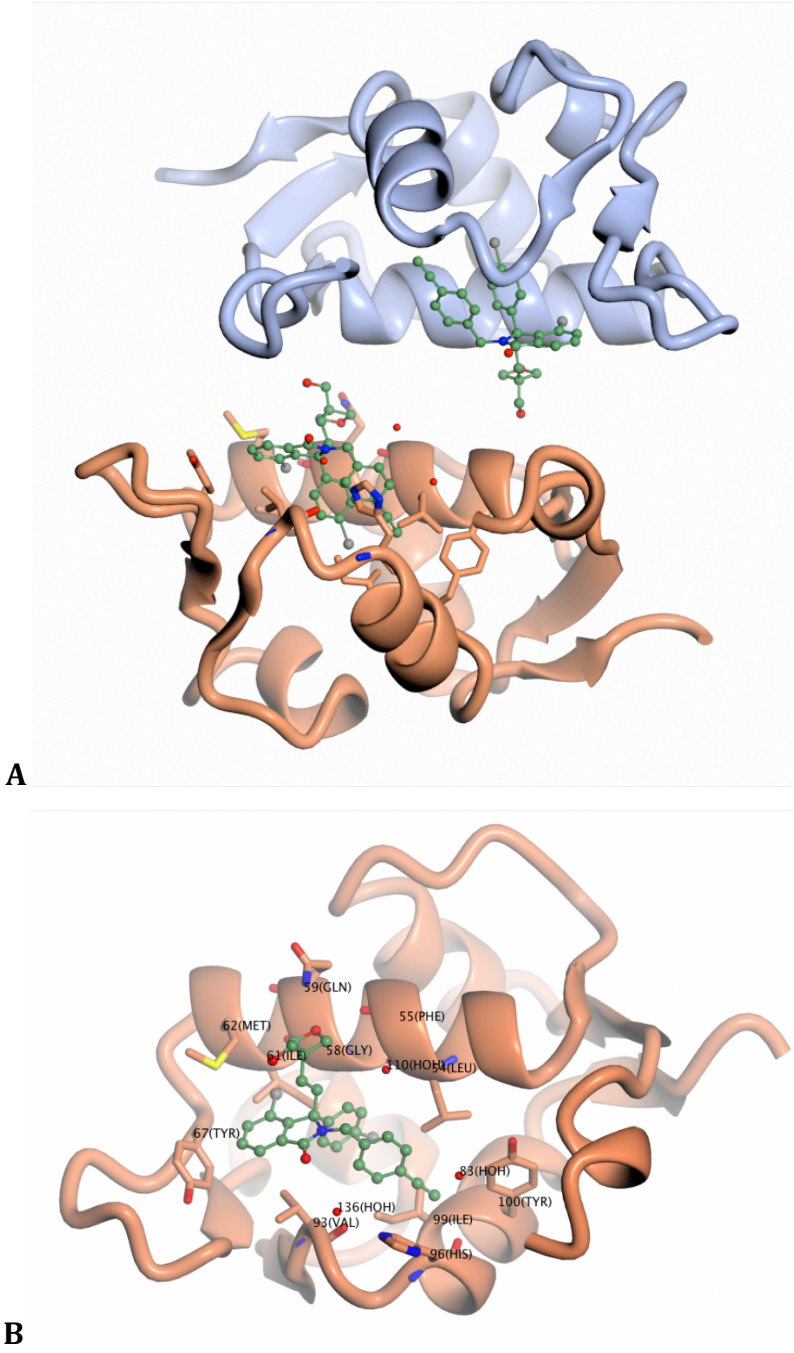


Figure 2.12 Structure of NCL00018710 bound to MDM2 (A) MDM2- NCL00018710 complex**.** Two copies of MDM2 and NCL00018710 were in the unit cell. **(B)** NCL00018710 occupies the p53 binding pocket and inserts its side groups into the tryptophan, leucine and phenylalanine pockets, contacting polar and non-polar residues within the binding pocket. The MDM2 copies are shown in light blue and coral ribbon form, with NCL00018710 in green ball-and-stick representation. Figure prepared in CCP4MG.

Here, addition of NCL00018710 does not change the overall fold (when compared to 1YCR). Again, in this structure two copies of MDM2 and two copies of NCL00018710 were present in the unit cell. In this case strand exchange did not occur. NCL00018710 occupies the p53 binding pocket and inserts its side groups into the tryptophan, leucine and phenylalanine pockets as expected. The isoindolinones pack back-to-back between the two MDM2 molecules. The binding mode of NCL00018710 is illustrated in Figure 2.15.

2.3.7 Structure of MDM2 bound to NU8406A

The structure of the NU8406A-MDM2 complex (residues 17-125, K51A) is illustrated in Figure 2.13.

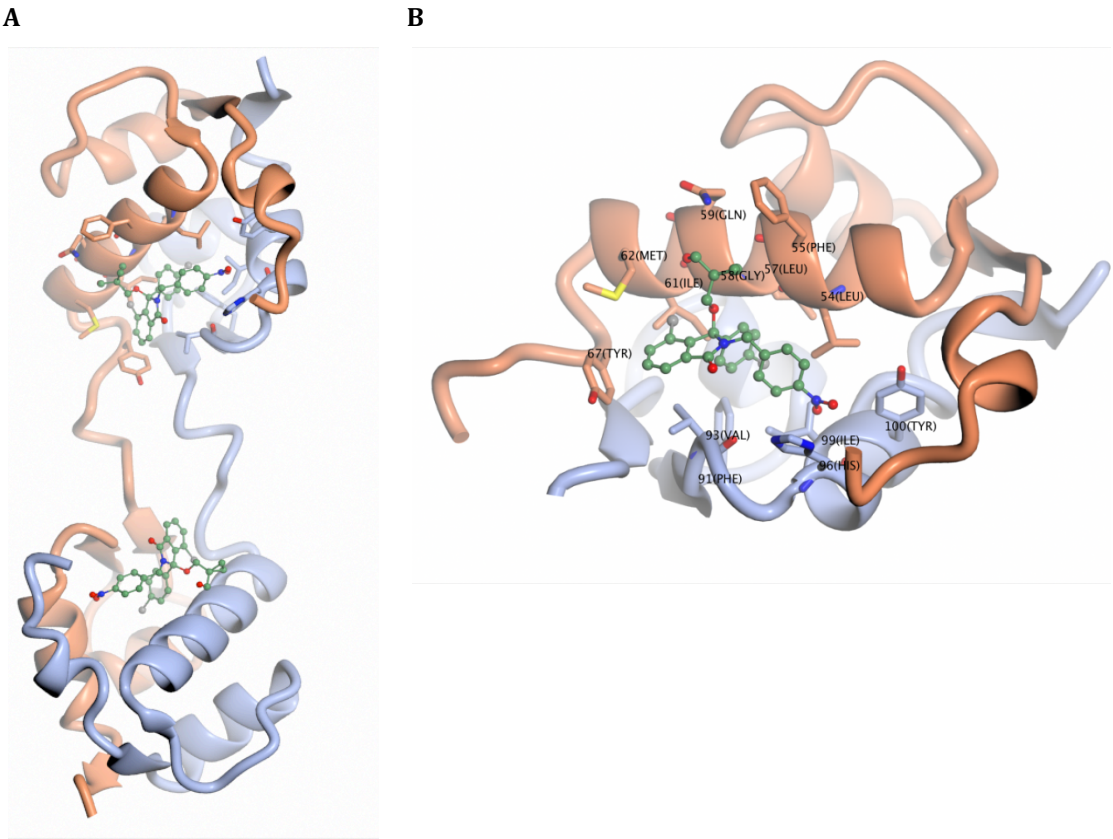


Figure 2.13 Structure of NU8406A bound to MDM2. (A) MDM2-NU8406A complex. Two copies of MDM2 and two copies of NU8406A were in the unit cell. The two MDM2 molecules formed a dimer with a strand exchange, whereby each MDM2 fold was made up of part of each copy of MDM2. **(B)** NU8406A occupies the p53 binding pocket and inserts its side groups into the Trp, Leu and Phe pockets, contacting polar and non-polar residues within the binding pocket. The MDM2 copies are shown in light blue and coral ribbon form, with NU8406A in green ball-and-stick representation. Figure prepared in CCP4MG.

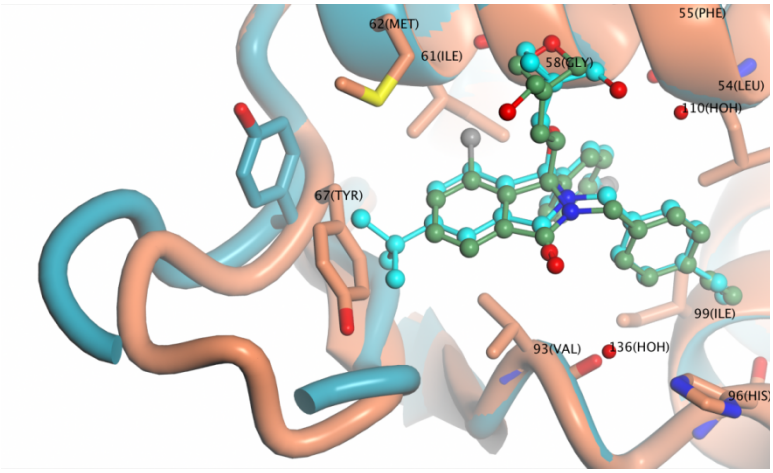
In this structure with NU8406A bound, two copies of MDM2 and two copies of NU8406A were present in the unit cell, and they form a dimer by strand exchange, although with a different organisation than in the structure with NCL00018934 bound. NU8406A occupies the p53-binding pocket and inserts its side groups into the tryptophan, leucine and phenylalanine pockets as expected. The binding mode of NU8406A is shown in Figure 2.15.

2.3.8 Review of isoindolinone binding modes

2.3.8.1 Comparison with the binding modes observed in the nutlin series

Since the nutlins now form a potent and specific class of MDM2 inhibitors, our aim is to mimic the interactions formed by these compounds with our isoindolinone compound series and to exploit the same structural features within the binding site in order to achieve gains in potency. In particular we would like to mimic the binding mode of Nutlin 3a, the most potent of the inhibitors. Our structures show that the compound series is progressing more and more closely to a fit with the binding site, reflected in the gains in potency and specificity exhibited with the latest compounds. In particular, the addition of a *t*-Butyl side group at position 7 on the A ring (in compound NCL00018934) exploits the full depth of the F19 pocket. As observed in the Nutlin 3a-bound structure, the position of Y67 is flipped by several angstroms through a combination of rotation and local translation of the loop. However, Nutlin 2 and our other inhibitors (NU8406A and NCL00018710) do not have sufficient bulk at this location to provide enough steric interference to cause this 'flipped out' tyrosine conformation, which appears to make a significant contribution to potency. This key interaction is illustrated in Figure 2.14.

A



B

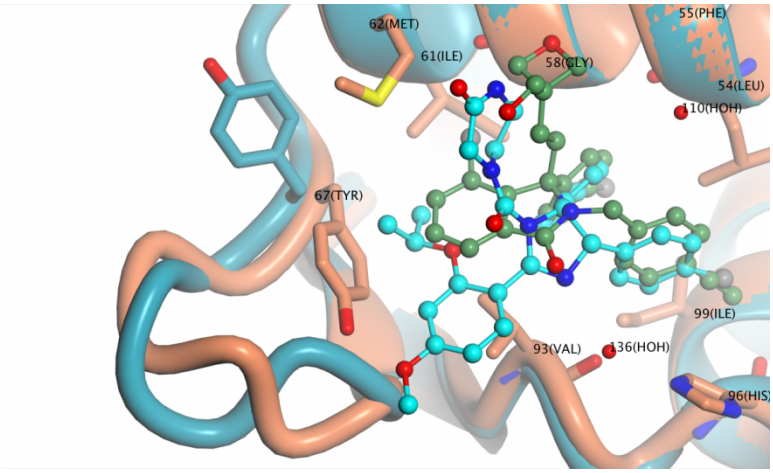


Figure 2.14 Effect of addition of *t*Butyl group on the position of the side chain of Y67 (A) MDM2-NCL18710 co-crystal structure (protein: coral ribbons, inhibitor: green) overlaid with the MDM2-NCL00018934 co-crystal structure (protein: blue ribbons, inhibitor: cyan). The addition of the *t*Butyl group on the inhibitor precludes the Y67 side chain adopting its authentic position, so the side chain 'flips' out. **(B)** MDM2-NCL18710 co-crystal structure (protein: coral ribbons, inhibitor: green) overlaid with MDM2-Nutlin 3a co-crystal structure (protein: blue ribbons, inhibitor: cyan). The 'flipped out' Y67 side chain is in almost the same place in the isoindolinone inhibitor (NCL00018934) and Nutlin 3a bound structures.

The overall binding mode, showing which chemical group on the isoindolinone inserts into each of the phenylalanine, tryptophan and leucine pockets is illustrated in Figure 2.15. This Figure highlights which groups may be altered and allows us to suggest further developments to the series.

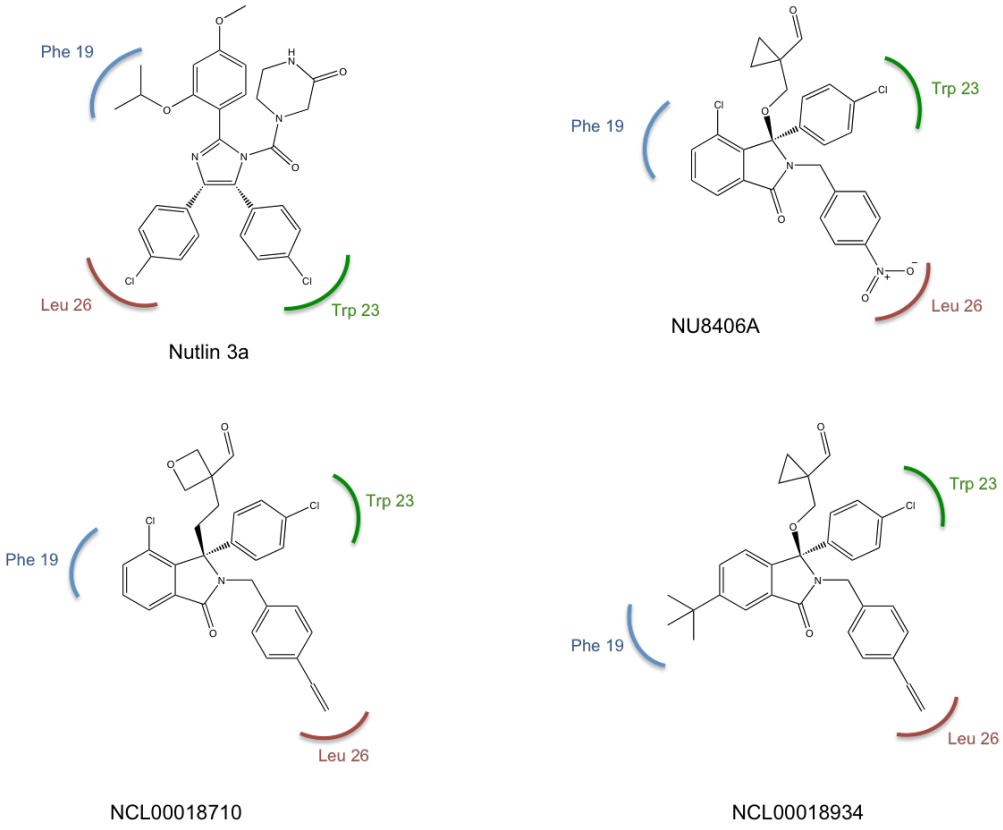


Figure 2.15 Summary of isoindolinone binding modes when compared to Nutlin 3a The position of the groups within the F19, W23 and L26 binding pockets is indicated.

In each case a chlorophenyl ring nicely fits the tryptophan pocket and this has therefore not been altered within our compound series. However the substituents filling the L26 and F19 pockets may be altered. The F19 pocket is best fitted by reasonably large and hydrophobic groups or by the A ring itself. However, a wide range of polar or non-polar groups can fill the L26 pocket, as long as the group is positioned in the correct location. This relaxed structural requirement enables us to explore different options on the B ring. The D substituent may also change conservatively as long as it retains some polar character, since it extends above the binding site and does not play a direct role in the interaction.

2.4 MDMX expression and purification

2.4.1 Cloning

MDMX cloning and mutagenesis were carried out by Jane Endicott. Five constructs were made (Table 2.4), based on the published structure of a MDMX-p53 peptide complex (Popowicz *et al,* 2008).

MDMX residue range	Vector	Add N-terminal residue
18-111	pGEX6P-1	GPLGS
22-111	pGEX6P-1	GPLGS
26-111	pGEX6P-1	GPLGS
18-111	pET46Ek	M
19-111	pET46Ek	-

Table 2.4 MDMX mutants produced

2.4.2 Expression

Expression was tested using the five MDMX constructs. Three (covering residues 18-111, 22-111 and 26-111) were expressed as GST fusion proteins in pGEX6P-1 plasmids and two (covering residues 18-111 and 19-111) were expressed with an N-terminal 6-His tag in pET46A plasmids. All proteins were expressed following induction with 0.1 mM IPTG in Rosetta (DE3)/pLYS-S cells overnight at 19^{o}C.

Since the expression of the His-tagged constructs was so poor, I have so far concentrated on the GST-tagged constructs for crystallisation. However, in order to identify further possible crystallisation conditions, further work on the His-tagged constructs should be performed, perhaps repeating the expression with different media or a different strain of *E. coli*. The yield was quite low for all constructs so the expression should be scaled up in order to obtain enough protein to screen more conditions and attempt co-crystallisation with other NCL inhibitors.

2.4.3 Purification

Purification of MDMX was carried out from pellets frozen at -80°C. Cell pellets expressing MDMX as a GST fusion were resuspended in HBS (50mM HEPES, pH 7.5, 250mM NaCl, 0.02% MTG, supplemented with Complete protease inhibitor cocktail, Roche), and after sonication and clarification (100,000 x g), the cell lysates were applied to a glutathione sepharose 4B gravity flow affinity chromatography column. GST-MDMX bound to the glutathione was eluted by addition of glutathione (10 mM in HBS) and then the GST tag was cleaved by 3C protease digestion. The mixture of MDMX and GST was subsequently buffer-exchanged and further purified by size-exclusion chromatography (S75 HR26/60 column, GE Healthcare). Fractions were pooled and the purity of the final product was checked using SDS-PAGE.

Cells expressing His-tagged MDMX were pelleted (5000 x g) then resuspended in HBS (50mM HEPES pH7.5, 250mM NaCl, supplemented with Complete protease inhibitor cocktail, Roche), sonicated and then following clarification (100,000 x g) the lysate was applied to a Ni-NTA affinity column pre-equilibrated in HBS. MDMX was eluted by addition of 250 mM imidazole. An additional step involved cleavage of the His tag by enterokinase (1:50 w/w) for 3-4 days at 4°C (in a buffer containing 50 mM NaCl, 20 mM Tris-HCl pH 7.5, 2 mM $CaCl_2$), before further purification by size-exclusion chromatography. This process appeared to be particularly inefficient. The purity of the final product was checked using SDS-PAGE. The progress of a typical MDMX prep is shown in Figure 2.16.

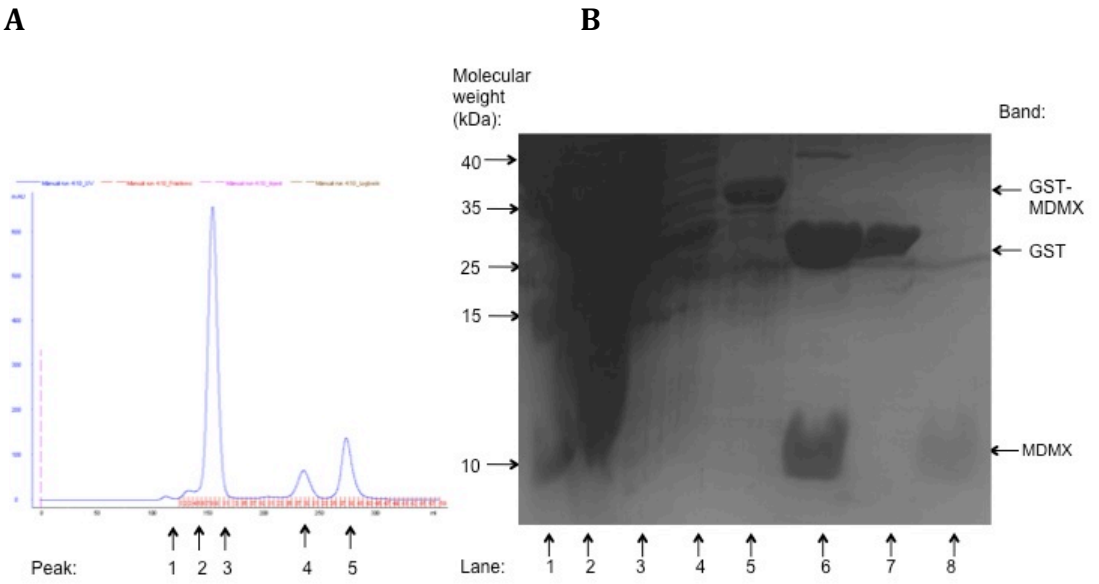


Figure 2.16 Purification of MDMX A: chromatogram from gel filtration. Peak 1, 2: Aggregates; Peak 3: GST; Peak 4: MDMX; Peak 5: salt peak. B: Photograph of SDS-PAGE gel. The positions of the GST and MDMX bands are marked, and the number of the lane is marked below. Lane: 1 = protein standards; 2 = lysate; 3 = flow-through (affinity column); 4 = wash (affinity column); 5 = elution (affinity column); 6 = elution (affinity column) cleaved by 3C protease; 7 = pooled fractions from peak 3 of gel filtration; 8 = pooled fractions from peak 4 of gel filtration.

The progress of expression and purification of these MDMX constructs is shown in Table 2.5.

Construct	Expression (2L)	Purification yields (mg/L)
1: pGEX6P-1 residues 18-111	✓	✓ (Yield: 2mg)
2: pGEX6P-1 residues 22-111	✓	✓ (Yield 5mg)
3: pGEX6P-1 residues 26-111	✓	✓ (Yield 2mg)
4: pET46A residues 18-111	×✓ (minimal)	× (no protein recovered)
5: pET46A residues 19-111	×✓ (minimal)	× (minimal protein recovered)

Table 2.5 Expression and purification of MDMX constructs

2.5 MDMX crystallisation attempts

2.5.1 Broad screening

The protein was not soluble at 15 mg/ml as published (Popowicz *et al,* 2008) so trays were set up at just under 10 mg/ml. It is unclear how this modification will affect the crystallisation process, and whether this was due to the effect of the inhibitor binding, compared to that of p53 peptide binding.

After purification was complete, a three-fold excess of inhibitor was added once the protein was at 3 mg/ml and the mix was incubated overnight, with the final concentration to 10 mg/ml the following day, before the protein was filtered and trays set up. Crystallisation trials were performed with those constructs for which purification was successful, using NCL-00017766 as the ligand. Yields were low and this limited the number of trays that could be set up. Commercially available broad screens (Wizard, PACT, PEG/ION, Morpheus and JCSG+) were utilised to identify possible crystal conditions. Crystal trays were set up in 96-well trays (SwissSci, 3 sub-wells) using a Tecan robot to pipette well solution and a Mosquito robot to pipette protein and mix well solution, using the sitting drop method. Trays were monitored over the course of several weeks to identify conditions with crystals or other interesting results.

2.5.2 Narrow screening

Narrow screens were also set up around the crystallisation conditions identified by Popowicz *et al* using the enterokinase-digested MDMX bound to a 15-residue p53 peptide (Popowicz *et al*, 2008; PBD ID: 3DAC). Crystal trials were set up by hand, in 24-well hanging and sitting drop trays. No crystals were obtained using the conditions previously identified. However, (i) these crystals were of a slightly different construct and (ii) this may also be due to the presence of our compound rather than the p53 peptide. Some narrow screens were set up using the same method, investigating conditions around those identified from commercial broad screens. The phase separation was reproducible, however no diffraction-quality crystals were obtained.

2.5.3 Identification of possible crystallisation conditions

It would be useful to initiate crystallisation trials with other inhibitors (perhaps the potent NCL-00018006 or NCL-00018007) as currently only NCL-00017766 has been used due to poor MDMX yields post purification. However, new crystallisation conditions were identified in which microcrystals grew (Figure 2.17). Some wells with phase separation were also identified, so screening of conditions around these may identify further crystal conditions.

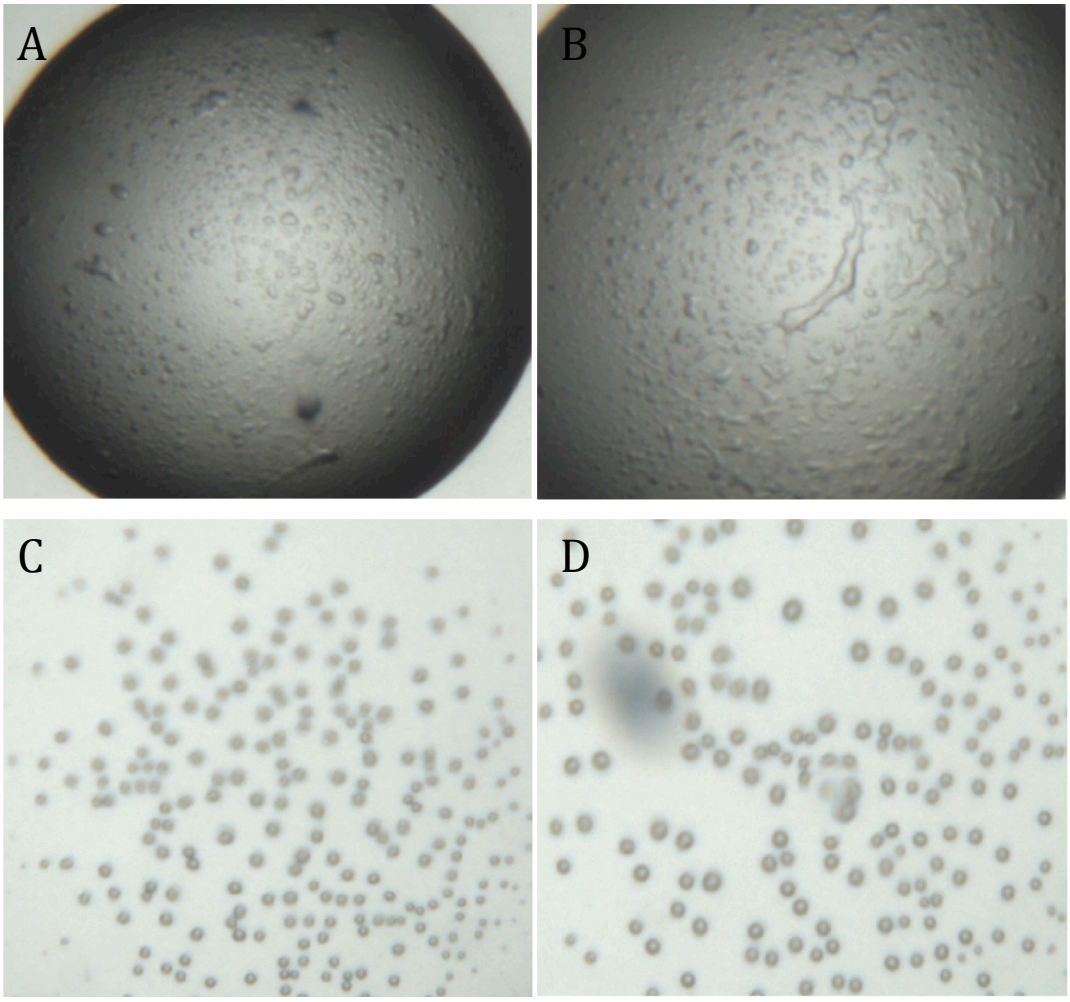


Figure 2.17 New possible crystal conditions identified for MDMX (A) Phase separation. Construct: residues 22-111. Conditions: (0.1M Mes pH 6.5, 15% w/v PEG6k, 5% w/v MPD) **(B)** Phase separation. Construct: residues 22-111. Conditions: (0.1M $MgCl_2$, 0.1M Na citrate pH 5, 15% w/v PEG 4k) **(C)** Microcrystals. Construct: residues 22-111. Conditions: (0.2M ammonium chloride, 20% w/v PEG 3.35k); **(D)** Microcrystals. Construct: residues 22-111. Conditions: (0.2M ammonium chloride, 20% w/v PE 3.35k).

Further screening of crystallisation conditions, possibly using other compounds within the NCL compound series, and optimisation and narrow screening around those identified here, should allow progress to a crystal structure of MDMX bound to NCL compounds. These structures would inform our understanding of the binding modes of NCL compounds, and would help to explain why some compounds show significant selectivity for MDM2 and retain specificity.

2.6 Conclusions and further experiments

The determination of these co-crystal structures of isoindolinones bound to MDM2 have been extremely important for the development of the inhibitor series. The information regarding the binding modes allows us to make more informed choices about which compounds to synthesise next and in some cases has directly led to an improvement within the series. For example, after obtaining the structure of NCL-00018934 bound to MDM2, we suggested the addition of a hydroxyl group on the t-Butyl. Very recently this led to the development of NCL-00019583 (Figure 2.18).

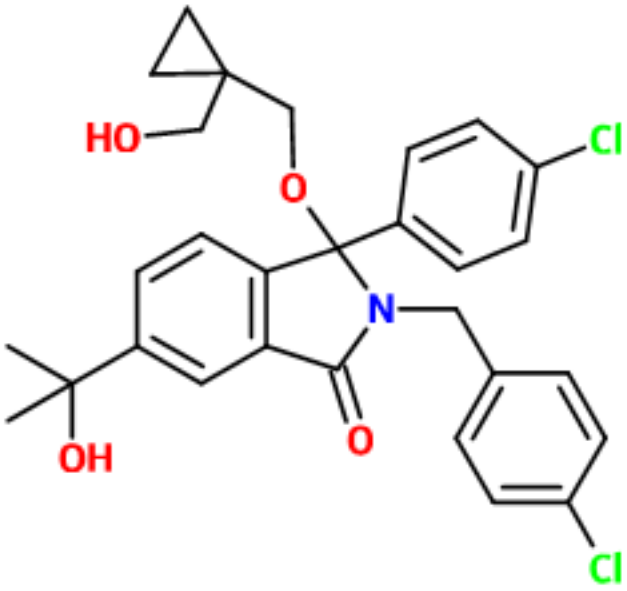


Figure 2.18 Structure of NCL-00019583

The rationale for the design of compound NCL-00019583 was the following: in the NCL-00018934 co-crystal structure, we observed favourable interactions

between a methyl hydrogen from the t-butyl group, and the pi system of a nearby tyrosine. However these methyls were also contacting carbonyl groups from a nearby glutamine. We therefore felt that keeping some methyl groups, but substituting one of them for a hydrogen bond donor would improve potency. This strategy was very successful since the new compound has an IC_{50} of 11 nM against MDM2 before chiral resolution, suggesting that the more active enantiomer will have an IC_{50} of ~5nM, and it retains specificity.

The determination of further co-crystal structures of MDM2 with other compounds in the series will provide additional SAR information regarding the contribution of individual chemical groups to the interaction, and further inform development of this compound series. Further work by the team in Newcastle University has shown that the isoindolinones that specifically and potently inhibit MDM2 are able to enter cells. Crystallisation of different constructs of MDM2 with the most potent and soluble isoindolinones is now possible and the structures determined so far with compounds NCL00018934, NCL00018710 and NU8406A show the variety of postures the ligands can make within the binding site. Key structural features within the pockets that can be exploited by drug molecules have been identified, and we now have a greater understanding of the observed flexibility within the p53-binding pocket. Structures of some of these compounds bound to MDMX are very desirable since this would help us to understand the key differences between binding of the compounds to the MDM2 vs. MDMX binding sites, and therefore help us understand how drug selectivity between these two targets may be achieved.

3 The Plk1 Polobox: the search for a co-crystal structure

3.1 Background and introduction

3.1.1 Plk1: regulating mitosis

3.1.1.1 Plk1: sequence, structure and function

Plk1 is a major regulator of mitotic progression and participates in the G2/M transition by phosphorylating multiple cellular targets. Overexpression of Plk1 in HeLa cells causes mitotic delay (Mundt *et al,* 1997; Jang *et al,* 2002); on the other hand inhibition causes defects in centrosome maturation, spindle formation and separation of sister chromatids (Sumara *et al*, 2004). For this reason, Plk activity must be tightly controlled within the cell cycle.

The Plk family was first identified from a screen of *Drosophila* maternal effect lethal mutations in which mutants with abnormal spindle poles were called polo mutants (Sunkel and Glover, 1998). Homologous proteins were later found in mouse (Clay *et al*, 1999), *S. cerevisiae* (named Cdc5: Kitada *et al*, 1993) and *S. pombe* (called Plo1: Okhura *et al*, 1995). Higher eukaryotes have multiple Plk family members (4 in mammals), of which Plks 1 and 4 are essential, but Plks 2, 3 and 5 do not appear to be essential since knockout mice are viable. Plk1 is probably the most extensively studied.

Members of the Plk family are composed of an N-terminal Ser/Thr protein kinase domain (KD) and a C-terminal tandem repeat of an inhibitory substrate binding domain called the Polobox domain (PBD), which is unique to this protein family. The domain architecture of Plk1 is illustrated in Figure 3.1.

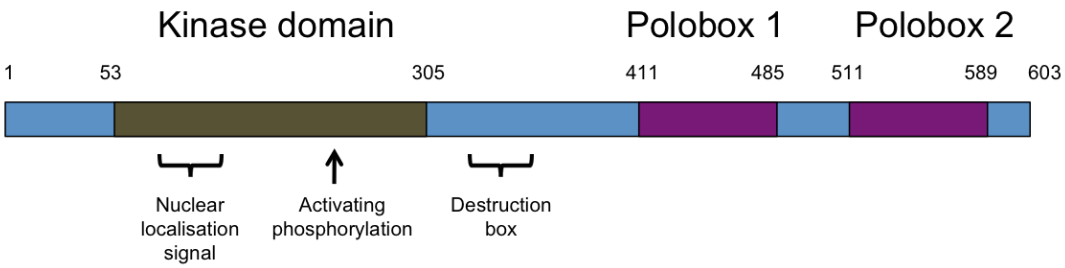


Figure 3.1 Primary structure of PLK1 The nuclear localisation signal is located from residues 134-148, and consists of the motif RRRSLLELHKRRK; activating phosphorylation occurs on T210; the destruction box is located between residues 337 to 340 and consists of the motif RKPL.

The kinase domain of Plks is highly conserved, with at least 40% sequence identity between family members. A conserved glycine-rich loop (sequence GK|EGG|SFA) plays a key role in coordinating phosphorylation. ATP binds to the kinase domain and is coordinated by residues K82, E101 and D194. The kinase requires phosphorylation of T210 and S137 for full activation. The kinase domain is linked to the PBD via a flexible linker sequence. The PBD consists of two Polobox motifs, which despite having only 12% sequence identity, have an identical fold, composed of a six-stranded beta-sheet and an alpha-helix (see Figure 3.4 on page 107).

The PBD has several functions: firstly it acts as an inhibitor of Plk kinase function by binding unphosphorylated Plk over the top of the kinase domain and blocking access to the active site (Jang *et al*, 2002). T210 phosphorylation releases the Polobox block on kinase activity, and the mechanism allows phosphorylation of Plk to be linked to Plk activity. In addition, the PBD acts as a regulator of Plk subcellular localisation (Lee *et al*, 1999; Jang *et al*, 2002) directing Plk activity to where it is required. Finally it binds and recognises a phosphopeptide with the recognition motif SpS|pTP|X. This motif is found in many substrates of Plk and so acts to recruit them to the kinase domain for subsequent phosphorylation.

A structure of the entire Plk1 has not been determined, presumably because the flexible linker between the domains is too mobile and inhibits crystallisation. However, the structures of the kinase and Polobox domains have been

determined separately. The structure of the kinase domain exhibits several interesting points (Kothe *et al*, 2007). The construct used for crystallisation encodes residues 37-330, which encompasses the entire kinase domain, and the mutation T210V, which aided crystallisation and was compatible with an active kinase conformation. The structure of the kinase domain was determined bound to AMPPNP, and to PHA-680626, a potent Aurora A, and less potent Plk1 inhibitor.

The kinase domain exhibits a typical protein kinase fold, with an N-terminal lobe of mostly antiparallel β sheet, and a mostly α helical C-terminal lobe, connected by a hinge region. The ATP binding site is located in a cleft between the two lobes, making contact with residues in the hinge region and loops. The activation loop is unphosphorylated but unexpectedly was found to adopt the extended active confirmation, possibly due to its interaction with a Zn^{2+} ion. The AMPPNP ligand adopted the expected conformation but there appears to be significant flexibility in the phosphate groups. PHA-680626 mimicked the ATP pose, including key H-bonds to E131 and C133. Several unique structural features of the binding site compared to related kinases may help inform drug design at this site. The structure of the PBD is discussed further in Section 3.1.2.2.

3.1.1.2 Regulation, binding partners and substrates

Plk activity varies greatly during the cell cycle. Its activity is the result of several contributing mechanisms, a balance of expression and degradation, Polobox-mediated inhibition, phosphorylation state and changes in localisation.

Plk expression levels have been monitored using a luciferase-Plk fusion construct, and using this system Plk1 appeared to increase in G2, be highest through M phase, and then decrease in G1 (Uchimi *et al*, 1997). Phosphorylation of Plk1 on Ser137 and Thr210 activates Plk1 and reduces PBD inhibition of Plk activity. Plk then phosphorylates its target substrates and promotes progression through mitosis. The flip side of Plk expression is the rate of degradation, which is also highly regulated. Anaphase promoting complex/cyclosome (APC/C)-mediated Plk1 degradation at the end of mitosis is promoted by the destruction box following Plk1 ubiquitination (Lindon and Jonathon, 2004). Chaperones such as Hsp90 interact with the PBD to stabilise Plk1 and inhibit proteosomal degradation (Blank *et al*, 2003). The sub-cellular localisation of Plk1 changes during the cell cycle. During prophase it is located at the centrosomes, it then migrates in metaphase to the spindle poles, in anaphase to the spindle midzone and in telophase to the post-mitotic bridges (Donaldson *et al*, 2001).

The main functions of Plk1 are illustrated in Figure 3.2.

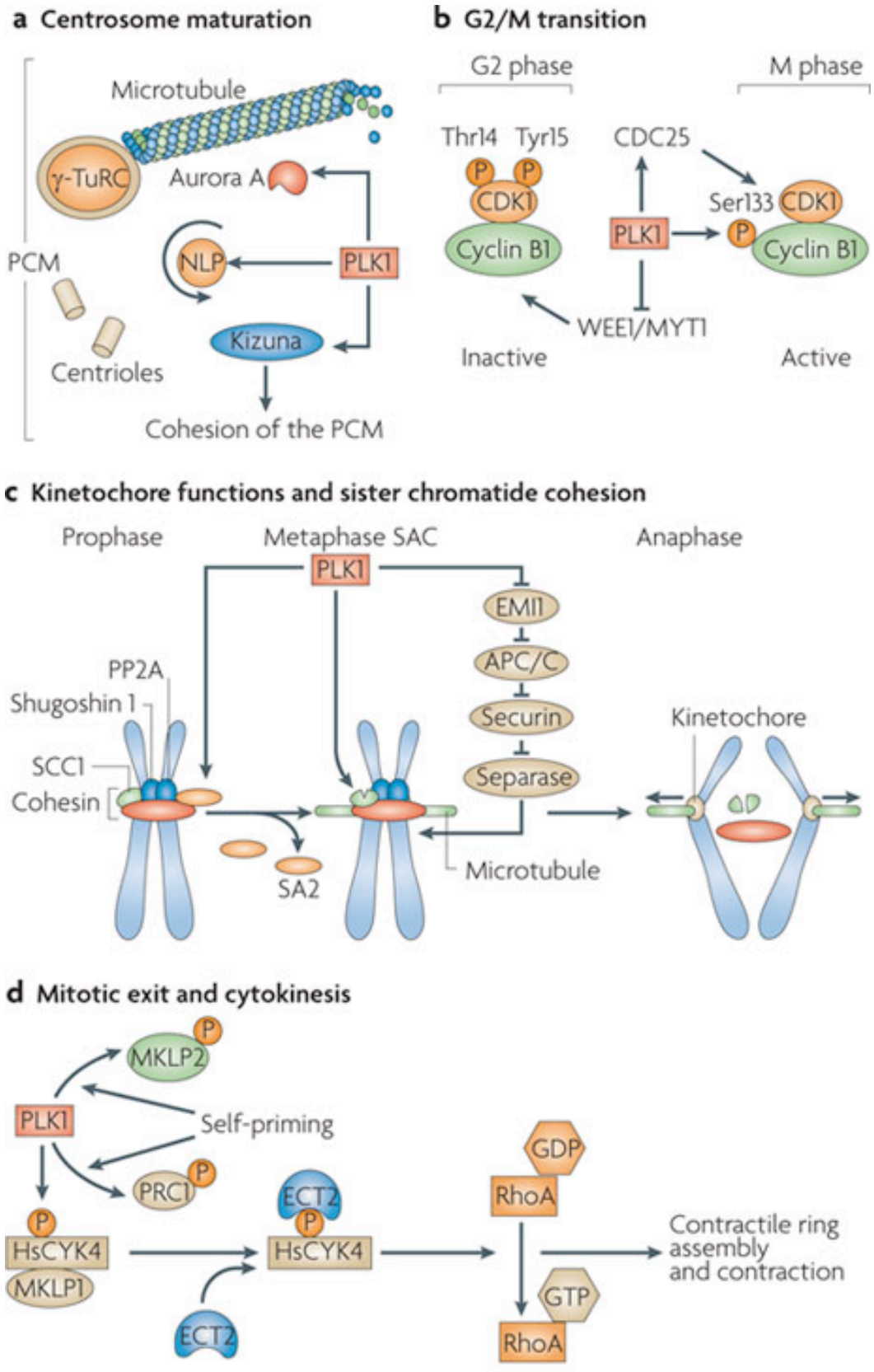


Figure 3.2 Cellular functions of Plk1 Plk1 is involved in key processes often disrupted in cancer, including centrosome maturation, progress of the G2/M transition, sister chromatid cohesion and mitotic exit and cytokinesis. Reproduced from Strebhardt, 2010. *Figure legend continues on page 100.*

(a) Centrosome maturation The role of Plk1 is to phosphorylate ninein-like protein (NLP), which then in turn recruits the γ-tubulin ring complex (γ-TuRC), which itself promotes the nucleation of new microtubules. Plk1 also controls the activity of Aurora A , which is a key player in the maturation of centrosomes, and Kizuna, a components which organisespericentriolar material.
(b) G2/M transition Plk1 promotes activation of CDK1 (to give the cell-cycle complex CDK1/cyclin B1, necessary for entry into M phase) by phosphorylating the positive regulator CDC25, and inhibiting the CDK1 inhibitors Wee1 and Myt1.
(c) Kinetochore functions and cohesion of sister chromatids Plk1 phosphorylates and activates the APC/C, which inhibits securin, which in turn inhibits separase. Separase cleaves the SCC1 protein and allows removal of the cohesin ring, allowing sister chromatids to be pulled apart.
(d) Mitotic exit and cytokinesis Here, Plk1 is localised to the spindle midzone by docking with PRC1 and MKLP2. It then phosphorylates HsCYK4 to allow ECT2 docking and indirectly results in formation of the actmyosin ring and the cell cortex, which will contract and form a cleavage furrow.

Plk1 is quite a promiscuous kinase, with wide substrate specificity. Many substrates within the Plk1 interactome have been identified from a proteomic screen (Lowery *et al*, 2007). Many of the cellular effects of Plk1 are due to its control of CDK1/cyclin B activity. Among its most important substrates is Cdc25, an activator of CDK1/cyclin B. Plk1 phosphorylates S198 of Cdc25 resulting in Cdc25 translocation to the nucleus (Toyoshima-Morimoto *et al*, 2002). Cyclin B is also phosphorylated by Plk1 on S133, resulting in CDK1 activation at centrosomes (Jackman *et al*, 2003); and on S147, causing nuclear translocation (Toyoshima-Morimoto *et al*, 2002). Another substrate is Wee1, a key regulator of mitotic entry. Phosphorylation of Wee1 by Plk1 and also by CDK1/cyclin B leads to its degradation in early mitosis and release of inhibition of CDK1/cyclin B (Watanabe *et al*, 2005). Plk1-mediated phosphorylation of Myt1, a CDK1-inhibitory kinase that like Wee1 phosphorylates CDK1 within the glycine-rich motif, leads to its inhibition and therefore the activation of CDK1/cyclin B.

Inhibitors of the APC/C such as the early mitotic inhibitor 1 (Emi1) are also phosphorylated by Plk. This modification results in their degradation and therefore the subsequent triggering of APC/C activity, which in turn leads to destruction of mitotic cyclins and other targets, including Plk1 itself, causing mitotic exit (Eckerdt and Strebhardt, 2006). Other characterised Plk substrates include Aurora A and Kizuna, that are involved in centrosome organisation and promote development of a mature centrosome.

Two models to explain how Plk1 is recruited to the correct subcellular location, to regulate its activity throughout the cell cycle, were proposed by Lowery *et al.*, (Lowery *et al,* 2005), and illustrated in Figure 3.3.

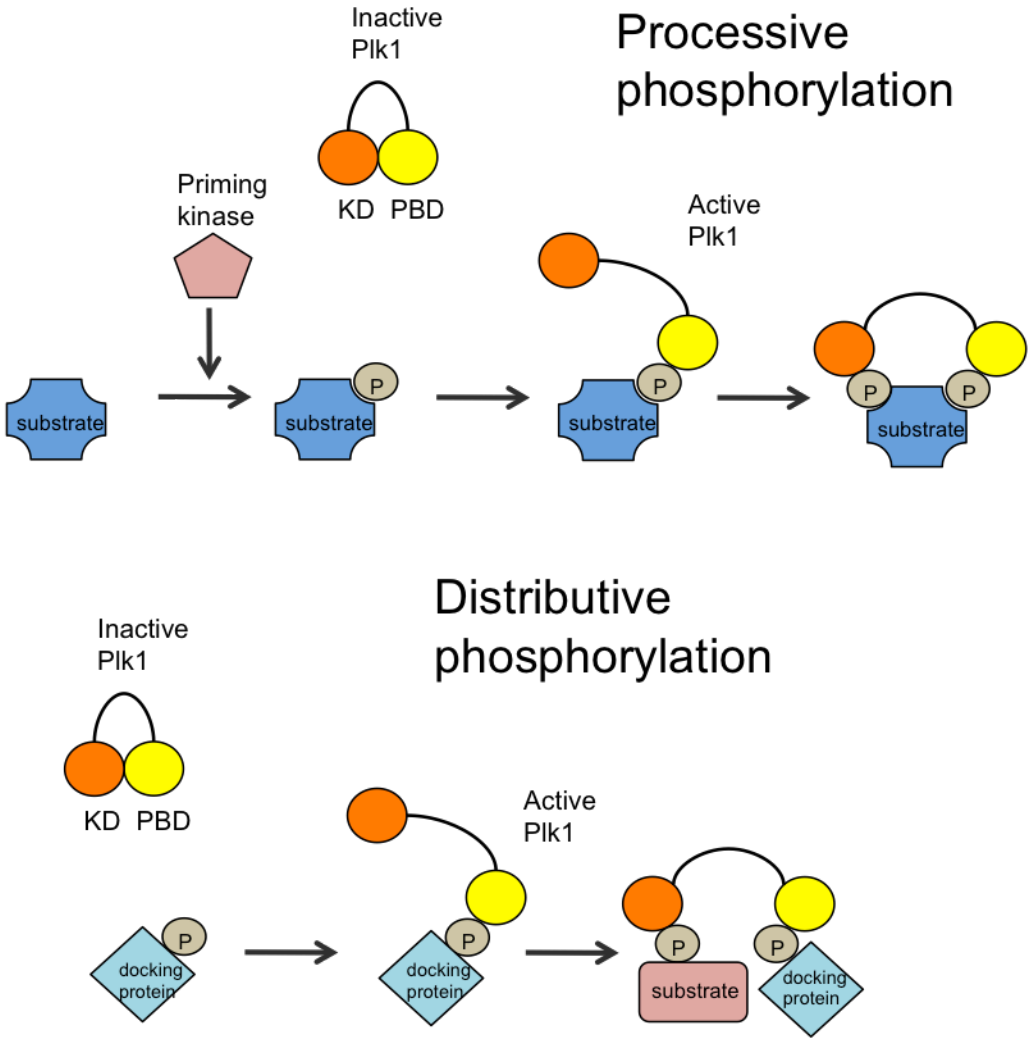


Figure 3.3 Models for Plk1 action, proposed by Lowery *et al*, 2005. Plk1 is represented as an orange kinase domain and a yellow Polobox domain.

Processive phosphorylation involves a priming kinase and Plk1 acting in concert to phosphorylate a substrate. The priming kinase first generates the pThr/pSer epitope recognised by the Plk1 PBD. Binding of this epitope to the PBD activates Plk1 kinase activity and the kinase domain then phosphorylates the substrate at another site (or on multiple sites, for example at microtubules). Distributive phosphorylation involves the recruitment of Plk1 by PBD binding to a pThr/pSer

on a docking protein. This localises Plk1 to the correct subcellular location and the kinase domain then phosphorylates other substrates in the vicinity of the docking protein.

3.1.1.3 Plk1 and carcinogenesis

Plk1 is a proto-oncogene and is over-expressed in glioma (Dietzmann *et al,* 2001), endometrial carcinoma (Takai *et al,* 2001), ovarian cancer (Takai *et al,* 2001b) and colorectal cancer (Takahashi *et al,* 2003). High Plk levels are associated with a poor prognosis for treatment (Knecht *et al,* 2001) and may increase the likelihood of tumour metastasis, increase proliferation rates and impair tumour response to treatment (Wolf *et al,* 1997).

Down-regulation of Plk1 activity by antisense oligonucleotides causes growth inhibition of cancer cells and mitotic abnormalities *in vitro*, and significant anti-tumour activity in mouse xenografts (Spankuch-Schmitt *et al*, 2002; Elez *et al*, 2003). The effect appears to be due to induction of apoptosis, possibly mediated by defects in spindle formation (Spankuch-Schmitt *et al*, 2002b). In addition, the effect of Plk1 down-regulation appears to be specific to tumour cells, with no effect on normal cells. This may be due to deficient or abnormal p53 activity in tumour cells (Guan *et al*, 2005) and makes Plk1 a particularly attractive target for chemotherapy because it should limit therapy-related side effects.

Mutation of the Plk1 gene causes both mitotic spindle abnormalities and chromosome separation to fail after duplication, which is implicated in carcinogenesis as it can result in aneuploidy (Sunkel and Glover, 1988; Llamazares *et al,* 1991). Plk1 also phosphorylates microtubules directly, which may be the driver of the effects on mitotic spindle morphology (Feng *et al,* 1999). Plk1 appears to be dispensible for mitotic entry in some cell types, possibly due to compensation by Aurora kinase or other Plks. Plk1 deletion stops mitotic exit completely and is therefore a lethal mutation. Since Plk1 activity is necessary for early embryonic development, homozygous Plk1 knockout mice are not viable (Lu *et al,* 2008).

However, upregulation or elevated Plk1 activity is also implicated in carcinogenesis due to its involvement with desensitisation of cells to the DNA damage checkpoint (Smits *et al,* 2000; Syljuasen *et al,* 2006), regulation of transcription during mitosis (Fu *et al,* 2008), vimentin-mediated invasion into the extra-cellular matrix (ECM) (Rizki *et al,* 2007) and induction of aneuploidy. The latter effect results from its role in regulation of sister chromatid cohesion and centrosomal functionality (Meraldi *et al,* 2002). Due to the large number of Plk1 substrates within mitosis and the wide-ranging effects of its activity, Plk1 is a major target for cancer therapy.

3.1.2 The Polobox domain: substrate binding

3.1.2.1 Function of the PBD

The PBD is critical for Plk activity, performing several key regulatory functions (reviewed in Park *et al,* 2011). These include controlling Plk phosphorylation, Plk localisation, inhibiting Plk function and binding substrates to promote their phosphorylation by the Plk kinase domain. Its control over Plk localisation is essential for Plk function, as demonstrated by the fact that a single point mutation in PB1 of Plk1 (W414F) delocalises Plk1 from microtubules and kinetochores and significantly impairs its function, whilst leaving kinase activity intact (Lee *et al,* 1998). Although the exact mechanism by which the PBD controls the temporal and spatial localisation of Plk1 is still unknown, it is clear that this is mediated by PBD binding to specific substrates. Dominant negative mutations of the PBD of Plk1 cause spindle checkpoint arrest and failure of cytokinesis (Seong *et al,* 2002).

Localisation of Plk1 to the centrosomes and kinetochores in late interphase (until telophase) and to the spindle midzone in anaphase allows it to phosphorylate the correct substrates and is key to its activity (Hanisch *et al,* 2006). Mutations in the PBD block Plk1 complementing temperature-sensitive mutations of Cdc5 in yeast, its localisation to spindle poles and its ability to organise cytokinetic structures (as visualised by a GFP fusion) (Song *et al,* 2000). As no structure of full-length Plk1 exists, the details of PBD binding to the kinase domain and the intra-molecular contacts formed remain unknown.

The PBD also exerts an auto-inhibitory effect on Plk function by binding unphosphorylated Plk, possibly over the top of the kinase domain and blocking access to the active site. C-terminal deletions of Plk1 or mutation of T210 results in increased kinase activity, showing that the PBD has an inhibitory function (Mundt *et al,* 1997). The PBD cannot bind phosphorylated Plk (Jang *et al,* 2002). T210 phosphorylation by Aurora A (mediated by Bora) releases the Polobox

block on kinase activity (Macurek *et al,* 2008), and the mechanism allows phosphorylation of Plk to be linked to Plk activity.

The PBD recruits Plk to substrates by binding phosphoepitopes within them and and therefore targets the kinase activity to these substrates. The molecular basis of phosphoepitope binding has been characterised using proteomic screens with short pSer/pThr-containing peptides. These studies established a consensus binding motif composed of residues [P/F]-[Φ/P]-Φ-[T/Q/H/M]-S-[pT/pS]-[P/X] (where Φ is a hydrophobic residue and X is any residue). The optimal phosphopeptide was MQSpTPL (Elia *et al,* 2003; Elia *et al*, 2003b).

3.1.2.2 Structures of the PBD

The crystal structure of the PBD from Plk1 was determined by Cheng and colleagues (Cheng *et al,* 2003). The full PBD consisting of PB1 and PB2 from residues 367-603, as well as PBD (residues 345-603) in complex with a phosphopeptide, show that PB1 and PB2 have an identical fold, despite having only 12% sequence identity. Each Polobox motif consists of six β strands and one α helix. The two Poloboxes stack on each other to form the PBD, generating a cleft between them, which forms the phospho-peptide binding site (Figure 3.4). A small α helical 'cap' is believed to stabilise the fold. There is no conformational change in the PBD upon phospho-peptide binding. This result suggests that the PBD mainly plays a role in substrate recognition and recruitment rather than using a signal transduction mechanism to initiate phosphorylation by the kinase domain.

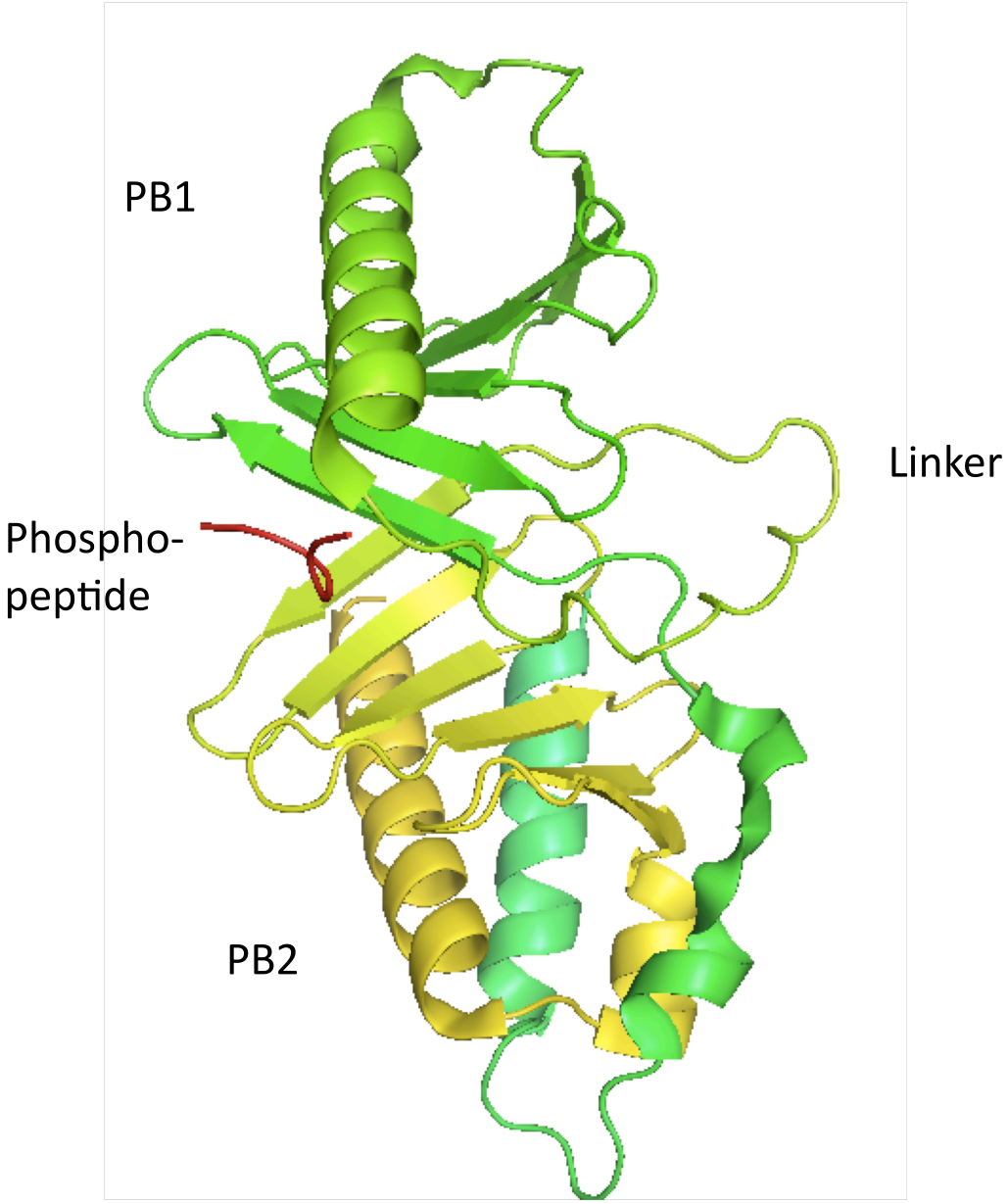


Figure 3.4 Structure of the Plk1 Polobox domain PBD ID 1Q4K Illustrated in ribbon form, Polobox domains 1 and 2 are shown as ribbons, with colouring in a spectrum from yellow-green along the protein chain to illustrate the geometry of the formation. The flexible linker and the phophopeptide MQSpTPL (red) are indicated. Figure made in Pymol.

Comparison of the structure of this MQSpTPL-bound PBD to the PLHSpT and PPHSpT-bound PBD structures determined by Yun *et al.* (Yun *et al,* 2010) illuminates further the means by which PBD substrate selectivity is conferred. The phospho-peptide forms contacts with key residues W414, L490, H538 and K540 (Figure 3.5), all of which were previously identified as functionally important through analysis of the activity of mutants.

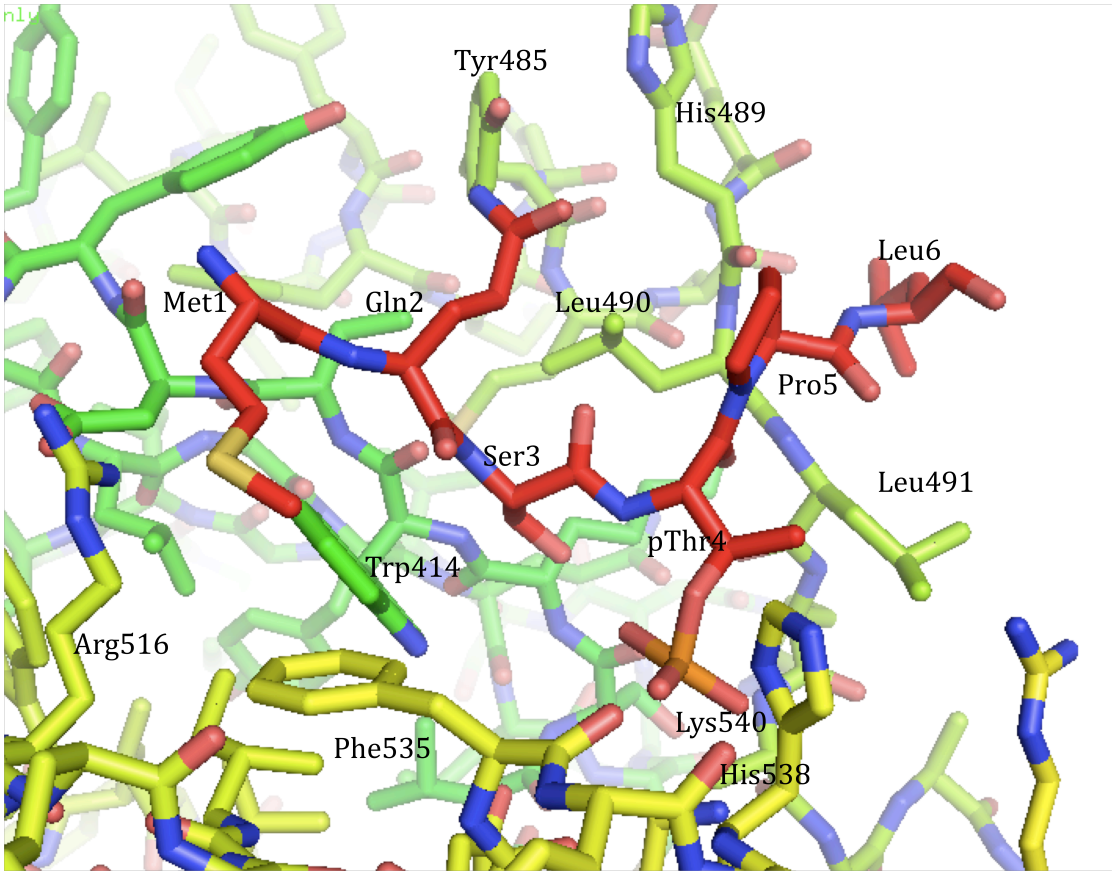


Figure 3.5 Interactions between the phospho-peptide MQSpTPL and residues of Plk1 PBD PBD ID 1Q4K Illustrated in stick form, Polobox domains 1 and 2 are shown with colouring in a spectrum from yellow-green along the protein chain and the phospho-peptide is in red (the same colouring as in Figure 3.4). Residues involved in making interactions with the phospho-peptide are labelled. Figure made in Pymol.

The residue in position P-1 (where position P-0 is the phosphoserine or phosphothreonine residue) is strongly selected as a serine, because this side chain points down towards the Plk surface and forms H-bonds with W414 and a water molecule which bridges with a main-chain carbonyl. The residue in position P+1 showed a modest selection as a proline, which appears to be because it introduces a kink in the peptide, which directs the amino group of the P+2 residue back towards the PBD to form H-bonds.

In the PLHSpT-PBD complex structure reported by Yun *et al.,* the N-terminal (P-4) proline H-bonds with R516 and the pyrridoline ring docks into a hydrophobic pocket formed by F414 and F535. However, the proline residue in the

comparable P-4 position in the PPHSpT-PBD complex could not form these contacts because the additional proline in the peptide sequence alters the backbone conformation. In MQSpTPL, the methionine at position P-3 mimics the N-terminal proline by docking its side-chain into the proline-binding pocket. However, it lacks the histidine at position P-2, which appears to confer Plk1 specificity, explaining the low specificity of the MQSpTPL peptide.

The structure of an unusual PBD from Sak, a Plk4 variant that contains only one Polobox, has also been solved (Leung *et al*, 2002). This structure has a complicated topology in which the Polobox from each of two molecules forms a strand-exchanged dimer to reconstitute the PBD. Although some of the Sak PBD can overlay with the Plk1 PBD structure, there are large differences within the structure due to the change in topology, and as a result the phospho-peptide binding site may be quite different. The structures of the Plk1 PBD can inform drug development, as it is now possible to search for drugs that bind the phospho-peptide binding site and displace substrate since the interactions made by the peptide with the PBD are now understood in detail. Using the information about interactions at each position within the peptide we can design peptidomimetic drugs, which replace these interactions to displace cellular PBD substrates.

3.1.3 Plk1 as a drug target

3.1.3.1 Effects of inhibiting Plk1 for cancer therapy

Plk1 is an important target for cancer therapy due to its wide-ranging activity in cell cycle regulation, including its control of microtubule dynamics, spindle formation and regulation of CDK1/cyclin B. Because Plk1 has so many cellular targets, inhibition affects multiple pathways within the cell. Inhibition of Plk1 has been demonstrated using antisense oligonucleotides in cell culture and in mouse xenograft models, in several studies. Elez and colleagues (Elez *et al,* 2003) used antisense oligonucleotides in combination with electroporation of tumours *in vivo* to promote specific uptake and reduce overall toxicity. Inhibition of Plk1 caused tumour shrinkage and produced a synergistic effect when administered with antisense oligonucleotides against Bcl-2, in cell lines overexpressing Plk1 and Bcl-2.

The use of siRNA to deplete Plk1 levels has also been shown to have wide-ranging growth-inhibitory effects, to promote apoptosis and cause abnormal spindle formation in tumour cells in a manner partly reliant on p53 depletion in cells (Guan *et al,* 2005). Normal human cell lines with wild-type p53 levels showed very little response to Plk1 depletion, showing that targeting Plk1 activity using siRNA, small molecule inhibitors or interfering peptides, can be a viable strategy for cancer therapy. This specificity for tumour cells may be partly due to the lower likelihood of siRNA uptake by normal cells (Spankuch-Schmitt *et al,* 2002b), but even after siRNA uptake, the anti-proliferative, pro-apoptotic and microtubule-disrupting effects of Plk1 depletion seen in tumour cells are greatly reduced or completely absent in normal cells.

Overall, the depletion of Plk1 using RNAi technology appears to cause dramatic effects in tumour cells and in mouse xenograft models, including cellular growth inhibition, with an increased percentage of cells in G2/M, abnormal spindle formation, multiple centrosomes or centrosomes that cannot nucleate microtubules, and anti-tumour activity *in vivo* (Spankuch-Schmitt *et al,* 2002).

These results validate the use of Plk1 as a drug target. Since our understanding of the activity of Plk1 is improving, with the availability of its interactome and the molecular basis of Plk activity being illuminated by structural studies, Plk1 inhibition is becoming an increasingly important drug target.

3.1.3.2 Strategies for inhibiting PLK1

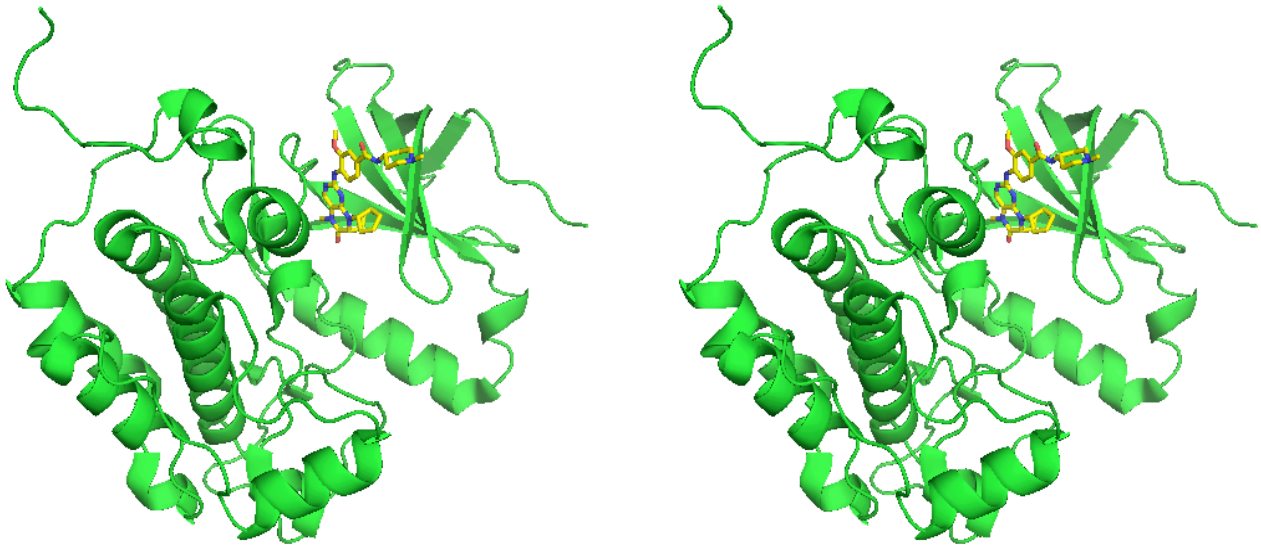
There are two main strategies for inhibition of Plk1 using small molecules: either target them to the kinase domain and exploit features of the active site to achieve potency and selectivity, or target the PBD and compete for phosphoepitope binding, therefore interfering with Plk1 localisation. siRNA has been shown to be effective as a technology for disrupting Plk1 by depletion of its cellular levels, so use of siRNAs as a drug provides an alternative possibility for drug development (Guan *et al,* 2005). Generally speaking, however, the use of small molecules provides a more stable platform for inhibition. The main inhibitors of Plk1 in both classes are reviewed by Murugan (Murugan *et al,* 2011).

There are several inhibitors already available which target the kinase domain of Plk1. The first to be discovered was the natural product scytonemin, an extremely non-selective kinase inhibitor (McInnes *et al,* 2005). ON01910, a benzyl styryl sulfone analogue, was initially thought to be potent and selective for Plk1 (Gumireddy *et al,* 2005) but actually appears to be a relatively non-selective kinase inhibitor (Zhong *et al,* 2010), with other targets including PI3K. It certainly has microtubule-disrupting effects and appears to work synergistically with other chemotherapeutic agents, so is currently in clinical trials jointly with other therapies.

Over the last five years, the co-crystal structures of a range of Plk1 inhibitors bound to the kinase domain have been published and reveal details of the interactions that confer specificity and potency. Wortmannin is a PI3K and Plk inhibitor, which covalently modifies the Plk1 kinase domain at K68, but also inhibits Plks 2 and 3 without covalent modification. The structure of Wortmannin bound to the Plk1 kinase domain was determined (Elling *et al,* 2008) and shows how the active site residues interact with the drug to confer potency. Aside from the covalent modification, several H-bonds form, to K68, C119 and other residues indirectly via bridging water molecules.

BI2536, a potent inhibitor selective for Plks 1-3, is currently showing some efficacy in Phase II clinical trials for various types of solid tumours, including as a monotherapy in patients with relapsed NSCLC (Sebastian *et al,* 2010). The structure of BI2536 bound to the Plk1 kinase domain (Kothe *et al,* 2007; Figure 3.6) shows that a methoxy group of the inhibitor fits into the small pocket formed by L132 in the hinge region.

A



B

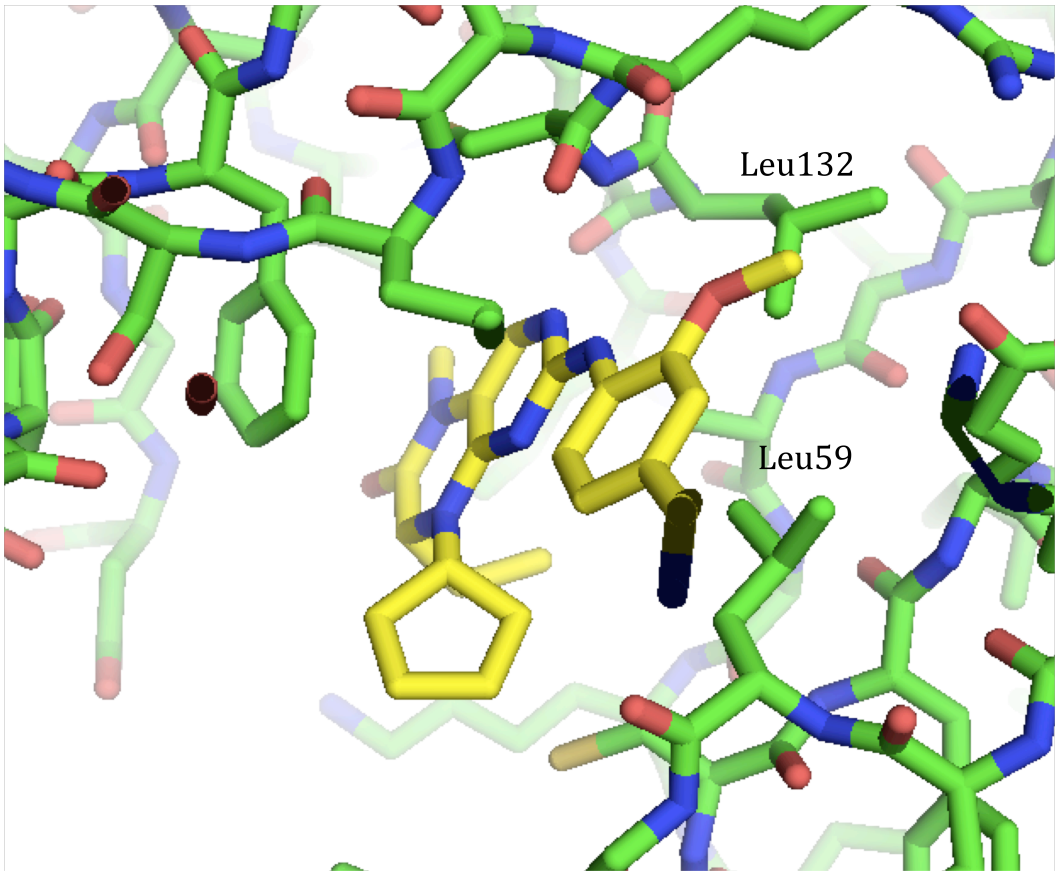


Figure 3.6 BI2536, a specific inhibitor of Plks Structure of BI2536 bound to the Plk1 kinase domain (PDB ID 2RKU). Figure prepared in Pymol. **(A)** Stereo view showing overall position of BI2536. Protein is shown in green ribbons and the inhibitor is is yellow stick representation. **(B)** Close up view of the binding pose of BI2536 showing the methoxy group of the inhibitor inserted into the small hydrophobic pocket made by the residues Leu 59 and Leu 132.

This explains the specificity of the inhibitor against Plk1 since in Plk2 the residue is a tyrosine, which is too bulky to replicate this interaction. The identity of the amino acid at this position may be a key determinant of selectivity. The inhibitor PHA-680626 also exploits this pocket, but also forms key interactions with a phenylalanine at the bottom of the ATP pocket, a cysteine in the roof of the binding site and positive residues adjacent to the hinge region (Kothe *et al*, 2007b). Observation of these Plk1-specific features shows avenues that may be exploited to develop drugs specific to Plk1.

Additional inhibitor-kinase domain co-crystal structures demonstrate similar interactions, including the structure of the kinase domain bound to a DARPin (designed ankyrin repeat protein) (Bandeiras *et al,* 2008); BI6727 (a BI2536 derivative) (Rudolph *et al,* 2009); a quinazoline mimetic (Beria *et al,* 2010) and a purine mimetic (Elling *et al,* 2008b). Although the design and development of drugs to target the kinase domain is more advanced than those targeting the PBD, there may be significant advantages to the latter approach. As with all PPI inhibitors, the unique nature of the interaction (as the PBD is not found in any other proteins, but only in Plks 1-3) means that side effects should be minimised.

3.1.3.3 Known PLK1 inhibitors targeting the PBD

Thymoquinone, a natural product from the black seed *Nigella sativa,* is a potent inhibitor of Plk1, binds to the PBD and displaces substrates containing the pThr/pSer binding motif (Figure 3.7). Thymoquinone has previously been observed to have anti-neoplastic activity, and was identified as an inhibitor of Plks using a fluorescence polarisation assay, which observes phospho-peptide displacement (Reindl *et al,* 2008). Its derivative poloxin demonstrates specific activity against the Plks, with an IC_{50} of 4.8±1.3 μM against Plk1, 18.7±1.8 μM against Plk2 and 53.9±8.5 μM against Plk3 and with no activity against other phospho-peptide binding domains. Thymoquinone is even more potent, with an IC_{50} of 1.14±0.04 μM but a less desirable specificity profile, binding the Chk2 FHA domain, the Pin1 WW domain and the SH2 domain of STAT3.

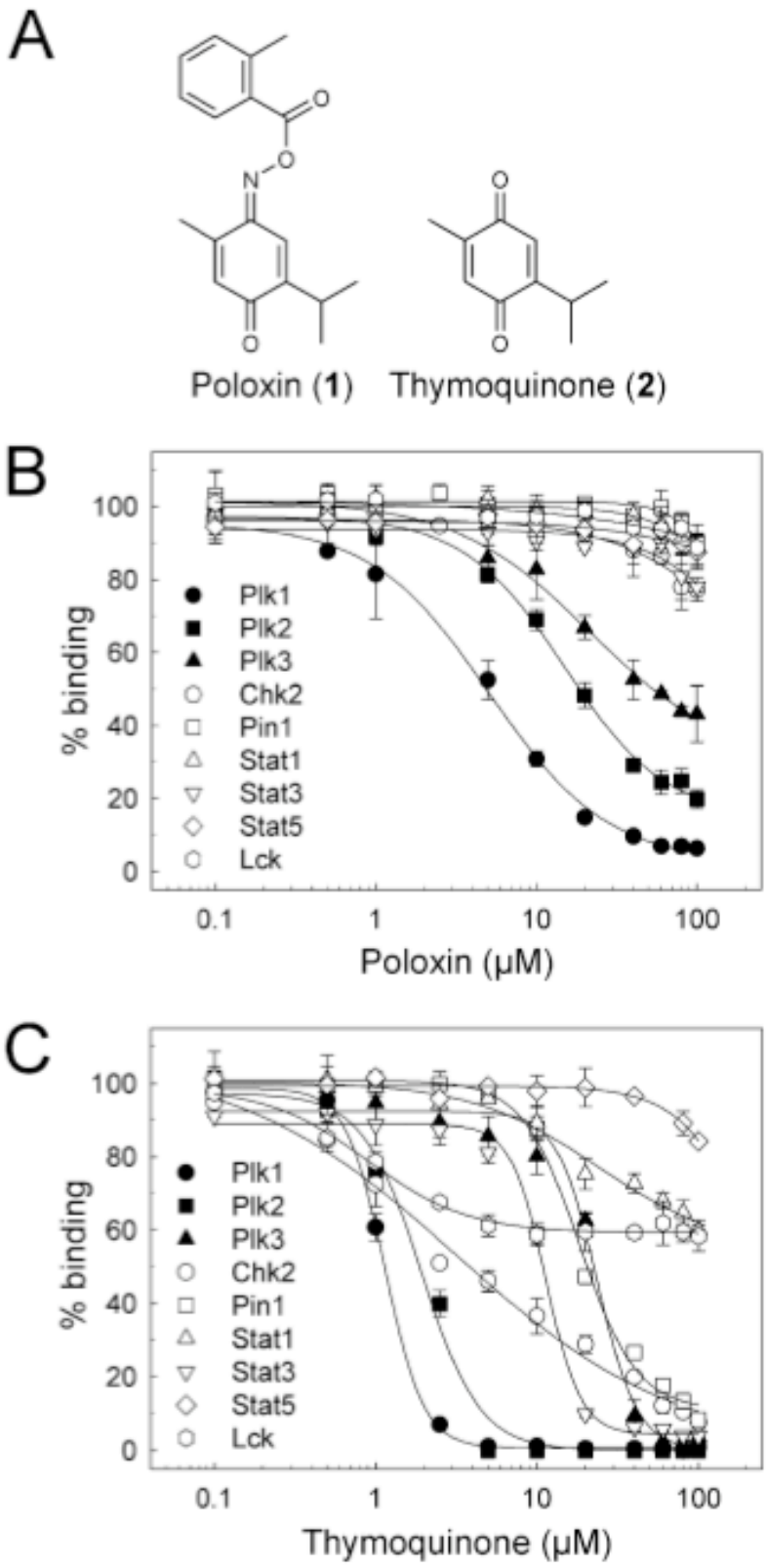


Figure 3.7 Poloxin and thymoquinone bind the PBD The inhibitors poloxin and Thymoquinone **(panel A)** were tested for binding against the Plk PBDs and several other Thr/pSer binding domains using a fluorescence polarisation assay **(B and C)**. Reproduced from Reindl *et al,* 2008.

Thymoquinone and poloxin were both able to induce mitotic arrest in HeLa cells, cause mislocalisation of Plk1 away from centrosomes and kinetochores into the cytoplasm, induce chromosome congression defects and cause apoptosis (Reindl *et al,* 2008). Likewise, poloxipan, another PPI inhibitor of Plks1-3 that binds with IC_{50} values of 3.2 μM for Plk1 PBD, 1.7 μM for Plk2 PBD and 3.0 μM for Plk3 PBD, can cause mitotic arrest, Plk1 mis-localisation and chromosome congression in HeLa cells (Reindl *et al*, 2009). Cells treated with poloxipan also exhibited centrosome fragmentation, similar to that seen with the Plk1 PBD inhibitor purpurogallin (Watanabe *et al,* 2008). This fragmentation may possibly be due to inhibition of the interaction of Plk1 with the centrosomal substrate Kizuna, involved in maintaining centriole structure.

3.1.4 Aims of this project

The identification of thymoquinone and poloxin as potent inhibitors of the Plk1 PBD provides the basis for this project, in which we aimed to identify compounds that bind to the PBD from a selection related to thymoquinone, as well as from compounds known to inhibit Plk1. In addition, a second aim was to obtain a co-crystal structure of thymoquinone or derivatives bound to the Plk1 PBD, in order to understand how these compounds are able to displace phosphopeptides bound to the PBD. This information would be useful to inform drug design when developing compounds targeting the PBD, a relatively new approach for Plk1 inhibition. All compounds were kindly provided by Shudong Wang (Nottingham University).

3.2 Expression and purification of the Plk1 PBD

3.2.1 Full-length Plk1 PBD

The PBD of human Plk1 (residues 345-603) was expressed and purified as described in Appendix A2.2 (after the method of Cheng *et al*, 2003). Briefly, the Plk1 PBD was expressed as a GST fusion protein using the vector pGEX6P-1. The plasmid pGEX6P-1-Plk1 PBD was transformed into BL21(DE3)/pLYS-S competent cells and colonies were selected using ampicillin and chloramphenicol agar plates. The colonies were picked and grown overnight in starter cultures of LB containing ampicillin (50 μg/ml) and chloramphenicol (35 μg/ml). Cultures were scaled up to 6L using the same media and protein was expressed overnight at 19^{o}C following induction with 0.1 mM IPTG. Cells were pelleted by centrifugation (5000 x g), flash frozen using liquid nitrogen and kept at -80^{o}C until required.

To purify Plk1 PBD, cell pellets containing the GST-Plk1 PBD fusion were resuspended in HEPES-buffered saline (HBS: 50 mM HEPES, pH 7.5, 250 mM NaCl, 0.02% MTG, supplemented with Complete protease inhibitor cocktail), and after sonication and clarification (100,000 x g), the cell lysates were applied to a glutathione sepharose 4B gravity flow affinity chromatography column. The complex was eluted by addition of glutathione (10 mM in HBS) and then the GST tag was cleaved by 3C protease digestion overnight. The mixture of Plk1 PBD and GST was subsequently buffer-exchanged and further purified by size-exclusion chromatography (HR26/60 S200). Fractions containing Plk1 PBD were pooled and the purity of the final product was checked using SDS-PAGE.

The progress of a typical purification of the full-length Plk1 PBD is shown in Figure 3.8.

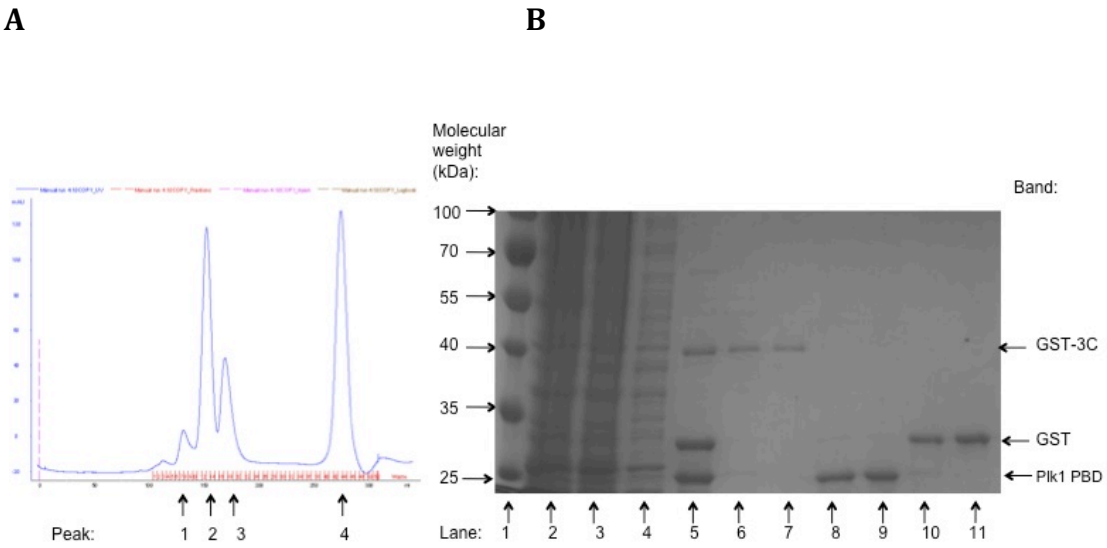


Figure 3.8 Purification of Plk1 PBD full-length (A) Size exclusion chromatogram with peaks labelled. Peak 1: Aggregates; Peak 2: GST-3C; Peak 3: GST; Peak 4: Plk1 PBD **(B)** SDS-PAGE gel of selected fractions. Lanes on the gel: (1) PageRuler ladder, (2) lysate, (3) affinity column flow-through, (4) affinity column wash, (5) affinity column elution after GST cleavage by 3C-GST, (6-7) 2x fractions from peak 2 (contains GCT-3C), (8-9) 2x fractions from peak 4 (contains Plk1 PBD), (10-11) 2x fractions from peak 3 (contains GST)

A typical yield using the above method was approximately 8 mg/L culture, although some of this was subsequently lost in the process of concentrating the protein for crystallisation.

3.2.2 Truncated PBD

Plk1 PBD was also purified in a truncated form, using subtilisin, a serine protease with broad specificity. This cleaves the PBD at two points, between residues 366 and 367, and between residues 494 to 495. Both of these cleavage points are in unstructured regions. The first causes removal of the N-terminal unstructured region 345-367, and the second occurs between PB1 and PB2. The purpose of the subtilisin cleavage was to aid crystallisation by decreasing flexibility and movement within the protein. After cleavage with subtilisin, two fragment of 10 and 12 kDa are obtained (reflecting each of the two separate polo boxes). These co-migrate during size-exclusion chromatography to reconstitute the full PBD 367-603 and therefore appear to have a high affinity for one another. The points of cleavage within the secondary structure of Plk1 PBD are illustrated in Figure 3.9.

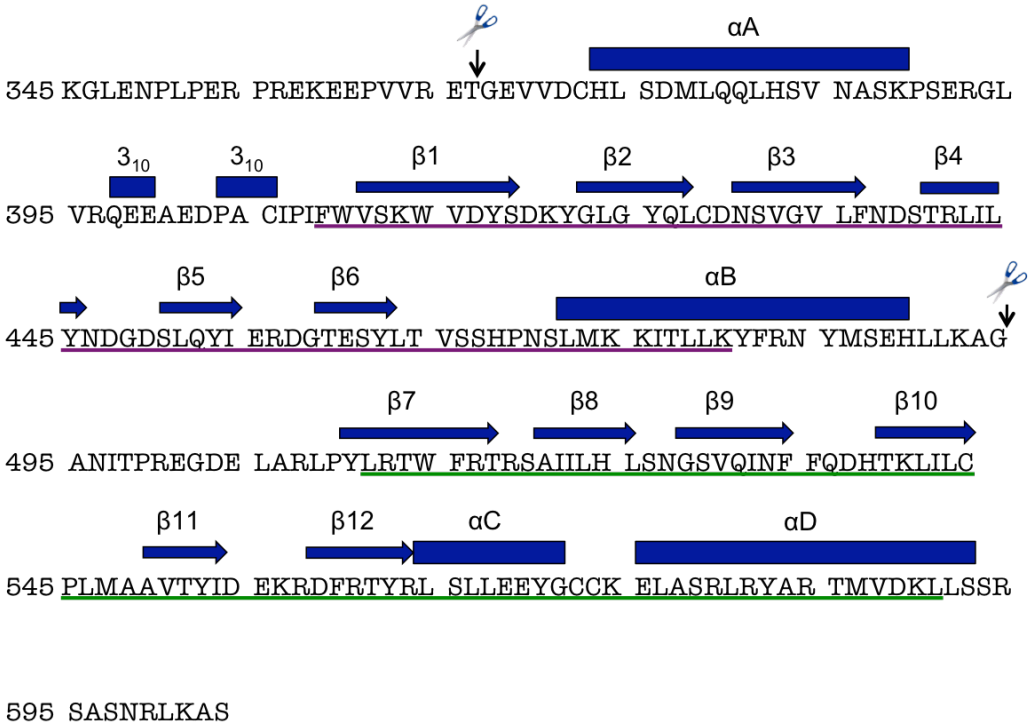


Figure 3.9 Secondary structure of the Polobox domain Residue numbers are given. The secondary structure (as identified in the crystal structure by Cheng *et al*, 2003) is shown: α helices are shown as rectangles, β strands as arrows. The subtilisin cleavage points are illustrated. PB1 comprises the residues underlined in purple; PB2 comprises the residues underlined in green.

Plk1 PBD, obtained from the process of expression and purification detailed in Section 3.2.1, was incubated with 1:200 (w/w) subtilisin and the progress of PBD digestion followed using SDS-PAGE. Digestion went to completion within 30 minutes and the reaction was stopped by adding 0.1mM phenylmethylsufonyl fluoride (PMSF). The resulting mixture of PB1, PB2, the peptide representing residues 345-366 and subtilisin, was purified further to remove the enzyme and peptide in order to obtain pure PBD for crystallisation. Size exclusion chromatography showed that PB1 and PB2 co-migrated, re-forming the PBD complex.

3.3 Fluorescence anisotropy: SAR of compounds

3.3.1 Thymoquinone and poloxin derivatives

3.3.1.1 Confirmation of IC_{50} value for thymoquinone

After the observation of Plk1 inhibition by thymoquinone and poloxin using a fluorescence polarisation assay reported by Reindl (Reindl *et al,* 2008), the same method was employed to probe compound binding to PBD. This technique measures displacement of a fluorescent peptide and its use when performing a competition assay is illustrated in Figure 3.10.

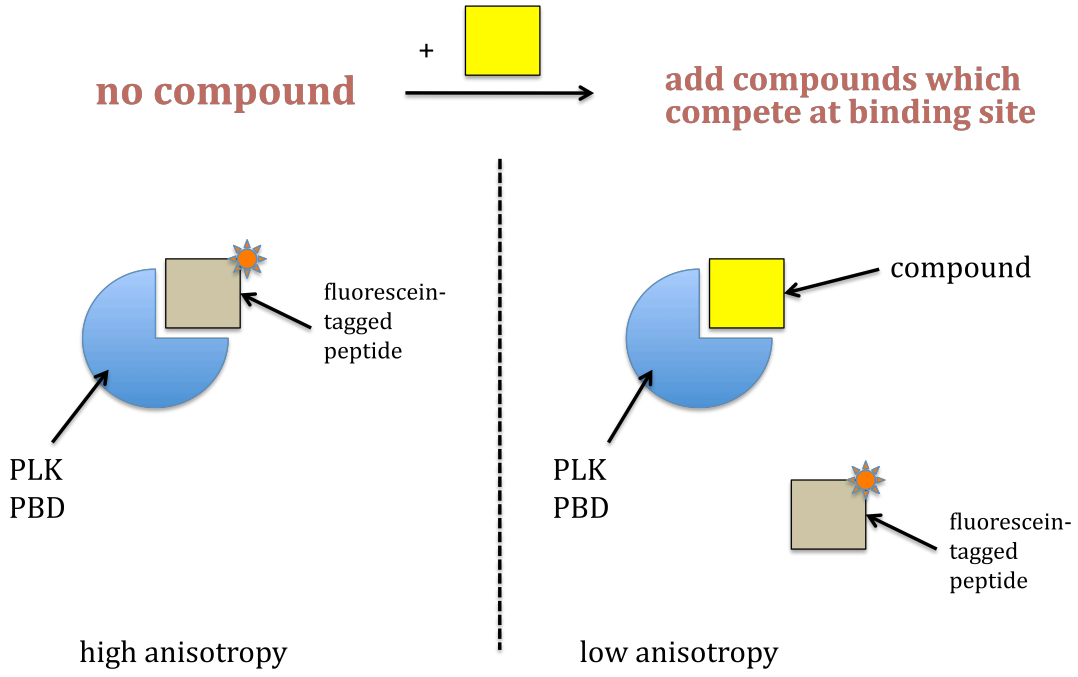


Figure 3.10 Basis of fluorescence anisotropy competition assay Compounds which are able to displace the fluorescein-tagged peptide by binding competitively cause a change in the fluorescence anisotropy as the tumbling rate of the fluorescein tag changes, due to the size difference in the molecular species. Titration of compounds into the PBD bound to fluorescent peptide gives a sigmoidal curve, which can be analysed to obtain an IC_{50} value.

Initially, the method was tested using the same conditions as outlined in the paper: 5-carboxyfluorescein-MQSpTPL (from Severn Biotech) at 10 nM, Plk1 PBD at 65 nM and potential PBD-binding compounds at a range of

concentrations from 0-200 μM were pre-incubated at room temperature for 30 minutes in black 384-well plates in order to allow equilibration and to eliminate bubbles. Fluorescence anisotropy was determined using a BMG Pherastar FS microplate reader and inhibition curves were fitted using Prism. Several controls were included in the experiments: a Plk1 PBD control with no fluorescent peptide, a Plk1 PBD + fluorescent peptide control (with no compound added) and a control with unlabelled 200 μM MQSpTPL peptide. After trying different conditions, the concentration of Plk1 PBD used was doubled to 130 nM, and the 5-carboxyfluorescein-MQSpTPL concentration was also doubled to 20 nM, since this increased the signal obtained from the experiment.

The assay was first used to calculate the IC_{50} value for thymoquinone against the Plk1 PBD, in order to check that the assay was working properly and agreed with the published value (Reindl *et al,* 2008). Some problems with reproducibility were encountered. Repeating the assay daily revealed that these problems were due to problems with freezing and thawing aliquots of Plk1 PBD, presumably due to degradation, aggregation or denaturation of the protein (Figure 3.11).

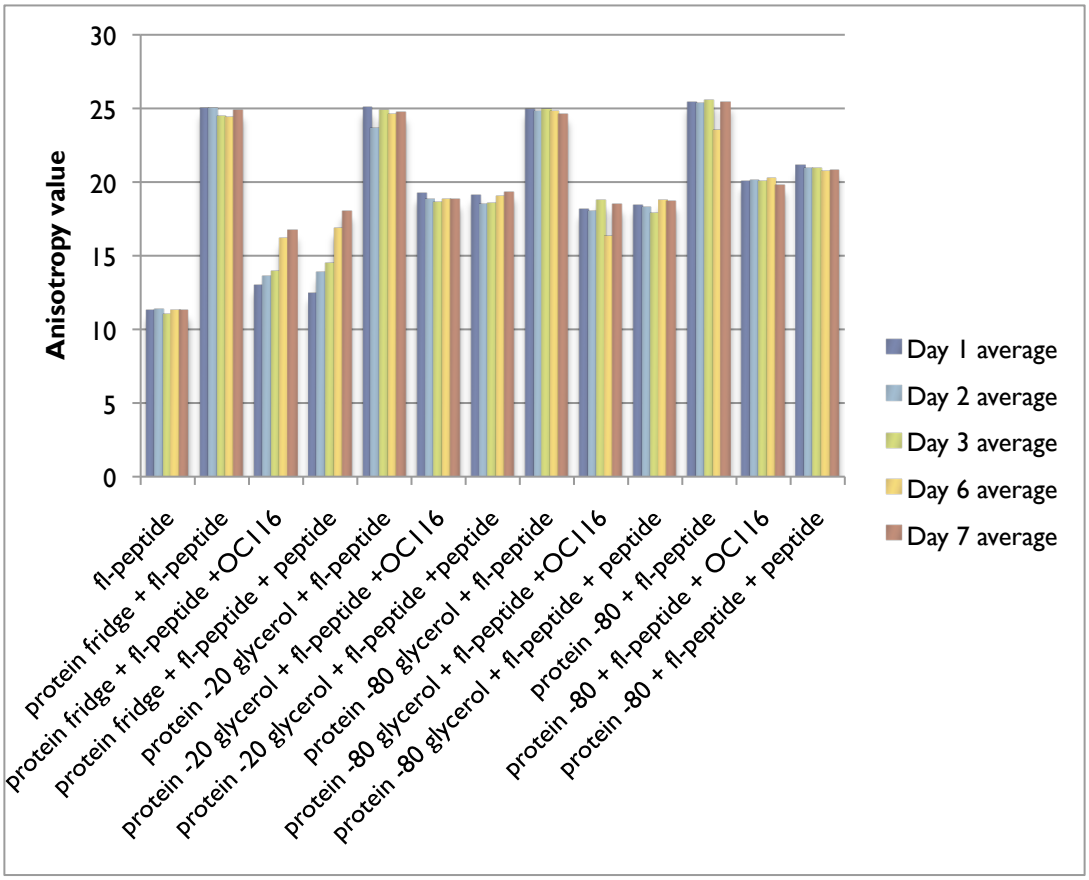


Figure 3.11 The response of Plk1 PBD to a freeze/thaw cycle After freezing and thawing, or keeping in the fridge for several days, the ability of Plk1 PBD to bind a fluorescent peptide as determined by fluorescence anisotropy is impaired. The response is slightly improved by freezing the Plk1 PBD with glycerol.

Due to these problems with freezing and thawing the protein, or using protein from the fridge, a new preparation was used every time an assay was carried out, and the protein was used for assays for a maximum of 3 days. The fluorescence anisotropy curve obtained for titration with thymoquinone is shown in Figure 3.12.

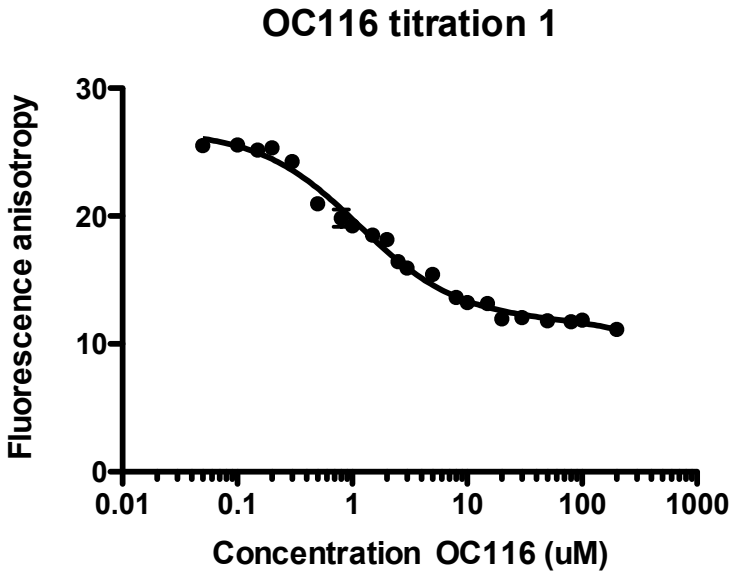


Figure 3.12 Fluorescence anisotropy competition assay Vertical error bars show standard deviation. The experiment was carried out in triplicate.

The IC_{50} was calculated as 1.72±0.28 μM, not too dissimilar to the 1.14±0.04 μM calculated by Reindl *et al*. This experiment validates the methodology used by acting as an internal control. Error bars were calculated as the average of three measurements from experiments performed from the same protein prep, but on different days.

3.3.1.2 Determination of IC_{50} values for thymoquinone derivatives

The fluorescence anisotropy assay was then used to investigate the binding of the "OC" compound series, a set of derivatives of thymoquinone developed at Nottingham University, the structures of which are shown in Figure 3.13. These compounds were designed as PBD binders using the structures of poloxin and thymoquinone as starting points.

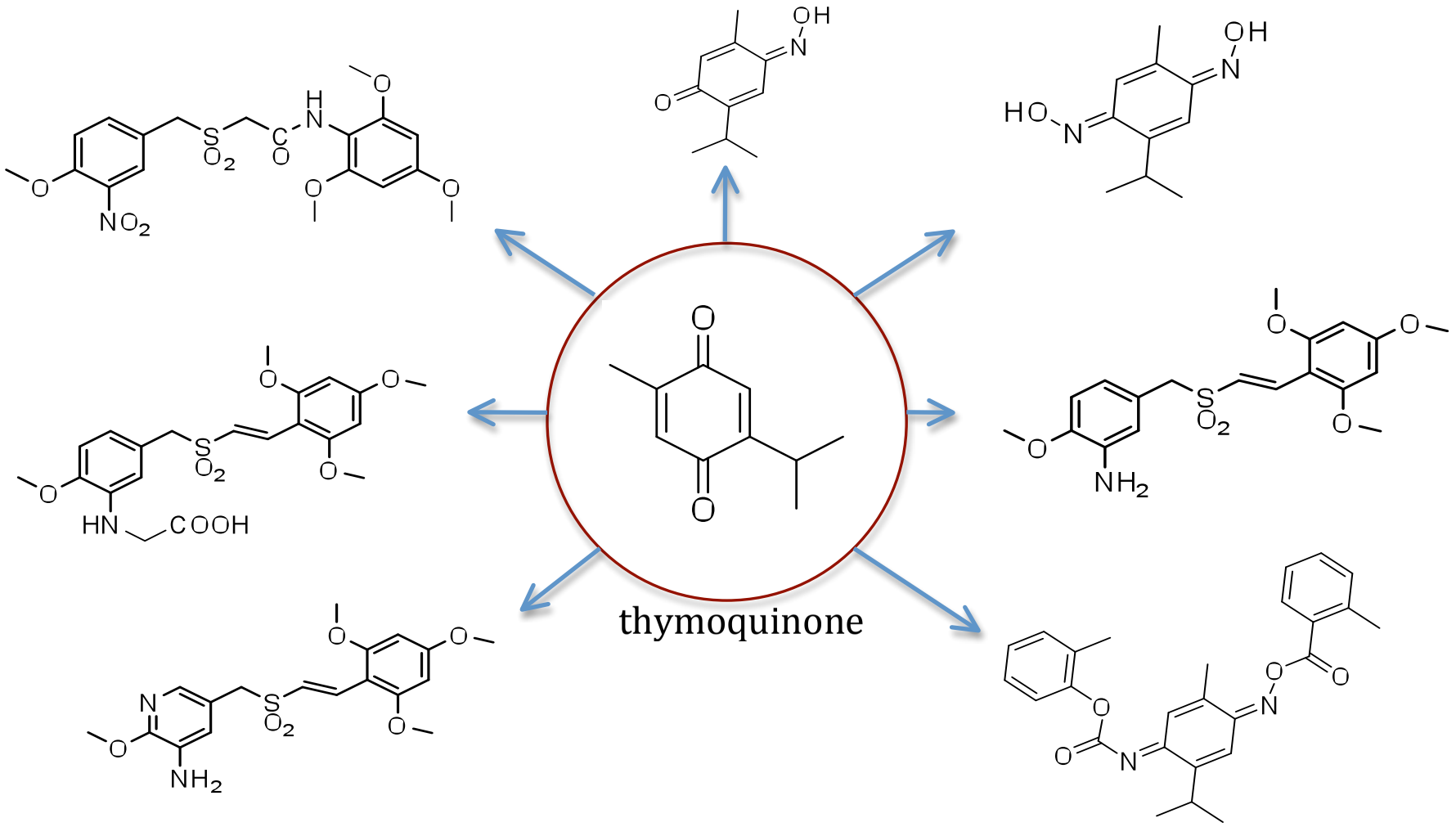


Figure 3.13 Derivatives of thymoquinone Compounds were synthesised by Shudong Wang. Clockwise from top left: OC74, OC108, OC109, OC21, OC111, OC107, OC23

The results of the fluorescence anisotropy assay with the "OC" compounds are shown in Figure 3.14. The results are not shown for OC13, OC21, OC23 and OC74 since these titrations did not produce any significant alteration in the anisotropy value up to 200 μM and therefore do not appear to bind to the PBD to any appreciable extent.

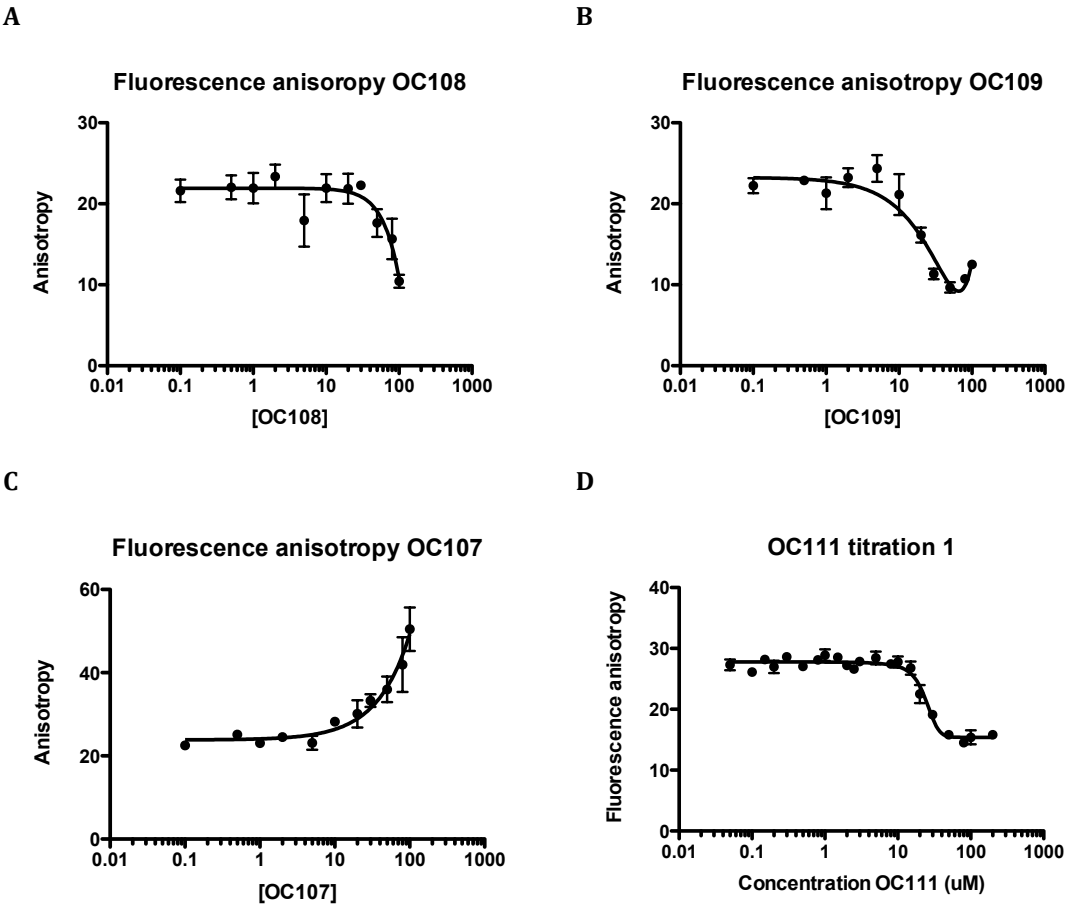


Figure 3.14 OC series compounds bind to the Plk1 PBD with varying affinity. Fluorescence anisotropy competition assay measuring titrations of OC compound series into the Plk1 PBD. Vertical error bars represent standard deviation. The assay was carried out in triplicate.

The IC_{50} for OC111 binding to Plk1 PBD was calculated as 23.92 ±1.93 μM (Figure 3.14, Panel **D**). Since none of the other compounds appeared to saturate binding at 200 μM (the maximum achievable concentration due to solubility issues) an accurate IC_{50} could not be calculated for these compounds but is likely to be >50 μM in each case. Error bars were significantly higher for compounds

OC107, OC108 and OC109, and in addition the curve for OC107 appeared to show increasing anisotropy with a higher concentration of compound. This may be because the compound bound non-specifically and caused aggregation and non-specific binding of fluorescein-tagged MQSpTPL. Likewise the small increase in anisotropy observed with high concentrations of OC109 might be due to protein aggregation. In all cases the IC_{50}s may be lower than suggested here, because there were significant solubility problems for all compounds in the OC series. Solubility was especially problematic with OC21, OC107 and OC111 and the effective concentration in solution for these compounds was likely to be lower than measured.

3.3.2 Screening of other known Plk1 inhibitors for PBD binding

The fluorescence anisotropy assay was also used to test known Plk1 inhibitors for PBD binding. ON01910 is a Plk1 inhibitor that has shown activity in xenograft models and is currently in clinical trials as an anti-cancer therapy (as a single agent and in combination therapy) (clinicaltrials.gov). It has broad specificity and also inhibits PDGFR, ABL, FLT1, CDK2, Plk2, Src and Fyn (Murugan *et al,* 2011). The mode of interaction of ON01910 with Plk1 has not been elaborated, but it is known that it is not ATP competitive but is substrate competitive (Gumireddy *et al,* 2005). These results suggested that it might bind to the PBD and disrupt its interaction with substrates, hence disrupting Plk1 activity and therefore microtubule organisation. TL77 and AH111 are related compounds that had been suggested to inhibit Plk1 by binding the PBD.

The results of the fluorescence anisotropy assay are shown in Figure 3.15.

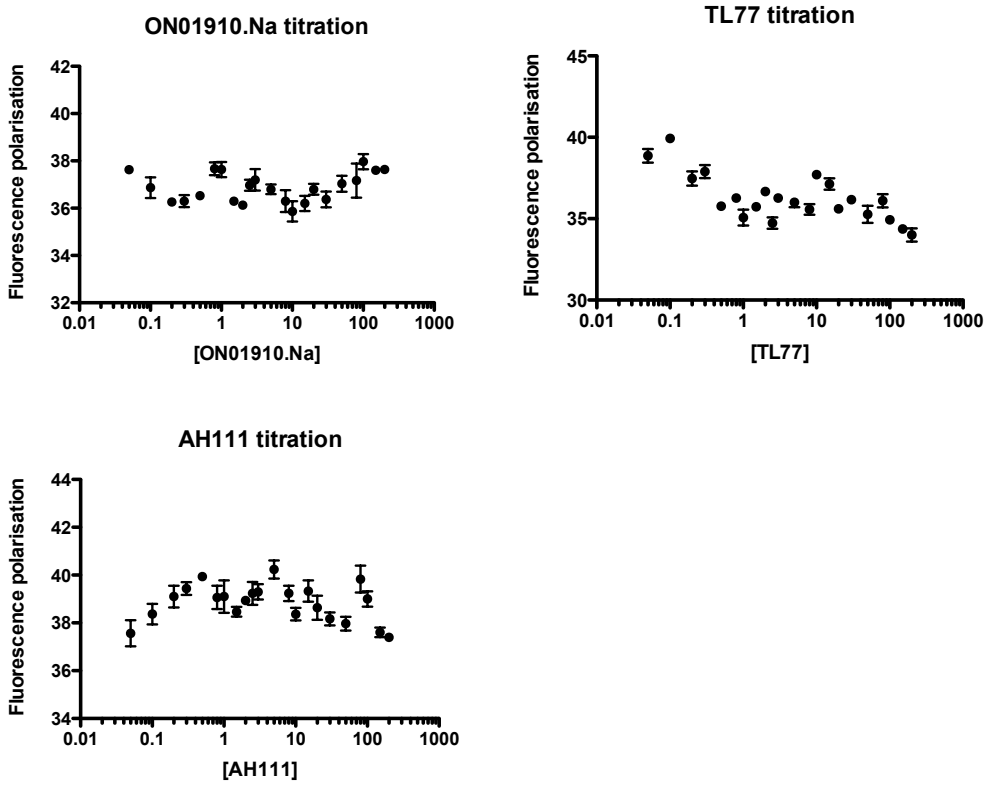


Figure 3.15. ON01910.Na and related compounds do not bind to the Plk1 PBD Fluorescence anisotropy assay carried out as before. Vertical error bars show standard deviation. Experiments were carried out in triplicate.

We have been able to establish a fluorescence-based assay to screen PBD binding compounds. This assay was first validated using a known PBD binding compound. Unfortunately, a subsequent screen of a small set of potential PBD binding molecules did not yield any more potent compounds, so thymoquinone remains the best candidate for structural investigation.

3.4 Crystallisation attempts: soaking the crystals

3.4.1 Soaking protocols and strategies

The crystal conditions previously identified by R Cheng are shown in Table 3.1. These included truncated polobox (subtilisin-cleaved) at 5 mg/ml final concentration, 5-10% (v/v) PEG 4000, 0.1 M sodium citrate pH 6.0 and 0.1 M ammonium acetate (condition 2).

Condition	**Polobox construct length**	**MSGpTPL peptide included?**	**Appearance of crystals**	**Do crystals diffract?**
1	Untruncated (residues 345-603)	No	Thin plates growing in clusters	Yes, but low resolution (only about 5.5 Å)
2	Truncated (residues 367-603)	No	Thin plates growing in clusters	Yes, to 2.2 Å
3	Untruncated (residues 345-603)	Yes	Thicker plates in clusters, some single crystals	Yes, to 2.3 Å

Table 3.1 Previously identified Plk1 crystal forms as detailed by R. Cheng, DPhil thesis 2003

Using these conditions crystals of the same appearance as those previously grown were reproduced (Figure 3.16). These conditions were also used to grow crystals without peptide or compounds added, which were subsequently used for compound soaks, in an attempt to obtain a structure of Plk1 PBD bound to thymoquinone or derivatives.

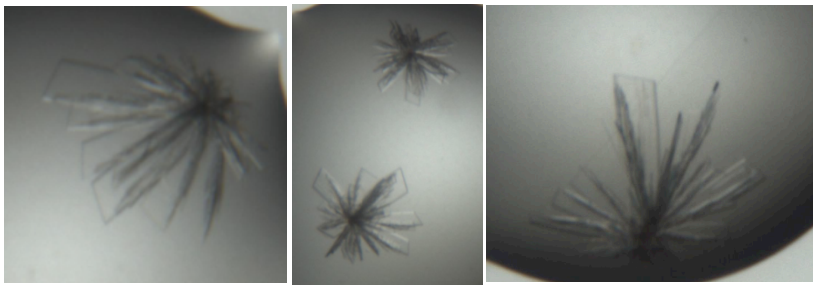


Figure 3.16 Crystals of apo PBD (subtilisin-cleaved) used for compound soaks The plates were extremely thin and fragile and always grew in clusters of plates or needles, to a maximum length of 400μm.

Several different protocols for compound soaks were employed to maximise the chance of success. An overnight soak of the compounds at 100 μM final concentration (in well solution) followed by cryoprotection with 8M sodium formate or 35% glycerol containing 100 μM compound, proved too harsh and crystals dissolved or became severely cracked and degraded. These crystals did not diffract. A soak at 100 μM final concentration (in well solution) for 2 hours produced much less distortion or degradation of the crystals, and these were sent for data collection at the Diamond Light Source on beamline I04. Disappointingly, although they diffracted to about 2.5 Å, upon analysis of the data, using 1Q4K as a search model for molecular replacement, it became apparent that the soak had been unsuccessful as no difference density identifying a bound compound was visible.

The soaking protocol was further modified to improve exposure to the compound whilst minimising shock to the crystal. A sequential soaking procedure in which the concentration of compound was increased sequentially by 20 μM, up to a maximum of 100 μM, for 30 minutes at each step, appeared to produce less disintegration of the crystal. However, again there was no difference density identifying a bound inhibitor.

3.5 Crystallisation attempts: co-crystallisation

3.5.1 Co-crystallisation: full-length PBD

Attempts to crystallise the full-length (undigested) PBD focused on obtaining co-crystals with either MQSpTPL (in order to try and soak in compounds, displacing the peptide) or directly with thymoquinone or derivatives. This approach was taken because the apo crystals previously obtained with the full-length protein did not diffract to high resolution. No co-crystals of thymoquinone-bound (or OC compound-bound) Plk1 PBD were obtained using a panel of commercially available screens (MD1+2, JCSG+, MIDAS, Morpheus, Wizard, Proplex, PACT and PEG/Ion). However, co-crystals of full-length Plk1 PBD bound to the optimal phosphopeptide MQSpTPL were obtained in narrow screens based on the conditions identified by Cheng *et al* (1-10% PEG 20000, 0.1M MES pH 6.5) and set up either using hanging- or sitting-drop mode (Figure 3.17).

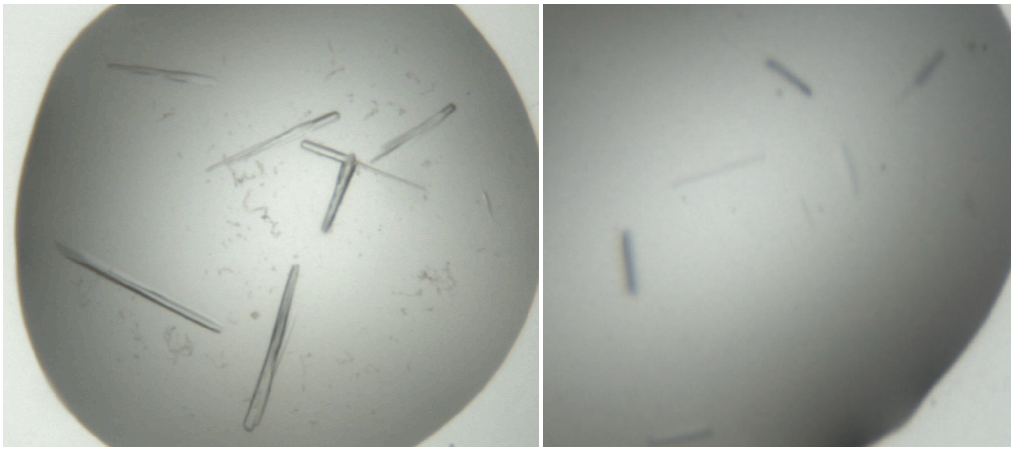


Figure 3.17 Plk1 PBD co-crystals with the MQSpTPL optimal phosphopeptide Crystals were thin rods of maximum 500 μm in length

These crystals were subsequently soaked with thymoquinone or the "OC" derivatives using the protocols described in Section 3.4.1. However, they did not diffract to high resolution after soaking. These crystals proved quite delicate and due to their thinness were prone to breakage, which may have compromised data quality.

3.5.2 Co-crystallisation: truncated PBD

No co-crystals of thymoquinone-bound (or "OC" compound-bound) Plk1 PBD were obtained in any of the commercially available broad screens tried (MD1+2, JCSG+, MIDAS, Morpheus, Wizard, Proplex, PACT and PEG/Ion). Narrow screens were set up using some conditions in which phase separation was identified, however no crystals were obtained.

3.6 Further experiments

Due to the lack of success with obtaining a co-crystal structure of our compounds bound to the PBD after screening thousands of conditions and employing various techniques to improve crystallisation, it would probably be better to spend some time optimising the construct rather than working further on co-crystallisation with the current construct. The PBD prep could benefit from some optimisation. The final product obtained after purification is acceptably pure, (100% for the full-length PBD; <80% when subtilisin-treated) and the yield is good; however crystallisation in a form which diffracts and allows for soaking in compounds, or the growth of co-crystals, has proved extremely problematic. It is possible that engineering the PBD for better crystallisability, possibly by truncating some loops or introducing surface mutations at key sites, may improve the situation with regards to crystallisation.

Due to the relocation of our collaborator further stocks of the "OC" compound series were not available. However, Thymoquinone is commercially available so it would be possible to consider experiments soaking crystals of Plk1 PBD and performing crystallisation trials to obtain co-crystals using this compound and new constructs. It is possible that the "OC" thymoquinone derivatives, or similar derivative compounds, may become commercially available. Further derivatives may also enable more sophisticated SAR to be performed, in order to better understand the interactions made between PPI inhibitors and the PBD.

It is not clear whether thymoquinone would be expected to bind in the phosphopeptide binding site or work via an indirect effect, since there is still very little information about the interaction. It is somewhat surprising that such a potent inhibition is found in such a small compound, especially for a PPI disrupter. However this may be a reflection of the 'hot spots' idea, whereby the inhibitor is binding in key places within the binding site and disrupting interactions that contribute most of the potency to the PPI. In the absence of a crystal structure it would be interesting to perform molecular docking to try and make some predictions as to where the compound might bind in the PBD. However the small size of the compound makes this quite difficult, as it is extremely unlikely that any program would be able to narrow the solutions down to find the correct one.

It is still important to continue attempts to obtain a crystal structure of the PBD bound to thymoquinone or poloxin in order to understand the molecular basis by which these compounds displace the phosphopeptide and inhibit the protein-protein interaction. This appears to be a rather more difficult task than originally anticipated due to the difficulties in working with Plk crystals, particularly in crystal handling, compound soaking and data collection and processing. However if the structure were to be determined it would be very significant for our understanding of drug development at the phosphopeptide binding site.

Chapter 4: CDK2/cyclin A: targeting the recruitment site

4.1 Background and introduction

4.1.1 CDKs, the role of CDK2 in the cell cycle and control of CDK2 activity

Cyclin-dependent kinases (CDKs) are the master regulators of the cell cycle (Figure 4.1). The cyclins, which control their activity, vary cyclically in expression and activate different CDKs at different points in the cell cycle resulting in the phosphorylation of diverse targets. CDK inhibition has long been an area of interest due to the central importance of the CDKs for cell cycle regulation, their interaction with oncogenes and tumour suppressors and their frequent deregulation in human tumours (reviewed in Shapiro, 2006).

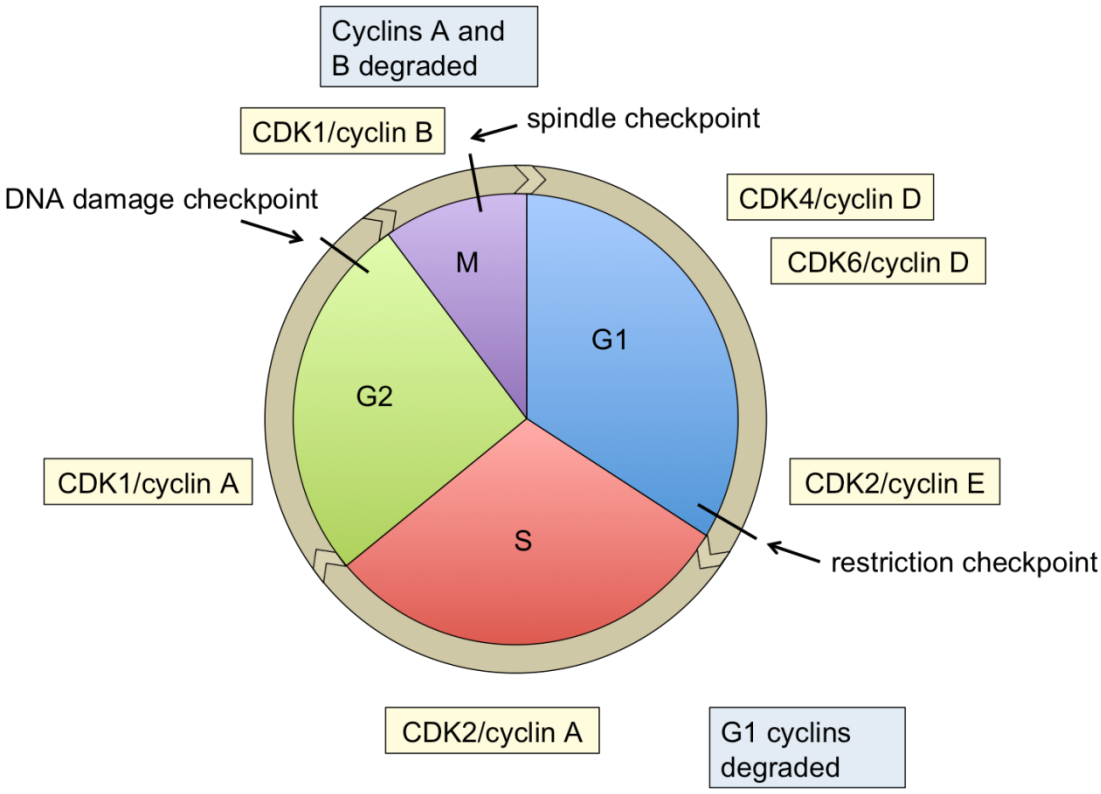


Figure 4.1 CDKs and cyclins control the cell cycle. G1, S and G2: interphase; M: mitotic phase (consists of mitosis [prophase, metaphase, anaphase, telophase] and cytokinesis). The control of cell cycle progression is regulated by cyclins and CDKs. CDK2 in complex with the G1 cyclins (such as cyclin E), drives G1 progression. After degradation of the G1 cyclins in S phase, CDK2 binds to cyclin A, which promotes S phase

progression. *(Fig 4.1 legend continued on page xx)*

During G2, CDK1 binds to cyclins A and B, but is inactivated by phosphorylation. Upon dephosphorylation of the CDK1/cyclin B complex the cell can progress from G2 to M phase. The APC (anaphase-promoting complex) drives M phase progression. Checkpoints active throughout the cell cycle ensure correct progress, monitor DNA integrity and provide a restriction on growth, protecting against cancer development. Examples of checkpoints include the restriction point (Start point in yeast), at the G1/S boundary; the G2 checkpoint, which prevents cells with DNA damage from undergoing mitosis; and the spindle checkpoint, which happens in metaphase and ensures that chromosomes are correctly aligned at the mitotic plate. If bipolar tension is sensed anaphase entry is initiated.

The CDK4-CDK6/cyclin D-Rb pathway is frequently mutated in human cancer: either through loss of Rb or inactivation of $p16^{INK4A}$ by gene deletion, point mutation, or transcriptional silencing by methylation (Ruas and Peters, 1998). Cancer cells with increased activity of cyclin D-dependent kinases then obtain a growth advantage, may escape from senescence and progress towards malignancy. Generally, CDK dysfunction in cancer cells may weaken the robustness of checkpoints and allows escape from growth control. This may be one reason for the large amount of redundancy and compensatory mechanisms between CDKs (Satyanarayama and Kaldis, 2009). Mouse models with loss of function mutations in the different CDKs and cyclins illustrate that there is some compensation of function between the CDK/cyclin complexes in mature cells, although knockouts in an early developmental stage can be embryonic lethal. Knocking out CDKs 2, 3, 4 or 6 is viable but loss of function of CDK1 or CDK11 is embryonic lethal and loss of function of CDK2, 4 and 6 is embryonic lethal (reviewed in Malumbres, 2011). Knockouts of the cyclins show that only cyclin A2 and B1 are required for viability. Knockouts of other cyclins show more mild phenotypes (Gong and Ferrell, 2010).

CDK2/cyclin A is important for driving S phase progression in humans and phosphorylates many substrates involved in DNA replication (Walker and Maller, 1991; Girard *et al,* 1991). Its activity is tightly controlled by members of the Kip/Cip family of inhibitory proteins; by an activating phosphorylation of T160 in CDK2 (mediated by CDK7/cyclin H); and by an inhibitory phosphorylation on Y15 (Figure 4.2). The complex is deactivated when cyclin A

is destroyed by the anaphase promoting complex/cyclosome at the G2/M transition. CDK2 is also activated by cyclin E in late G1 phase and required for the G1 to S transition. CDK2/cyclin E phosphorylates $p27^{Kip1}$, leading to its degradation. Removal of this CKI creates an environment in which CDK2/cyclin A activity can rise. The CDK2/cyclin E complex also phosphorylates Rb and stimulates E2F-mediated transcription, which in turn promotes the expression of cyclin A and entry into S phase (reviewed in Sherr and Roberts, 2004).

4.1.2. CDK2 structural biology

CDK2/cyclin A is the best-characterised CDK/cyclin complex at the structural level. Several X-ray crystal structures in different states as well as multiple crystal structures of CDK2/cyclin A/ATP-competitive inhibitor complexes have been determined. Early structural work is reviewed in Pavletich, 1999. CDK2 is only fully active as a kinase when phosphorylated on Thr160 and in complex with cyclin A (Figure 4.2).

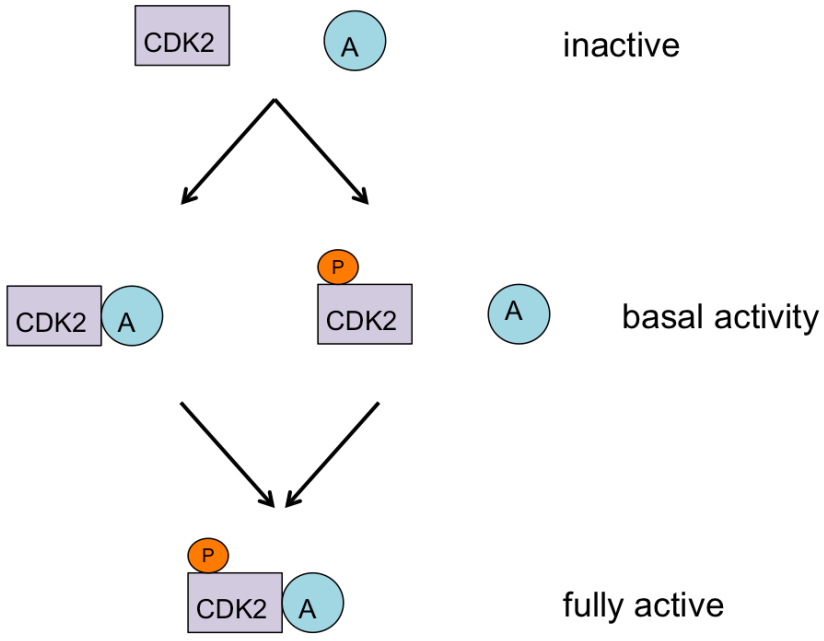


Figure 4.2 Control of CDK2 activity The binding of cyclin A (or E) and phosphorylation on residue T160 are both required for full activity of CDK2. Upon phosphorylation, the activation segment of CDK2 changes configuration to a more open state and reveals the active site. Cyclin binding to CDK2 alters the conformation of CDK2 and recruits substrates for phosphorylation. Destruction of cyclin A in M phase suppresses CDK2 activity and favours cell cycle progression.

A structure of the CDK2/cyclin A complex allowed elucidation of the mechanism of activation of CDK2 by binding to cyclin A: large conformational changes occur to the PSTAIRE helix and the activation segment, which remodel and remove the block on the catalytic cleft (Jeffrey *et al,* 1995). A structure of the fully active phosphorylated (T160p) form of CDK2 bound to cyclin A (phosphoCDK2/cyclin A) subsequently illustrated how phosphorylation activates the complex: the phosphate group acts as an organising centre, binding three arginine residues and rearranging the catalytic cleft further (Russo *et al*, 1996b). Thirdly, structures of $p16^{INK4a}$ and $p27^{Cip2}$ bound to pCDK2/cyclin A illuminated the mechanism of action of the Cip and INK4 families of inhibitors (Russo *et al*, 1998; Russo *et al*, 1996). The structure of the CDK2/cyclin A complex is shown in Figure 4.3 and the main structural elements that could potentially be targeted by ATP non-competitive inhibitors are marked.

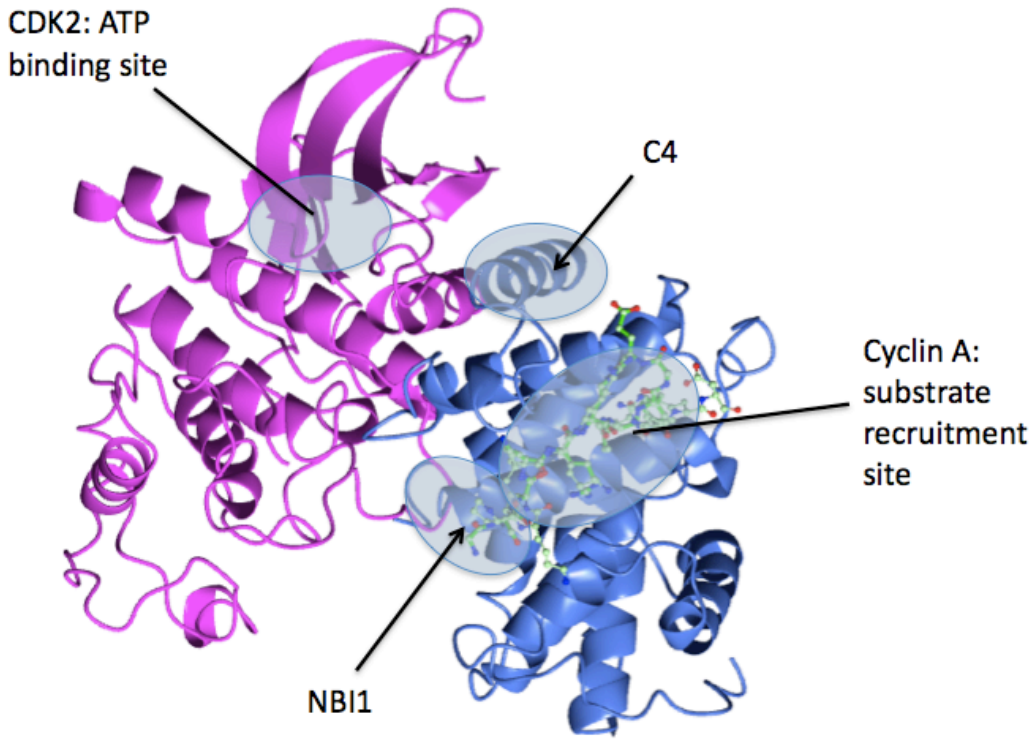


Figure 4.3 pCDK2/cyclin A complex structure. Potential points where small molecules could bind to CDK2/cyclin A to inhibit activity are labeled on the crystal structure. (PDB code 1H28 that includes a p107 11-mer peptide bound at the recruitment site). CDK2 is shown in purple and cyclin A is shown in blue. Translucent blue ellipses show the areas that have been targeted by small molecules. C4 and NBI1 target respectively the α5 helix of cyclin A (which interacts with the PSTAIRE motif of CDK2) and the α3–α5 cleft,

implicated in the interaction of CDK2 and cyclin A (reviewed in Orzaez *et al*., 2009). Other inhibitors have targeted the cyclin binding groove or the ATP binding site in CDK2. (Adapted from Orzaez *et al*., 2009).

Though CDK2 has a preference for substrates that contain the short SPXK/R motif around the site of phospho-transfer, early studies by Adams (Adams *et al*, 1996) also recognised that a number of good substrates including p107, pRb, E2F and p53 as well as the CKIs all contain the motif RXL. RXL works to assist substrate recognition, vastly increasing the efficiency of phosphorylation for substrates including this motif. The RXL motif binds to a pocket termed the recruitment site that is conserved in a number of cyclins (Figure 4.4).

The mechanism of recruitment is still not fully understood but has been partially elucidated by structural studies. Lowe *et al* obtained a crystal structure of the pCDK2/cyclin A complex interacting with a short p107-derived RXL-containing recruitment peptide ((Lowe *et al*, 2002), Figure 4.4). Several ion pairs were observed between the RXL peptide and the recruitment site, which explain why this sequence is required. Some of the key interactions are formed between: E220 and the residue in position P-2, W217 and the main chain carbonyl of the residue in position P-4, Q254 and the backbone of the recruitment peptide (where position P is the leucine in the RXL motif).

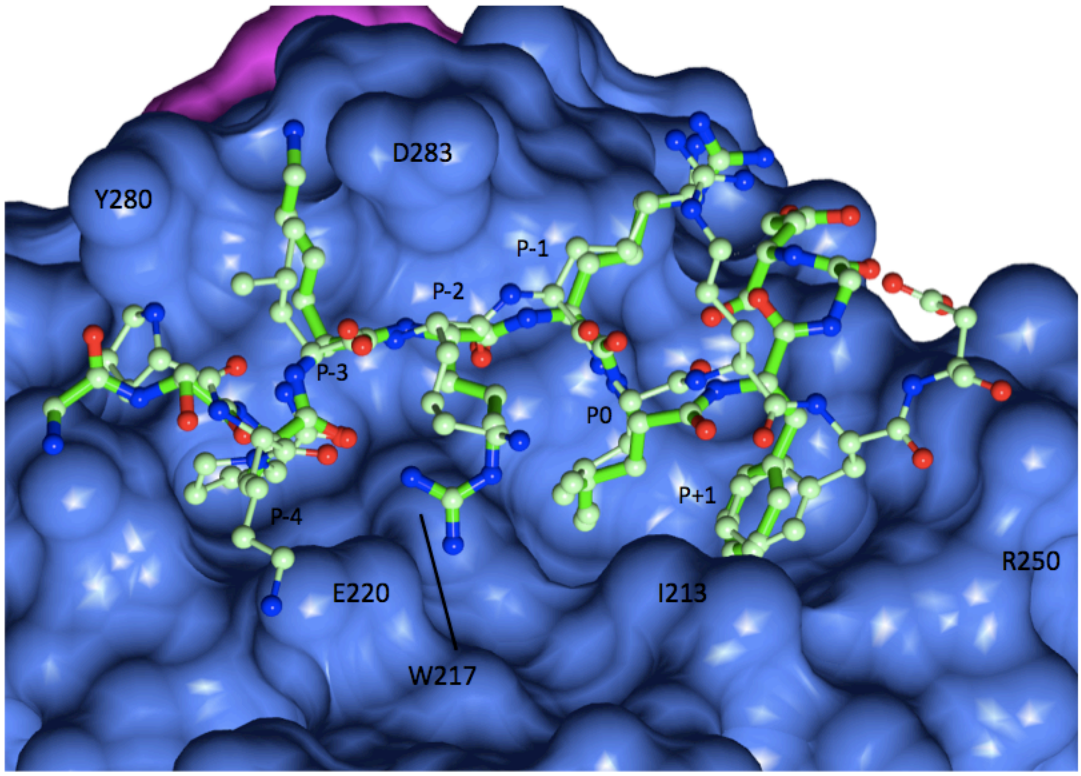


Figure 4.4 The cyclin A substrate recruitment site. The RXL-containing peptides (11-mers) from E2F and p107 are overlaid and shown in ball-and-stick representation. (E2F (dark green) and p107 (light green)). Cyclin A is shown as a surface in blue (CDK2 in purple at back). Residues in the recruitment site making important contacts with the peptides are labelled. The residues of the peptides are labeled by position number. Figure prepared in CCP4MG using the PDB files 1H25 and 1H28.

Lowe *et al* used X-ray crystallography and isothermal titration calorimetry to study the interactions present at the recruitment site when the recruitment peptide from E2F1, or 11-mer peptides from p53, pRb, p27, E2F or p107 were provided as substrates (Lowe *et al,* 2002). An additional hydrophobic residue C-terminal to the leucine of the RXL motif (or in the adjacent position) was shown to insert into the recruitment site and to be an important contributor to recruitment peptide recognition. This structure further elucidated the contacts between the RXL peptide and the recruitment site, and illustrated that a diverse array of substrates may be accommodated because variations in sequence can be tolerated provided the shape is conserved.

Kinetic studies using a peptide substrate derived from Cdc6 designed to observe simultaneous binding of a substrate to the recruitment site and the catalytic site, with the two regions joined by a flexible linker region determined a K_m for the interaction that was 18 times lower when the RXL motif was covalently linked to the substrate peptide (Cheng *et al,* 2006). Other studies demonstrated that the RXL motif could be located either N- or C-terminal to the site of phosphotransfer and provided that the optimal length for the intervening sequence is 12-18 residues. This means that the RXL motif and the SP(X)K motif (the site of phosphorylation) can be simultaneously bound to the cyclin to promote phosphorylation by the CDK (Takeda *et al,* 2000). The Cheng *et al* structure also suggested how specificity can be achieved between substrates because by varying the length and content of the flexible linker region, the substrate affinity can be markedly altered.

4.1.3. CDK2 as a drug target

The rationale for choosing CDKs as drug targets relies on the fact that dysfunction in these pathways often leads to escape from growth controls and malignancy. There is preliminary evidence that CDK inhibitors may act as anti-tumour agents in selected cases *in vivo.* Two classes of dual CDK1 and CDK2 inhibitors, the phenylamino pyrimidines (PAPs) and tri-substituted purines, have shown differences in their cellular effects in normal cells compared to tumor cells (Ruetz *et al*., 2003). Normal cells are reversibly blocked by these CDK inhibitors in either the G1/S-phase or at G2/M. However, in transformed cells lacking functional G2/M control checkpoints, the compounds produced a prolonged proliferation block at the G2/M transition. G2/M arrest also appears to be less well tolerated by transformed cells, leading to apoptosis. These dual CDK1/ CDK2 inhibitors are also significantly more active on proliferating cells, compared to quiescent cells. Additional recent *in vitro* work has shown proof of principle for a pan-CDK inhibitor: flavopiridol was shown to inhibit the growth of anaplastic large cell lymphoma cells and lead to apoptosis (Bonvini *et al*., 2009).

However, there is some reason to be cautious about the use of CDK inhibition for cancer therapy. Firstly, the central role of CDKs in control of the cell cycle has been widely questioned in recent years as many CDKs have been shown to be dispensible (reviewed in Malumbres and Barbacid, 2009). CDK1 appears to be required for the cell cycle but interphase CDKs are only essential for proliferation of certain specialized cells (and for meiosis). CDK2 knockout mice are viable (Berthet *et al*., 2003), although the combined loss of CDK2 and CDK4 is embryonic lethal (Berthet *et al*., 2006) but tolerated when induced pharmacologically in mature cells after development (Barriere *et al,* 2007). It is now becoming clear that there is a high level of functional redundancy within the CDK family. CDK4 can phosphorylate pRb even at CDK2-preferred sites such as Thr831 (Zarkowska and Mittnacht, 1997) and Cip/Kip proteins are still able to induce G_1 arrest in CDK2 -/- cells, again suggesting that CDK2 is dispensable and redundant.

These data suggest that selective inhibition of CDK2 may not be a useful therapeutic strategy as a monotherapy. However, one exception may be for malignant melanoma, since in melanocytes, transcription of CDK2 is regulated by the melanocyte lineage transcription factor microphthalmia-associated transcription factor (MITF) and expression of MITF and CDK2 is highly correlated only in primary human melanomas. Low MITF and CDK2 expression in melanoma cell lines leads to increased susceptibility to G_1 arrest when CDK2 is removed by siRNA targeting, or decreased proliferation with roscovitine, a CDK2 inhibitor (Du *et al*., 2004). A selective CDK2 inhibitor may therefore be useful for some types of melanoma.

Additionally, recent evidence has suggested that some CDK inhibitors also affect the stability and activity of the tumour-suppressor protein p53, a major drug target. Administration of the potent CDK2 inhibitors roscovitine and olomoucine caused nuclear accumulation and increased activity of p53 (David-Pfeuty, 1999). This might increase the efficacy of CDK inhibitor therapy, and could explain why differing results are sometimes observed with these therapies: the effect would depend on the cell's p53 status. Furthermore, CDK2 inhibitors have been

suggested to protect normal cells from the effects of chemotherapy (Wesierska-Gadek and Schmid, 2006). For all these reasons, CDK2 appears to be a valid drug target.

Previous attempts to develop a selective CDK2 inhibitor have varied in success (reviewed in Malumbres and Barbacid, 2009). CDK inhibitors which were introduced into the clinic as selective CDK1 or CDK2 inhibitors were subsequently found to also inhibit other members of the family. A particular example is flavopiridol, which could be more accurately described as a pan-CDK inhibitor. However, some of the problems with specificity are attendant to the strategy previously used, (targeting the catalytic site), since this is more similar between CDKs than the recruitment site, which is still conserved but has more flexibility. Targeting the recruitment site to inhibit interaction with the substrate could therefore prove much more promising.

4.1.4. Aims of the project

The general aim of this project is to study the interactions between cyclin A and recruited CDK substrates, with particular reference to identifying compounds that bind and inhibit the interaction. Structural biology and biophysical assays will be combined with molecular biology to probe details of the structural interface, and the kinetics of inhibitor binding. In addition the project aims to address open questions about the basic mechanism by which substrate recruitment works and the exact roles of different residues within the hydrophobic patch on cyclin A in substrate recruitment. The initial aim was to screen libraries of compounds for inhibitors of CDK2/cyclin A that act by binding at the recruitment site.

4.2 Preparation for library screening

4.2.1 Protein expression and purification

pCDK2/cyclin A was expressed and purified following the method first described in (Brown *et al*., 1999) and covered in detail in Appendix A2. Briefly, CDK2 was co-expressed as a GST fusion protein with GST-CIV1, the *S.cerevisiae* CDK-activating kinase from the vector pGEX6P-1. Human and bovine cyclin A were both expressed from the vector pET21d. p107 S32A was expressed as a GST fusion protein using the vector pGEX6P-1. All proteins were expressed following induction with 0.1 mM IPTG in BL21(DE3)/pLYS-S cells overnight at 18^{o}C.

Lysates from cell pellets expressing CDK2 and CIV1, and untagged human cyclin A were sequentially applied to a glutathione sepharose 4B gravity flow affinity chromatography column. By this method, the GST-CDK2 bound to the glutathione generates an affinity column for the untagged cyclin A. The complex was eluted by addition of glutathione (10 mM in HBS) and then the GST tag was cleaved by 3C protease digestion (1:50 w/w). The mixture of CDK2/cyclin A and GST was subsequently buffer-exchanged and further purified by size-exclusion chromatography (HR26/60 S200 equilibrated in HBS). Pooled fractions were finally applied to a glutathione sepharose affinity column to separate pCDK2/cyclin A from the GST dimer. The purity of the final product was checked using SDS-PAGE. The activity of the complex was measured using a standard gel shift assay using pRb as a substrate (Brown *et al*., 1999).

Cyclin A was prepared following the method described in Brown *et al,* 1999. Cells expressing His-tagged bovine cyclin A were pelleted (5000 x g) then resuspended in modified TBS (50mM Tris pH 8.5, 300mM NaCl, 20mM $MgCl_2$, 0.02% MTG, supplemented with Complete protease inhibitor cocktail), sonicated and then following clarification (100,000 x g) the lysate was applied to a Ni-NTA affinity column pre-equilibrated in HBS. Cyclin A was eluted by applying an imidazole gradient (0-500 mM in HBS). Fractions containing cyclin A were

pooled and further purified by size-exclusion chromatography. The purity of the final product was checked using SDS-PAGE.

4.2.2 Mutagenesis of p107: Producing an optimal substrate for CDK2/cyclin A assays

p107 (RBL1) is a major substrate for CDK2/cyclin A at the end of S phase, and it was used in the radioactive kinase assays performed here as a substrate in order to study the binding of inhibitors to CDK2/cyclin A. In order to simplify the subsequent analysis of kinetic and binding experiments, residues 1-55 of the p107 sequence in the vector pGEX6P-1 (pGEX6P-1p107) was first engineered to contain only a single CDK2/cyclin A phosphorylation site. The major CDK2 phosphorylation site at Ser23 in the sequence SPIR was retained and a secondary site at Ser32 located between SXS and the RXL motif starting at R40 was removed. This mutation was introduced into both the wild-type p107 sequence and also a mutated sequence in which the RXL motif had been removed by mutation to alanine (pGEX6P-1p107ΔRXL), as described in Leng *et al,* 2002. Mutagenesis was performed using the QuikChange II Site-Directed Mutagenesis Kit from Stratagene, and primers from VHBio (Appendix 1). Successful mutagenesis was confirmed by DNA sequencing.

The mutated p107 sequence in the vector pGEX6P-1 (pGEX6P-p107MUT) was then transformed into BL21(DE3)/pLYS-S cells, expressed and purified using sequential glutathione sepharose affinity and size-exclusion chromatography columns (without cleavage of GST), and with other steps as described above (Figure 4.5). pGEX6P-1p107 containing the wild-type p107 sequence (pGEX6P-p107) was also successfully expressed and purified (Figure 4.5).

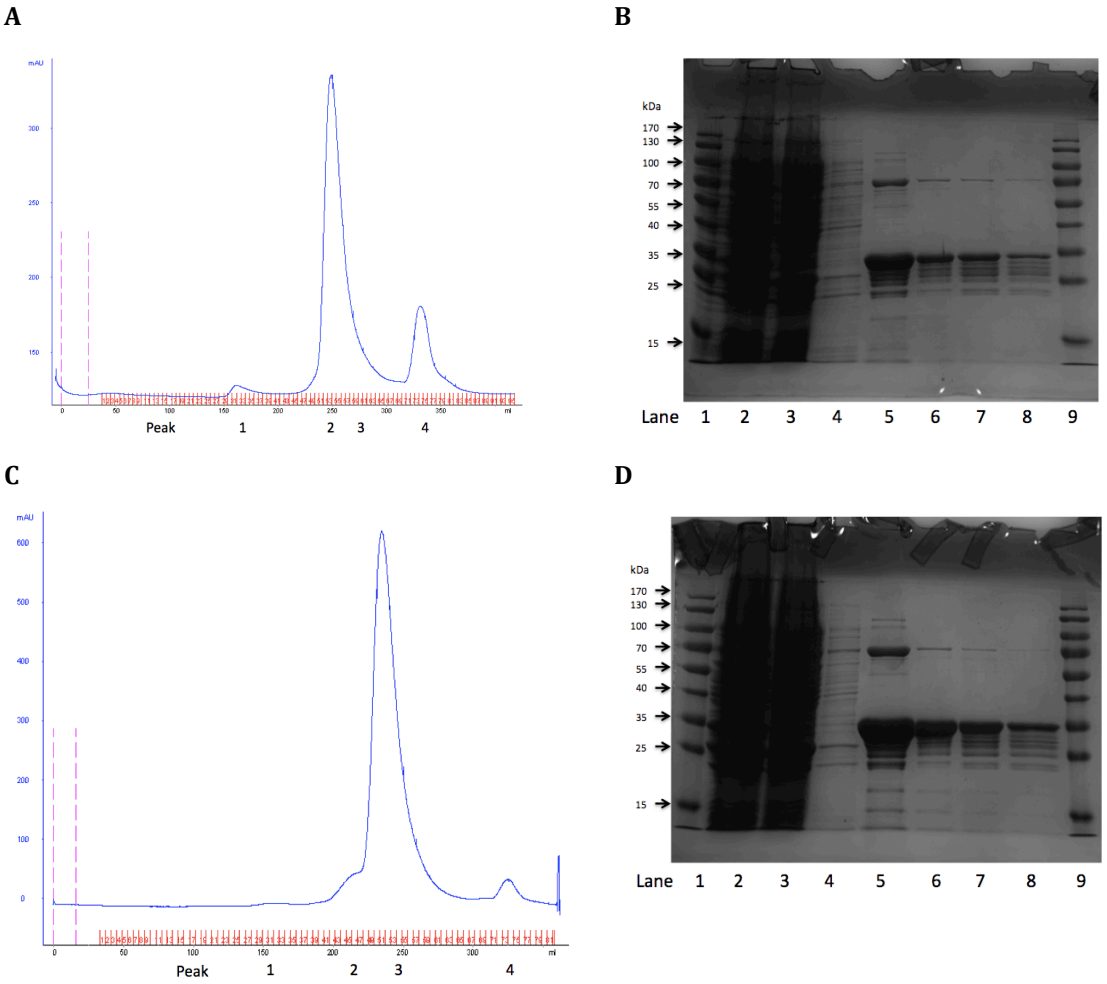


Figure 4.5. p107 purification. Purification of mutated (S23A) p107: **(A), (B)** with wild-type RXL motif**, (C), (D)** with deleted RXL motif. **(A), (C)** = size exclusion chromatograms; **(B), (D)** = associated SDS-PAGE. Peaks in the size-exclusion chromatograph are: 1 = aggregate; 2 = GST dimer; 3 = GST-p107; 4 = glutathione. Lanes in the gels show samples: 1 = molecular weight markers; 2 = whole-cell lysate; 3 = flowthrough glutathione sepharose affinity column; 4 = wash glutathione sepharose affinity column; 5 = elution glutathione sepharose affinity column; 6 = Fraction 52 from gel filtration; 7 = Fraction 56 from gel filtration; 8 = pooled fractions from gel filtration (final product); 9 = molecular weight markers. The ~70 kDa band represents aggregates and the ~35 kDa band represents GST-p107. Fragments running ahead of the full-length protein in the final product are breakdown products of GST-p107. The size of molecular weight markers is given.

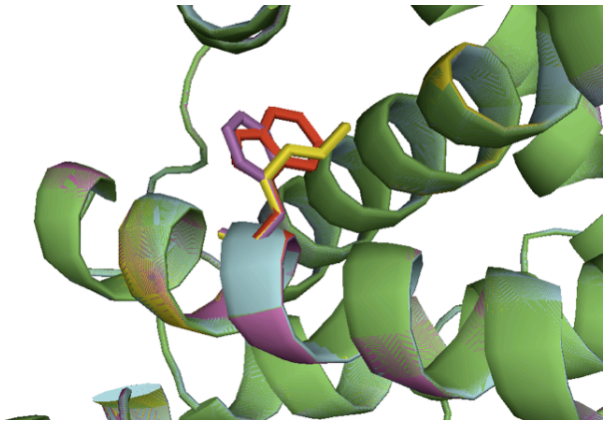
As observed by SDS-PAGE, p107 is prone to degradation (e.g. Figure 4.5, panel B, lane 8). This purification protocol could therefore benefit from further optimisation. However, the final p107 product was active as a substrate in

radioactive kinase assays, and so was deemed to be of sufficient quality for screening assays.

4.2.3 Mutagenesis of Cyclin A: investigation of Trp217 function in substrate recognition

Trp217 is found in the cyclin A recruitment site and has been shown in various cyclin A/peptide structures to interact with the leucine and (if present) another hydrophobic residue in the RXL(F) motif. To provide a reagent to investigate substrate binding this residue was mutated to isoleucine, phenylalanine and lysine. The introduction of each mutation was modelled using SwissModel (Figure 4.6). In each case the amino acid change could be accommodated in the structure and was not predicted to substantially alter the overall fold.

A



B

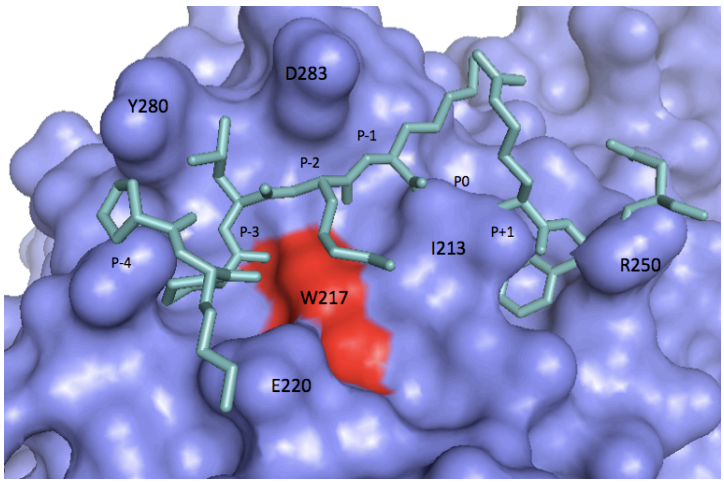


Figure 4.6 Modelling cyclin A mutations. Representation of the W217X mutated structure as calculated by SwissModel, shown in Pymol. *Figure legend continued on page 152.*

(A) Structure of mutated cyclin A (W217I/F/K) in ribbon view, with W217X shown in stick representation to highlight the change. Trp = red; Phe = purple; Lys = yellow; Ile = cyan. **(B)** The substrate recruitment site

in cyclin A (shown as surface in blue), with E2F 11-mer RXL-containing peptide bound (as in Figure 1.2) shown in cyan. Residues in the recruitment site making important contacts with the peptides are labelled. Peptide residues are labeled by position number. The W217X mutation site is shown in red.

Mutagenesis of the pET21d-HscyclinA (residues 175-432) or pET21dBtcyclin A (residues 175-432) vectors containing respectively the human (untagged) or bovine (C-terminal His_6-tagged) cyclin A sequences was attempted using the QuikChange II Site-Directed Mutagenesis Kit from Stratagene, and primers from IDT (Appendix 1). Mutagenesis of pET21dBtcyclinA, but not pET21dHscyclin A, was successful. pET21dBtcyclinA constructs each encoding one of the three mutations were then transformed into BL21(DE3)pLYS-S cells for protein expression.

4.2.4 Protein production and expression: cyclin A W217I/F/K

Protein expression and purification of cyclin A W217X, was carried out as described previously for the wild-type protein. The results of the purifications are shown in Figure 4.7:

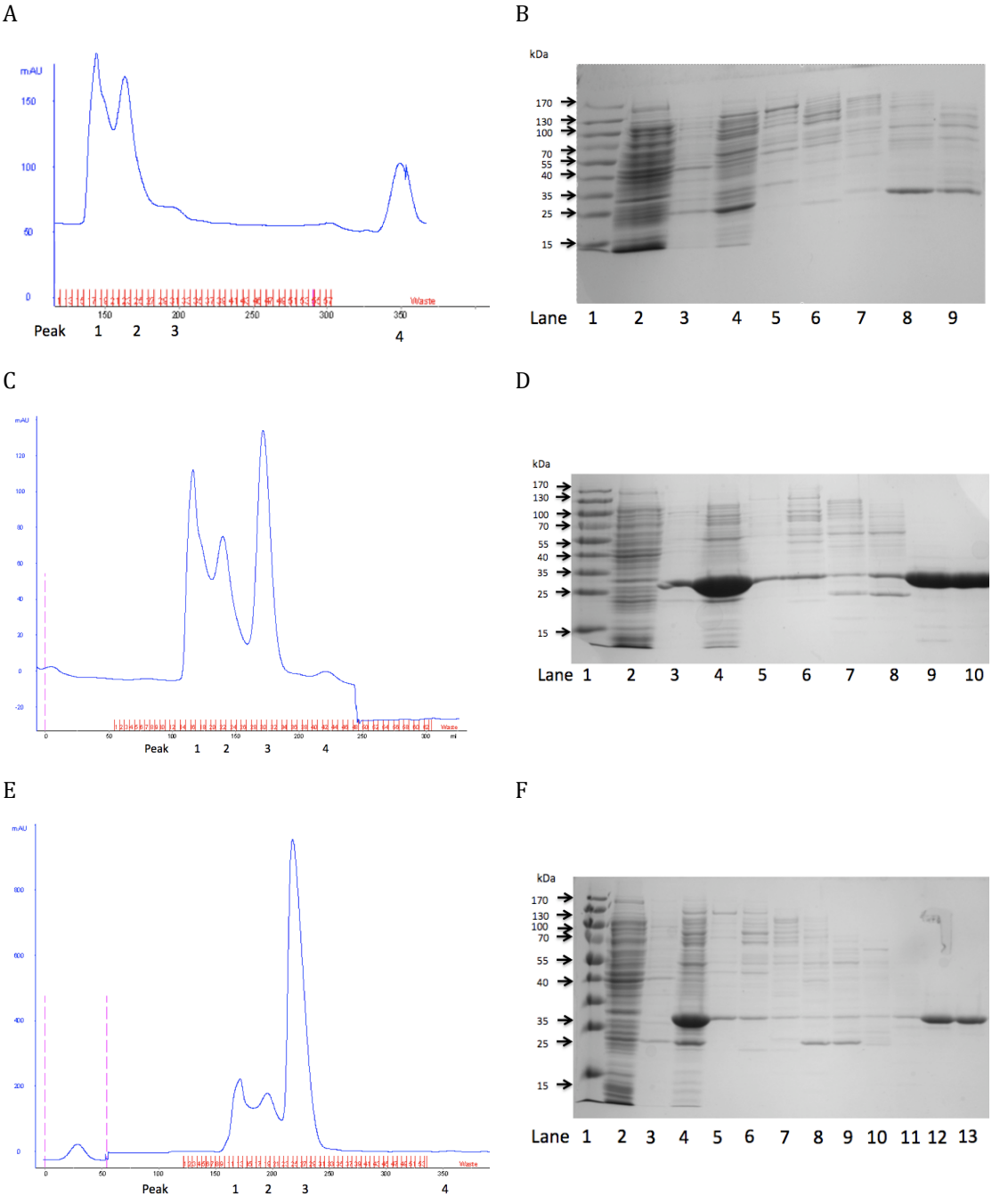


Figure 4.7 Purification of mutated (W217I/F/K) bovine cyclin A. **(A, B)** W217I purification; **(C, D)** W217F; **(E, F)** W217K. A, C and E panels are the respective size-exclusion chromatograms, B, D and F panels, accompanying SDS-PAGE analysis. In the size-exclusion chromatograms, peaks are labelled. Peaks 1 and 2 = cyclin A aggregates; 3 = cyclin A monomer; 4 = imidazole. In SDS-PAGE images, lanes are numbered. Lane 1 = molecular weight markers; 2 = flowthrough Ni-NTA affinity column; 3 = wash affinity column; 4 = elution affinity column. In a) 5 = fraction 18; 6 = fraction 21; 7 = fraction 24; 8 and 9 = pooled fractions (final product). In b) 5 = fraction 16; 6 = fraction 21; 7 = fraction 29; 8 = fraction 31; 9 and 10 = pooled fractions (final product). In c) 5 = fraction 11; 6 = fraction 14; 7 = fraction 17; 8 = fraction 20; 9 = fraction 23; 10 = fraction 26; 11 = fraction 29; 12 and 13 = pooled fractions (final product). The ~30 kDa band represents cyclin A.

The observation of one band in the W217F and W217K purification SDS-PAGE results (Fig 4.7, panels D and F, lanes 9 and 12 respectively) shows pure product. The W217I monomer peak was very small and no significant amount of product was visible by SDS-PAGE.

Cyclin A W217I does not appear to express as a monomer in sizeable quantities, but appears only as aggregates. However, the purification of cyclin A W217F and W217K was successful and these mutants may now be used to investigate the function of this tryptophan residue in the recruitment of substrate.

4.2.5 Characterisation of cyclin A W217 mutants

A pulldown was performed to check that the cyclin A mutants are properly folded, and a bandshift assay using pRb as a substrate was employed to check that they are able to activate CDK2 (Figure 4.8) (experiment performed as in Mayo and Mayol, 2009).

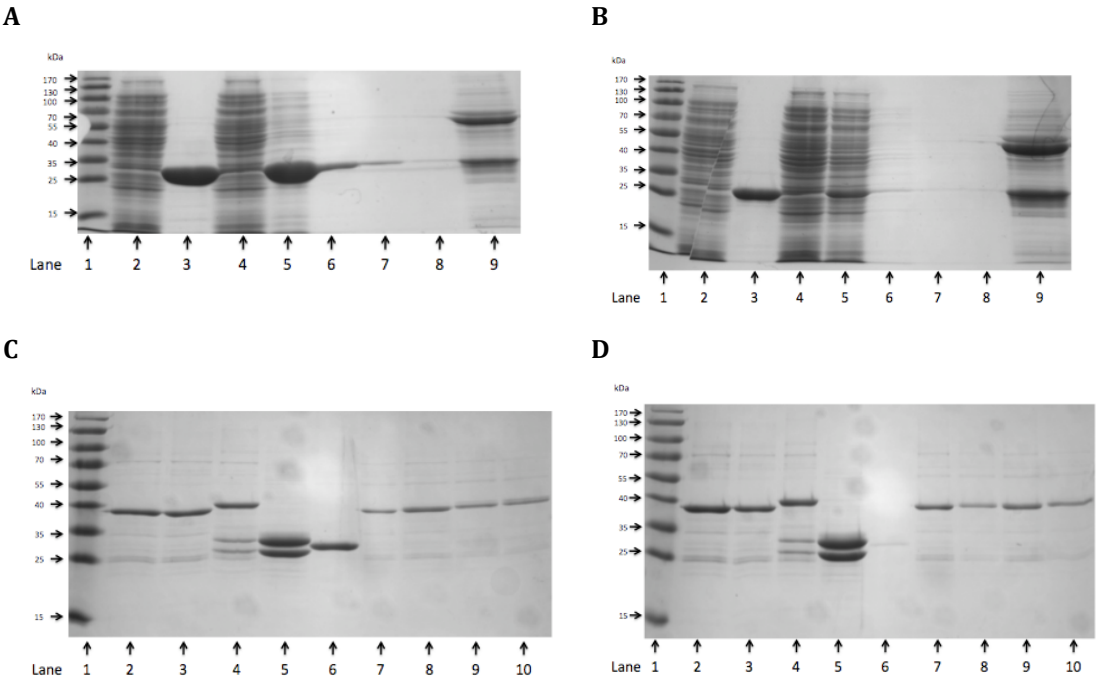


Figure 4.8 Validation of mutated (W217X) cyclin A: pulldowns and bandshifts. **(A)** Small-scale pulldown of W217F cyclin A using GST-CDK2 and glutathione sepharose 4B beads. **(B)** Small-scale pulldown of W217K cyclin A using GST-CDK2 and glutathione sepharose 4B beads. Both: lanes are numbered. Lane 1 = molecular weight markers; 2 = cyclin A in cell lysate; 3 = unmutated cyclin A alone; 4 = pCDK2-GST in cell lysate; 5-8 = washes; 9 = beads with GST-CDK2 and cyclin A bound. **(C)** and **(D)** Bandshift experiments using pRb as a substrate. Upon phosphorylation, pRb changes its electrophoretic mobility markedly and this can be observed on the gel. Lanes are numbered: Lane 1 = molecular weight markers; 2-3 = pRb alone (unphosphorylated); 4 = 60 minute reaction with validated (unmutated) CDK2/cyclin A; 5 = unmutated pCDK2/cyclin A complex; 6 = mutated (W217X) cyclin A alone; 7 = 0 minute timepoint with W217X pCDK2/cyclin A complex; 8 = 20 minute timepoint; 9 = 40 minute timepoint; 10 = 60 minute timepoint. Bands on the gel represent: (gels A and B): ~30 kDa band = cyclin A, ~55 kDa band = GST-CDK2; (gels C and D): ~40 kDa band = GST-pRb, ~30 kDa band = cyclin A, ~35 kDa band = pCDK2.

The two bands shown in lane 9, panels A and B are GST-CDK2 (upper band) and untagged W217F/K cyclin A (lower band). The presence of two bands illustrates that the mutated cyclin A is able to bind CDK2 and is therefore correctly folded.

The bandshift experiments showed that mutation of W217 markedly ablates the ability of cyclin A to activate CDK2 towards pRb as a substrate. The reaction proceeds slowly and after an hour is not yet complete; compared to the unmutated pCDK2/cyclin A complex, for which the reaction goes to completion in less than 20 minutes.

The CDK2/cyclin A W217 mutants were also tested using the fluorescence polarisation assay to identify the effect of the mutation of this residue on recruitment peptide binding. The assay format described in Appendix A6 was used, with a range of concentrations of fluorescent peptide and protein tested. The wild-type and mutated CDK2/cyclin A were titrated at concentrations up to 300nM, into a sample of 20 nM fluorescent peptide 5-carboxyfluorescein-RRLFGE (Severn Biotech). Using this assay, no significant binding of the recruitment peptide was detected to CDK2/cyclin A (Figure 4.9).

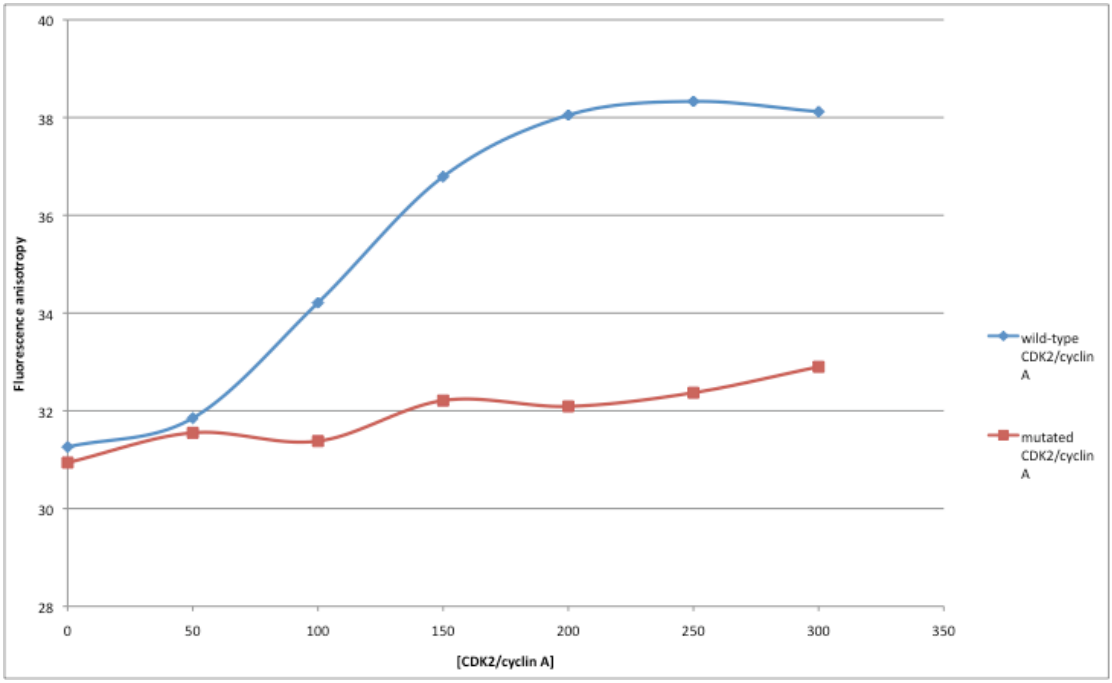


Figure 4.9 Mutation of W217F in the cyclin A recruitment site ablates ability to bind recruitment peptide Red line: results with mutated (W217F) cyclin A. Blue line: results with wild-type cyclin A. CDK2/cyclin A was titrated in and the fluorescence anisotropy measured, reflecting whether the recruitment peptide was bound to CDK2/cyclin A. A fluorescence anisotropy reading of 31 represents completely free recruitment peptide and a reading of 38 represents 100% bound recruitment peptide.

Since the Trp217 mutation appeared to have such a profound effect on the binding of recruitment peptide, this precluded the further use of these mutants in binding assays, as originally intended, since this would almost certainly prevent the CDK2/cyclin A complex having enough enzyme activity to effectively evaluate the inhibitor action. However, this result was extremely interesting since this confirms that Trp217 performs a key role in mediating substrate recruitment and therefore suggests that during inhibitor development to target the recruitment site, it would be advantageous to maximise interactions between the inhibitor and this residue.

4.3 Selection of compound libraries

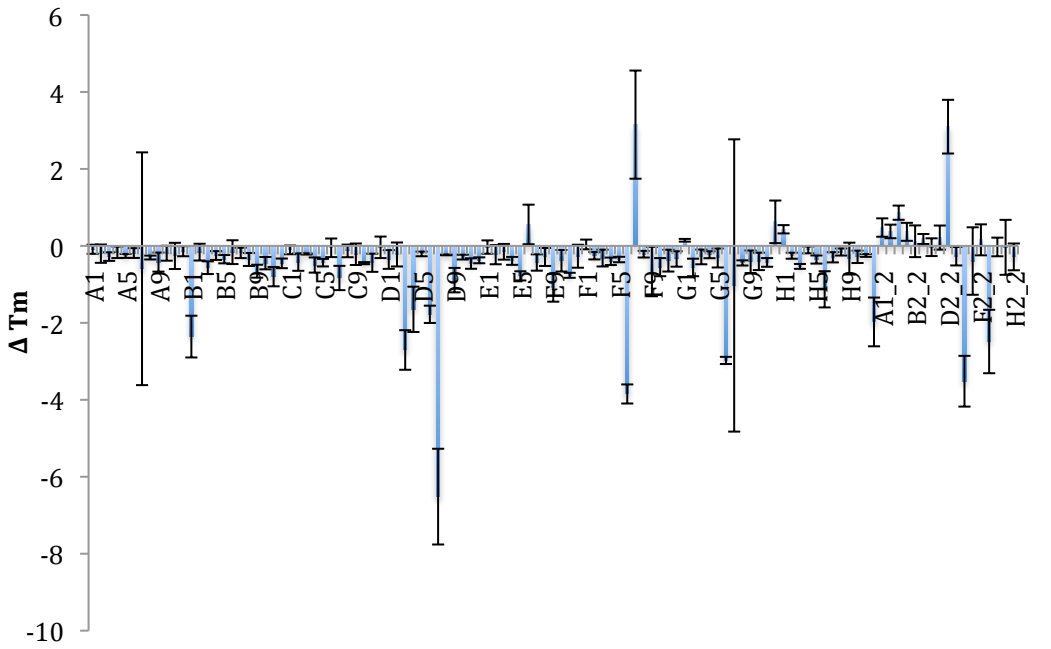
The first compound set is a collection of 116 compounds ordered from ChemDiv that were identified in a virtual screen (M.E.M. Noble and A. Echalier, unpublished results). Virtual screening was carried out with the docking platform FlexX**.** This program keeps the protein rigid and the ligand as flexible. It fragments the ligand, places them and then re-assembles the fragments. It was used here to screen all the compounds available in the ChemDiv catalogue (600,000 compounds): pre-screening with Lipinski's rule of five, with energy minimisation carried out by Corina. Finally the program calculated the top binding pose and accepted this as the solution. The second screening set contains 80 extracts from natural (marine) sources (from the EC Prokinase Consortium, unpublished results).

4.4 Medium throughput screening of libraries: Thermal denaturation assays

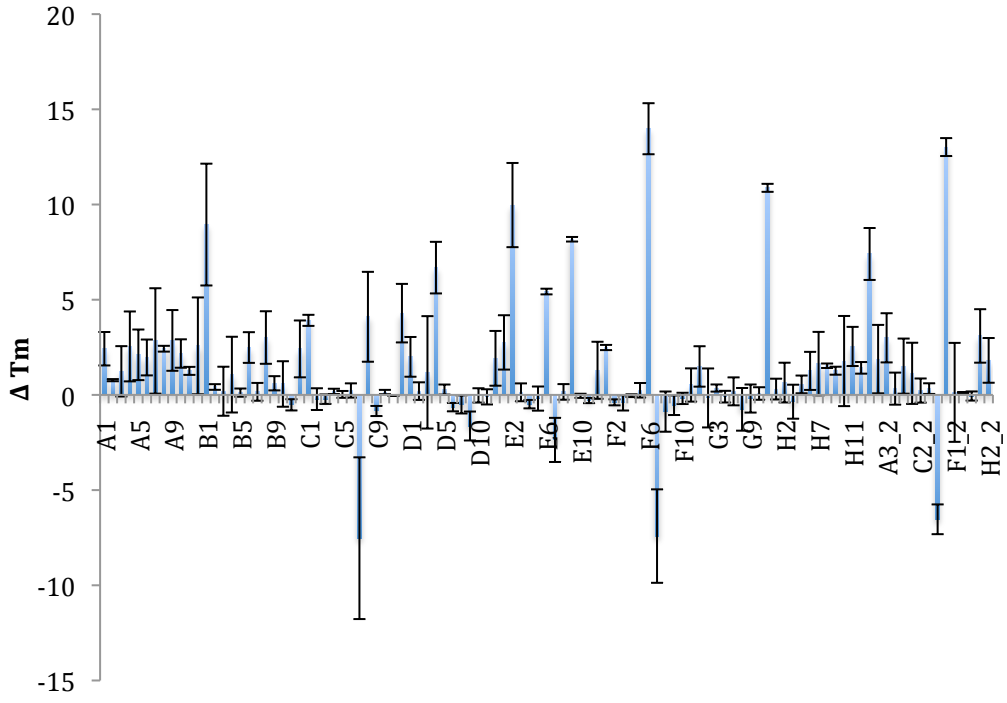
The thermal denaturation assay follows the denaturation of a protein over a temperature range, observing stabilisation of the protein by binding partners. In this assay, a protein is incubated with the dye SYPRO®-Orange and is then gradually denatured by raising the temperature in small steps in an RT-PCR machine. The dye binds to hydrophobic residues that are exposed as the protein becomes denatured and as a result the dye undergoes a shift in its fluorescent spectrum upon excitation at 465 nm (Merckx *et al*, 2008; Senisterra *et al,* 2012). Compound binding can change the protein's stability and this is monitored as a change in the protein's Tm value as compared to a control sample. A potential inhibitor would be expected to stabilise the protein.

In this case the assay was employed to screen compound sets for binding to CDK2/cyclin A (or monomeric cyclin A). The thermal denaturation assay was performed, as detailed in Appendix A7, in triplicate for both compound sets, using both pCDK2/cyclin A and cyclin A. The results of the two thermal denaturation screens are shown in Figure 4.10. A hexapeptide containing the RXL substrate recruitment motif (sequence RRLFGE) was included in each case as a positive control, to allow a comparison with the magnitude of the Tm shift expected for a peptide substrate that binds tightly to the recruitment site (1-5 °C; G. Robertson, unpublished results).

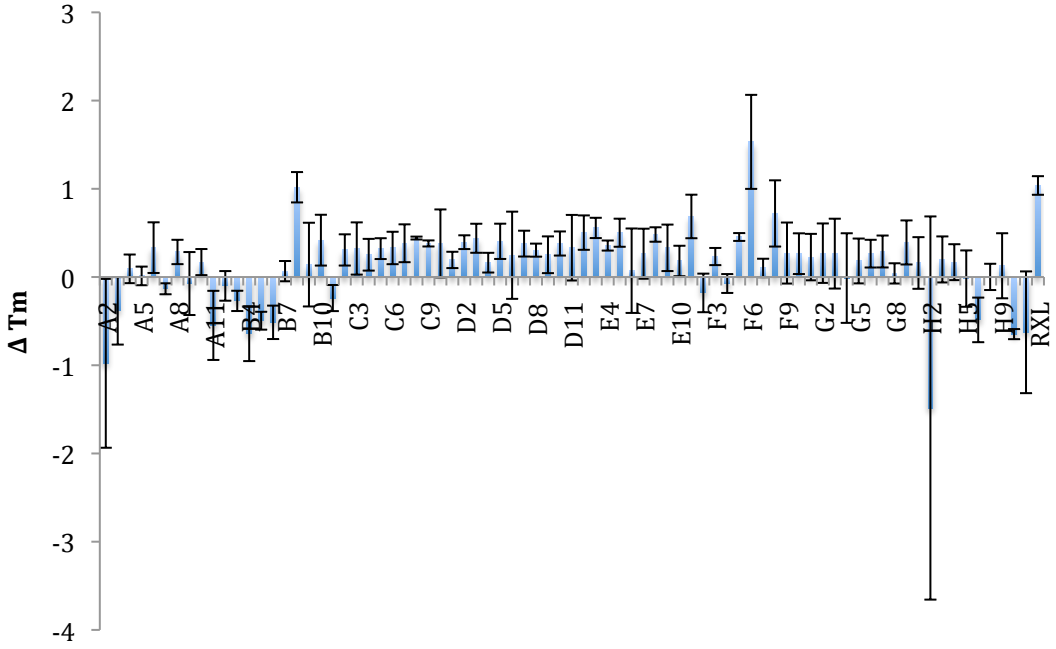
A



B



C



D

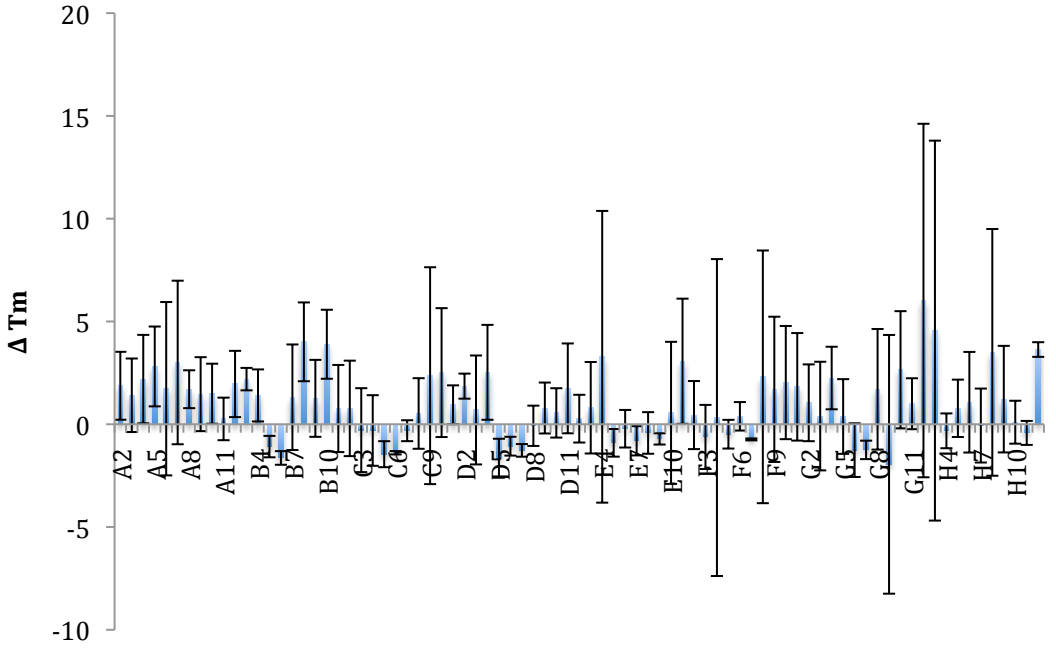


Figure 4.10 Thermal denaturation assay. (A) ChemDiv screen against cyclin A; **(B)** ChemDiv screen, against pCDK2/cyclin A; **(C)** Natural product inhibitors against cyclin A; **(D)** Natural product inhibitors against pCDK2/cyclin A. The figures show the change in melting temperature (ΔTm) on the *x*-axis, upon addition of each compound (along the *y*-axis), compared to the reference cells with DMSO only added.

The assay was somewhat inconclusive due to the large errors observed. The shifts observed were small (1°C at most) and therefore were insignificant compared to the errors in the measurements. In comparison, the RXL hexapeptide produced a 3°C increase in thermal stability (measured against the T160pCDK2/cyclin A complex) with a much smaller error than the average error for the compound set. Many of the compounds produced a negative Tm shift. This result suggests that the inhibitors were destabilising the protein possibly by causing aggregation or denaturation, or by binding non-specifically to the complex. However, some compounds were identified from the screen as of particular interest, namely B8, B10, C11, E11, F6, F8 and F10 from the natural products screen and B1, G6, G11 and D2_2 from the ChemDiv screen (the structures of these compounds are given in Appendix 3). To verify the activity of these compounds, the libraries were rescreened using alternative assays: a gelshift assay (measuring substrate phosphorylation), radioactive kinase assay (measuring ^{32}P incorporation) and a fluorescence polarisation assay (measuring binding by displacement of a fluorescent peptide) (see Sections 4.5-4.6).

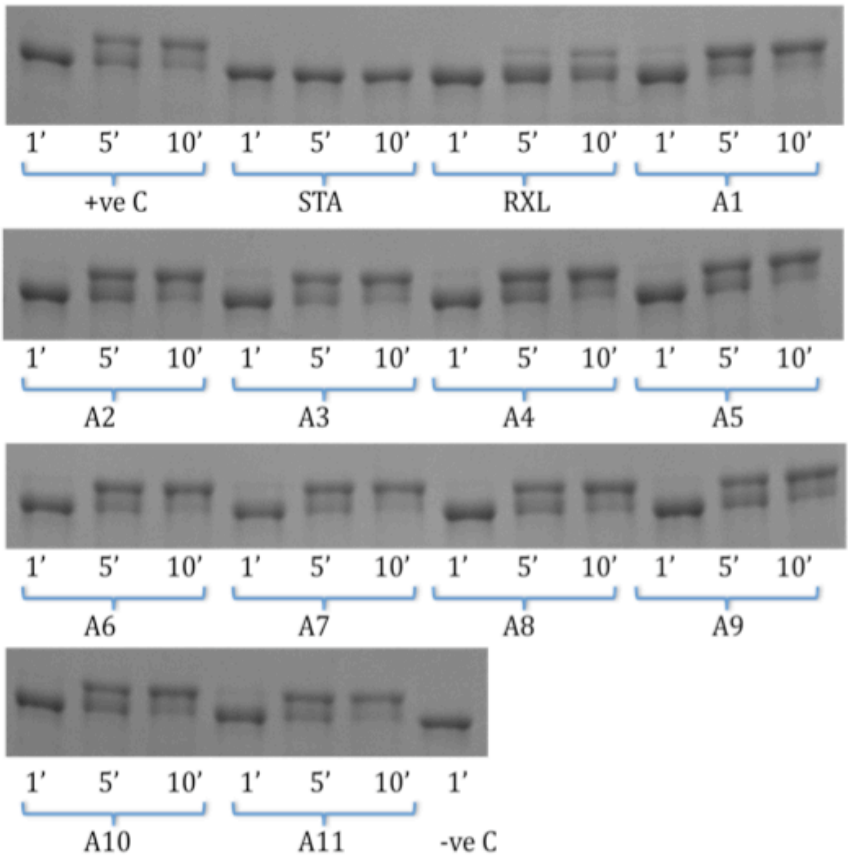
4.5 Screening of ChemDiv and natural product inhibitors by bandshift experiments

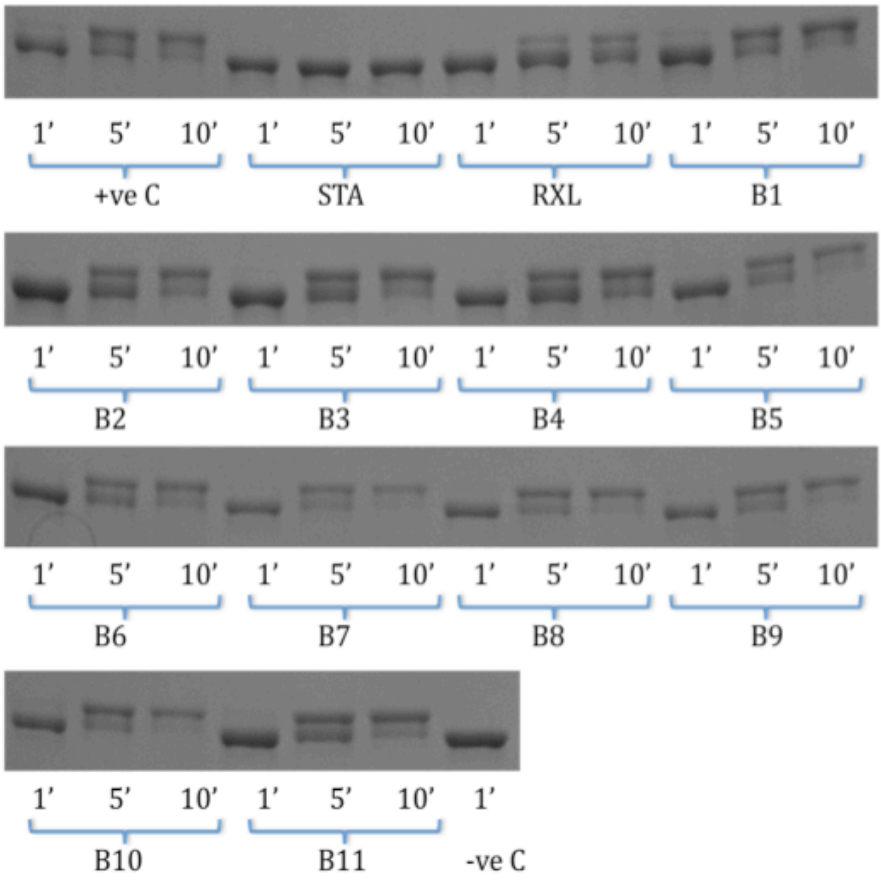
This assay looks at the progress of the phosphorylation reaction of pRb-GST. When hyperphosphorylated, pRb-GST changes electrophoretic mobility and in this way the compounds can be screened for their effect on the activity of CDK2/cyclin A. The controls used were a negative control (no CDK2/cyclin A), a positive control (with DMSO instead of compounds) and inhibitor controls (with Staurosporine or RXL peptide instead of compounds). Reactions were carried out at room temperature. The method is as follows: first 10mM stocks of compounds to be tested were made up, as well as aliquots of 10x kinase buffer (5mM ATP, $MgCl_2$, 0.5M Tris-HCl pH 7.5), 200nM CDK2/cyclin A, and 3.5mg/ml pRb-GST for use each day. Into each tube I pipetted 6μl of the compound stock (or RXL, Staurosporine or DMSO control stock), 9μl pRb-GST stock, 6μl 10x kinase buffer and 33μl ddH_2O. The reaction was started by adding 6μl CDK2/cyclin A stock. The reactions were stopped after each timepoint by removing 20μl and adding to

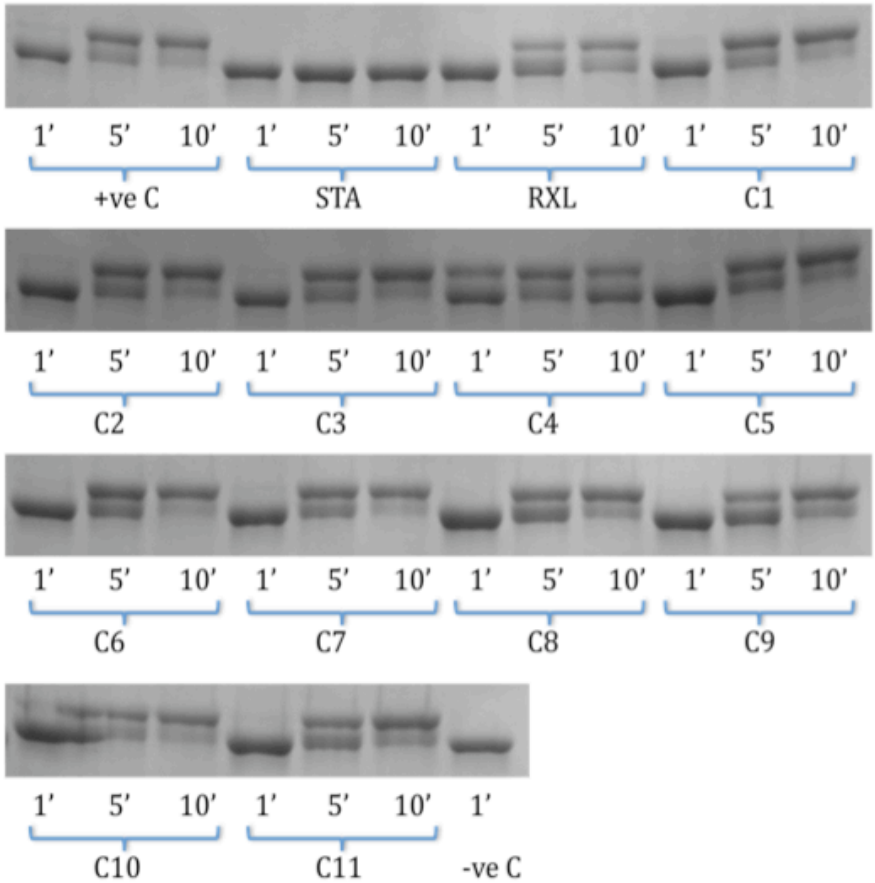
SDS loading dye after each timepoint, and the samples were analysed by SDS-PAGE.

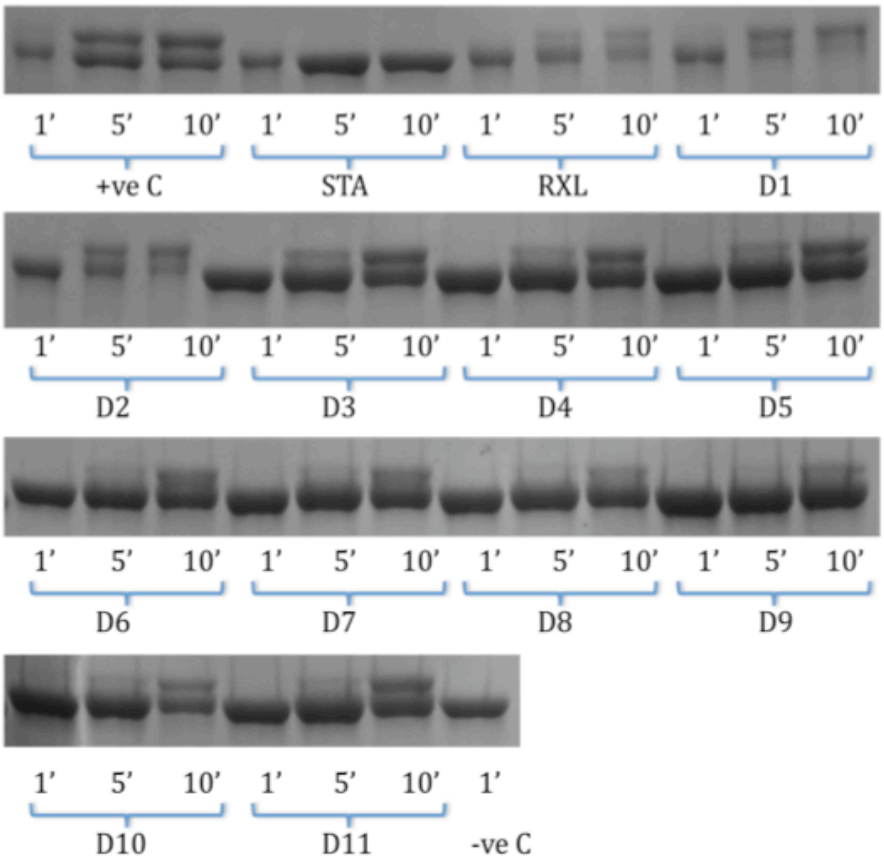
The results for the natural product inhibitor set are shown in Figure 4.11:

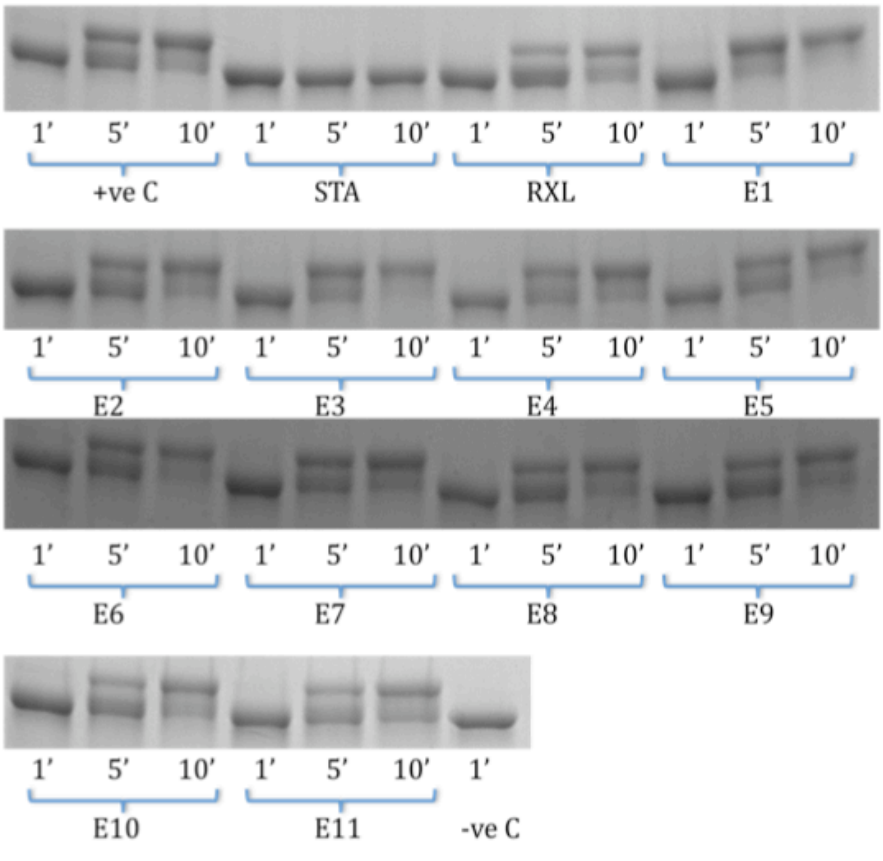
Figure 4.11 Screening of natural product inhibitors by bandshift assay The lower band on the gel represents unphosphorylated pRb-GST and the upper band represents hyperphosphorylated pRb-GST. The timepoints (1, 5 and 10 minutes) and the identity of the compound or control inhibitor added are shown below the bands. RXL: RRLFGE recruitment peptide; STA: staurosporine; +ve C: positive control (DMSO instead of inhibitor) and –ve C: negative control (no CDK2/cyclin A added)). *Figure continues overleaf, over pages 161-166.*

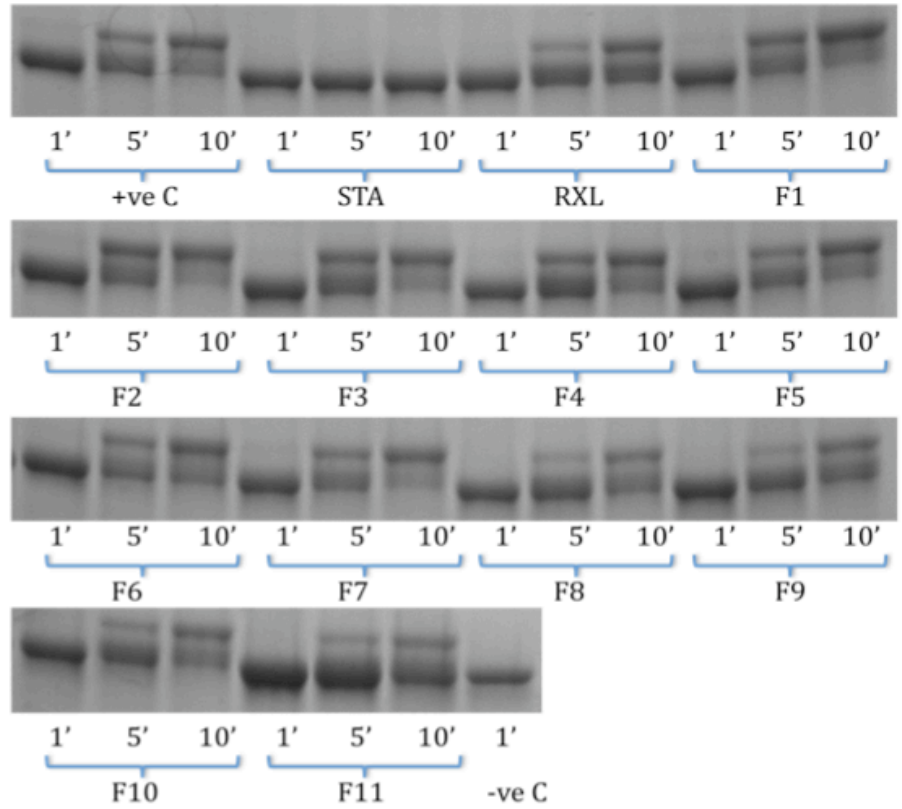


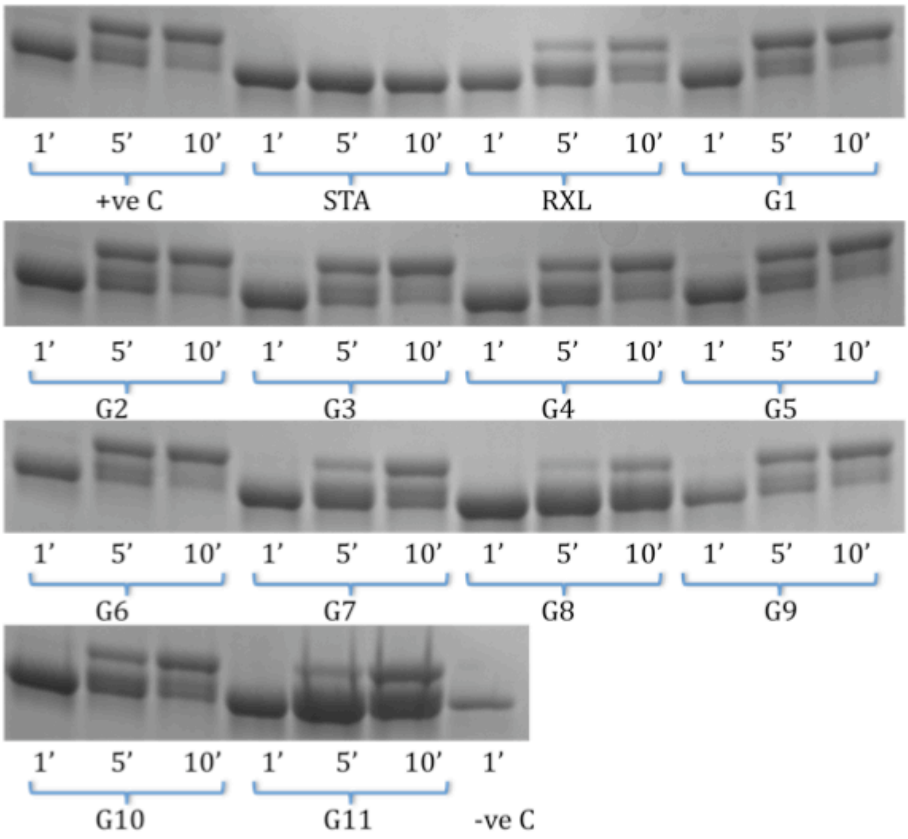

1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
+ve C
STA
RXL
B1
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
B2
B3
B4
B5
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
B6
B7
B8
B9
1' 5' 10' 1' 5' 10' 1'
B10
B11
-ve C

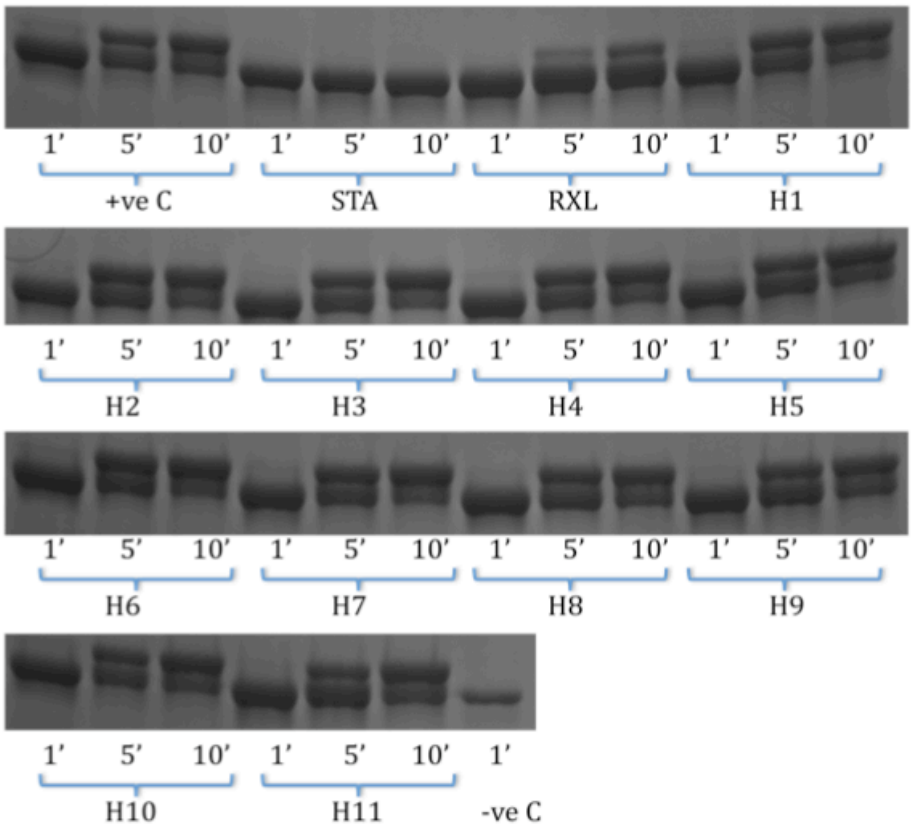

1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
+ve C
STA
RXL
C1
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
C2
C3
C4
C5
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
C6
C7
C8
C9
1' 5' 10' 1' 5' 10' 1'
C10
C11
-ve C


1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
+ve C
STA
RXL
D1
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
D2
D3
D4
D5
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
D6
D7
D8
D9
1' 5' 10' 1' 5' 10' 1'
D10
D11
-ve C


1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
+ve C
STA
RXL
E1
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
E2
E3
E4
E5
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
E6
E7
E8
E9
1' 5' 10' 1' 5' 10' 1'
E10
E11
-ve C


1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
+ve C STA RXL F1
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
F2 F3 F4 F5
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
F6 F7 F8 F9
1' 5' 10' 1' 5' 10' 1'
F10 F11 -ve C


1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
+ve C STA RXL G1
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
G2 G3 G4 G5
1' 5' 10' 1' 5' 10' 1' 5' 10' 1' 5' 10'
G6 G7 G8 G9
1' 5' 10' 1' 5' 10' 1'
G10 G11 -ve C



Compounds showing some activity in this assay were B4, D3-D11, F11, G7, G8, G11, H9 and H11. These were promising hits and should be investigated further using the fluorescence polarisation assay. However, none of these showed activity approaching that of the RXL peptide.

The results of the bandshift assay against the ChemDiv compound set showed activity against CDK2/cyclin A phosphorylation of pRb for compounds D8, G6, E1.2, E2.2 and E3.2 (results shown in Figure 4.12).

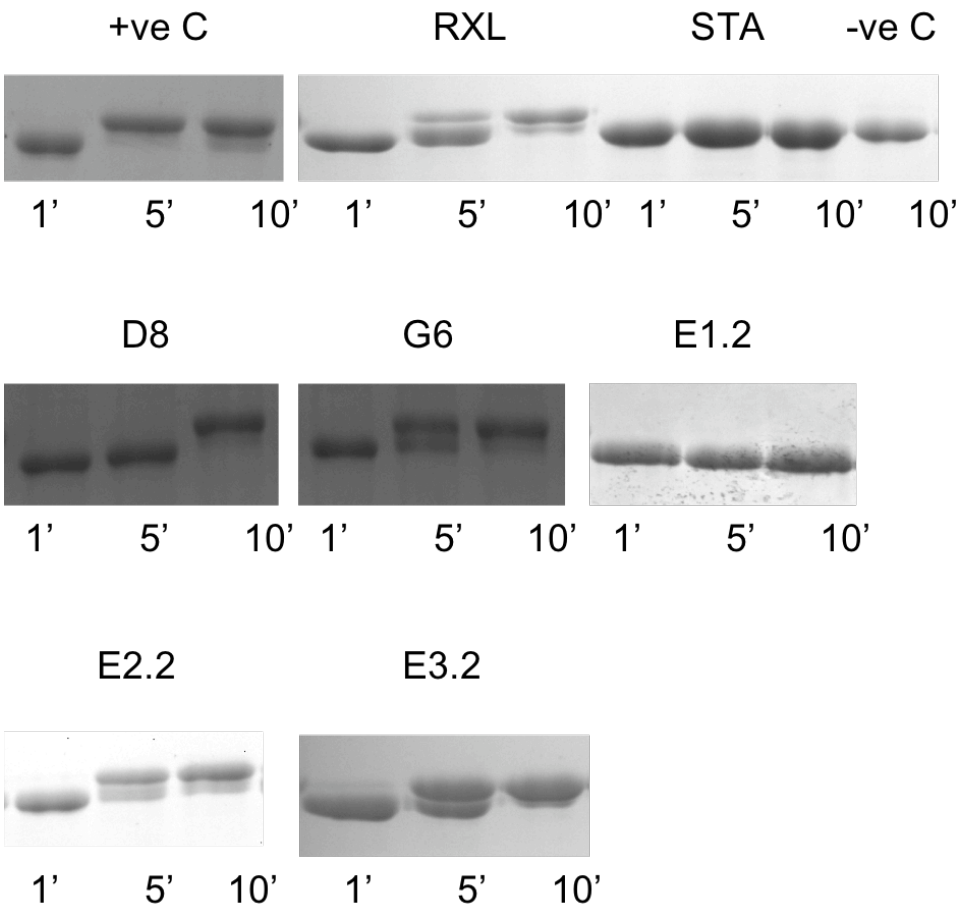


Figure 4.12 Bandshift assay: hits from ChemDiv library screening Results are shown for compounds where the inhibition seen was significant, compared to a positive control (with DMSO), inhibitor controls with RXL peptide (RRLFGE), or staurosporine, and a negative control (no CDK2/cyclin A added). The reaction was stopped after 1, 5 and 10 minutes and aliquots were run on SDS-PAGE. The lower band represents unphosphorylated GST-pRb and the upper band represents hyperphosphorylated GST-pRb.

The ChemDiv set of inhibitors was then tested again using a radioactive kinase assay.

4.6 Radioactive kinase assays: screening compound sets for binding to CDK2/cyclin A

This assay measures the amount of phosphorylation of p107 by CDK2/cyclin A, when γ-^{32}P-ATP is provided as a substrate, and has been used to screen compound libraries quickly to identify inhibitors of CDK2/cyclin A, as well as to obtain kinetic information for ‘hit’ compounds. The assay was carefully optimised in order to allow it to be used successfully for screening the two

compound sets. When the reproducibility was judged to be satisfactory, the substrate concentrations in the assay were then optimised. Using these conditions the calculated K_m (p107) was 15 μM and the T160pCDK2/cyclin A V_{max} was 10 nmoles ^{32}P incorporated/mg/minute.

Radioactive kinase assays were performed using the P81 paper-based method essentially as described in Hastie *et al,* with minor modifications (Hastie *et al,* 2006). The radioactive kinase assay was performed by setting up reactions with 4nM CDK2/cyclin A, 15μM p107 (or other substrate), 100μM inhibitors (if applicable) and kinase buffer (containing 0.5M $MgCl_2$, 0.5M Tris pH 7.0) and initiating the reaction by adding a mix of ^{32}P-γ-ATP (5μCi) and cold ATP (200μM). Reactions were incubated at 22^oC, a timecourse was performed and at each timepoint 10μl of the reaction mixture was spotted onto the centre of a circular P81 paper, which was then dropped into a solution of 1% phosphoric acid to stop the reaction. The papers were washed three times with water to remove excess unreacted γ-^{32}P-ATP, and the results were imaged by exposing the papers on a blanked image cassette overnight and developing using a Storm 860 imaging system. Phosphoincorporation (in picomoles) was quantified by analysing the image from the P81 papers, using the program Quantity One (from BioRad), with reference to an ATP dilution control, after subtraction of the background radioactivity levels. More detail about the method as well as an example of the format of the results for this assay, is given in Appendix A5.

The response of the assay to a selected set of known CDK2/cyclin A inhibitors was determined, to clarify what magnitude of change in reaction rate would be expected for inhibitors with a range of IC_{50} values (Figure 4.13). By using these controls an approximate IC_{50} value can be estimated for any 'hits' during screening.

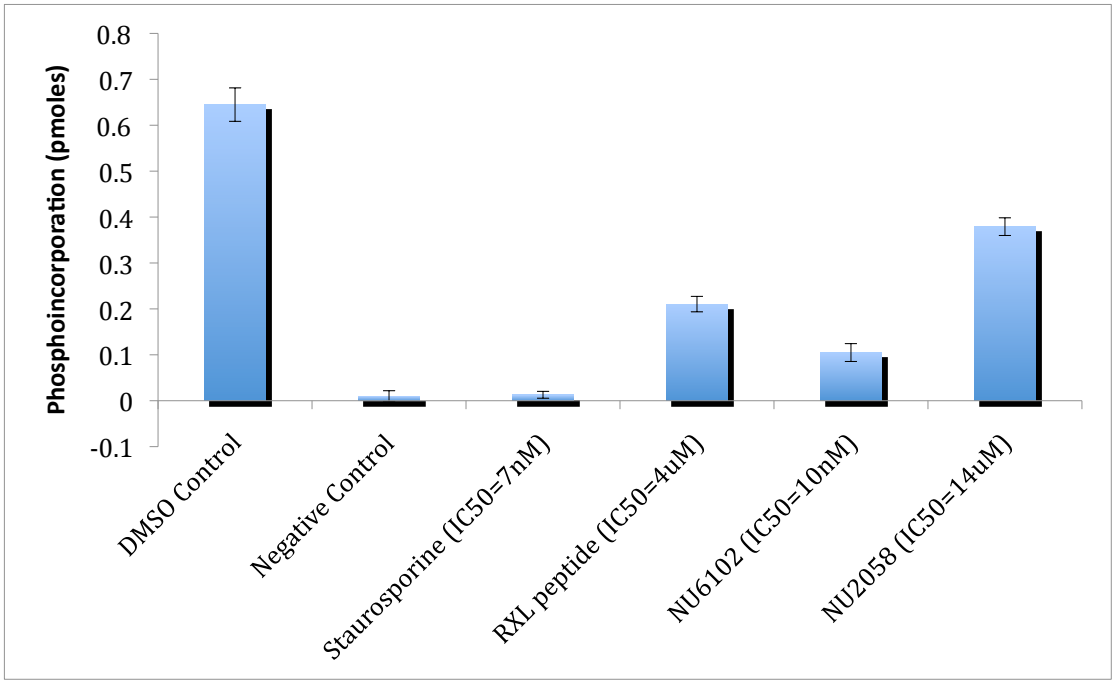
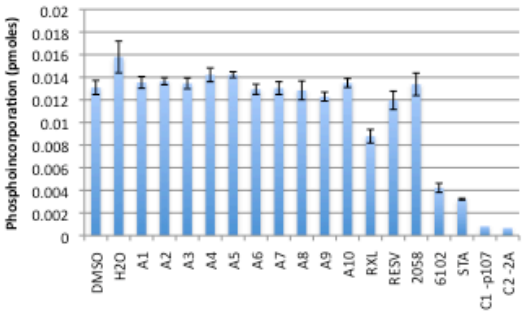


Figure 4.13 Radioactive kinase assay calibration. Results of the radioactive kinase assay for reactions with different inhibitors added at 100μM (and all other reactants at concentrations as previously stated). The reaction was allowed to proceed for 3 minutes, with 15μM p107 and 4nM pCDK2/cyclin A.

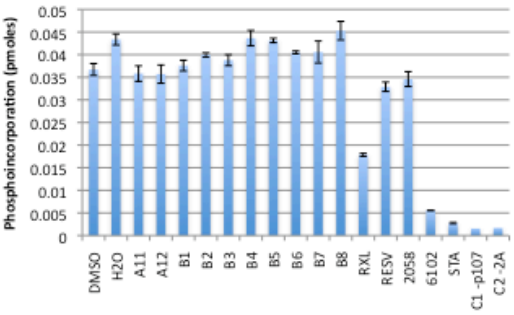
Since the signals observed for known inhibitors were significant, and the assay gave reproducible results for these controls, it was judged to be sufficiently optimised to allow medium throughput screening to identify 'hit' compounds and generate estimates of IC_{50} values.

The results of the screening of the ChemDiv inhibitor set (identified from virtual screening) using the radioactive kinase assay are shown in Figure 4.14:

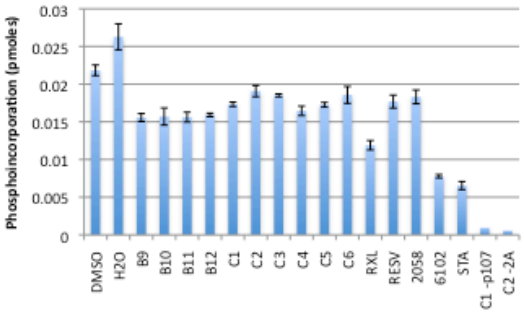
Radioactive kinase assay: ChemDiv screen A1-A10



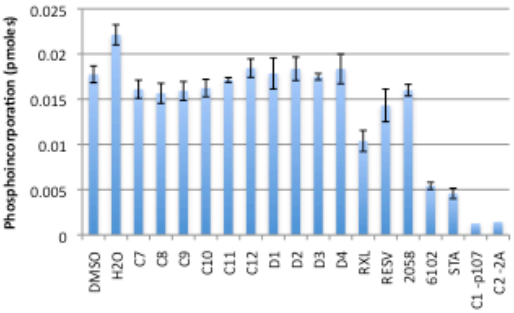
Radioactive kinase assay: Chemdiv screen A11-B8



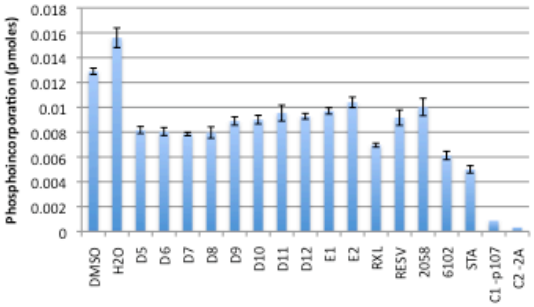
Radioactive kinase assay: ChemDiv screen B9-C6



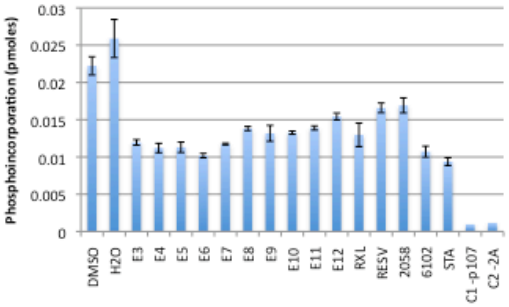
Radioactive kinase assay: ChemDiv screen C7-D4



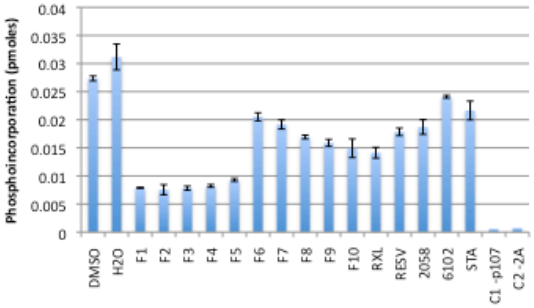
Radioactive kinase assay: ChemDiv screen D5-E2



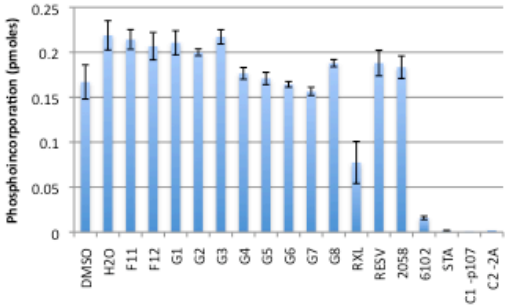
Radioactive kinase assay: ChemDiv screen E3-E12



Radioactive kinase assay: ChemDiv screen F1-F10



Radioactive kinase assay: ChemDiv screen F11-G8



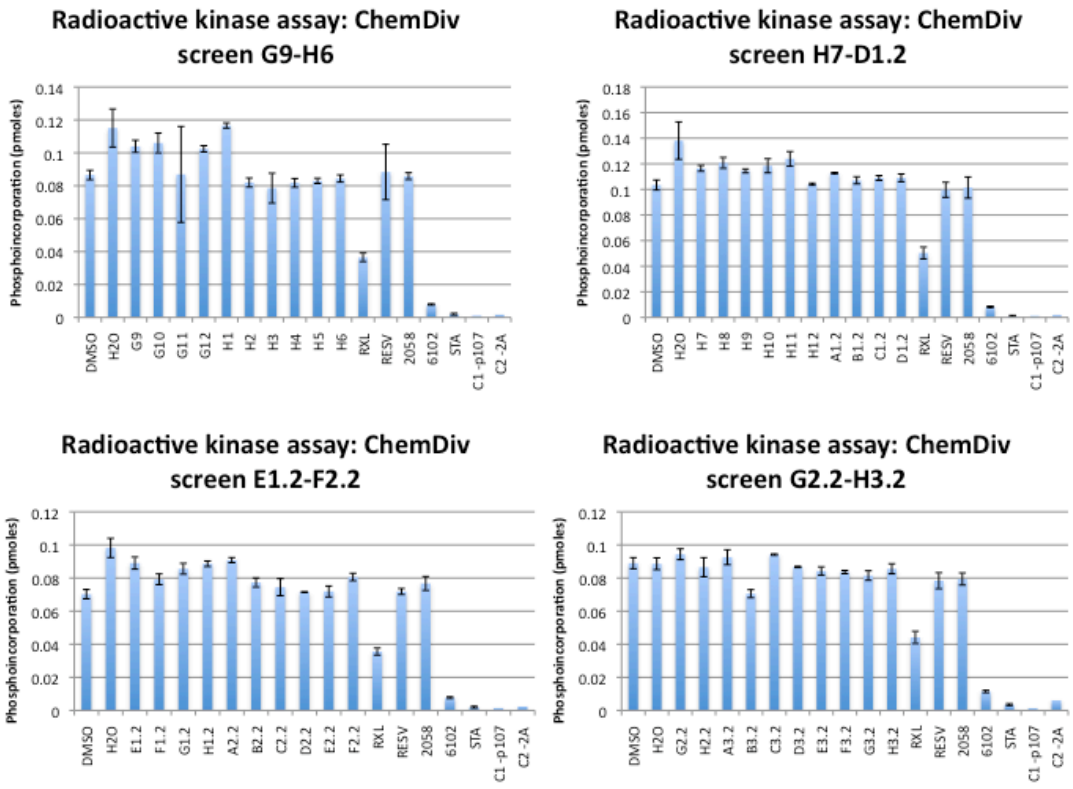


Figure 4.14 Results of the radioactive kinase assay The assay was carried out in triplicate. Error bars shown are standard deviation. *Figure continues over pages 170-171.*

The hits from this assay were: B10, E6, F10 and B3.2. F1-5 also appeared initially to show activity against CDK2/cyclin A phosphorylation of p107-GST but further testing eliminated these compounds from the investigation. The hit compounds were validated using further radioactive kinase experiments and fluorescence polarisation experiments.

4.7 Summary of screening results

The results of screening from the thermal denaturation, bandshift and radioactive kinase assays, combined with previous results from SPR (by Aude Echalier), are shown in Figure 4.15.

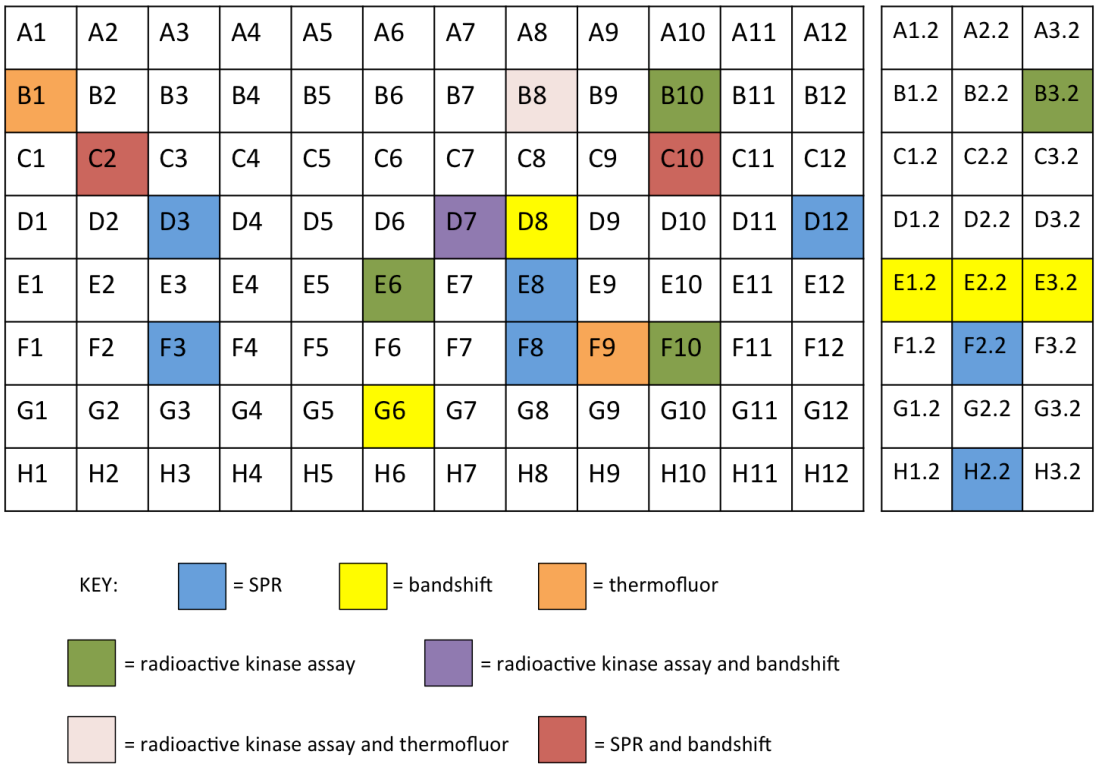


Figure 4.15 Summary of results from screening ChemDiv inhibitor set with different techniques: positive results from each of the techniques used are highlighted in the relevant colours as shown in the key.

The difference in results from each assay reflects the difference in outcomes being measured: SPR and thermal denaturation measure binding to CDK2/cyclin A, whereas the bandshift and radioactive kinase assays measure effect on CDK2/cyclin A activity against recruited and non-recruited substrates.

4.8 Validation of hits from ChemDiv screening

Further kinetic experiments on these 'hit' compounds were then performed to check the mode of action of each one, to eliminate those which act non-specifically or at the ATP binding site in CDK2, and to identify those that bind at the cyclin A recruitment site. The mode of action can be elucidated by comparing the effect of the inhibitor on the phosphorylation of recruited substrates (such as p107) compared to non-recruited substrates (such as histone H1), as shown in Figure 4.16:

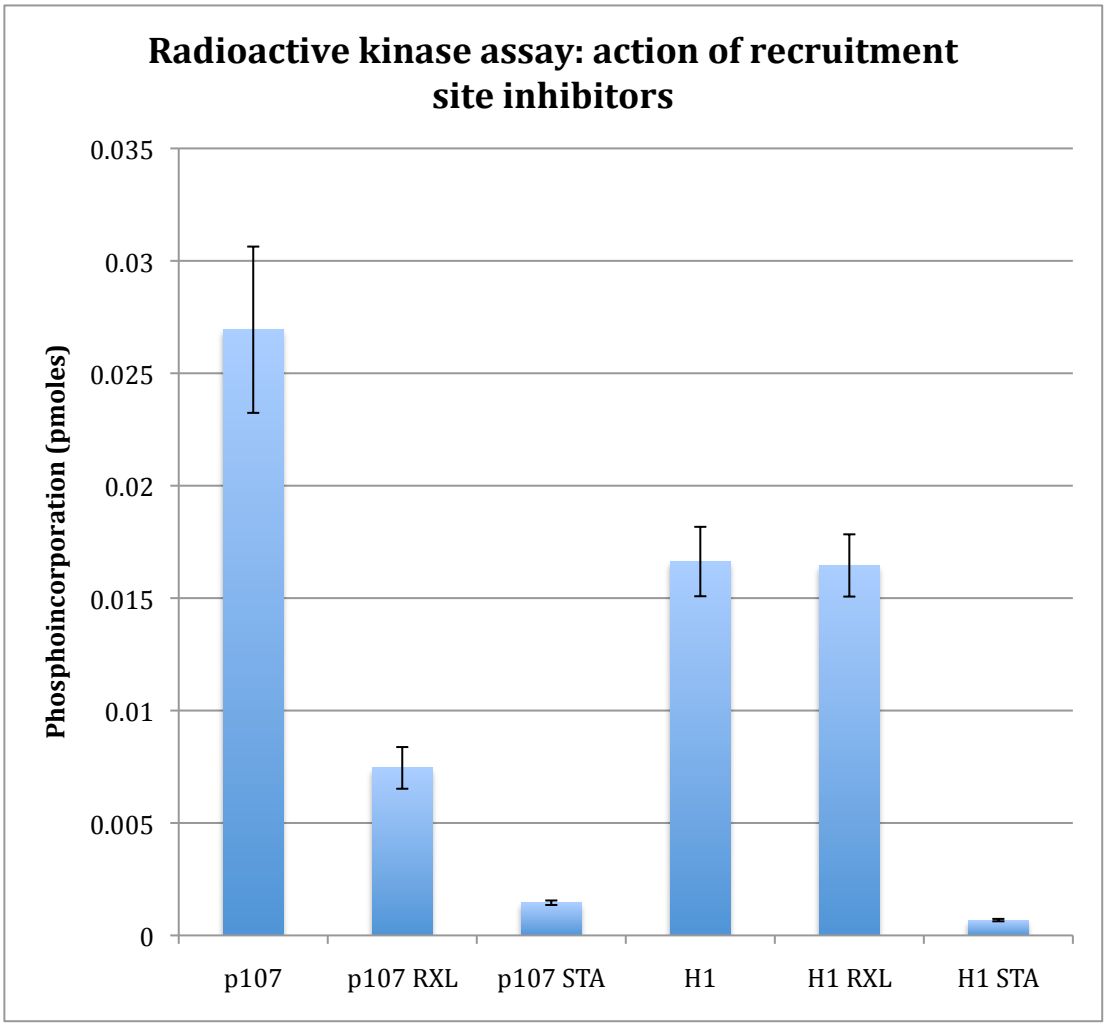


Figure 4.16 The radioactive kinase assay shows that recruitment site inhibitors affect phosphorylation of p107 but not histone H1. The radioactive kinase assay was used to check whether an RXL recruitment peptide (RRLFGE) affected the phosphorylation of a recruited substrate (p107) compared to a non-recruited substrate (histone H1); staurosporine (STA) was used as a control, since this ATP-competitive inhibitor affects the phosphorylation of recruited and non-recruited substrates equally.

After optimisation of this assay using the two different substrates, it was used to re-test and validate that all the hit compounds did bind at the recruitment site (Figure 4.17).

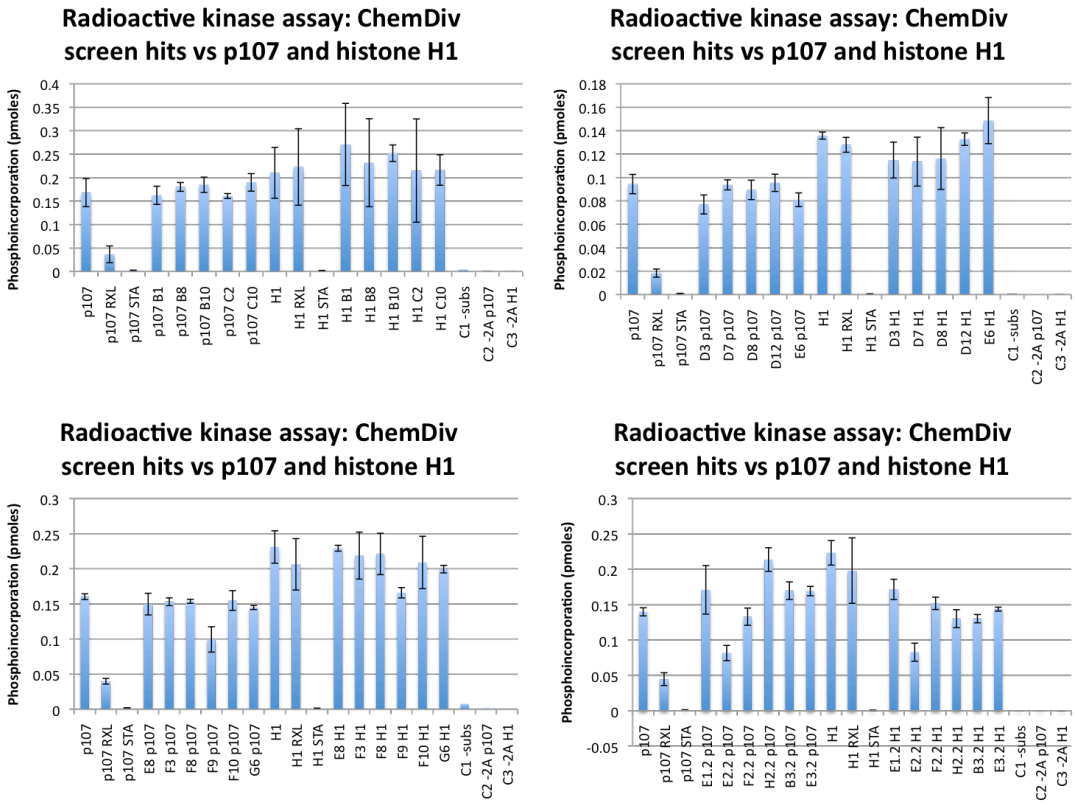


Figure 4.17 Validation of hits: the compounds identified as hits from the thermal denaturation, gelshift or radioactive kinase assays were re-screened using the radioactive assay with p107 as a substrate, as well as histone H1 (a non-recruited substrate) for comparison.

The results of the validation assay showed that most leads previously established showed activity towards histone H1 as well as p107, (for example, F9) and must therefore act non-specifically, or at a different site. However, compounds C2 and E6 showed mild inhibition of p107 but no inhibition of histone H1. These were therefore re-tested using a fluorescence polarisation assay to probe the mode of action.

4.9 Fluorescence polarisation assay: checking for displacement of a recruitment peptide

A fluorescence polarisation assay measuring displacement of a fluorescently-tagged recruitment peptide bound to the CDK2/cyclin A target by putative PPI inhibitors was employed to confirm the mode of action of the hit compounds and to obtain accurate IC_{50} values for these compounds. This is the same assay employed in Chapter 3 to screen Plk1 PBD binding inhibitors and the same method was used (see Appendix A6 for details of the method). Within the ChemDiv compound set selected by virtual screening, there are approximately 20 sulphonamides closely related structurally to E6. Of these five compounds have a 2,4-dimethylbenzyl group at the sulphonamide end of the molecule (henceforth called A) and variable chemical groups at the amide end of the molecule (henceforth called B). These compounds are shown in Figure 4.18.

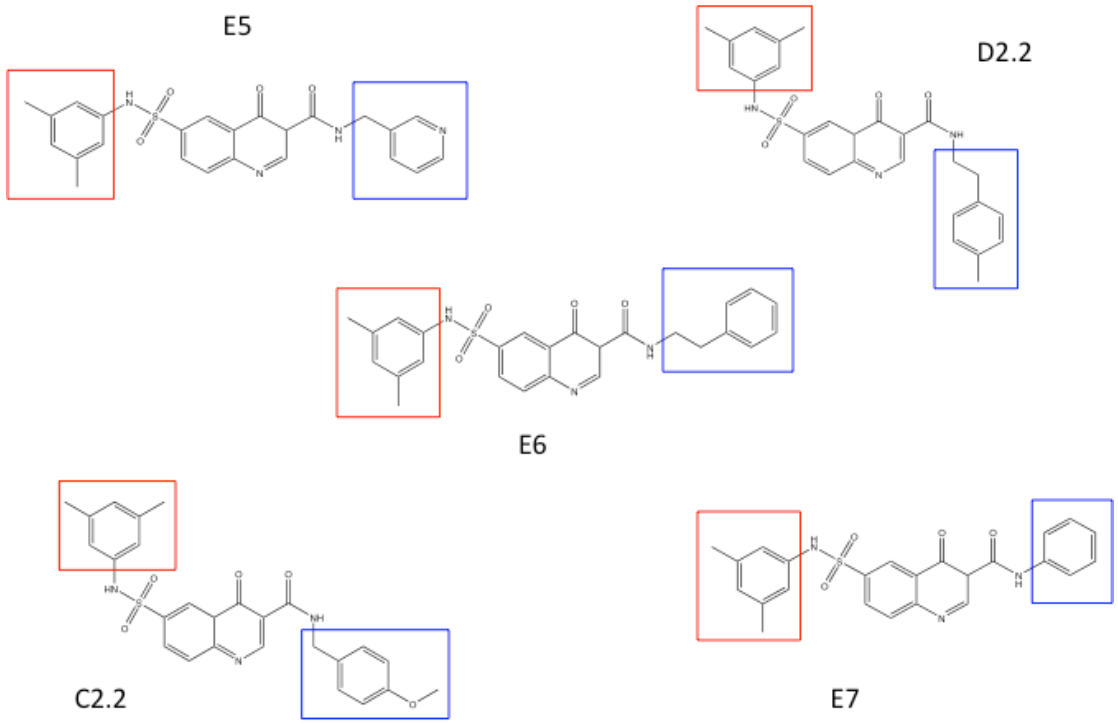


Figure 4.18 E6 and structurally related compounds. Illustrates structures of compounds chosen for use in the fluorescence anisotropy assay (to calculate IC_{50} values and determine SAR between compounds). These were chosen because they are closely related structurally to E6. The 2,4-dimethylbenzyl group is at the sulphonamide end of the molecule (A, highlighted by a red box) and various different chemical groups are at the amide end of the molecule (B, highlighted by a blue box).

The fluorescence anisotropy assay was used to calculate IC_{50} values and determine SAR between these compounds. The results of this assay are shown in Figure 4.19.

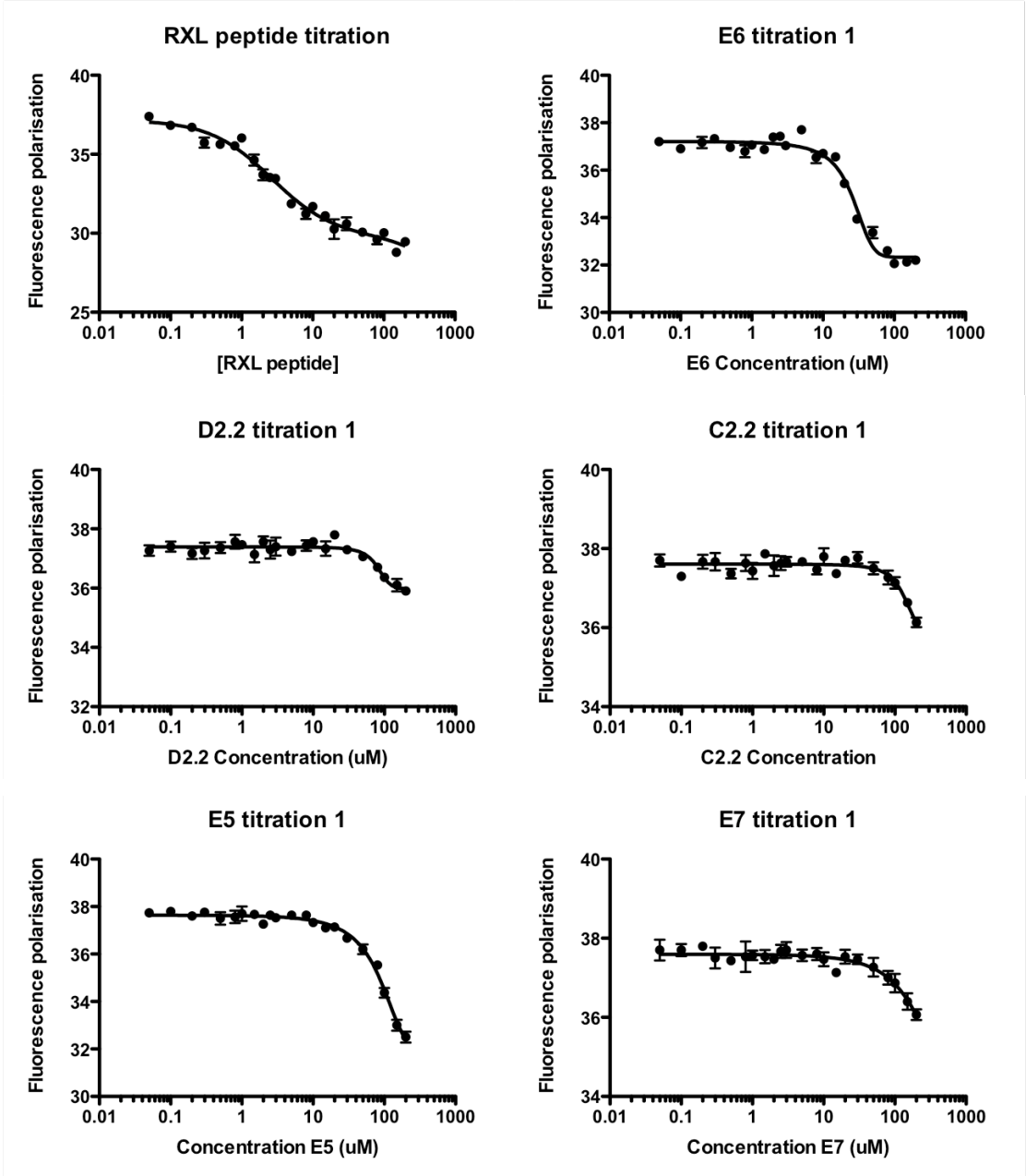


Figure 4.19 Fluorescence polarisation: displacement of a labelled recruitment peptide by RXL (unlabelled recruitment peptide), compound E6 and structurally related compounds. IC_{50} values: RXL: 2.62±0.744μM; E6: 24.9±3.4μM; D2.2: 80.5±17.5μM; C2.2: 135.9±60μM; E5: 72.6±17.5μM; E7: 114±77μM Compound C2 did not show displacement of RXL peptide from the recruitment site within the titration up to 200μM. The assay was carried out using 5-carboxyfluorescein-RRLFGE (from Severn Biotech) at 20nM, CDK2/cyclin A at 130nM and potential recruitment site-binding compounds at a range of concentrations from 0-200μM. The assay was performed in black 384-well plates after 30 minutes pre-incubation at 20°C, in a BMG Pherastar FS microplate reader and inhibition curves were fitted using Prism.

A comparison between the results of the fluorescence anisotropy assay and the structures of the compounds is shown in Table 4.1

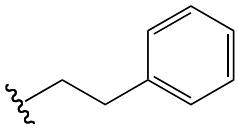
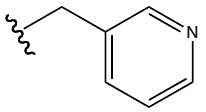
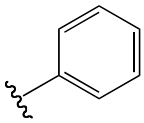
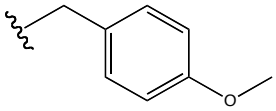
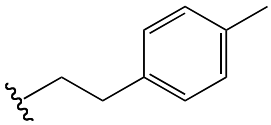
Compound number	Chemical group at B	IC_{50} vs CDK2/cyclin A
E6		24.9±3.4μM
E5		72.6±17.5μM
E7		114±77μM
C2.2		135.9±60μM
D2.2		80.5±17.5μM

Table 4.1 Results of fluorescence anisotropy: effect of changing chemical group at B

Compound C2 however, showed no displacement of the labelled peptide. Analysis of the SAR of the reactions, following these experiments, shows that longer side chains are not well tolerated within the binding site. Shortening the link at B also decreases the potency of the compounds. This suggests that exploring links of the same length but with different polar or non-polar groups added on the ring might improve potency for these compounds. However, it is difficult to interpret this because of the large errors on some of the IC_{50} values due to the fact that the curves achieved are not perfectly sigmoidal, since 100% displacement was not achieved up to 200μM.

4.10 Co-crystallisation of CDK2/cyclin A with hit compounds

Crystal trials were performed in 96-well plates (SwissSci, 3 subwells) using commercial broad screens, including Wizard, PEG/Ion, PACT, Morpheus, JCSG+, MD1+2 and Midas. These were set up using a Tecan robot to dispense well solution, and a Mosquito robot to dispense protein and mix with well solution. In addition custom screens around the previously identified crystal conditions for CDK2/cyclin A ('closed recruitment site': 0.8M KCl, 1.1M $(NH_4)_2SO_4$, 0.1M Hepes pH 7; 'open recruitment site': 1M Li_2SO_4, 0.1M Hepes pH 7) were performed. Some of the crystals obtained using CDK2/cyclin A bound to compounds C2 and E6, hits from the ChemDiv screen, are shown in Figure 4.20.

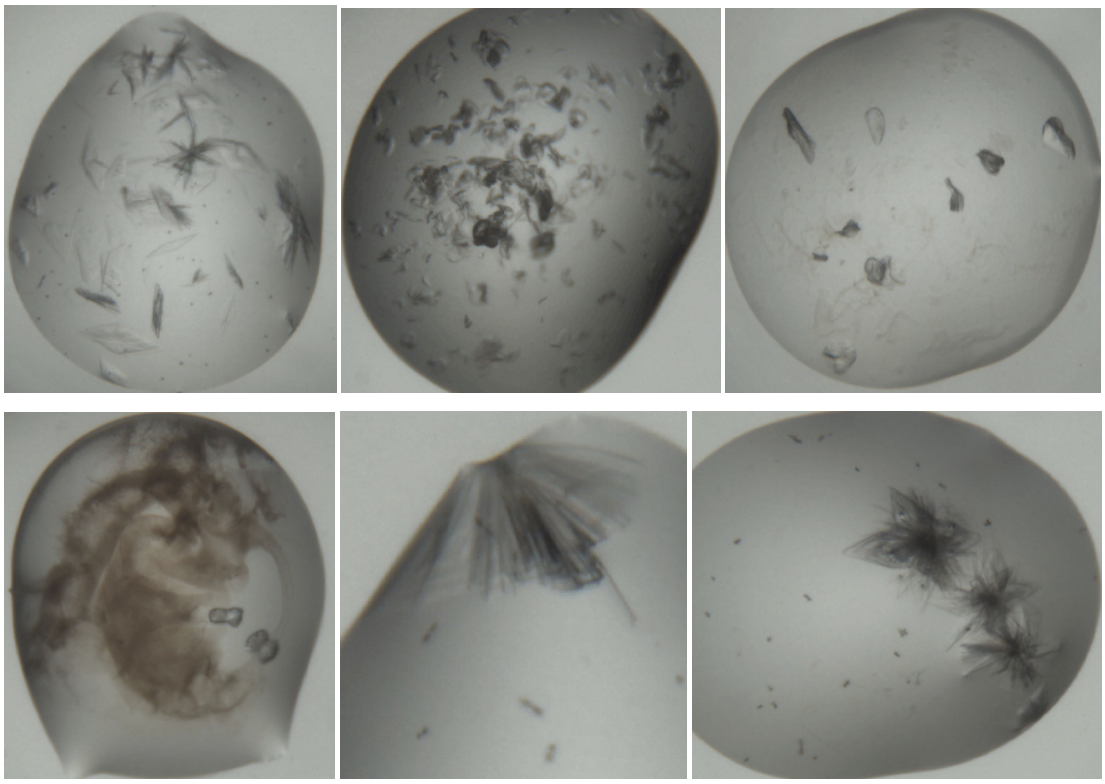


Figure 4.20 Co-crystals of CDK2/cyclin A with hit compounds: top row: grown with C2, bottom row: grown with E6. Conditions (clockwise from top left): 0.1M CHES pH 9.5, 1.0M sodium citrate, final pH 9.7; 0.2M Sodium iodide, 0.1M Bis Tris propane pH 8.5, 20% w/v PEG 3350; 0.09M NPS, 0.1M Tris Bicine pH 8.5, EtGlycol, PEG 8K 30.0%; 0.1M Sodium malonate pH 4.0, 12% w/v PEG 3,350; 0.2M NaCl, 0.1M Imidazole pH 8.0, 1.0M K/Na tartrate, final pH 8.4; 0.2M NaCl, 0.1M Imidazole pH8.0, 0.4M NaH_2PO_4/1.6M K_2HPO_4 Final pH = 7.6.

The crystals were taken to Diamond Light Source for data collection (on beamline I04), where strategy for data collection was determined using Mosflm.

Some of the crystals diffracted, but only poorly and to low resolution. Further attempts at co-crystallisation of CDK2/cyclin A with lead compound E6 or derivatives should focus on optimising these crystals to obtain a crystal form that diffracts to high resolution and allows structure determination.

4.11 Conclusion and further work

The successful identification of a lead compound targeting this PPI from libraries of commercially available compounds, using a combination of virtual screening and biochemical assays to investigate binding and activity, suggests that this is a valid general strategy for identification of PPI inhibitors. My experience conducting several different types of assay on this project indicates that a fluorescence polarisation assay as used here is particularly useful for identification of PPI inhibitors, since it can probe the binding mode and check that a potential inhibitor is able to displace the interacting partner. However, the radioactive kinase assay and thermal denaturation assay were somewhat problematic as they required substantial periods of time for optimisation before they could be used, and even after optimisation the errors remained substantial, limiting the usefulness of these assays. The improvement of signal-to-noise on these assays took a long time and one of the main problems facing this project was the lack of solubility of some of the compounds, which meant that the assays could not be carried out at concentrations higher than 200μM. Since these assays were employed for initial screening purposes and the compound all acted at low potency, this made it extremely difficult to obtain a good enough signal from these compounds. However, the combination of the different techniques allowed effective screening as it circumvented the different drawbacks of each method of screening.

E6 now forms the lead compound for this interaction: with preliminary screening I have demonstrated binding and inhibition of the CDK2/cyclin A complex; follow-up screening using fluorescence anisotropy then proved that the compound disrupts interactions with recruited substrates by blocking access of an RXL motif to the recruitment site. Since the IC_{50} of E6 against CDK2/cyclin A is

25 μM, some work will need to be done to improve the potency of this compound. The solubility must also be improved before it can be used as a drug. Further investigation of the activity of structurally related compounds will provide further insight into binding modes and will inform inhibitor design. It is suggested that a good starting point would be to order further compounds with different groups at the A end of the molecule and the same group at the B end as in E6; as well as compounds with the same group at the A end and groups at the B end of the same length and size as in E6 but with more polar groups. Testing to check for cellular uptake, binding to other cyclins and the effect on cell cycle progression must be undertaken. This work develops the hope that a combination of virtual screening and *in vitro* selection of compound libraries can be used to isolate specific inhibitors of PPIs, although this approach may necessitate a significant period of further work to optimise the potency and pharmaceutical properties of the hit compounds before the testing phase of drug development can begin.

5 Other drug design strategies targeting CDK2/cyclin A

5.1 Background and introduction

5.1.1 Functions of CDK2

The action of the CDKs in driving cell cycle progression has been explained in Chapter 4. However, it is becoming clear that the CDKs have functions in addition to regulating the cell cycle. These functions include amongst others the control of senescence by regulation of Myc activity. CDK2-mediated phosphorylation of Myc on S62 suppresses Ras-induced senescence (discussed in Hydbring and Larsson, 2010). Suppression of cellular senescence by CDK2 can be antagonised by cellular binding partners such as $p27^{Kip1}$ or $p21^{Cip1}$, via phosphorylation of T14 and Y15 (in the glycine-rich loop of CDK2) by the kinases Wee1 and Myt1, or by pharmacological intervention. In each case this will cause senescence in tumour cells and therefore slow tumour growth (Hydbring *et al,* 2010; Campaner *et al,* 2010).

Due to its effect on Myc, CDK2 may also be involved in other Myc-mediated processes such as apoptosis, self-renewal, immortalisation and the development of stem cells. These functions normally assist in tissue regeneration but are also involved in tumorigenesis (Hydbring and Larsson, 2010). The Myc/CDK2 relationship may therefore form a molecular switch that alters the balance between senescence and cellular renewal, by altering the balance of expression between hTERT (human telomerase reverse transcriptase) and Bmi-1 (a regulator of p16 and p19 activity), and p21 and p16.

5.1.2 Structural basis for CDK2 activity and functions

The structure of inactive monomeric CDK2 was determined by De Bondt and colleagues and revealed a typical bilobal kinase fold (De Bondt *et al,* 1993). The N-terminal lobe (residues 1-79) is mostly β-sheet with one α-helix (the C helix, containing the PSTAIRE motif). The glycine-rich loop is between β sheets 1 and 2

and the N- and C-terminal lobes are joined by a hinge region (residues 80-85). The much larger C-terminal lobe (residues 86-297) is mostly α-helical, and contains the activation segment, responsible for substrate binding. This segment is defined as the sequence between the conserved DFG and APE motifs and includes T160, the residue that must be phosphorylated for full activity.

The molecular basis for activation is now well understood (Pavletich, 1999; Brown *et al,* 1995; Brown *et al,* 1996). In the unphosphorylated form, CDK2 is inactive and the two lobes form an open conformation, with the activation segment partially covering and occluding the catalytic site. Phosphorylation of T160 causes displacement of the activation segment and further disorder within this region, but only allows basal kinase activity (0.3% full activity). On the other hand, binding of cyclin A without phosphorylation causes major changes in the CDK2 fold, including rotation and translation of the C-helix, so that it can form more contacts with residues from the N-terminal lobe. Key ATP-binding residues, especially K33 and E51, are then orientated correctly to allow ATP binding; the activation segment moves out of the way and the two lobes form a more closed conformation. However no significant structural changes occur within cyclin A, and cyclin A binding does not significantly enhance CDK2 activity. Phosphorylation on T160 as well as binding of cyclin A causes full activity.

5.1.3 ATP-competitive inhibitors

The ATP binding site of CDK2 is located between the two lobes and consists of residues from both lobes. In particular, the adenine moiety forms H-bonds to residues from the hinge region, the ribose moiety H-bonds with residues D86 and Q131, and the triphospate moiety H-bonds with K33 and D145 (via the bridging Mg^{2+} ion). Any ATP-competitive inhibitor must therefore be designed to replicate these interactions with the ATP-binding site as closely as possible. The pose of ATP within the binding site, compared to the pose taken by some ATP-competitive inhibitors of CDK2, is shown in Figure 5.1. The ATP-binding site can be split in distinct possible target areas.

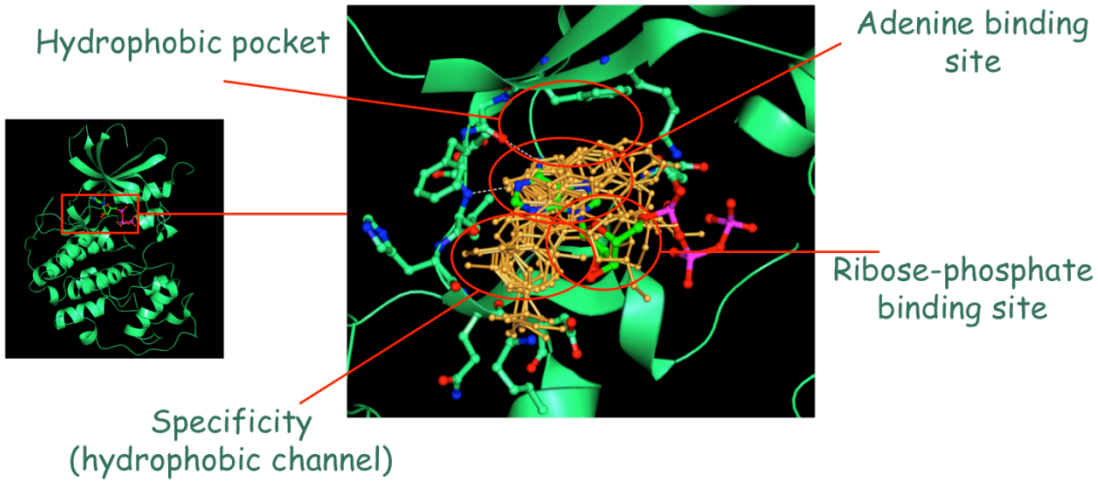
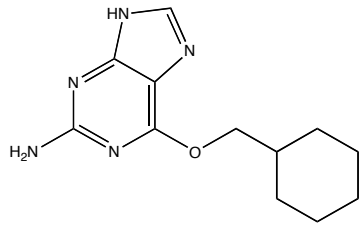
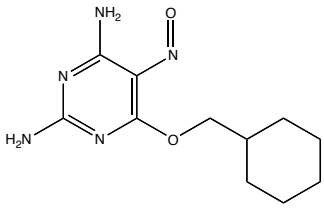


Figure 5.1 Features of the ATP binding site CDK2 bound to ATP (PDB ID: 1HCK) with overlay of ATP-competitive inhibitors in the binding site, showing the space taken up, compared to ATP. Regions which can be targeted include the adenine binding site and the ribose-phosphate binding site. Specificity may be achieved by targeting the hydrophobic channel

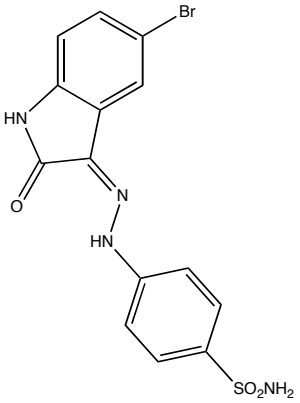
The structures of some ATP-competitive kinase inhibitors are given in Figure 5.2. The molecular basis for inhibition of Staurosporine, a non-specific kinase, compared to the NU compounds, which are specific for CDK1 and CDK2 inhibition, is described further in Section 5.3.1. Generally, selective inhibitors make contacts with the hydrophobic channel (see Figure 5.1) and non-selective inhibitors make most contacts in the adenine or ribose-phosphate binding areas within the binding site. Potency and specificity of inhibition are also mediated by flexibility within the binding site.



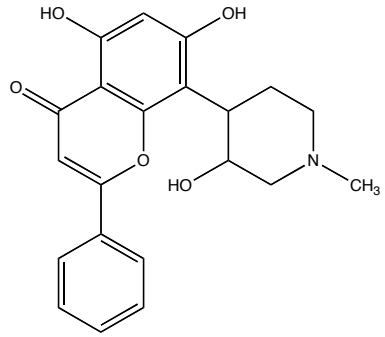
NU2058



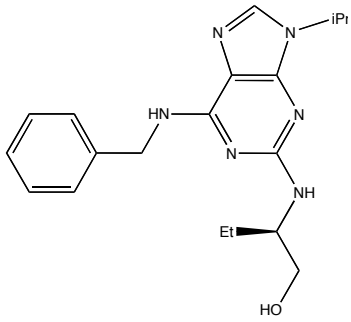
NU6027



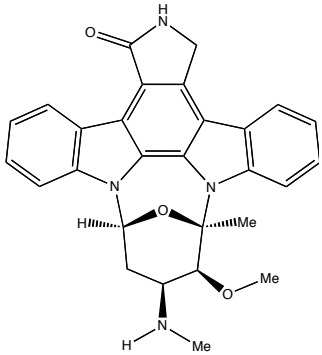
Oxindole 3



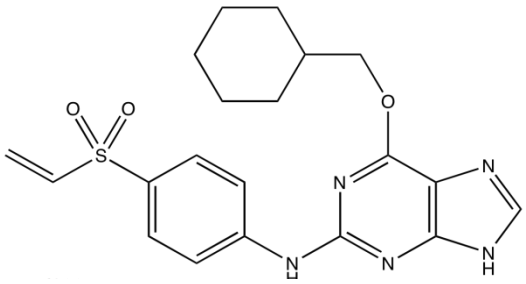
Flavopiridol



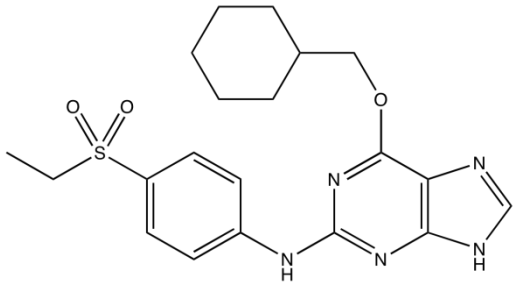
Roscovotine



Staurosporine



NU6300



NU6310

Figure 5.2 Structures of key ATP-competitive CDK2 inhibitors.

In each case, a potent inhibitor must replicate as closely as possible the interactions made by ATP with residues in the binding site, and its shape and hydrophobicity profile should be a good fit for the binding site (Noble *et al*, 2005). Structural information regarding binding is therefore vital for drug development and design in order to achieve potency and specificity (Pratt *et al,* 2004).

5.1.5 Covalent inhibitors: recent successes, issues and development

Examples of covalent inhibitors include oxymorphazone, an opioid analgesic that covalently modifies the μ opioid receptor. This drug shows a long-lasting effect but exhibited tolerance problems in patients (Ling *et al,* 1984). Another covalent inhibitor is the NSAID Zomepirac/Zomax, which was withdrawn from use in 1983 due to unexpected anaphylactic reactions, which caused several deaths. This compound shows covalent binding to several plasma proteins (Smith *et al,* 1986).

Exploitation of covalent interactions can be useful for achieving high potency early in drug development process, but has traditionally been avoided due to the risk of hepatotoxicity, for example as seen in 14C-bromobenzene and acetaminophen, reviewed in Smith *et al,* 2009. Examples of successful covalently acting drugs include omeprazole, a proton pump inhibitor (and related compounds) (Lindberg *et al,* 1986; Lorentzon *et al,* 1997) and the beta lactam antibiotics (Waxman and Strominger, 1983). The toxicity noted for covalent drugs may not inherent feature, and a better understanding of the exact modifications which cause toxicity compared to those that can be tolerated, may enable us to develop these drug candidates further (Baillie, 2006; Liebler, 2008; Said *et al,* 2004).

There are some advantages of covalent action: firstly, it helps to mitigate against high cellular concentrations of ATP, which can be a problem for ATP-competitive drugs especially kinase inhibitors. Additionally, covalent inhibition can lead to an extremely sustained inhibition. Irreversible inhibition can be highly selective, perhaps by targeting rare amino acids. Even if covalent binding is lost due to the target residue being mutated, some inhibition may remain because the reversible binding (that occurs before covalent reaction) remains (Huth *et al,* 2007). Lastly, this approach can minimise side effects and off-target effects if the drug is cleaved rapidly after the irreversible reaction has taken place (Robertson, 2005).

There are some disadvantages to choosing this method of approach: it can be difficult to assess the potency of irreversible inhibitors, because the IC_{50} is extremely dependent on incubation time. Toxicities occur due to off-target modifications, since this route usually relies on electrophilic modification of multiple proteins, sometimes by biologically reactive intermediates. Traditionally this has been a major concern but can be alleviated by excluding known 'toxicophores' from the design, and using more potent drugs at lower doses (Dennehy *et al,* 2006). Another issue is that this is an extremely difficult discovery pipeline and often discovery is by serendipity initially rather than target-based approaches, due to the difficulty of predicting a covalent mechanism of action. For example, aspirin and the penicillins were discovered first and only later realised to be covalently acting.

5.1.6 Aims of the project

The aims of this project were to investigate the binding of NU6300 to CDK2 and to identify the mode of action, including whether or not the binding is covalent, and if so, which residue is the site of covalent modification. Structural and biochemical techniques were employed to probe the binding mode. Understanding the action of NU6300 is critical to inform further development of this compound.

Additionally, structures of CDK2/cyclin A bound to non-covalent ATP-competitive purine analogues were determined in order to provide information which could suggest further modifications that might be made to this series of compounds in order to increase potency.

5.2 NU6300: irreversible inhibition

5.2.1 Preparation of CDK2 mutants for testing NU6300

5.2.1.1 Mutagenesis

CDK2 K88EK89V double mutant, and D86A, K88A, K88E, and K89V single mutants were created in pGEX6P-1-CDK2 plasmids using the QuikChange kit according to the manufacturer's instructions and mutations were confirmed by DNA sequencing. The mutant plasmids were transformed into DH5α cells and large quantities of pure plasmid were obtained using a QiaPrep MiniPrep kit.

5.2.1.2 Expression

Mutant or wild-type human CDK2 mutants and wild-type human cyclin A were expressed and purified as described in Lowe *et al*, 2002. The CDK2 GST fusion protein was co-expressed with the kinase Civ1 (which phosphorylates CDK2 on T160 during expression) using the vector pGEX6P-1. The plasmids pGEX6P-1-CDK2 and pET-cyclin A were transformed into BL21(DE3)/pLYS-S competent cells and colonies were selected using ampicillin and chloramphenicol agar plates. The colonies were picked and grown overnight in starter cultures of LB containing ampicillin and chloramphenicol. Cultures were scaled up using the same media, so up to 6L could be grown at a time. Protein was expressed overnight at 19^{o}C following induction with 0.1 mM IPTG in BL21(DE3)/pLYS-S cells. Cell pellets were obtained using centrifugation (5000 x g), flash frozen using liquid nitrogen and kept at -80^{o}C until required.

5.2.1.3 Purification

To purify the CDK2/cyclin A complex, cell pellets containing the CDK2-GST fusion were resuspended in HBS (50mM HEPES, pH 7.5, 250mM NaCl, 0.02% MTG, supplemented with Complete protease inhibitor cocktail), and after sonication and clarification (100,000 x g), the cell lysates were sequentially

applied to a glutathione sepharose 4B gravity flow affinity chromatography column that binds GST-CDK2. Lysate containing cyclin A were then applied to the affinity column, reconstituting the CDK2/cyclin A complex. The complex was eluted by addition of glutathione (10 mM in HBS) and then the GST tag was cleaved by 3C protease digestion overnight. The mixture of CDK2/cyclin A and GST was subsequently buffer-exchanged and further purified by size-exclusion chromatography (HR26/60 S200). Fractions containing CDK2/cyclin A were pooled and the purity of the final product was checked using SDS-PAGE. Yield of a prep using the above method was approximately 10mg/L culture.

5.2.2 Does NU6300 covalently modify CDK2?

5.2.2.1 Mass spectrometry: checking for covalent modification

Wild-type CDK2/cyclin A was first used to check whether the enzyme is covalently modified. 200 μM CDK2/cyclin A was incubated overnight with 1mM DMSO, NU6300 or NU6310. Mass spectrometry was performed on the samples to identify whether the inhibitors were bound covalently or not. A covalent addition should show up in the spectrogram as an additional mass, however inhibitors that are not bound covalently but only form electrostatic interactions will separate from the enzyme in the process of ionisation (during electrospray). The results are shown in Figure 5.3. The unmodified CDK2 has a mass of 34426 Da, which agrees well with the mass predicted from sequence (34430). Covalent addition of NU6300 was confirmed by detection of a species with an additional mass of 414 Da in the sample incubated with NU6300 but not with the other samples. The spectrograms also showed that only the CDK2 was modified: in each chromatogram the cyclin A has a mass equivalent to the sequence–predicted mass (30072), although other peaks were seen and were difficult to identify.

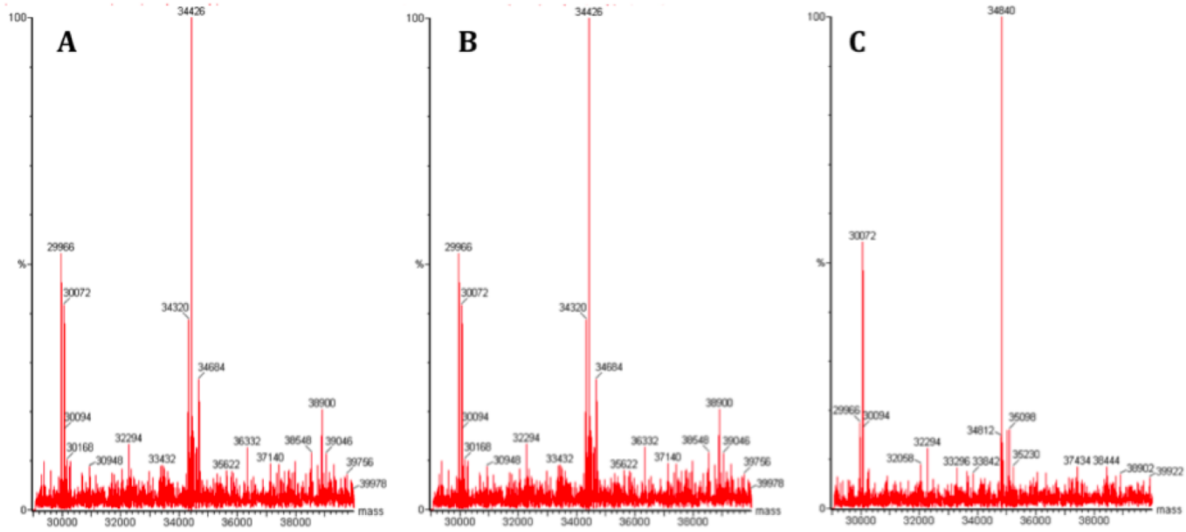


Figure 5.3 NU6300 can covalently modify CDK2/cyclin A. After overnight incubation with NU6300, the mass of the complex is increased by 414, showing modification. Mass spectrogram A: CDK2/cyclin A

incubated with DMSO; B: CDK2/cyclin A incubated with NU6310, a non-covalent inhibitor; C: CDK2/cyclin A incubated with NU6300

5.2.2.2 Methodology: pull-down bandshift assay

The activity of CDK2 can be measured using a bandshift assay, as employed previously in Chapter 2 to screen potential inhibitors. The substrate pRb-GST, which can be phosphorylated on multiple residues, shows a substantial shift in electrophoretic mobility on phosphorylation. In this experiment, CDK2/cyclin A was first incubated overnight with NU6300 or other inhibitors, and then the enzyme was dialysed to remove non covalently-bound compounds. The bandshift assay was then carried out using the dialysed enzyme and pRb substrate. If the inhibitor has been removed, the activity of the enzyme should be recovered and a bandshift will be observed. However, if the enzyme has been covalently modified, the inhibitor will not be removed by dialysis and therefore no modification of pRb will occur. This experiment can be used to compare the activity of different mutants of CDK2 and therefore to identify the binding mode of NU6300.

The covalent action of NU6300 was first confirmed using this method. The results shown in Figure 5.4 demonstrate that after incubation with NU6300, CDK2/cyclin A is unable to phosphorylate pRb (NU6300 cannot be removed by dialysis and the inhibition must therefore be covalent).

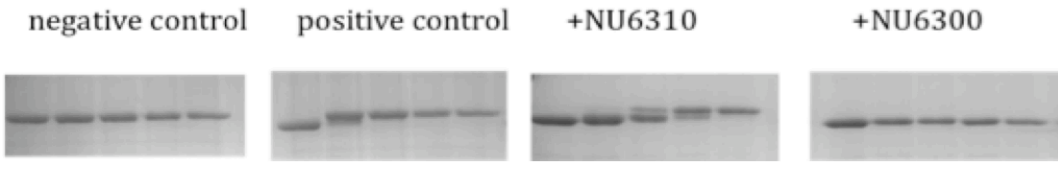


Figure 5.4 NU6300 covalently inhibits wild-type CDK2. Negative control: Reaction mix from which CDK2/cyclin A has been omitted. pRB does not undergo a gel shift. Positive control, CDK2/cyclin A pre-incubated with DMSO; +NU6310: pre-incubated with NU6310, +NU6300: pre-incubated with NU6300. The timepoints shown are (left to right): 0, 2, 5, 10 and 15 minutes. All incubations were overnight, and dialysis was performed over a period of 8 hours before the experiment was carried out.

5.2.2.3 Timescales of modification

In order to determine the timescale of modification, mass spectrometry was carried out as described above, following different lengths of incubation. The shortest incubation period attempted was 30 minutes, after which CDK2 showed complete modification. It is likely that the timescale for the covalent modification is therefore in the range of seconds or a few minutes.

5.2.2.4 Degradation of compound in solution

During the course of experiments to probe the mechanism of action of NU6300, it became apparent that the compound is unstable. We observed that experiments performed with older stock solutions showed much less or no activity. The exact timescale and nature of the degradation NU6300 undergoes has not been determined exactly but it is likely to be degraded completely after about 4 hours. The proposed mechanism for this reaction is illustrated below (Figure 5.5).

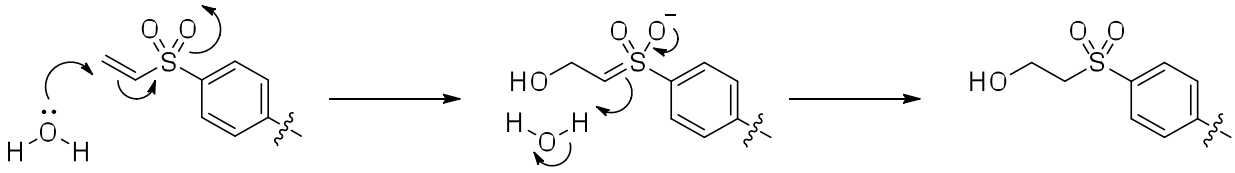


Figure 5.5 Hydrolysis of NU6300 in solution

The compound formed in this reaction may still be able to bind CDK2 and inhibit non-covalently, but has completely lost the capacity for covalent addition. For the purpose of these investigations, therefore, a new aliquot of NU6300 suspended in DMSO was prepared immediately before use.

5.2.3 Identification of the covalently modified residue

5.2.3.1 Testing mutants for covalent modification

Based on a proposed NU6300 binding mode, the site of modification was suggested to be D86, K88, or K89. Firstly, a CDK2 K88EK89V double mutant was produced, to distinguish between a modification at D86, and a modification of either K88 or K89. The effect of incubation with NU6300 was tested on this mutant again using mass spectrometry and bandshift after dialysis as previously described (Figures 5.6 and 5.7 respectively). These experiments identified either or both K88 and K89 as sites of covalent modification.

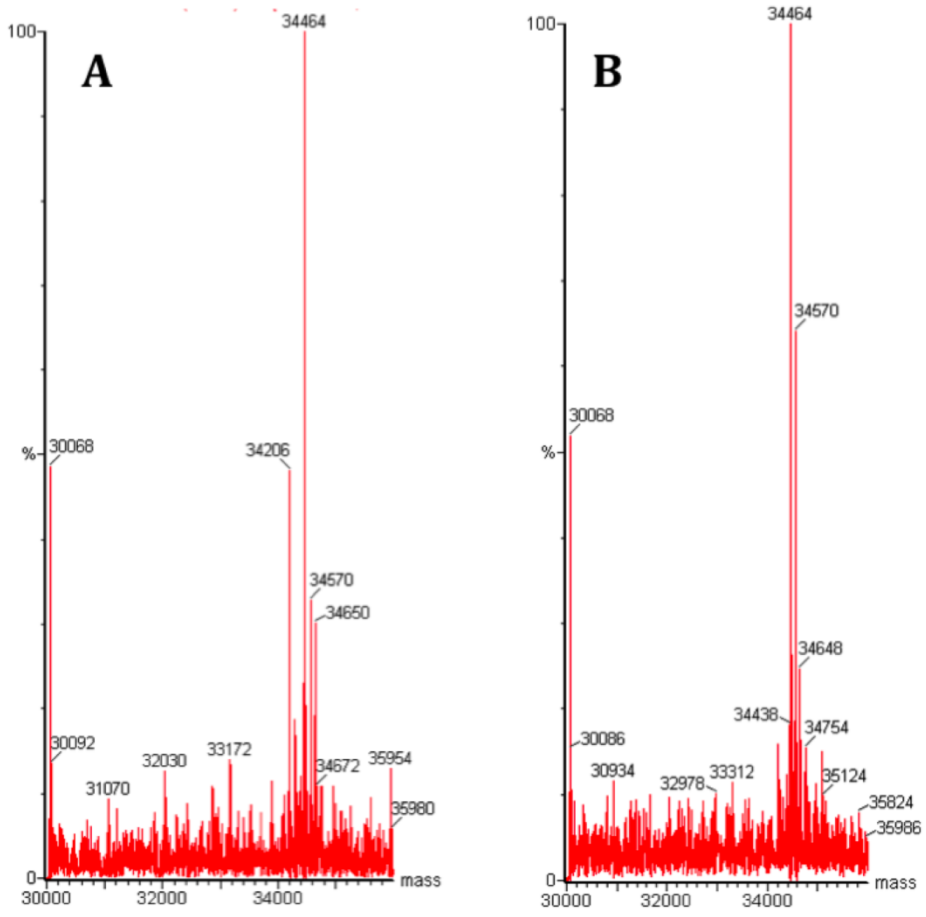


Figure 5.6 NU6300 modifies either K88 or K89. Mass spectrogram of CDK2/cyclin A mutant K88EK89V (A) Incubated overnight with DMSO. (B) incubated overnight with NU6300. NU6300 binding to CDK2 is abolished on mutation of K88 and K89, showing that one of these residues is the primary site of modification.

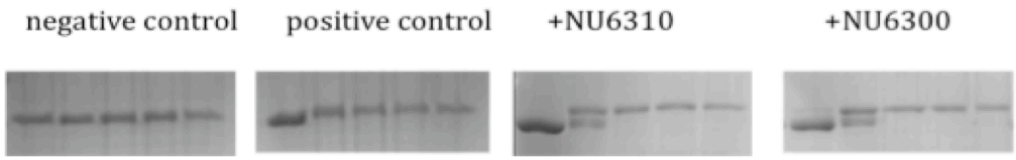


Figure 5.7 NU6300 can be removed by dialysis from CDK2 K88EK89V, showing that the normal residue of modification is one of these. Negative control: lacks enzyme; Positive control, CDK2/cyclin A pre-incubated with DMSO; +NU6310: pre-incubated with NU6310, +NU6300: pre-incubated with NU6300. The timepoints shown are (left to right): 0, 2, 5, 10 and 15 minutes. All incubation was overnight, and dialysis was performed over a period of 8 hours before the experiment was carried out.

Single mutants were produced to probe the identity of the modified residue. These mutants were again tested using mass spectrometry and a bandshift assay after dialysis. The results of the mass spectrometry are shown in Figures 5.8-5.9

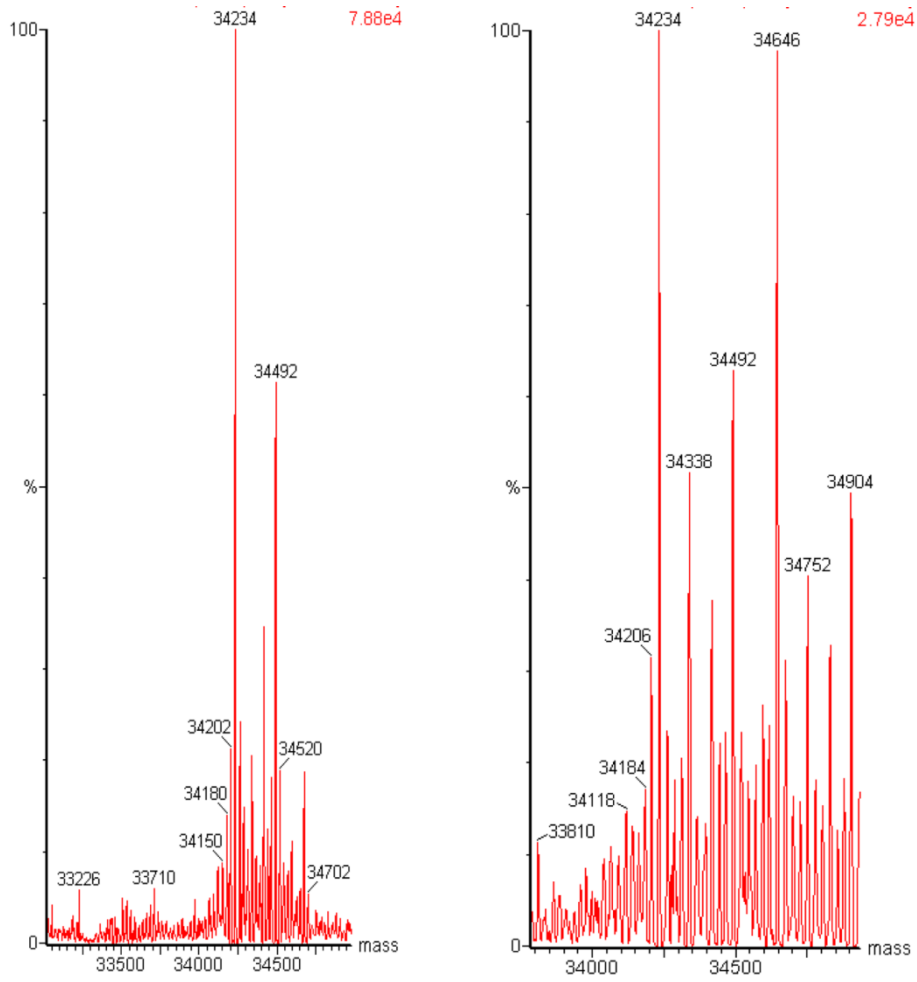


Figure 5.8 Identification of the CDK2 residue covalently modified by NU6300 CDK2 K89V incubated overnight with: Left: DMSO Right: NU6300. A large peak with an additional +414 in mass is observed but there is still a large peak of unreacted protein, therefore K89 has a major role in the reaction.

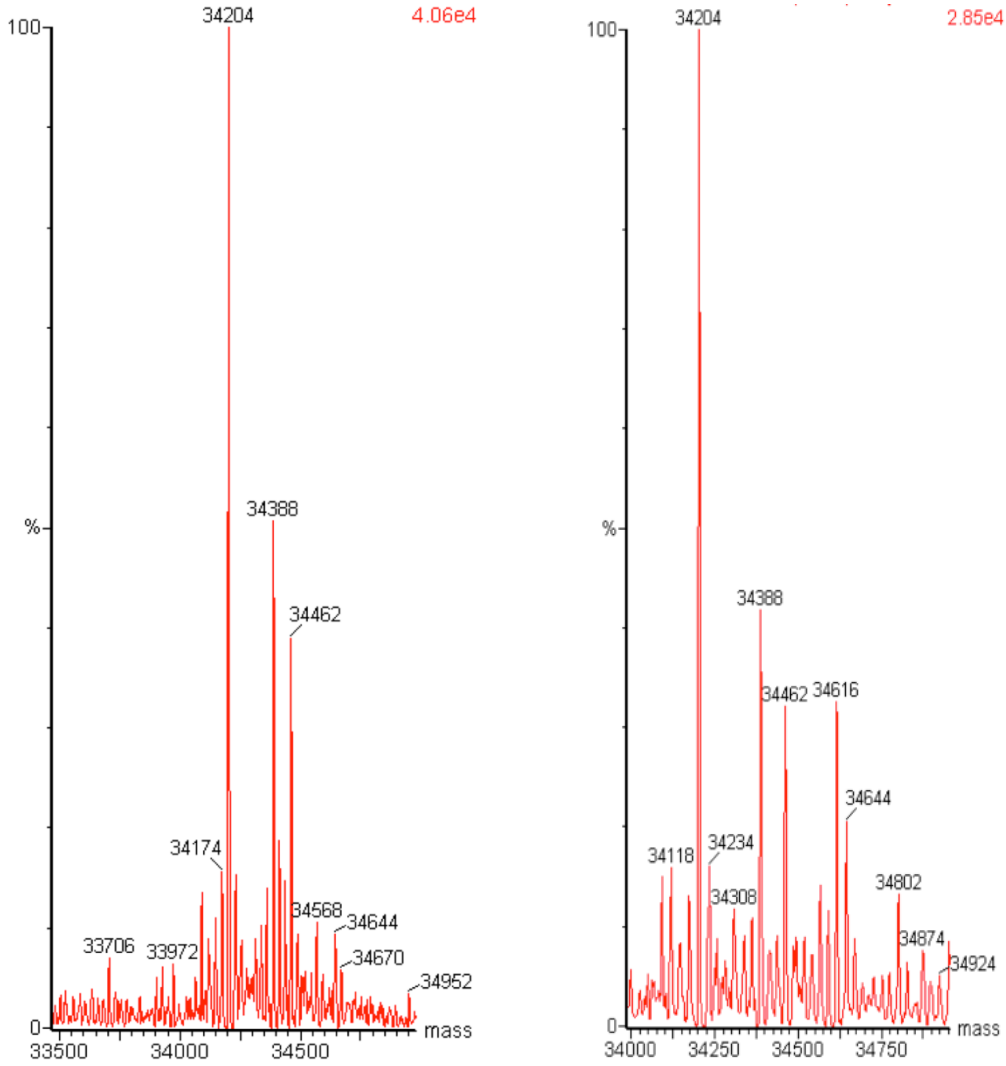


Figure 5.9 K88 is a minor reaction site of NU6300 CDK2 K88E incubated overnight with: Left: DMSO Right: NU6300. A very small peak with an additional +414 in mass is observed but almost all protein is unreacted.

The results of the mass spectrometry show that K89 is the preferred residue for covalent modification, although the results from the K89V mutant, in which a small amount of reaction still occurs, suggest that it is possible but less likely for K88 to be modified.

The results from the bandshift assays corroborate this conclusion, as shown in Figure 5.10.

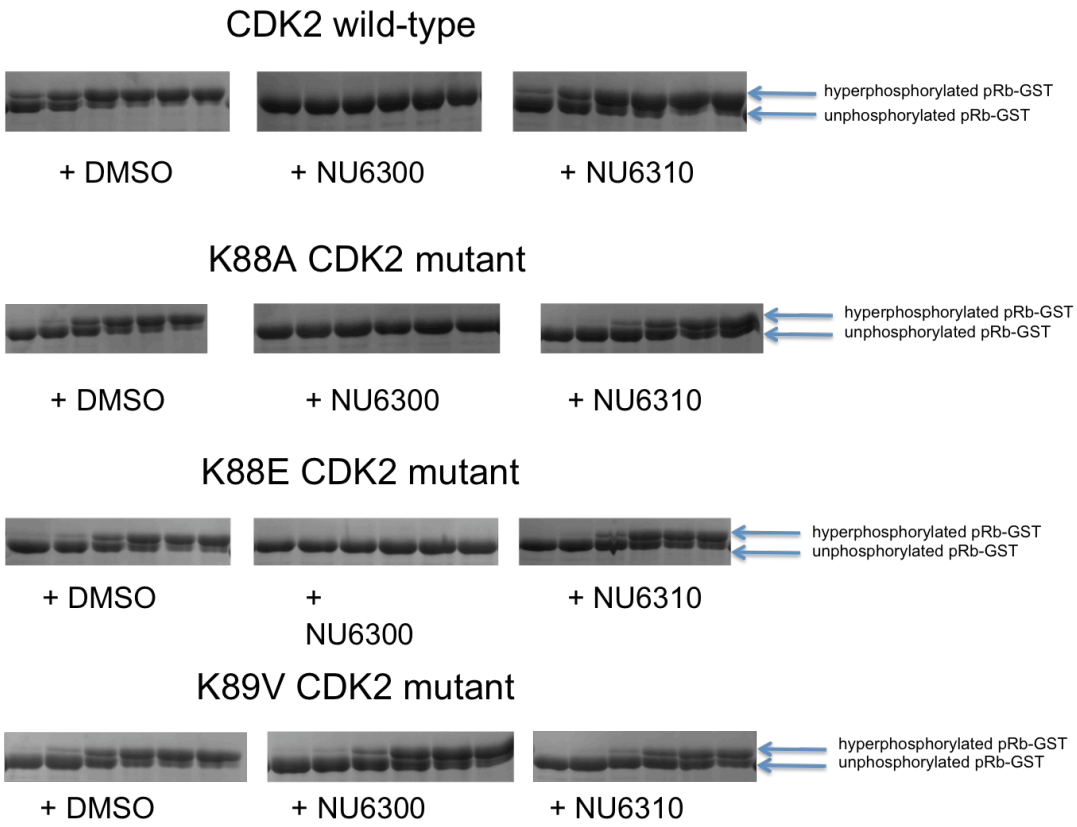


Figure 5.10 Mutation of K89 abolishes NU6300 covalent modification. Activity is recovered after dialysis, as shown in this bandshift assay carried out with different CDK2 mutants. CDK2/cyclin A was incubated with DMSO, NU6300 or NU6310 as shown. The timepoints shown are (left to right): 0, 1, 2, 5, 10 and 15 minutes. All incubation was overnight, and dialysis was performed over a period of 8 hours before the experiment was carried out.

The results of the final bandshifts using single mutants confirm that K89 is the main site of covalent modification, as K88 mutants are still covalently modified by NU6300 and do not recover their activity after dialysis. However, the K89 mutants do recover their activity after dialysis.

5.2.4 Determination of the structure of CDK2/cyclin A modified by NU6300

5.2.4.1 Crystallisation

Crystallisation was carried out using standard CDK2/cyclin A crystallisation conditions (0.8M KCl, 1.1M $(NH_4)_2SO_4$, 0.1M Hepes pH 7) in hanging drop trays at 4°C, as used previously in Chapter 4.10. CDK2/cyclin A was at 10 mg/ml and a 4-fold excess of NU6300 was added. Co-crystals of a cubic shape, up to 400 μm long, grew within one week. The co-crystals were cryo-protected using 8M sodium formate and flash-frozen in liquid nitrogen.

5.2.4.2 Data collection

The data were collected on beamline X at the European Synchrotron Radiation Facility (ESRF), using a robotic sample changer. Strategy was calculated using Mosflm from three test images collected at 90° separations. The maximum resolution obtained was 1.99 Å.

5.2.4.3 Data analysis and structure refinement

The data were integrated using Mosflm and scaled using Scala. Molecular replacement was performed by MolRep using the structure of CDK2/cyclin A (PDB ID code 1H28) with the peptide removed as a starting model. Restrained and rigid body refinement was carried out in REFMAC5 and manual model building and water identification was performed using Coot. All methods were the same as in Appendix B. The statistics obtained for data collection and analysis are shown in Table 1 on page 198 (along with the statistics for the other structures determined here).

5.2.4.4 Crystal structure of NU6300-modified CDK2/cyclin A

Good continuous and unambiguous difference density was observed, representing NU6300 bound to CDK2. As expected, the inhibitor is bound within the ATP binding site and makes the triplet of hydrogen bonds with the CDK2 hinge region (Figure 5.11). As reported for other N2-substituted derivatives of this inhibitor series, the aniline group points out of the active site cleft and packs against the CDK2 forming a π–π stacking interaction with the peptide backbone between N85 and D86. The covalent bond between the vinyl sulfone substituted at the ρ-position and the side chain K89 is clearly visible. This observation confirms the results of the mass spectrometry and gelshift assays with the CDK2 mutants. Figure 5.11 shows the pose of the inhibitor within the binding site.

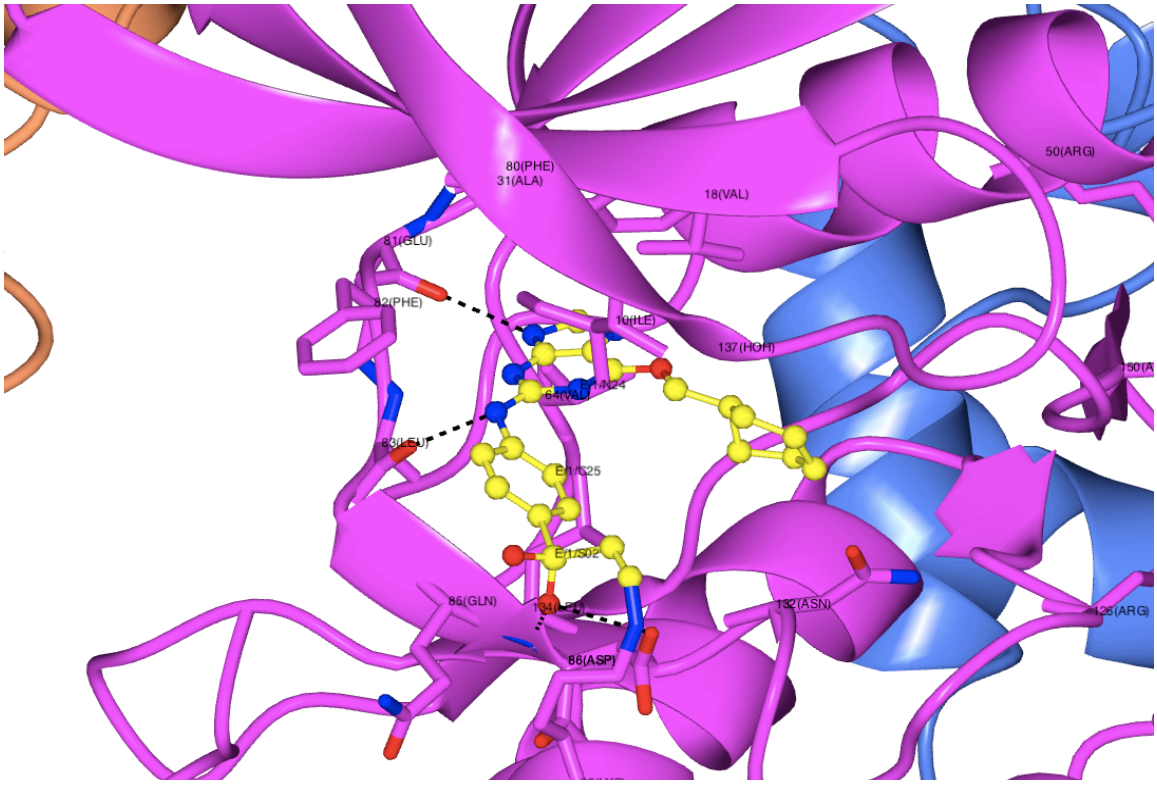


Figure 5.11 The crystal structure of NU6300 bound to CDK2/cyclin A NU6300 covalently modifies K89 as the primary site of modification. The inhibitor is drawn with carbon atoms in yellow. CDK2/cyclin A is represented in ribbon form with CDK2 and cyclin A coloured magenta and blue respectively. The three hydrogen bonds between the adenine ring and the CDK2 peptide backbone within the hinge region are included as dotted lines. Figure prepared in CCP4MG.

In addition to the hydrogen bonds made with the hinge region, NU6300 also interacts with both the main chain amino and side chain carboxyl of D86. These hydrogen bonds may be important for initial inhibitor binding and may help to orientate the compound correctly to allow the covalent modification to take place. This could also explain why D86 was observed to have a minor effect on the activity of NU6300 against CDK2, since mutation of this residue would not directly affect the reaction, but would contribute to inhibitor binding to the CDK2 active site. Absence of D86 could therefore slow down the reaction but not completely stop it.

A possible mechanism for the reaction of NU6300 with Lys89 of pCDK2 is shown in Figure 5.12:

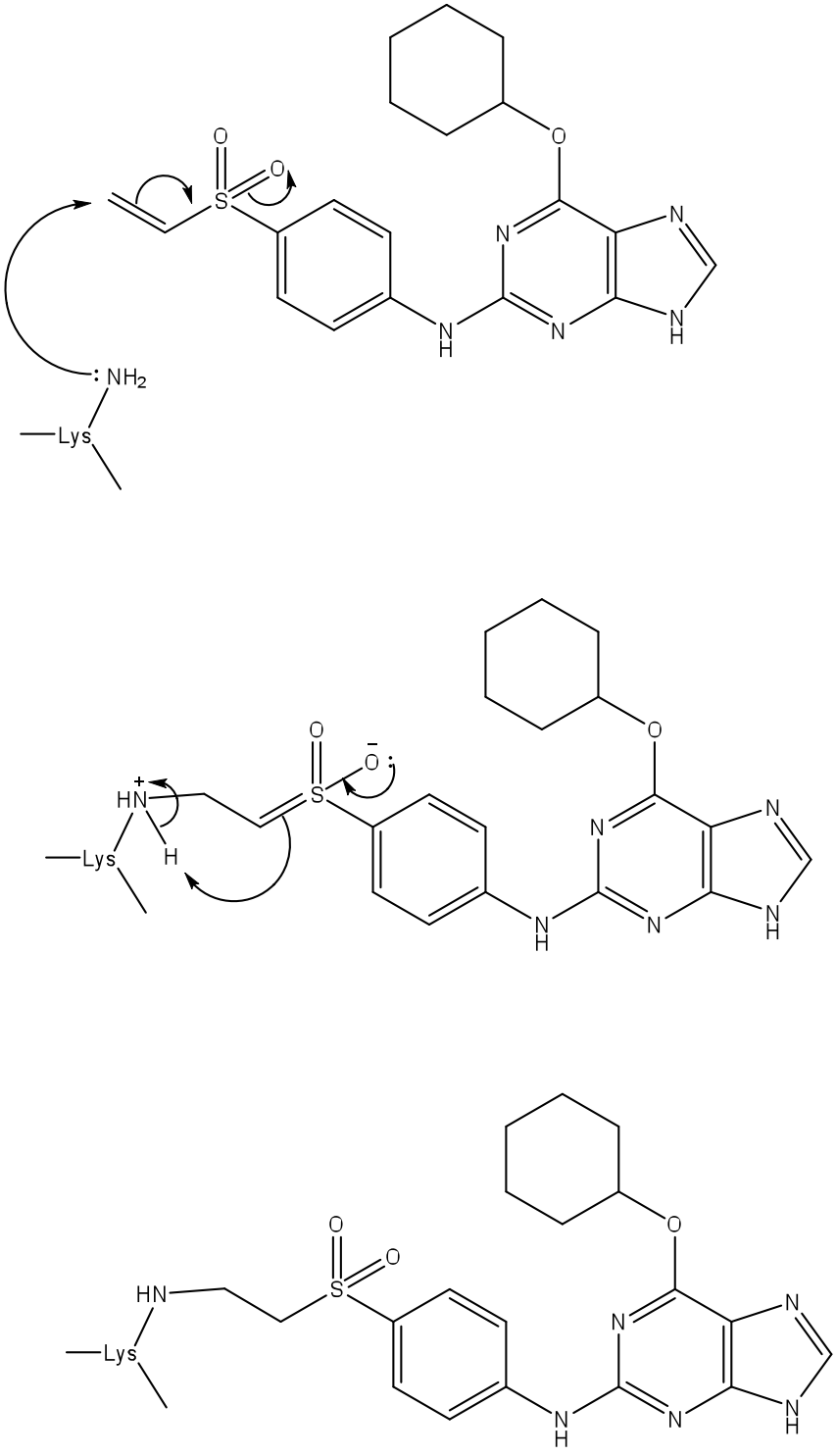


Figure 5.12 Suggested possible mechanism of modification of pCDK2 by NU6300. A possible mechanism driven by the attack of the lone pair on the amino group of Lys89 is shown.

5.3 Co-crystal structures of CDK2/cyclin A with traditional ATP-competitive drugs

5.3.1 Background and introduction

5.3.1.1 Compound series development

The NCL compound series was developed at Newcastle University from the O(6)-cyclohexylmethylguanines (inhibitors of CDK1 and CDK2), initially resulting in the discovery of O(6)-cyclohexylmethyl-2-(4'-sulfamoylanilino)purine. A process of iterative structure-based design has been used to optimize the ATP-competitive inhibition of CDK1 and CDK2 by these new inhibitors, so that they are 1,000-fold more potent than the parent compound. The structures of human pT160-CDK2/cyclin A in complex with the inhibitors NU2058 and NU6102 (PBD ID codes 1H1P and 1H1S, respectively) showed that the increased potency is mainly due to the formation of two additional hydrogen bonds between the inhibitor and Asp 86 of CDK2, which allow optimum hydrophobic packing of the inhibitor's anilino group in the region that particularly determined specificity (Davies *et al,* 2002). In addition, the shape complementarity to the pocket is excellent and NU6102 provides a 'perfect fit' for the hydrophobic patches on CDK2, in a way that it does not for other kinases (shown in Figure 5.13).

A **B**

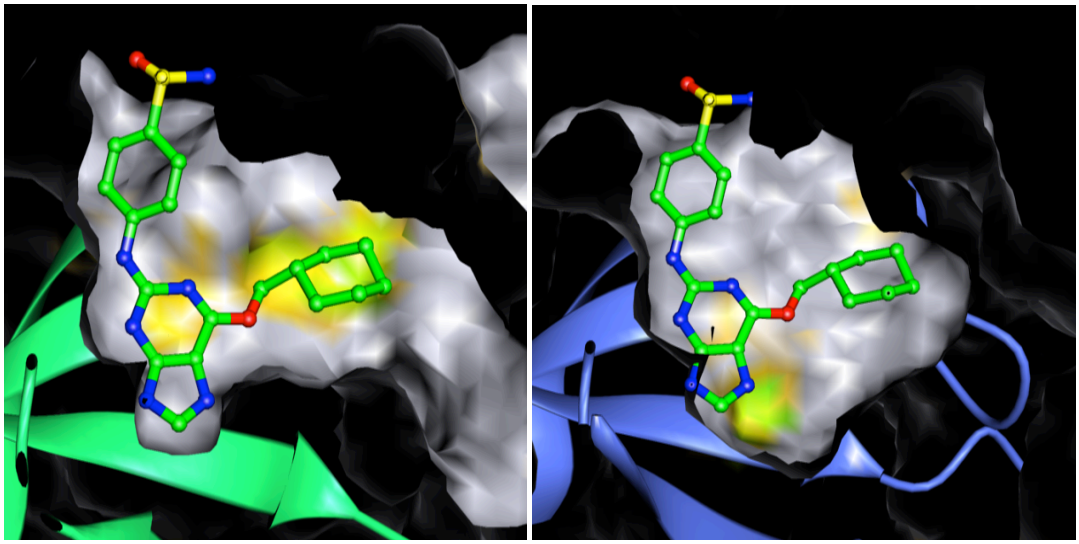


Figure 5.13 Binding of NU6102 to CDK2 (A) or PfPk5 (B). The inhibitor clashes more with groups in the PfPK5 ATP binding site, and a different hydrophobic patch means that only binding to CDK2 will be strongly stabilised by hydrophobic packing.

Knowledge of how we can exploit differences within the ATP binding site between different CDKs to achieve some level of specificity, has therefore guided the development of this inhibitor series.

5.3.1.2 Other ATP-competitive kinase inhibitors

Staurosporine is a very non-selective ATP-competitive kinase inhibitor, which can induce apoptosis effectively. A comparison of the more selective NCL inhibitors with a compound such as this, shows the advantages of both approaches: it is easier to produce a general kinase inhibitor that to select for the exact CDK, and taking this approach may have a stronger therapeutic effect, since multiple targets may be hit and this could work synergistically. However this does also significantly raise the possibility of encountering damaging side effects. The co-crystal structures of Staurosporine bound to CDK2, and to cAMP-dependent protein kinase (cAPK) have been determined (PBD ID codes 1AQ1 and 1STC) which shows why it acts non-specifically but is a potent inhibitor against a wide range of kinases (Lawrie *et al,* 1997; Prade *et al,* 1997). Staurosporine binds in the position of the adenosine moiety of ATP, one of the parts of the binding site that does not really confer specificity. The ATP binding site itself is extremely flexible, and with some flexibility on the part of the inhibitor, Staurosporine – side chain interactions can be recreated (Toledo and Lydon, 1997). Residues that make a particularly large conformational changes on Staurosporine binding are: Lys 89, His 84, Lys 33 and Gln 131 in CDK2 and Phe 54, Met 120, Thr 183 and Phe 327 in cAPK. In this way the Staurosporine binding pose is almost identical and it will inhibit both proteins almost equally.

5.3.1.3 Aims of this project

Crystal structures of several compounds from this series bound to CDK2/cyclin A were determined. The intention was to provide specific information regarding the binding mode and interactions formed in each case, in order to inform drug design within this compound series.

5.3.2 Structure determination

5.3.2.1 Crystallisation

NCL14550, NCL18046 and NCL18365 were co-crystallised with T160pCDK2/cyclin A as previously described using standard CDK2/cyclin A crystallisation conditions (1.1M ammonium sulphate, 0.75M KCl) in hanging drop trays at 4 ^{o}C {Davies, 2002}. CDK2/cyclin A was at 10 mg/ml and a 4-fold excess of each inhibitor was added. After a 20-minute incubation period on ice, the incubation mix was filtered using a filtration unit with a MWCO of 5 kDa. Crystallisation trays were set up by hand using stock solutions of ammonium sulphate and potassium chloride. Crystals were obtained after two weeks, they were all of a square shape and grew up to 300μm in the longest direction. Before data collection, crystals were briefly immersed in cryo-protectant (8 M sodium formate) before freezing.

5.3.2.2 Data collection

Data collection was carried out at the ESRF using beamline ID23. We performed a remote collection, using a robotic sample changer, and strategy was calculated using Mosflm.

5.3.2.3 Data analysis and refinement of structure

Data processing was carried out using MOSFLM (Leslie, 1992), SCALA (Evans, 1993) and other programs of the CCP4 suite (Winn *et al,* 2011). The structures of NCL14550, NCL18046 and NCL18365 bound to CDK2/cyclin A were solved by molecular replacement using Phaser (McCoy, 2007), using as the search model a high-resolution structure of T160pCDK2/cyclin A bound to a recruitment peptide (from Cheng *et al*, 2006; PDB accession code 2CCH). In each case, a single unambiguous solution was found with an inhibitor bound to each of the two copies of the binary complex in the asymmetric unit. This solution for each structure was then subjected to rigid body refinement in REFMAC version 5.5.0066 (Murshudov, 1997), in each case revealing unambiguous electron density in the CDK2 ATP-binding site, consistent with the expected shape of the inhibitor.

Models of NCL14550, NCL18046 and NCL18365 were created using the Elbow Builder utility, and fitted into the density, using Coot (Emsley, 2004). The inhibitor atoms were kept in all subsequent models during refinement. The structures were then refined further by rounds of manual rebuilding in Coot and restrained refinement in REFMAC5.5.0066, including TLS refinement. Towards the end of refinement, waters were added using the Coot water picking utility and manually verified. The statistics for the datasets from the structures of NCL14550, NCL18046 and NCL18365 bound to T160pCDK2/cyclin A and for the crystallographic refinement are presented in Table 5.1.

	CDK2/cyclin A + NCL14550	CDK2/cyclin A + NCL18046	CDK2/cyclin A + NCL18365	CDK2/cyclin A + NU6300
Data collection parameters:				
Wavelength (Å)	0.96	0.96	0.96	0.96
Space group	P212121	P212121	P212121	P212121
a	74.40	74.28	74.39	74.79
b	135.35	134.73	135.51	135.61
c	148.66	148.75	148.98	148.78
Resolution (Å)	45.5 – 2.5	41.44 – 2.7	41.30-2.95	40.03-2.4
Total reflections	26242	35660	30408	56743
Unique reflections	5259	12098	11358	20077
Completeness (%)	99.8	89.82	98.58	99.8
R_{merge}	0.221 (0.105)	0.235 (0.112)	0.228 (0.146)	0.217 (0.098)
Refinement parameters:				
Resolution (Å)	45.5-2.5	41.44-2.7	41.30-2.95	40.03-2.4
R_{work}	17.6	19.0	21.1	19.5
R_{free}	23.4	24.8	23.7	22.9
Number of atoms:				
Protein	9159	9254	8691	9144
Peptide/Ligand	64	70	66	65
Solvent	501	264	388	412
Average B-factor	18.1	13.61	20.25	16.93
Root mean square deviation				
Bond lengths (Å)	0.017	0.017	0.016	0.021
Bond angles (°)	1.745	1.732	1.684	1.855

Table 5.1. X ray data collection and refinement statistics for CDK2/cyclin A/inhibitor complexes

5.3.3 Comparison of binding mode with other compounds in the series

5.3.3.1 Co-crystal structure of NCL-00014550 with CDK2/cyclin A

The co-crystal structure of NCL-00014550 with CDK2/cyclin A is illustrated in Figure 5.14, shown in structural view and also in schematic form.

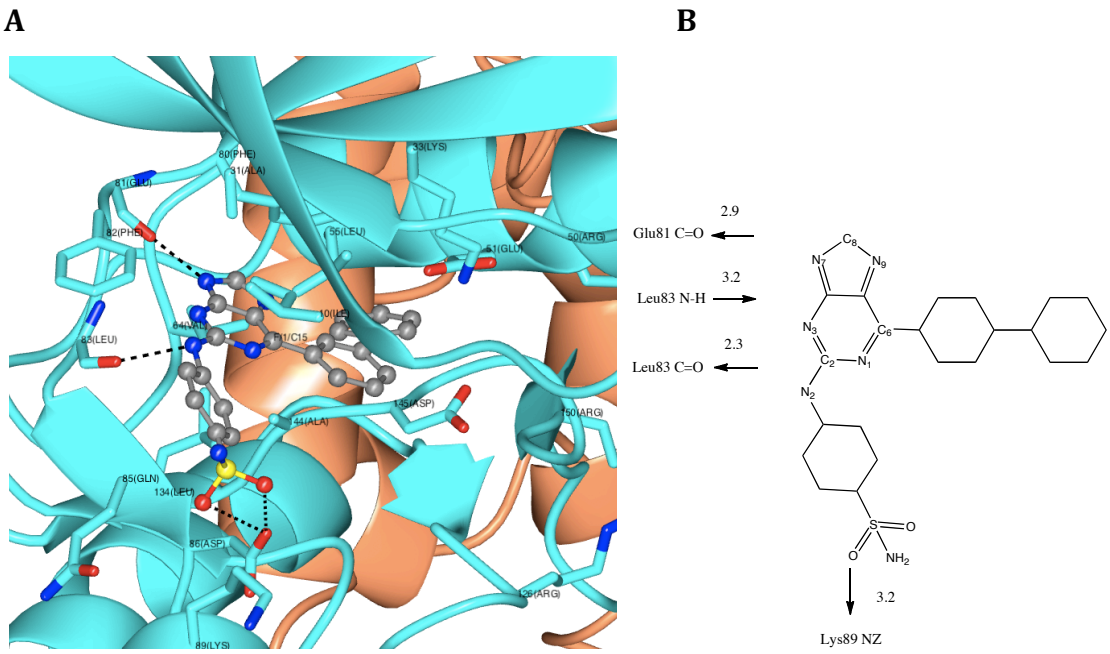


Figure 5.14: Binding pose of NCL-00014550 within the ATP binding site

(**A)** The binding pose of NCL-00014550 is shown in each of the two copies in the asymmetric unit. There is a difference in the position of the C6 substituent between the two copies, as one points towards the back of the pocket. The protein backbone is shown in ribbon form, with CDK2 shown in magenta or cyan (with cyclin A in blue or orange respectively in the background). Selected interacting residues and the inhibitor are shown in ball-and-stick representation, with the inhibitor yellow or grey. Hydrogen bonds are shown as dashed black lines and the residues involved are labelled. (**B)** Ligplot giving a schematic representation of the hydrogen bonds made between backbone atoms of CDK2 residues E81, L83 and K89 located in the hinge region and NCL-00014550, to highlight the binding mode. Hydrogen bonds are drawn as arrows, with bond lengths in Angstroms indicated.

This compound largely retains the H-bonding interactions made by the purine backbone with backbone N-H and C=O, and is also stabilised by a hydrogen bond from a sulphonamide oxygen to the side chain NZ of Lys 89.

5.3.3.2 Co-crystal structures of NCL-00018365 and NCL-00018046 with CDK2/cyclin A

The co-crystal structure of CDK2/cyclin A with NCL-00018365, and that with NCL-00018046 were determined and are shown in Figure 5.15.

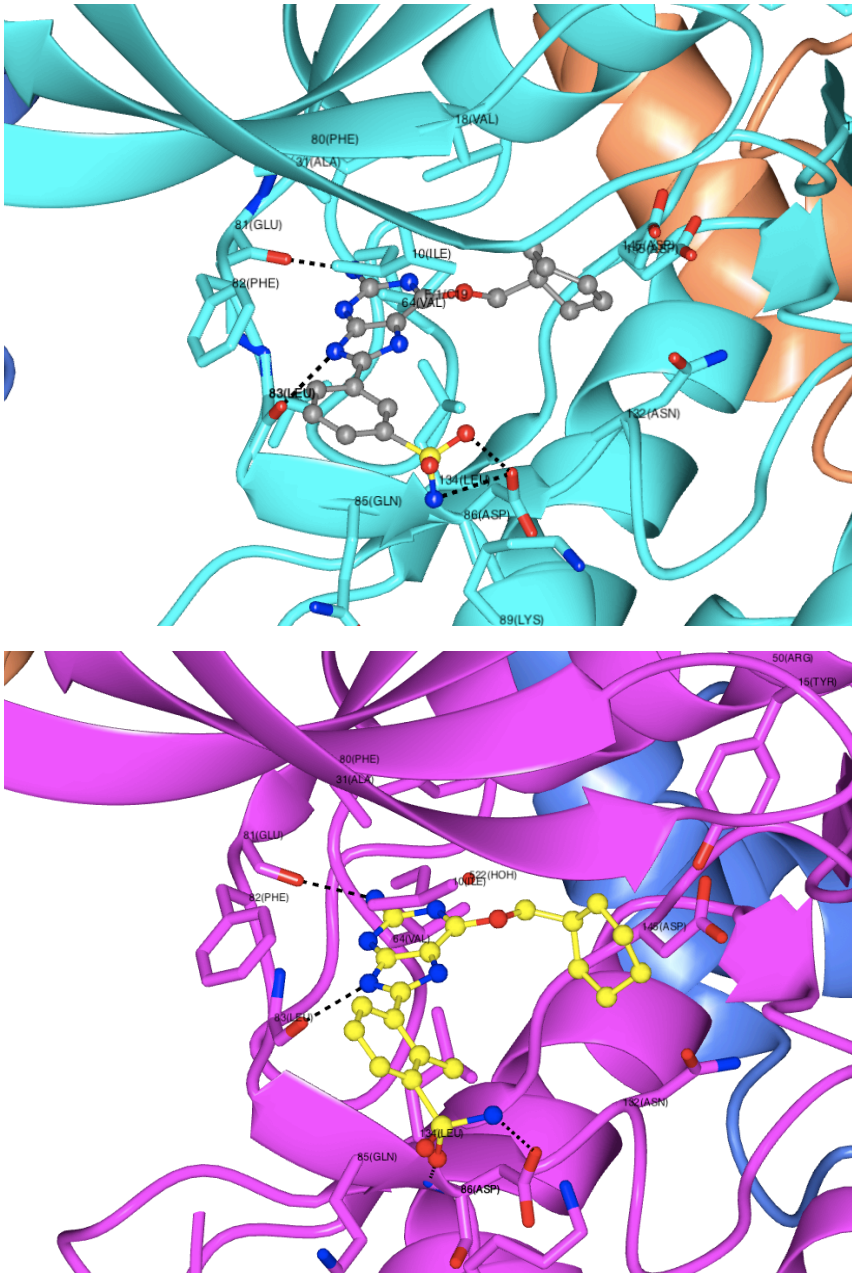
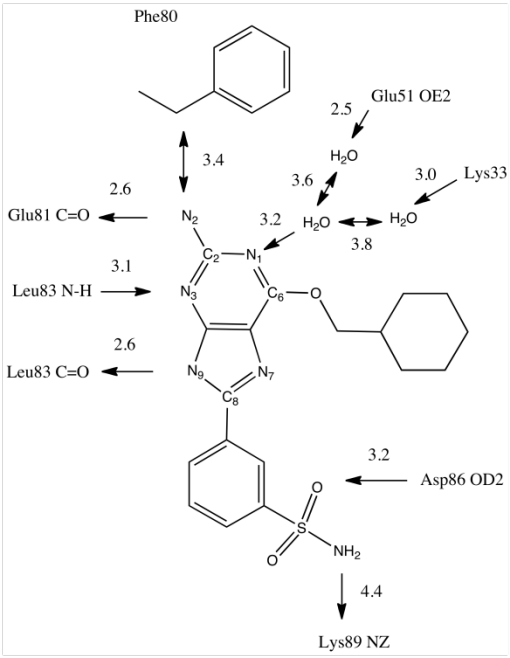
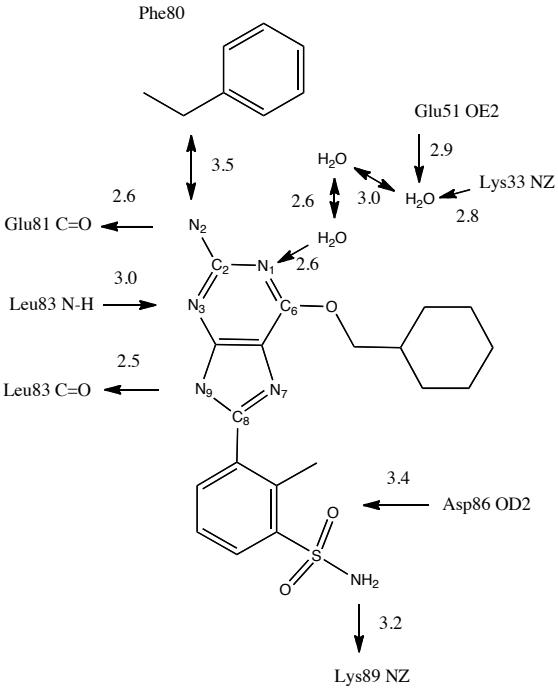


Fig 5.15 Binding poses of NCL-00018365 and NCL-00018046 within the ATP binding site The binding pose of NCL-00018365 (**A**) and NCL-00018046 (**B**) is shown. The protein backbone is shown in ribbon form, (upper: blue/orange; lower: magenta/blue). Selected interacting residues and the inhibitor are shown in ball-and-stick representation, with the inhibitor yellow or grey. Hydrogen bonds are shown as dashed black lines and the residues involved are labelled.

A ligplot is given for each of the ligand in the co-crystal structures, to provide a schematic representation od the binding mode and illustrate the key interactions made by the ligand (Figure 5.16).



A



B

Fig 5.16: Illustration of the binding mode of NCL-00018365 (A) and NCL-00018046 (B) to CDK2.
Ligplots: schematic representation of the hydrogen bonds made between backbone atoms of CDK2 residues E81, L83 and K89 located in the hinge region and NCL-00018365 and NCL-00018046, to highlight the binding mode. Hydrogen bonds are drawn as arrows, with bond lengths in Angstroms indicated.

These compounds largely retain the H-bonding interactions made by the purine backbone with backbone N-H and C=O, and are also stabilised by multiple additional interactions: a hydrogen bond from the sulphonamide nitrogen to the side chain NZ of Lys 89, and from the sulphonamide oxygen to the side chain of Asp86. In addition, a small group of water molecules mediates a longer distance interaction with Glu 51 and Lys 33. The addition of a methyl group on the C8-phenyl ring, to form NCL-00018046 (from NCL-00018365) was predicted to cause a rotation of this ring to minimise steric clashes, and strengthening the H-bonding network by decreasing the distances involved. This does indeed happen as predicted, and explains the slightly higher potency of NCL-00018046, although the effect is minor and the only bond distance that significantly decrease is that between the sulphonamide group and Lys 89.

5.4 Conclusions and further work

NU6300 is the first reported covalently-acting CDK inhibitor. The results from these experiments have deduced the mechanism of action. Experiments performed by our collaborators in Newcastle have also shown that this compound is a potent and specific inhibitor of CDK2, and that it is able to enter cells when provided in culture. There are a couple of points that emerge from our experiments that also suggest NU6300 may be a viable drug compound or lead. Firstly, the compound is degraded in solution and does not retain its ability to covalently modify proteins for more than a few hours. On the other hand, it reacts very quickly (as shown in our mass spectrometry experiments). These properties are ideal for a covalent drug in order to avoid some of the toxicities associated with covalent action. The drug should bind tightly to CDK2, the covalent modification will occur and then that ability will be lost, no this minimises cross-reactivity. The other point that came out of our experiments is the fact that NU6300 usually modifies K89, but if K89 is mutated, it still retains some ability to modify covalently K88. This could be useful in a drug molecule as it may help avoid problems associated with drug resistance.

The co-crystal structures of the other ATP-competitive inhibitors bound 'as expected' in most cases. The hydrogen-bonding network is retained quite predictably in this series. We also saw that a prediction that had been made regarding the effect of the addition of a methyl group onto a compound, was accurate and this should give us some confidence that we are beginning to understand the chemical space of the ATP-binding site of CDK2 quite well.

Further work with NU6300 is likely to focus on producing and testing derivatives that maintain a covalent action and have advantages in terms of potency, solubility or other pharmacological properties.

6 Concluding remarks

6.1 Advantages of alternative strategies

The use of ATP-competitive inhibitors, as explored in this thesis against CDK2, is a well-established method for inhibitor design. It has the advantage of employing a well-understood binding site, for which lots of inhibitors have been described, which could act as potential starting points or reference compounds for a drug design project. However, the disadvantages of this approach are that although the response is likely to be predictable and high-affinity ATP-competitive inhibitors can be identified, the problem of selectivity remains, since many proteins contain this structurally similar binding site. Staurosporine is an example of a very potent, but extremely non-selective kinase inhibitor. Although this might not always be a problem, it can mean that any drug developed is likely to have more unexpected side and off-target effects. In addition, resistance to ATP-competitive drugs is well-documented and may become a problem, although this can be overcome when used as a combination therapy. We have employed methods of structure-based drug design to increase the potency of our compounds and this has often allowed predictable SAR, since the binding site is so well understood.

The use of covalent inhibitors is a promising area of research and our compound NU6300, whilst early in development and testing, provides an interesting example of this class since it is the first reported covalently-acting CDK inhibitor. We have been able to identify the site of modification and determine the structure of the compound bound to CDK2: we therefore now understand much more about its mode of action.

Targeting PPIs is clearly a viable strategy for a wide range of interactions. We picked those targets with known structures, where the PPI site was clearly defined and presented a 'druggable' target with a deep cleft, as much like a traditional binding pocket as possible, rather than a flat interface. In each case we met with some success. For MDM2, the inhibitor series binds potently and

specifically, shows activity in cell-based assays and we understand the binding mode and interactions formed, from our co-crystal structures with the compounds bound to MDM2. We are now looking to expand our knowledge by probing the binding of these compounds to MDMX, in order to better understand the differences between interactions made with residues in each binding site.

Our investigation into using the Plk1 PBD as a target for PPI drug design replicates previous results in which the natural product thymoquinone was shown to be a potent inhibitor of this interaction. However, thymoquinone derivatives and related compounds, as well as other known Plk1 inhibitors of unknown binding mode proved much less potent. Inhibitor design for this project may therefore prove difficult, but would be guided by a structure of thymoquinone bound to the PBD. This structure could provide definitive and specific information regarding the interactions formed by this inhibitor in the binding site and suggest further alterations that might produce an inhibitor with a better fit to the binding site.

Our work using the cyclin A recruitment site as a PPI target is still at an early stage but the identification of a hit compound for inhibition of this interaction is interesting, and suggests that a sufficiently targeted approach using virtual screening to pre-select compounds from large numbers in libraries, is viable for PPI inhibitor identification. Although the lead compound is currently of a low affinity, SAR, structural analysis and further testing of derivatives may allow identification of a more potent compound and the identification of compounds with any activity at such an early stage in the project is a positive step. The structure of the compound identified fits with previous observations regarding the type of molecule likely to be a PPI inhibitor.

6.2 Future direction for the projects

The promising lead compound E6, selected from our library after screening, should be developed by testing E6 analogues and derivatives with further alterations to the molecule at points A and B, possibly using the fluorescence anisotropy assay that appears to be successful with this system. Further information regarding the specificity of binding is also important, and structure determination of E6 bound to CDK2/cyclin A will help us to understand the binding mode and inform further inhibitor design. This is complicated by the fact that the only crystal forms available for CDK2/cyclin A have a dimer of CDK2/cyclin A in the asymmetric unit where either one or both of the two recruitment sites are occluded. As a result, the crystals are not amenable to soaks or co-crystallisation. In order to maximise the likelihood of successfully obtaining a structure, two possible strategies present themselves: either to re-screen CDK2/cyclin A or the less soluble cyclin A monomer for new crystal forms, or to optimise around the conditions and crystal form already identified.

Our attempts to co-crystallise or soak in inhibitors to the Plk1 PBD were unsuccessful, and the protein appears to present a particularly tough crystallisation target. Since it is small enough for NMR this could provide an alternative strategy to obtain information regarding inhibitor binding. HSQC experiments could be carried out to calculate chemical shifts upon titration (as previously performed in our group to look at MDM2/inhibitor binding), and later a full structural analysis may be warranted.

The MDM2 project is at a very exciting phase. Development of the series is now focusing on those areas identified through the crystal structures described in this thesis. The synthesis of a new extremely potent compound NCLbased on the information available from our structures, demonstrates the efficacy of the structure-guided drug design process in this case. Further improvements to specificity, potency and solubility should enable these compounds to enter clinical trials.

Further development of our covalent CDK2 inhibitor may focus on developing analogues that last longer in solution but maintain the covalent binding mode.

References

Ackler, S. et al., 2012. Navitoclax (ABT 263) and Bendamustine ± Rituximab Induce Enhanced Killing of Non-Hodgkin's Lymphoma Tumors in Vivo. British journal of pharmacology. 167(4): pp.881-91

Aisen, P.S., 2006. Development of a disease-modifying treatment for Alzheimer's disease: Alzhemed. Alzheimer's & dementia : the journal of the Alzheimer's Association, 2(3), pp.153–4.

Aisen, P.S. et al., 2011. Tramiprosate in mild-to-moderate Alzheimer's disease - a randomized, double-blind, placebo-controlled, multi-centre study (the Alphase Study). Archives of medical science : AMS, 7(1), pp.102–11.

Akritopoulou-Zanze, I. & Hajduk, P.J., 2009. Kinase-targeted libraries: the design and synthesis of novel, potent, and selective kinase inhibitors. Drug discovery today, 14(5-6), pp.291–7.

Archambault, V. & Glover, D.M., 2008. Yeast Polo-like kinase substrates are nailed with the right tools. Genome biology, 9(1), p.203.

Arguello, F. et al., 1998. Flavopiridol induces apoptosis of normal lymphoid cells, causes immunosuppression, and has potent antitumor activity In vivo against human leukemia and lymphoma xenografts. Blood, 91(7), pp.2482–90.

Arkin, M.R. & Wells, J. a, 2004. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. Nature reviews. Drug discovery, 3(4), pp.301–17.

Barrière, C. et al., 2007. Mice thrive without Cdk4 and Cdk2. Molecular oncology, 1(1), pp.72–83.

Barten, M. et al., 2005. Dynamics of β-Amyloid Reductions in Brain , Cerebrospinal Fluid , and Plasma of β-Amyloid Precursor Protein Transgenic Mice Treated with a γ-Secretase Inhibitor. The Journal of Pharmacology and Experimental Therapeutics, 312(2), pp.635–643.

Baumli, S. et al., 2008. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. The EMBO journal, 27(13), pp.1907–18.

Bendris, N. et al., 2011. Cyclin A2 mutagenesis analysis: a new insight into CDK activation and cellular localization requirements. PloS one, 6(7), p.e22879.

Best, J.D. et al., 2006. In Vivo Characterization of Aβ (40) Changes in Brain and Cerebrospinal Fluid Using the Novel γ -Secretase Inhibitor N- [cis-. The Journal of Pharmacology and Experimental Therapeutics, 317(2), pp.786–790.

Bible, K.C. et al., 2012. A Phase 2 Trial of Flavopiridol (Alvocidib) and Cisplatin in Platin-Resistant Ovarian and Primary Peritoneal Carcinoma: MC0261. Gynecologic oncology, 127(1), pp.55–62.

Blazer, L.L. & Neubig, R.R., 2009. Small molecule protein-protein interaction inhibitors as CNS therapeutic agents: current progress and future hurdles. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology, 34(1), pp.126–41.

Bleicher, K.H. et al., 2003. Hit and lead generation: beyond high-throughput screening. Nature reviews. Drug discovery, 2(5), pp.369–78.

Bleicher, K.H. et al., 2002. Parallel solution- and solid-phase synthesis of spirohydantoin derivatives as neurokinin-1 receptor ligands. Bioorganic & medicinal chemistry letters, 12(18), pp.2519–22.

Block, P. et al., 2007. Strategies to Search and Design Stabilizers of Protein – Protein Interactions : A Feasibility Study. Proteins: Structure, Function and Bioinformatics, 186(September 2006), pp.170–186.

Buckingham, S., 2004. Picking the Pockets of Protein-Protein Interactions Charting Chemical Space, (April), pp.1–4.

Carter, M.D., Simms, G. A & Weaver, D.F., 2010. The development of new therapeutics for Alzheimer's disease. Clinical pharmacology and therapeutics, 88(4), pp.475–86.

Chao, S.H. et al., 2000. Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. The Journal of biological chemistry, 275(37), pp.28345–8.

Chen, R. et al., 2005. Transcription inhibition by flavopiridol: mechanism of chronic lymphocytic leukemia cell death. Blood, 106(7), pp.2513–9.

Cheng, K.-Y. et al., 2003. The crystal structure of the human polo-like kinase-1 polo box domain and its phospho-peptide complex. The EMBO journal, 22(21), pp.5757–68.

Cheng, K.-Y. et al., 2006. The role of the phospho-CDK2/cyclin A recruitment site in substrate recognition. The Journal of biological chemistry, 281(32), pp.23167–79.

Chun, A.W. et al., 2009. Preclinical pharmacokinetics and in vitro activity of ON 01910.Na, a novel anti-cancer agent. Cancer chemotherapy and pharmacology, 65(1), pp.177–86.

Clark, R.C. et al., 2009. The isolation, total synthesis and structure elucidation of chlorofusin, a natural product inhibitor of the p53-MDM2 protein-protein interaction. Nat Prod Rep, 26(4), pp.465–477.

Cochran, A G., 2000. Antagonists of protein-protein interactions. Chemistry & biology, 7(4), pp.R85–94.

Coma, I. et al., 2009. Process validation and screen reproducibility in high-throughput screening. Journal of biomolecular screening, 14(1), pp.66–76.

de Cárcer, G., Manning, G. & Malumbres, M., 2011. From Plk1 to Plk5: Functional evolution of polo-like kinases. Cell Cycle, 10(14), pp.2255–2262.

Davenport, J. et al., 2012. Gambogic acid, a natural product inhibitor of Hsp90. Journal of natural products, 74(5), pp.1085–1092.

David, E., Tramontin, T. & Zemmel, R., 2009. Pharmaceutical R&D: the road to positive returns. Nature reviews. Drug discovery, 8(8), pp.609–10.

David-Pfeuty, T., 1999. Potent inhibitors of cyclin-dependent kinase 2 induce nuclear accumulation of wild-type p53 and nucleolar fragmentation in human untransformed and tumor-derived cells. Oncogene, 18(52), pp.7409–22.

Day, P.J. et al., 2009. Crystal structure of human CDK4 in complex with a D-type cyclin. Proceedings of the National Academy of Sciences of the United States of America, 106(11), pp.4166–70.

Dietzmann, K. et al., 2001. Increased human polo-like kinase-1 expression in gliomas. Journal of neuro-oncology, 53(1), pp.1–11.

Do, T.N. et al., 2003. Preferential induction of necrosis in human breast cancer cells by a p53 peptide derived from the MDM2 binding site. Oncogene, 22(10), pp.1431–44.

Drees, M. et al., 1997. Flavopiridol (L86-8275): selective antitumor activity in vitro and activity in vivo for prostate carcinoma cells. Clinical cancer research : an official journal of the American Association for Cancer Research, 3(2), pp.273–9.

Earley, S. et al., 2012. In vivo imaging of drug-induced mitochondrial outer membrane permeabilization at single-cell resolution. Cancer research, 72(12), pp.2949–56.

Eckerdt, F. & Strebhardt, K., 2006. Polo-like kinase 1: target and regulator of anaphase-promoting complex/cyclosome-dependent proteolysis. Cancer research, 66(14), pp.6895–8.

Elez, R. et al., 2003. Tumor regression by combination antisense therapy against Plk1 and Bcl-2. Oncogene, 22(1), pp.69–80.

Elia, A.E.H. et al., 2003. The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain. Cell, 115(1), pp.83–95.

Emilien, G. et al., 2000. Impact of genomics on drug discovery and clinical medicine. QJM : monthly journal of the Association of Physicians, 93(7), pp.391–423.

Emsley P, Cowtan K (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D60, 2126-2132.

Emsley P, et al. (2010) Features and Development of Coot. Acta Crystallographica D 66, 486-501

Essmann, F. & Schulze-Osthoff, K., 2012. Translational approaches targeting the p53 pathway for anti-cancer therapy. British journal of pharmacology, 165(2), pp.328–44.

Filippakopoulos, P. et al., 2011. Selective inhibition of BET bromodomains. Nature, 468(7327), pp.1067–1073. Filippakopoulos, P. et al., 2010. Selective inhibition of BET bromodomains. Nature, 468(7327), pp.1067–73.

Fox, J.T. & Myung, K., 2012. Cell - based high - throughput screens for the discovery of chemotherapeutic agents ABSTRACT : Oncotarget, 3(5), pp.581–585.

Fricker, S.P., 2007. Metal based drugs: from serendipity to design. Dalton transactions (Cambridge, England : 2003), (43), pp.4903–17.

Genis, D., Kirpichenok, M. & Kombarov, R., 2012. A minimalist fragment approach for the design of natural-product-like synthetic scaffolds. Drug discovery today, 00(00), pp.1–5.

Gerwick, W.H. & Moore, B.S., 2003. Lessons from the Past and Charting the Future of Marine Natural Products Drug Discovery and Chemical Biology. Chaos, Solitons & Fractals, 15(5), p.II.

Girard, F. et al., 1991. Cyclin A is required for the onset of DNA replication in mammalian fibroblasts. Cell, 67(6), pp.1169–79.

Gong, D. & Ferrell, J.E., 2010. The Roles of Cyclin A2 , B1 , and B2 in Early and Late Mitotic Events. Molecular biology of the Cell, 21, pp.3149–3161.

Guan, R. et al., 2005. Small interfering RNA-mediated Polo-like kinase 1 depletion preferentially reduces the survival of p53-defective, oncogenic transformed cells and inhibits tumor growth in animals. Cancer research, 65(7), pp.2698–704.

Gumireddy, K. et al., 2005. ON01910, a non-ATP-competitive small molecule inhibitor of Plk1, is a potent anticancer agent. Cancer cell, 7(3), pp.275–86.

Guo, C. et al., 2012. Design of oxobenzimidazoles and oxindoles as novel androgen receptor antagonists. Bioorganic & medicinal chemistry letters, 22(7), pp.2572–8.

Huggins, D.J., Sherman, W. & Tidor, B., 2012. Rational approaches to improving selectivity in drug design. Journal of medicinal chemistry, 55(4), pp.1424–44.

Högberg, L.D., Heddini, A. & Cars, O., 2010. The global need for effective antibiotics: challenges and recent advances. Trends in pharmacological sciences, 31(11), pp.509–15.

Ito, T. et al., 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences of the United States of America, 98(8), pp.4569–74.

Jain, S.V. et al., 2012. 2D, 3D-QSAR and docking studies of 1,2,3-thiadiazole thioacetanilides analogues as potent HIV-1 non-nucleoside reverse transcriptase inhibitors. Organic and medicinal chemistry letters, 2(1), p.22.

Jang, Y.-J. et al., 2002. Functional studies on the role of the C-terminal domain of mammalian polo-like kinase. Proceedings of the National Academy of Sciences of the United States of America, 99(4), pp.1984–9.

Jordan, M. a et al., 1993. Mechanism of mitotic block and inhibition of cell proliferation by taxol at low concentrations. Proceedings of the National Academy of Sciences of the United States of America, 90(20), pp.9552–6.

Jordan, M.A. & Wilson, Leslie, 2004. Microtubules as a target for anticancer drugs. Nature reviews Cancer, 4(April), pp.253–265.

Kelland, L.R., 2000. Flavopiridol, the first cyclin-dependent kinase inhibitor to enter the clinic : current status. Expert Opinion on Investigational Drugs, 9(12), pp.2903–2911.

Kesselheim, A.S., 2007. Intellectual property policy in the pharmaceutical sciences: the effect of inappropriate patents and market exclusivity extensions on the health care system. The AAPS journal, 9(3), pp.E306–11.

Knecht, R. et al., 1999. Prognostic Significance of Polo-like Kinase (PLK) Expression in Squamous Cell Carcinomas of the Head and Neck Advances in Brief Prognostic Significance of Polo-like Kinase (PLK) Expression in Squamous Cell Carcinomas of the Head and Neck 1. Cancer research, 59, pp.2794–2797.

Kothe, M., Kohls, D., Low, S., Coli, R., Rennie, G.R., et al., 2007. Selectivity-determining residues in Plk1. Chemical biology & drug design, 70(6), pp.540–6.

Kothe, M., Kohls, D., Low, S., Coli, R., Cheng, A.C., et al., 2007. Structure of the catalytic domain of human polo-like kinase 1. Biochemistry, 46(20), pp.5960–71.

König, a et al., 1997. The novel cyclin-dependent kinase inhibitor flavopiridol downregulates Bcl-2 and induces growth arrest and apoptosis in chronic B-cell leukemia lines. Blood, 90(11), pp.4307–12.

Lam, L.T. et al., 2001. Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. Genome biology, 2(10)

Lauria, A. et al., 2010. Molecular Modeling Approaches in the Discovery of New Drugs for Anti-Cancer Therapy: The Investigation of p53-MDM2 Interaction and its Inhibition by Small Molecules. Current Medicinal Chemistry, 17(28), pp.3142–3154.

Lee, K S et al., 1998. Mutation of the polo-box disrupts localization and mitotic functions of the mammalian polo kinase Plk. Proceedings of the National Academy of Sciences of the United States of America, 95(16), pp.9301–6.

Lee, Semin et al., 2009. Structural interactomics: informatics approaches to aid the interpretation of genetic variation and the development of novel therapeutics. Molecular biosystems, 5, pp.1456–1472.

Liu, G. et al., 2004. A Phase II Trial of Flavopiridol (NSC #649890) in Patients with Previously Untreated Metastatic Androgen-Independent Prostate Cancer. Clinical cancer research, 10, p.924.

Liu, S. et al., 2012. Structural and functional analysis of cyclin D1 reveals p27 and substrate inhibitor binding requirements. ASC Chemical Biology, 5(12), pp.1169–1182.

Lowery, D.M. et al., 2007. Proteomic screen defines the Polo-box domain interactome and identifies Rock2 as a Plk1 substrate. The EMBO journal, 26(9), pp.2262–73.

Lu, L.-Y. et al., 2008. Polo-like kinase 1 is essential for early embryonic development and tumor suppression. Molecular and cellular biology, 28(22), pp.6870–6.

Luke, J.J. et al., 2012. The cyclin-dependent kinase inhibitor flavopiridol potentiates doxorubicin efficacy in advanced sarcomas: preclinical investigations and results of a phase I dose-escalation clinical trial. Clinical cancer research : an official journal of the American Association for Cancer Research, 18(9), pp.2638–47.

Mackey, T.K. & Liang, B. A, 2012. Promoting global health: utilizing WHO to integrate public health, innovation and intellectual property. Drug discovery today, 00(00), pp.10–13.

Malumbres, M., 2011. Physiological relevance of cell cycle kinases. Physiology Review, 91(3), pp.973–1007.

Malumbres M, & Barbacid M. 2009 Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer. 9(3):153-66.

Marine, J.-C., 2010. Pharmacological rescue of p53 in cancer therapy: widening the sensitive tumor spectrum by targeting MDMX. Cancer cell, 18(5), pp.399–400.

Matulis, D et al., 2005. Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor Biochemistry, 44, pp. 5258–5266

Mayer, A.M.S. et al., 2010. The odyssey of marine pharmaceuticals: a current pipeline perspective. Trends in pharmacological sciences, 31(6), pp.255–65.

Mayr, L.M. & Bojanic, D., 2009. Novel trends in high-throughput screening. Current opinion in pharmacology, 9(5), pp.580–8.

Mazanetz, M.P. et al., 2011. Prediction of cyclin-dependent kinase 2 inhibitor potency using the fragment molecular orbital method. Journal of cheminformatics, 3(1), p.2.

Merckx A, et al., 2008. Structures of P. falciparum protein kinase 7 identify an activation motif and leads for inhibitor design. 16(2):228-38.

Milletti, F., 2012. Cell-penetrating peptides: classes, origin, and current landscape. Drug discovery today, 00(00).

Morelli, X., Bourgeas, R. & Roche, P., 2011. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). Current opinion in chemical biology, 15(4), pp.475–81.

Mullard, A., 2012. Protein – protein interaction inhibitors. Nature reviews Drug discovery, 11(March), pp.173–175.

Mundt, K.E. et al., 1997. On the regulation and function of human polo-like kinase 1 (PLK1): effects of overexpression on cell cycle progression. Biochemical and biophysical research communications, 239(2), pp.377–85.

Murugan, R.N. et al., 2011. Plk1-targeted small molecule inhibitors: molecular basis for their potency and specificity. Molecules and cells, 32(3), pp.209–20.

Noble, M. et al., 2005. Exploiting structural principles to design cyclin-dependent kinase inhibitors. Biochimica et biophysica acta, 1754(1-2), pp.58–64.

Pammolli, F., Magazzini, L. & Riccaboni, M., 2011. The productivity crisis in pharmaceutical R&D. Nature reviews. Drug discovery, 10(6), pp.428–38.

Park, H., Jeong, Y. & Hong, S., 2012. Structure-based de novo design and biochemical evaluation of novel BRAF kinase inhibitors. Bioorganic & medicinal chemistry letters, 22(2), pp.1027–30.

Park, J. et al., 2011. Polo-Box Domain : a versatile mediator of polo-like kinase function. Cell Mol Life Sci, 67(12), pp.1957–1970.

Patterson, C. et al., 2008. Diagnosis and treatment of dementia: 1. Risk assessment and primary prevention of Alzheimer disease. Canadian Medical Association Journal, 178(5), pp.548–556.

Plescia, J. et al., 2005. Rational design of shepherdin, a novel anticancer agent. Cancer cell, 7(5), pp.457–68.

Popowicz, G.M., Czarna, A. & Holak, T. a, 2008. Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain. Cell cycle (Georgetown, Tex.), 7(15), pp.2441–3.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical
Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reindl, W. et al., 2009. A pan-specific inhibitor of the polo-box domains of polo-like kinases arrests cancer cells in mitosis. Chembiochem : a European journal of chemical biology, 10(7), pp.1145–8.

Reindl, W. et al., 2008. Inhibition of polo-like kinase 1 by blocking polo-box domain-dependent protein-protein interactions. Chemistry & biology, 15(5), pp.459–66.

Rew, Y. et al., 2012. Structure-based design of novel inhibitors of the MDM2-p53 interaction. Journal of medicinal chemistry, 55(11), pp.4936–54.

da Rocha, A.B., Lopes, R.M. & Schwartsmann, G., 2001. Natural products in anticancer therapy. Current Opinion in Pharmacology, 1, pp.364–369.

Rudin, C.M. et al., 2012. Phase II study of single-agent navitoclax (ABT-263) and biomarker correlates in patients with relapsed small cell lung cancer. Clinical cancer research: an official journal of the American Association for Cancer Research, 18(11), pp.3163–9.

Santonico, E., Castagnoli, L. & Cesareni, G., 2005. Methods to reveal domain networks. Drug discovery today, 10(16), pp.1111–1117. Scannell, J.W. et al., 2012. Diagnosing the decline in pharmaceutical R&D efficiency. Nature reviews. Drug discovery, 11(3), pp.191–200.

Schrump, D.S. et al., 1998. Flavopiridol mediates cell cycle arrest and apoptosis in esophageal cancer cells. Clinical cancer research : an official journal of the American Association for Cancer Research, 4(11), pp.2885–90.

Schwartz, G K et al., 2001. Phase II study of the cyclin-dependent kinase inhibitor flavopiridol administered to patients with advanced gastric carcinoma. Journal of clinical oncology: official journal of the American Society of Clinical Oncology, 19(7), pp.1985–92.

Senderowicz, A.M., 2003. Small-molecule cyclin-dependent kinase modulators. Oncogene, 22(42), pp.6609–20.

Senisterra, G., Chau, I. & Vedadi, M., 2012. Thermal denaturation assays in chemical biology. Assay and drug development technologies, 10(2), pp.128–36.

Shangary, S. & Wang, S., 2008. Targeting the MDM2-p53 interaction for cancer therapy. Clinical cancer research : an official journal of the American Association for Cancer Research, 14(17), pp.5318–24.

Shapiro, G.I. et al., 1999. Flavopiridol Induces Cell Cycle Arrest and p53-independent Apoptosis in Non-Small Cell Lung Cancer Cell Lines Flavopiridol Induces Cell Cycle Arrest and p53-independent. Clinical cancer research, 5, pp.2925–2938.

Shenoy, S.R. et al., 2010. Structural and functional analyses of minimal phosphopeptides targeting the polo-box domain of polo-like kinase 1. Nat Struct Mol Biol, 16(8), pp.876–882.

Shlaes, D., 2003. The abandonment of antibacterials: why and wherefore? Current Opinion in Pharmacology, 3(5), pp.470–473.

Skariyachan, S. et al., 2012. Structure based virtual screening of novel inhibitors against multidrug resistant superbugs. Bioinformation, 8(9), pp.420–5.

Smith, A.J.T. et al., 2010. Beyond picomolar affinities: quantitative aspects of noncovalent and covalent binding of drugs to proteins. Journal of medicinal chemistry, 52(2), pp.225–233.

Song, S et al., 2000. Essential function of the polo box of Cdc5 in subcellular localization and induction of cytokinetic structures. Molecular and cellular biology, 20(1), pp.286–98.

Spankuch-Schmitt, B. et al., 2002a. Downregulation of human polo-like kinase activity by antisense oligonucleotides induces growth inhibition in cancer cells. Oncogene, 21, pp.3162–3171.

Spänkuch-Schmitt, B. et al., 2002b. Effect of RNA silencing of polo-like kinase-1 (PLK1) on apoptosis and spindle formation in human cancer cells. Journal of the National Cancer Institute, 94(24), pp.1863–77.

Spencer, R.W., 1998. High-throughput screening of historic collections: observations on file size, biological targets, and file diversity. Biotechnology and bioengineering, 61(1), pp.61–7.

Stahura, F.L. et al., 1999. Molecular scaffold-based design and comparison of combinatorial libraries focused on the ATP-binding site of protein kinases. Journal of molecular graphics & modelling, 17(1), pp.1–9, 51–2.

Strebhardt, K., 2010. Multifaceted polo-like kinases: drug targets and antitargets for cancer therapy. Nature reviews. Drug discovery, 9(8), pp.643–60.

Takahashi, T. et al., 2002. Polo-like kinase 1 (PLK1) is overexpressed in primary colorectal cancers. Cancer science, 1.

Takai, N. et al., 2001a. Expression of polo-like kinase in ovarian cancer is associated with histological grade and clinical stage. Cancer letters, 164(1), pp.41–9.

Takai, N. et al., 2001b. Polo-like kinase (PLK) expression in endometrial carcinoma. Cancer letters, 169(1), pp.41–9.

Takeda, D.Y., Wohlschlegel, J. a & Dutta, a, 2001. A bipartite substrate recognition motif for cyclin-dependent kinases. The Journal of biological chemistry, 276(3), pp.1993–7.

Thal, L.J. et al., 2006. The role of biomarkers in clinical trials for Alzheimer disease. Alzheimer disease and associated disorders, 20(1), pp.6–15.

Thiel, P., Kaiser, M. & Ottmann, C., 2012. Small-molecule stabilization of protein-protein interactions: an underestimated concept in drug discovery? Angewandte Chemie (International ed. in English), 51(9), pp.2012–8.

Toledo, L.M. & Lydon, N.B., 1997. Structures of staurosporine bound to CDK2 and cAPK — new tools for structure-based design of protein kinase inhibitors Protein kinases are involved in a wide variety of signal. Current Biology, 5, pp.1551–1556.

Tse, C. et al., 2008. ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. Cancer research, 68(9), pp.3421–8.

Tsvetkov, L. & Stern, D.F., 2005. Phosphorylation of Plk1 at S137 and T210 is Inhibited in Response to DNA Damage. Cell Cycle, 4(1), pp.166–171.

Tzakos, A.G. et al., 2011. Targeting oncogenic protein-protein interactions by diversity oriented synthesis and combinatorial chemistry approaches. Molecules (Basel, Switzerland), 16(6), pp.4408–27.

Uetz, P. et al., 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature, 403(6770), pp.623–7.

Utomo, D.H., Widodo, N. & Rifa'i, M., 2012. Identifications small molecules inhibitor of p53-mortalin complex for cancer drug using virtual screening. Bioinformation, 8(9), pp.426–9.

Vallance, P., Williams, P. & Dollery, C., 2010. The future is much closer collaboration between the pharmaceutical industry and academic medical centers. Clinical pharmacology and therapeutics, 87(5), pp.525–7.

Vassilev, L.T., 2007. MDM2 inhibitors for cancer therapy. Trends in molecular medicine, 13(1), pp.23–31.

Walker, D.H. & Maller, J.L., 1991. Role for cyclin A in the dependence of mitosis on completion of DNA replication. Nature, 354, pp.314–317.

Wang, Jiabel et al., 2010. MDM2 antagonist can inhibit tumor growth in hepatocellular carcinoma with different types of p53 in vitro. Journal of Gastroenterology and Hepatology, 26(2), pp.371–377.

Waxman DJ, Strominger JL. 1983 Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics. Annu Rev Biochem. 52: 825-69.

Wells, J. A & McClendon, C.L., 2007. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature, 450(7172), pp.1001–9.

Wesierska-Gadek J & Schmid G. 2006 Dual action of the inhibitors of cyclin-dependent kinases: targeting of the cell-cycle progression and activation of wild-type p53 protein. Expert Opin Investig Drugs 15(1):23-38.

Whitty, A. & Kumaravel, G., 2006. Between a rock and a hard place? Nature chemical biology, 2(3), pp.112–8.

Wilson, W.H. et al., 2010. Safety, Pharmacokinetics, Pharmacodynamics and Activity of Navitoclax, a targeted high affinity inhibitor of BCL-2, in lymphoid malignancies. Lancet Oncology, 11(12), pp.1149–1159.

M. D. Winn et al. 2011 Overview of the CCP4 suite and current developments. Acta. Cryst. D67, 235-242

Wolf, G. et al., 1997. Prognostic significance of polo-like kinase (PLK) expression in non-small cell lung cancer. Oncogene, 14, pp.543–549.

Wu, S.-Y. & Chiang, C.-M., 2007. The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. The Journal of biological chemistry, 282(18), pp.13141–5.

Yap, T.A. & Workman, P., 2012. Exploiting the Cancer Genome: Strategies for the Discovery and Clinical Development of Targeted Molecular Therapeutics. Annual Review of Pharmacology and Toxicology, 52(1), pp.549–573.

Zhang, C. & Bollag, G., 2010. Scaffold-based design of kinase inhibitors for cancer therapy. Current opinion in genetics & development, 20(1), pp.79–86.

Zhang, Lei et al., 2011. Homology modeling, molecular dynamic simulation and docking studies of cyclin dependent kinase 1. Journal of molecular modeling, 17(2), pp.219–26.

Zhang, Lianru et al., 2010. Gambogic acid inhibits Hsp90 and deregulates TNF-α/NF-κB in HeLa cells. Biochemical and biophysical research communications, 403(3-4), pp.282–7.

Zuber, J. et al., 2012. RNAi screen identifies Br4 as a therapeutic target in acute myeloid leukaemia. Nature, 478(7370), pp.524–528.

APPENDIX A
MATERIALS AND METHODS

A0 Materials

Chemicals were ordered from Sigma. Columns and media for chromatography were supplied by GE Healthcare or Qiagen. Crystallisation solutions and trays were provided by Hampton Research, Emerald Biosystems, Qiagen and Molecular Dimensions. Peptides and fluorescent peptides were provided by Severn Biotech. DNA primers were ordered from IDT Technologies and DNA sequencing was provided by Source BioScience. Columns and media for chromatography were supplied by GE Healthcare or Qiagen.

A1 Molecular biology: cloning

A1.1 Cloning templates

Plasmid pGEX-6P1 (GE Healthcare) encoding the cDNA for human MDM2 (residues 16-109 or residues 16-125) was kindly provided by Christiane Riedinger. Plasmid pGEX-6P1 encoding cDNAs for: full-length human pRb, full-length human CDK2 and Civ1, and plasmid pET-46A encoding the cDNAs for full-length human or bovine Cyclin A with a C-terminal His tag were kindly provided by Aude Echalier Glazer. Plasmid pGEX-6P1 encoding the cDNA for human Plk1 (Polobox domain) was kindly provided by Rob Cheng. Plasmid pGEX-6P1 encoding the cDNA for human MDMX (MDM4) residues 18-111, 19-111, 23-111 and 22-111, were kindly provided by Jane Endicott.

A1.2 Purification of recombinant plasmid

In order to provide sufficient plasmid stocks for performing multiple transformations for each project, the plasmids were transformed into *E. coli* strain DH5α (competent cells grown in-house from original stocks from Clontech) and plated onto LB-agar plates supplemented with the appropriate

antibiotic (100μg/ml ampicillin and/or 35μg/ml chloramphenicol). To purify recombinant plasmid, single colonies were picked from the plates and inoculated into 5ml LB in Falcon tubes, supplemented with the approproate antibiotic to make starter cultures, then incubated at 37°C overnight with shaking at 250rpm. Plasmids were purified using the Qiagen mini-prep kit in accordance with the manufacturer's instructions. In brief, concentrated overnight cultures were pelleted by centrifuging at 13500 xg for 2 minutes at 4°C in a bench-top centrifuge. Cell pellets were re-suspended in the provided buffer containing RNAse A, then lysed with the provided SDS buffer, in alkaline conditions under which DNA and protein precipitate. The solution was then neutralised and buffer exchanged into a high salt buffer, spun at 13500 xg for 10 minutes to remove cell debris, and the supernatant was applied to a miniprep spin column. The plasmid bound to the silica membrane whilst high salt conditions remained, and after washing with ethanol, was eluted with a low salt buffer (Tris pH 7.5). Plasmid stocks were typically >100ng/μl and were stored at -20°C.

A1.3 Polymerase chain reaction

To create new constructs, polymerase chain reactions were carried out using high fidelity KOD HotStart polymerase (Millipore) or Phusion® polymerase (Thermo Scientific). PCR primers were ordered from IDT Technologies. Primers were designed using PrimerX. Forward and reverse primers flanking the cDNA region to be amplified contained a restriction endonuclease site and additional flanking sequences (1-6 nucleotides) to facilitate the restriction digest. PCR reactions were set up as follows: 4μmoles forward and reverse primers, 0.2mM dNTPs, 2mM $MgCl_2$, PCR buffer (120mM Tris HCl, 10mM KCl, 6mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 0.001% BSA, final pH 8.0), 0.5μg template DNA and 1μl of 2.5U/μl DNA polymerase. The total reaction volume was 50μl (made up with milli-Q water). PCR reactions were carried out in a PerkinElmer GeneAmp PCR 2400 thermocycler using the temperature cycle: 98°C for 2 minutes (for initial denaturation), then 30 cycles of: 2 minutes at 98°C (denaturation), 10 seconds at 50-55°C (annealing), 15 seconds/kb at 72°C (extension) and at the end, 10

minutes at 72°C (final extension). Reactions were then stored at -20°C for further use.

A1.4 Ligation

Ligation of PCR products into plasmid vectors was performed as follows: following digestion of the purified PCR product and plasmid using relevant restriction enzyme (for 3-4 hours at 37°C), samples were analysed using agarose gel electrophoresis as detailed in Appendix A1.10. The plasmid and PCR fragment were purified from the agarose gel bands using the Qiagen gel extraction kit (according to the Manufacturer's instructions). Ligation of the purified PCR fragment into the plasmid using T4 DNA ligase was carried out overnight at room temperature, using 1.5μl of the vector mix, 3μl of the PCR product insert, 1μl of 400U/μl T4 DNA ligase and T4 ligase buffer (50mM Tris HCl, 10mM $MgCl_2$, 1mM ATP, 10mM DTT, 25μg/ml BSA, final pH 7.5), with a total reaction volume of 15μl. The ligation product was transformed into DH5α cells, they were plated on LB agar plates containing the appropriate antibiotic, and incubated overnight at 37°C. Single colonies were picked and inoculated into 5ml LB supplemented with the appropriate antibiotics. Plasmids were purified using the Qiagen Mini-prep kit (as detailed in Appendix A1.2) and positive clones were identified by restriction digest.

A1.5 Restriction digest

DNA was digested using restriction endonuclease: 20μl DNA was incubated with restriction buffer (composition depending on endonuclease used), 0.1mg/ml BSA and 1μl restriction endonuclease. The progress of the reaction was checked using agarose gel electrophoresis (as detailed in Appendix A1.10)

A1.6 Alkaline phosphatase

After restriction digest, the vector was treated with 1U phosphatase (NEB) for 1 hour at 37°C, to remove terminal 5' phosphate groups and prevent preferential recircularisation. The plasmid was then purified using the Qiagen plasmid gel extraction kit.

A1.7 Transformation

Plasmids were transformed into DH5α cells for cloning purposes, or into BL21(DE3), Rosetta2, or B834 (DE3) pLysS cells for protein expression. 1μl of plasmid DNA was mixed with 50μl competent cells in a 13ml round-bottomed transformation tube (Greiner) and incubated on ice for 15 minutes. The mixture was heat-shocked at 42-45°C for 35-45 seconds (depending on strain) to cause plasmid uptake, returned to an ice bath for 3 minutes, and 500μl NZY+ broth at 37°C was added. Reactions were grown for 1 hour at 37°C, with shaking at 225rpm, before plating out on LB agar plates supplemented with the appropriate antibiotic.

A1.8 Preparation of competent cells

Competent cells (*E. coli* strains XL1-gold, XL1-blue or DH5α) were prepared for use in transformations using the following method. Fresh cells were streaked from a glycerol stock on LB-agar supplemented with antibiotics as appropriate and grown overnight at 37°C. A single colony was picked to grow an overnight culture at 37 °C in 5 mL of LB and appropriate antibiotics. The overnight culture was used to inoculate 200 mL of LB which was incubated with vigorous shaking at 37 °C. When the cells reached the exponential growth phase (OD_{600} ~ 0.5), they were harvested by centrifugation for 10 minutes at 3000×g and the supernatant discarded. The following steps were carried out at 4 °C. Cells were re-suspended in 40 mL of cold sterile TFB-I buffer (30 mM potassium acetate, 100 mM rubidium chloride, 10 mM calcium chloride, 50 mM manganese chloride,

15% glycerol and a pH adjusted to 5.8 with acetic acid) and incubated for 30 minutes on ice. Cells were then centrifuged for 10 minutes at 3,000×g and the supernatant discarded. The pellet was resuspended in 8 mL of cold sterile TFB-II buffer (10 mM MOPS, 10 mM rubidium chloride, 75 mM calcium chloride, 15% glycerol and a pH adjusted to 6.8 with NaOH) and then transferred to 1.5 mL Eppendorf tubes in 200 μL aliquots, flash-frozen in liquid nitrogen and stored at -80 °C.

A1.9 Site-directed mutagenesis

Forward and reverse primers were designed using PrimerX, and purchased from IDT technologies. Site-directed mutagenesis was performed using the QuikChange II kit (Stratagene), following the manufacturer's instructions. Use of PCR was as described in Appendix A1.3. DNA sequencing, using the Source Bioscience facility from the Department of Biochemistry, Oxford, confirmed introduction of the correct mutation.

A1.10 Agarose gel electrophoresis

Agarose gels were made using 0.35g agarose, 50 ml TBE (45mM Tris-borate, 1mM EDTA) heated in a microwave to dissolve the agarose. Once cooled to lukewarm, 1μl ethidium bromide was added (final concentration 0.5μg/ml). Gels were poured and allowed to set. DNA samples were mixed with 6× loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol) and loaded into the wells on the agarose gel. Electrophoresis was performed in TBE buffer at 30-70V. DNA bands were visualised with a ChemiDoc UV imaging system (BioRad).

A2 Protein expression and purification

Quantities and specific conditions for protein expression with different constructs are listed in each chapter during the discussion of the method, including diversions from this general method.

A2.1 Bacterial culture

Bacterial cultures were grown using one of the following media (Table A.1: recipes given for 1l of broth/agar).

Bacterial culture media	Recipe (1l)
Lysogeny broth (LB)	10g tryptone
	5g yeast extract
	10g NaCl
LB agar	10g tryptone
	5g yeast extract
	10g NaCl
	16g agar
Terrific broth (TB)	12g tryptone
	24g yeast extract
	4ml 87% glycerol
	100ml 10x salts
10x salts (for TB)	23.13g KH_2PO_4
	123.41g K_2HPO_4
NZY+ broth	10g NZ amine (casein hydrolysate)
	5g yeast extract
	5g NaCl
	(Adjust volume to 1L and pH to 7.5
	Just before use add):
	12.5ml 1M $MgCl_2$
	12.5ml 1M $MgSO_4$
	20ml 20% (w/v) glucose

Table A.1 Recipes for bacterial culture media Values given are per litre of media

Constructs for protein expression in pGEX6P-1 or pET vectors were transformed in BL21(DE3), Rosetta2, or B834 (DE3) pLysS *E. coli* competent cells and plated onto LB-agar with appropriate antibiotics (100 mg/L ampicillin or 35 μg/ml chloramphenicol). Bacterial cultures were started (in 5-50ml LB or TB) from picking single or multiple colonies from the streaked or spread LB-agar plates or from glycerol stocks of competent cells (kept at -80°C) and grown overnight at 37°C. Bacterial cultures were then scaled up by using the saturated culture to inoculate flasks containing LB or TB (supplemented with the appropriate

antibiotic) in flasks (typically 0.5l in 2l flasks or 1l in 5l flasks) for large-scale expressions. Bacterial growth was maintained at 37°C with shaking at 220rpm for 4-8 hours until an OD_{600} of 0.6-0.8 (depending on protein prep) was achieved. Growth was then stopped by reducing the temperature for 1 hour to 18-20°C, with shaking at 75rpm, and protein expression was induced using 20mM IPTG and maintained overnight for 16-22 hours (depending on protein prep) at 18-20°C with shaking at 200rpm. Cells were harvested by centrifugation (4°C, 20-25 minutes, 4000g) and cell pellets were flash frozen in liquid nitrogen and stored at -20°C or -80°C.

A2.2 Protein purification

A2.2.1 Preparation of cell lysate

Cell lysate was obtained from cell pellets as follows: pellets were thawed under running water and sonicated for 1-2 minutes (using pulses of: 20 sec on, 40 sec off, power 5.5, on ice) to lyse cells. Cell debris was then pelleted using centrifugation (at 18500rpm, 4°C, 1 hour) and lysates were removed and kept on ice for use.

A2.2.2 Affinity chromatography

Gravity flow affinity chromatography was performed as a first stage in the prep in order to recover the protein of interest, using columns with a bed volume of 2-3ml at 4°C. For GST-tagged proteins, Glutathione Sepharose 4B beads (GE Healthcare) were used; and for His-tagged proteins, Ni-NTA FastFlow beads (Qiagen) were used. All purifications carried out using a His-tag were performed in buffers containing <2mM DTT, to avoid reduction of the nickel. The procedure for use of affinity chromatography columns was as follows: lysates filtered with a 0.22 μm syringe filter were loaded and removed by gravity flow. During purification of the CDK2/cyclin A complex, lysate containing GST-CDK2 was loaded first, allowed to flow through, and then lysate containing cyclin A was passed down, so that a pull-down occurred. The column was then washed with

25-50ml of the appropriate buffer, and the proteins were eluted (with 20mM glutathione, pH 6.5 for GST-tagged proteins, and with 250mM imidazole, for His-tagged proteins). After purification by affinity chromatography, the GST tag was cleaved from proteins using 1:50 (w/w) 3C protease (3C-GST, prepared as detailed for other proteins in the appendices and kept as a 50% glycerol stock at -20°C. The resins were cleaned and regenerated (for glutathione sepharose: using 6 M urea, washed with water, stored in 20% ethanol and for Ni-NTA: using 100 mM EDTA (pH 8.0) and 6 M urea, rinsed in water, regenerated using 100 mM NiSO4 and stored in 20% ethanol) and re-used several times for the same protein.

A2.2.3 Gel filtration chromatography

Gel filtration chromatography was carried out using Akta FPLC chromatographic systems, controlled by the UNICORN software (GE Healthcare) in order to further purify proteins by size. Pumps were washed in the relevant buffers and columns (S75 or S200 preparatory grade) were pre-equilibrated in 1-1.5× column volume of the appropriate buffer. Proteins were concentrated to ~1-5 mg/ml, filtered and injected onto the column using a 5 or 10ml superloop. Proteins coming off the bottom of the column were detected using UV absorbance at 280nm, and 4ml fractions were collected using an automatic fraction collector. After use columns were stored in low-salt buffer or water, supplemented with 0.1% azide to inhibit bacterial growth. The protein content of fractions was checked using SDS-PAGE, following the method outlined in A3.2, and the desired fractions were pooled, concentrated to the desired value and either kept on ice or at 4°C and used immediately, or flash frozen in liquid nitrogen and stored at -80°C.

A2.2.4 Dialysis

For the preparation of some proteins, and in the course of investigations for Chapter 5 exploring covalent inhibition (explained in the main text) it was

necessary to perform dialysis to remove a contaminant or bound drug. This was performed for large volumes (2-20ml) in Snakeskin dialysis tubing of an appropriate MWCO (GE Healthcare) and for small volumes (<2ml) in pin dialysis units (Fisher). In each case, the membrane was soaked in the appropriate buffer for 15 minutes at 4°C, with gentle stirring. The protein was then loaded and dialysis was carried out over 24 hours, (at 4°C with gentle stirring) with the buffer changed at intervals of 4-6 hours (at least 4 changes).

A3 Molecular biology: protein characterisation

A3.1 Determination of protein concentration

When needed, proteins were concentrated using Vivaspin sample concentrators (GE Healthcare) of the appropriate volume (2 or 20ml) and Molecular Weight Cut-off (MWCO: 3-20kDa). Protein concentrations were determined using a NanoDrop to measure the UV absorbance at 280 nm, using extinction coefficients for each protein determined by Protparam (http://web.expasy.org/protparam/) from the amino acid sequence.

A3.2 SDS-PAGE

Tris-glycine gels ranging from 10-20% acrylamide content were used for separation and identification of proteins. For some experiments pre-cast gels of the relevant acrylamide content were purchased from Invitrogen, however most gels were made using the Bio-Rad Mini Protean 3 system, according to the recipes given in Table A2 (along with recipes for loading and running buffers). Samples for SDS-PAGE were mixed with 5x SDS loading buffer and heated at 90°C for 5 minutes before loading into the wells of the gel. Electrophoresis was carried out at 150-200V until the dye front reached the bottom of the gel. Gels were removed from their cases and stained using standard Coomassie stain (0.01% Coomassie G, 0.1% HCl, 10% ethanol).

Resolving gel (makes 15ml)	Stacking gel (makes 6ml)	Running buffer	SDS loading buffer (5x)
3.4-6.9ml H_2O	4.1ml H_2O	25mM Tris	250mM Tris pH 6.8
4.0-7.5ml 30% acrylamide mix	1ml 30% acrylamide mix	250mM glycine	500mM DTT
3.8ml 1.5M Tris (pH 8.8)	0.75ml 1M Tris (pH 6.8)	0.1% SDS	10% SDS
0.15ml 10% SDS	0.06ml 10% SDS		0.5% bromophenol blue
0.15ml 10% ammonium persulphate	0.06ml 10% ammonium persulphate		50% glycerol
0.009ml TEMED	0.006ml TEMED		

Table A2 Recipes for SDS-PAGE: gels and running buffer TEMED: Tetramethylethylenediamine

A3.3 Mass spectrometry

Protein mass was calculated by electrospray ionisation mass spectrometry (ESI-MS) after desalting of samples using Ziptips with C_4 resin (provided by Millipore). Samples used were at least 3.5mg/ml in concentration.

A4 SDS-PAGE bandshift assay

The SDS-PAGE bandshift assay follows the progress of a reaction using the shift in electrophoretic mobility observed upon addition of multiple phosphate groups or other groups of significant mass (in this case phosphorylation of the substrate pRb). SDS-PAGE was performed as detailed in Appendix A3.2, using samples from a timecourse following the phosphorylation of pRb by CDK2/cyclin A. In this way the effect of the presence of various inhibitors upon the reaction could be identified. The controls used were a negative control (no CDK2/cyclin A), a positive control (with DMSO instead of compounds) and inhibitor controls (with Staurosporine or RXL peptide instead of compounds).

A4.1 Preparation of reagents

The CDK2/cyclin A, pRb-GST, inhibitor, and kinase buffer stocks were all prepared in advance of the assay and frozen at -20°C in aliquots. The stocks were

as follows: CDK2/cyclin A at 200nM; pRb-GST at 3.5mg/ml; 10x kinase buffer contained 5mM ATP, 100mM $MgCl_2$ and 0.5M Tris-HCl pH 7.5.; compound stocks for inhibitors or control inhibitors were at 10mM.

A4.2 Timecourse of reaction

Each reaction tube was set up with 6μl of the compound stock (or RXL, Staurosporine or DMSO control stock), 9μl pRb-GST stock, 6μl 10x kinase buffer and 33μl ddH_2O. The reactions were performed on ice. The reaction was started by adding 6μl CDK2/cyclin A stock and stopped by removing a 10μl aliquot of the reaction mix and adding it to a pre-prepared tube containing a 10μl aliquot of 5x SDS-loading dye. Immediately after the timecourse was completed for all samples, the samples were heated at 90°C for 5 minutes and samples were analysed using SDS-PAGE, as described in Appendix A3.2. The results for these assays are given in Chapter 4, and in each case the upper band on the gel represents (hyper)phosphorylated pRb-GST and the lower band on the gel represents unphosphorylated pRb-GST.

A5 Radioactive kinase assay

Radioactive kinase assays still provide the 'gold standard' for analysis of kinase activity, since they directly follow phosphorylation. This assay measures the amount of phosphorylation of p107 by CDK2/cyclin A, when γ-^{32}P-ATP is provided as a substrate, and was used to screen compound libraries quickly to identify inhibitors of CDK2/cyclin A, as well as to obtain kinetic information for 'hit' compounds and confirm their action against recruited vs non-recruited substrates.

A5.1 Preparation of reagents and equipment

The CDK2/cyclin A, p107-GST or histone H1 substrates, inhibitors, and kinase buffer stocks were all prepared in advance of the assay and frozen at -20°C in aliquots. The stocks were as follows: 40nM CDK2/cyclin A, 150μM p107 (or

histone H1), 1mM inhibitors (or inhibitor controls) and 2x kinase buffer (containing 1M $MgCl_2$, 1M Tris pH 7.0). P81 papers were also cut before the assay was started using a 2cm diameter hole punch, and labelled in pencil with the reaction point to be applied. The bench was prepared with suitable equipment for a radioactive experiment to be undertaken.

Finally CDK2/cyclin A, p107-GST or histone H1 substrate and inhibitors or controls were pipetted into reaction tubes to give final reaction concentrations of 4nM CDK2/cyclin A, 15μM p107 (or other substrate), 100μM inhibitors (or control) and 1x kinase buffer (0.5M $MgCl_2$, 0.5M Tris pH 7.0) before the experiment, and a kinase buffer stock was prepared containing ^{32}P-γ-ATP (5μCi/10 reactions) and cold ATP (200μM).

A5.2 Timecourse of reaction

Each reaction was started with the ATP mixture and then a timecourse performed. After each timepoint a 10μl aliquot was removed and spotted onto the centre of a P81 paper circle, and the reaction was stopped by dropping the paper into a 1% phosphoric acid solution on a rocking platform.

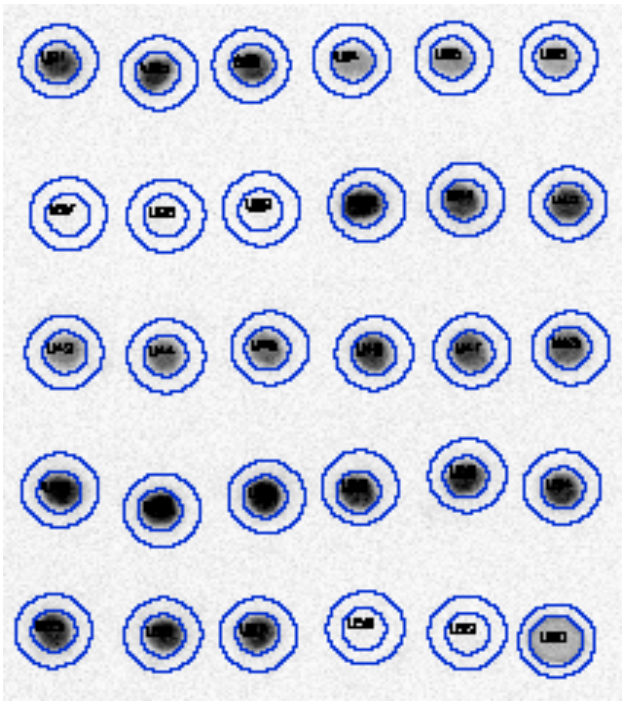
A5.3 Washing and mounting P81 papers

After the assay had been completed and all timepoints had been stopped, excess unreacted ^{32}P-ATP was removed by discarding the phosphoric acid solution and washing the P81 papers three times with water. This process appeared to be effective at removing background radiation from the papers.

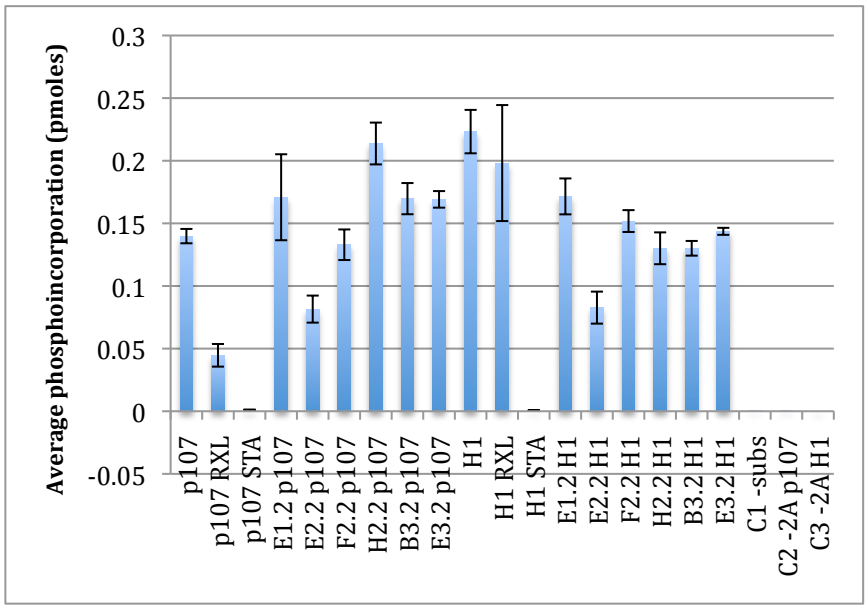
A5.4 Imaging and analysing P81 papers

The results were imaged by mounting the P81 papers using plastic wallets and tape, exposing the papers on a blanked image cassette overnight and developing the image using a Storm 860 imaging system. Phosphoincorporation (in picomoles) was quantified by analysing the image from the P81 papers, using the

program Quantity One (from BioRad), with reference to an ATP dilution control, after subtraction of the background radioactivity levels. An example of the format of the results for this assay is shown in Figure A.1.



A



B

Figure A.1 Format of results for radioactive kinase assay The assay was performed using P81 papers, the amount of phosphorylation incorporated imaged using a phosphorimager (Storm 860) after exposure of the plate overnight (example results shown in A), then the amount of phosphorylation incorporated was quantified using the program BioRad (results of quantification shown in B) **(A)** Image showing radioactive phosphate incorporation, from the exposed image cassette on completion of the reaction. Each spot represents one circle of P81 paper and one reaction time point. The density of the spot is proportional to the amount of radioactive phosphate incorporated. **(B)** Quantification of the image in **(A)**, showing the average

amount of phosphate incorporated for each reaction. The error bars represent the standard deviation between measurements repeated in triplicate.

The background was subtracted from the images by using Quantity One to calculate the pixel density within two circles, one only around the reaction spot, and one larger (each of fixed size for every reaction). The density in the outer ring was then calculated as: (*density larger circle–density smaller circle*) and then this background value was subtracted from the calculated density value in the smaller circle (spot), to give a background corrected density value. This was then used to compare standardised results between and within different experiments. An average was taken from reactions performed in triplicate and standard deviation calculated.

A6 Fluorescence anisotropy assay

A6.1 Preparation of reagents

Stock solutions of inhibitors and controls (40mM), protein (3μM) and fluorescent peptide (40nM) were prepared ahead of the experiments and frozen in appropriate aliquots for later use. HBS was prepared, containing 50mM Hepes pH 7.5, 250mM NaCl and 0.02% MTG. Just before performing the experiment, reaction tubes were made up from stocks, with each tube containing 10-20 nM fluorescent peptide, 65-130 nM protein and 0-200 μM inhibitors, with a 20 μl total volume (made up using HBS). In the experiments in Chapter 3, the protein used was PLK1 PBD and the fluorescent peptide used was 5-carboxyfluorescein-MQSpTPL (from Severn Biotech); in the experiments detailed in Chapter 4, the protein used was full-length human CDK2/cyclin A complex, and the fluorescent peptide used was 5-carboxyfluorescein-RRLFGE (from Severn Biotech).

A6.2 Preparation of completed plate

The fluorescent anisotropy well mix was transferred into rows of a black 384-well plate and incubated at room temperature for 30 minutes in order to allow equilibration and to eliminate bubbles. Several controls were included in the

experiments: a protein control with no fluorescent peptide, a protein + fluorescent peptide control (with no compound added) and a control with unlabelled peptide. In this way the 'all-bound' and 'all-unbound' states were characterised.

A6.3 Measurement of fluorescent anisotropy

Measurement of fluorescent anisotropy was made using the BMG PheraStar FS microplate reader, with temperature control and shaking. Anisotropy and intensity were output as a text file.

A6.4 Analysis of results

Results from fluorescence anisotropy assays were analysed using the program Prism, using the one-site binding displacement mode, to allow calculation of IC_{50} values from the binding curve.

A7 Thermal denaturation assay

The thermal denaturation assay uses changes in the thermal stability of a protein upon binding of a protein partner or a compound. In this assay, a protein is incubated with the dye SYPRO®-Orange and is then gradually denatured by raising the temperature in small steps in an RT-PCR machine. The dye binds to hydrophobic residues that are exposed as the protein becomes denatured and as a result the dye undergoes a shift in its fluorescent spectrum upon excitation at 465 nm (Merckx *et al*, 2008). The protein's stability can be changed by compound binding and this is monitored as a change in the protein's Tm value as compared to a control sample. The assay therefore provides a rapid method to identify compounds that bind to proteins, since a potential inhibitor would be expected to stabilise the protein and delay denaturation.

Thermal denaturation assays were performed as described in Merckx *et al*, 2008. The temperature dependence of the fluorescence occurring during the protein

denaturation process can be represented by the equation:

$$y(T) = y_F + \frac{y_U - y_F}{1 + e^{\Delta uG_{(T)} / RT}}$$

(where ΔuG is the difference in unfolding free energy between the folded and unfolded state, R is the gas constant and y_F and y_U are the fluorescence intensity of the probe in the presence of completely folded and unfolded protein respectively)

(Matulis *et al*, 2005). The baselines of the denatured and native state were approximated by a linear fit. The observed temperature shifts, ΔTm_{obs}, for each inhibitor were recorded as the difference between the transition midpoints of sample and reference wells containing protein without inhibitor in the same plate and determined by non linear least squares fit using the program R.

A7.1 Preparation of reagents

Stock solutions of inhibitors (40mM) and protein (3μM) were prepared ahead of the experiments and frozen in appropriate aliquots for later use. HBS was prepared, containing 50mM Hepes pH 7.5, 250mM NaCl and 0.02% MTG. Just before performing the experiment, 10.1μl 5000x concentrate SYPRO®-Orange dye (Sigma Aldrich) was mixed with 6.35ml buffer to create a buffered stock of dye. 4.74ml of this dye stock was then mixed with 2.5ml CDK2/cyclin A (3μM) stock. Assays were performed in triplicate on purified bovine cyclin A or pCDK2/cyclin A at 1μM final concentration buffered in modified HBS (50mM HEPES pH 7.5, 250mM NaCl, 0.02% MTG), with 100μM SYPRO®-Orange dye (Molecular Probes, CA) and 100μM compounds in 96-well PCR plates.

A7.2 Preparation of completed plate

14.5μl aliquots of the protein/dye mixture were then transferred into a white, non skirted, low profile, optical reading-compatible 96 well plate (96 well polypropylene plates #410088, Stratagene) covered with foil seals and the plates

were spun in a room-temperature centrifuge at 1000rpm for 1 minute to remove bubbles from the wells. Using the Mosquito pipetting system (with an adaptor to fit the PCR plate into the machine), 0.5μl aliquots of inhibitors were added to the protein/dye mix in the 96-well plate, using a protocol that transferred the aliquots from the 96-well plate in which these inhibitors were stored as 40mM stocks. After addition of the inhibitor stock, the plate was re-covered using an optically clear plastic seal.

A7.3 Measurement of thermal denaturation curves

After setting up the plate, the thermal denaturation curve was assessed between 25°C and 75°C over one hour, with increments of 1.5°C using a Stratagene RT PCR 305 machine (Stratagene, CA) that had been pre-equilibrated to 25°C. At each increment the fluorescence was measured using excitation and emission filters of 465 and 590 nm respectively, for each well.

A7.4 Analysis of results

Results were initially analysed in R (R Development Core Team (2008) using a script written by Martin Noble. R calculated thermal denaturation curves for the protein in each well using a method of fitting which relies on the least squares method and iteratively improves the fit to each curve. Curves and Tm values for each well were output as a text file, and these were transferred to Excel for further analysis, including calculation of averages and errors, since each result was repeated in triplicate. The curves were also examined by eye for signs of non-specific binding and to check that denaturation curves looked as expected. For the CDK2/cyclin A complex, denaturation curves sometimes showed two separate denaturation events, possibly representing disintegration of the complex (and denaturation of one binding partner) and denaturation of the second binding partner.

APPENDIX B

X-RAY CRYSTALLOGRAPHY

B1 Protein crystallisation

B1.1 Conditions for crystal growth

The process of crystallisation involves a protein coming out of solution in an ordered manner, to form a regular lattice of repeating units. The rules governing which conditions will favour crystal growth are not well understood and thus the task of growing diffraction-quality protein crystals remains difficult and unpredictable. The aim of the crystallisation process has to be to bring the protein into a supersaturated, metastable state, after which crystallisation occurs around a nucleation centre (formed of protein aggregate). In order to maintain the protein in a supersaturated state above its solubility curve and overcome the entropy barrier to crystallisation, the correct well solution must be used and a sealed environment created. By a process of vapour diffusion, water evaporates from the drop and effectively increases the protein and precipitant concentration, increasing the supersaturation and favouring crystallisation (Figure B.2).

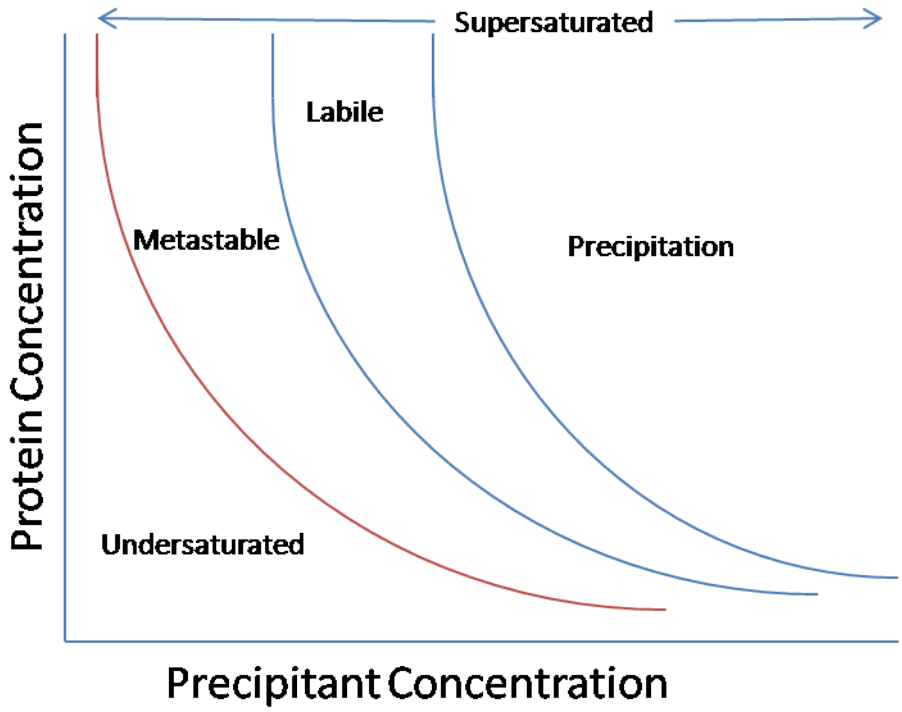


Figure B.1 Phase diagram showing the progress of a vapour diffusion experiment The red line is the solubility curve above which the protein solution is supersaturated and crystallisation is favoured. Figure reproduced from P212121.com

B2 Unit cell, symmetry and space group

The unit cell is the basic parallelopiped-shaped repeating unit from which crystals are constructed (Figure B.2). The unit has length a, b, c and angles α, β, γ. It is futher subdivided into planes referred to be the Miller indices (h, k, l)

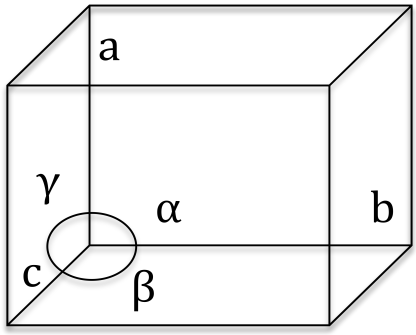


Figure B.2 The unit cell The vectors a, b and c describe the edges of the unit cell, and the angles α, β, γ describe the angles formed by these edges.

The regular array of unit cells within the crystal lattice is characterised by translational and rotational symmetry about planes. The combination of these types of symmetry gives rise to 32 point groups (Blundell and Johnson, 1976), only 11 of which can occur in a protein crystal. The rules determining the identity of the unit cell within the crystal give rise to a classification system in which protein crystals are subdivided into primitive (P), face-centred (F) or body-centred (I). Crystals are then categorised as one of seven crystal systems, and one of 14 Bravais lattices (listed in Table B.1).

Crystal system	Bravais lattice	Geometric constraints	Minimum point-group symmetry
Triclinic	P	None	1
Monoclinic	P, C	$\alpha = \gamma = 90°$ (where b is the unique axis – a 2-fold or screw axis in proteins) **or** $\alpha = \beta = 90°$ (where c is the unique axis – a 2-fold or screw axis in proteins)	2
Orthorhombic	P, C, I, F	$\alpha = \beta = \gamma = 90°$	222
Tetragonal	P, I	$a = b$; $\alpha = \beta = \gamma = 90°$	4
Trigonal	P	$a = b$; $\alpha = \beta = 90°$; $\gamma = 120°$ **or** $a = b = c$; $\alpha = \beta = \gamma$	3
Hexagonal	P	$a = b$; $\alpha = \beta = 90°$; $\gamma = 120°$	6
Cubic	P, I, F	$a = b = c$; $\alpha = \beta = \gamma = 90°$	23

Table B.1 The seven crystal systems and the fourteen Bravais lattices Table adapted from Blundell and Johnson, 1976 and Drenth, 1999)

The addition of other symmetry operations (mirror planes, screw axes, centres of inversion and 'glide' planes) into the classification system denotes 230 possible space groups, only 65 of which occur in protein (see International table of crystallography, Volume A). The unit cell is made of asymmetric units, the basic repeating unit for which operation of symmetry elements can generate an identical object within the unit cell (with the two objects related by the symmetry axis). A unit cell contains more than one asymmetric unit if it has high rotational symmetry, and an asymmetric unit can contain more than one protein molecule.

B3 Theory of diffraction

B3.1 The electron as an X-ray scatterer

X-rays hitting a protein crystal are scattered by a process called Thompson's, or coherent, scattering. This effect is inversely proportional to the mass of the scatterer. In practice, therefore, scattering in atoms is almost entirely due to the interaction of X-rays with electrons. An X-ray incident on an electron will call an oscillation of the electron (at the same frequency as the X-ray), and the electron then re-emits the X-ray at the same wavelength but 180° out of phase. Less than 2% of an incident X-ray will be scattered, but the repeating regular array of protein with a crystal means that the total scattering from the crystal is significant.

B3.2 Representation of a wave as a Fourier series

Since X-rays are a wave, they can be approximated and represented by a Fourier series, a series of cosine and sine functions, represented by the equation:

$$f(x) = \sum_{h=0}^{n} F_h[\cos 2\pi(hx) + i\sin 2\pi(hx)]$$

(where F_h is the amplitude of the wave and h is the frequency)

Expansion and rearrangement of this equation gives the Fourier series for a three dimensional function:

$$f(x,y,z) = \sum_{h}\sum_{k}\sum_{l} F_{hkl}\exp[2\pi i(hx + ky + lz)]$$

(describing the situation where a three-dimensional wave has frequency h in the x direction, k in the y direction, and l in the z direction)

The actual mathematical operation giving Fourier transformation of the function F(h) into f(x) is represented by the following equation:

$$F(h) = \int_{-\infty}^{+\infty} f(x)\exp[2\pi ihx]dx$$

The Fourier transform is a reversible operation, so F(h) could be recovered from f(x) by performing the reverse:

$$f(x) = \int_{-\infty}^{+\infty} F(h)\exp[-2\pi ihx]dh$$

B3.3 Bragg's law

Bragg's law describes the circumstances under which a diffraction pattern can be observed from a protein crystal. This is illustrated in Figure B.3, where two X-ray beams are incident upon two crystal lattice planes separated by distance d, at an angle θ. Both beams will be also reflected from the lattice planes by θ and two outcomes can occur: if the path differences between the two reflected beams are any multiple of λ/2, the beams will be cancelled out by destructive interference. Therefore in order to observe an X-ray diffraction pattern, the distance between the two rays must allow scattering in phase to give constructive interference.

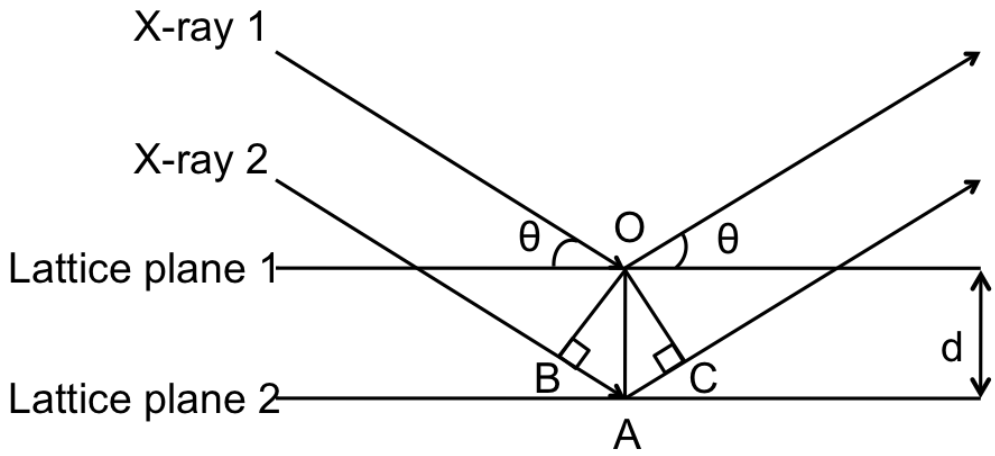


Figure B.3 Bragg's law: scattering from two lattice planes

Bragg's law is normally stated as

$$AB + AC = 2d\sin\theta = n\lambda$$

(where d is the interplanar spacing between the two lattice planes, θ is the angle of the incident X-ray beam and λ is the wavelength of the X-ray beam).

B3.4 The reciprocal lattice and the Ewald sphere

The reciprocal lattice and the Ewald sphere are imaginary spaces in which we can visualise diffraction from a crystal. The reciprocal lattice is defined as the set of imaginary points at which the direction of a vector from one point to another corresponds to the direction of a vector normal to the real space planes and the separation of the points in the reciprocal lattice corresponds to the reciprocal of the real interplanar distance. Imagining a crystal with an incident X-ray beam of wavelength λ, the Ewald sphere is drawn with its centre on the crystal and with a radius equal to 1/λ. At points where the reciprocal lattice and the Ewald sphere cross, Bragg's law is satisfied, $2d_{hkl}\sin\theta = \lambda$, and a diffraction spot may be observed. This concept is illustrated in Figure B.4:

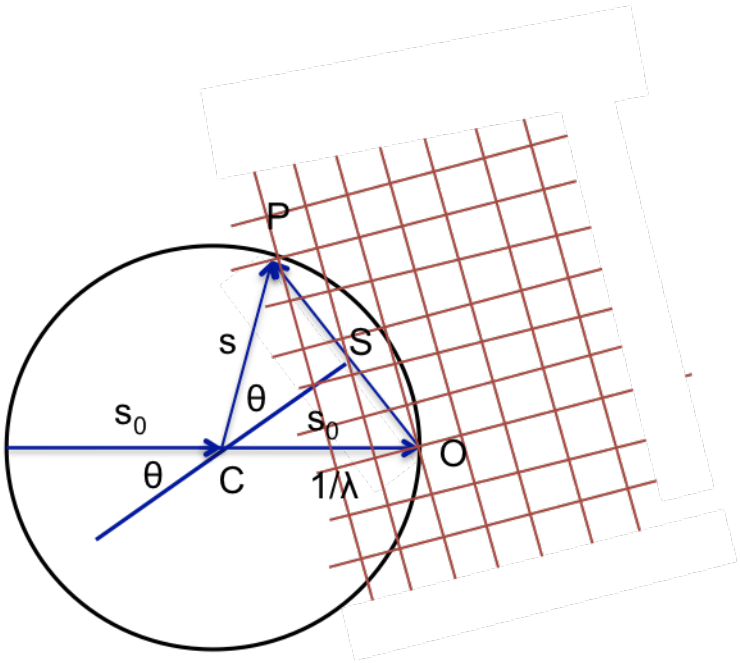


Figure B.4 The Ewald sphere The incident X-ray beam s_0, directed towards the reciprocal unit cell origin O, is diffracted by the crystal at point C, to pass through point P on the Ewald sphere. Since P is also a reciprocal lattice point, Bragg's law is satisfied, so diffraction occurs.

As is obvious from Figure B.4, only a small proportion of reciprocal lattice points cross the Ewald sphere in any one orientation. In order to sample a larger proportion of reciprocal space, the crystal is therefore rotated during an X-ray scattering experiment.

B3.5 Atomic scattering factor

The atomic scattering factor describes the intensity of scattering for a particular atom (dependent on the electron cloud within the atom). The atomic scattering factor is defined as:

$$f = \int_r \rho(r)\exp[2\pi i\mathbf{r}\cdot S]dr$$

(where f is the atomic scattering factor and ρ(r) is the density of electron at position r)

When the atomic scattering factor for an element has been determined, it is expressed as a function of $2\sin\theta/\lambda$ (Blundell and Johnson, 1976) and the total

scattering from a unit cell is the sum of the scattering from each electron in the unit cell, given by the equation:

$$G(S) = \sum_{j=1}^{n} f_j \exp[2\pi i \mathbf{r}_j \cdot S]$$

(where G(S) is the total scattering from the unit cell, f_j is the atomic scattering factor and j = 1, 2, 3, …… n are the number of scattering electrons in the unit cell)

B3.6 The Laue condition for diffraction

Because scattering from a crystal is the sum of the scattering from a large number of unit cells, the scattering will cancel except when conditions are such that unit cells scatter constructively in phase. These are known as the Laue conditions for diffraction and state that scattering occurs in discrete direction and only if **a**, **b** and **c** (the unit cell translation vectors) hold certain values. If the total scattering from a protein crystal is given by the equation:

$$F(S) = G(S) \cdot \sum_{t=1}^{n1} \exp[2\pi i t \mathbf{a} \cdot S] \cdot \sum_{u=1}^{n2} \exp[2\pi i u \mathbf{b} \cdot S] \cdot \sum_{v=1}^{n3} \exp[2\pi i v \cdot \mathbf{c} \cdot S]$$

(where G(S) is the total scattering from the unit cell; n1, n2 and n3 are the number of unit cells; t, u and v are integers and a, b and c are the unit cell translation vectors)

then the Laue conditions under which scattering is observed occur when:
$\mathbf{a}\cdot S = h$, $\mathbf{b}\cdot S = k$, $\mathbf{c}\cdot S = l$

B3.7 Fourier transform and the structure factor equation

The structure factor equation F(S) corresponds to the total diffraction of an X-ray beam by all electrons in a protein crystal:

$$F(S) = \sum_{j} f_j \exp[2\pi i \mathbf{r}_j \cdot S]$$

(where f_j is the atomic scattering factor for an electron at position r_j).

Taking into account the Laue conditions, this equation can be rewritten as:

$$F(hkl) = \sum_{j=1}^{n} f_j \exp 2\pi i(hx_j + ky_j + lz_j)$$

and instead of taking the sum of the contribution to diffraction from each individual electron, we instead represent the structure factor as an integration of the whole electron density in a unit cell, we can rewrite it as:

$$F(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \exp[2\pi i(hx + ky + lz)] dxdydz$$

(where V is the volume of the unit cell).

We then use a Fourier transform to calculate the electron density from the structure factor, so that:

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} \mathbf{F}(hkl) \exp[-2\pi i(hx + ky + lz)]$$

(where ρ(xyz) is the electron density at the coordinate (x, y, z)).

Finally, we interpret **F**(hkl) as a vector with an amplitude $|\mathbf{F}(hkl)|$ and phase, of angle α(hkl). Adding this information into the equation, we get:

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |\mathbf{F}(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

(where $|\mathbf{F}(hkl)|$ is the amplitude of reflection (h, k, l) and α(hkl) is the phase of the structure factor F(h k l).

B3.8 Friedel's law

Upon diffraction against a lattice plane, there are two reflections that result from diffraction on opposite faces of the same lattice plane, the *hkl* and *-h-k-l* reflections. In each case the amplitude is the same (as well as the intensity) but the phases are opposite. These reflections are known as a Friedel's pair. Friedel's

law explains why 100% complete diffraction data can be collected by a 180° rotation of the crystal (assuming no anomalous scatterers).

B4 Solving the phase problem

During the calculation of the electron density, $\rho(xyz)$, we need the amplitude, $|\mathbf{F}(hkl)|$ and the phase $\alpha(hkl)$ of a reflection *hkl*. We can calculate the amplitude since this is equal to the square root of the measured reflection density; however, we cannot record the phase during the experiment and this information in lost. This is often known as the 'phase problem' and necessitates the use of one of several techniques designed to recover this information. These are:

1. Direct methods: a mathematical approach by which estimates of the initial phase can be limited to a smaller number of possibilities by information about the relative sizes of atoms compared to the spaces in between. However this method is only likely to be successful for molecules with <150 non-hydrogen atoms, and it therefore has limited use in protein structure determination.

2. Single or multiple isomorphous replacement with heavy atoms (SIR/MIR): In isomorphous replacement, add a heavy atom, with a large number of electrons, which will perturb the structure factors and allow us to narrow down possible phase values.

3. Single or multi-wavelength anomalous dispersion: This method relies on the breakdown on Friedel's law, which occurs when a heavy atom (for which the wavelength of the X-ray is close to the resonance frequency of inner shell electrons) absorbs and re-emits the X-ray beam out of phase. This anomalous dispersion means that the heavy atoms can be located within the structure, providing the first step to structure solution.

4. Molecular replacement: where an existing closely related structure is available for use as a model, we can align our experimental and model information, and

use phase information from the previous completed structure to provide a starting phase close enough to work with. This method only works well if the model chosen is a valid close homologue.

For the purpose of this thesis, I shall only discuss molecular replacement, since this was the only technique used in the course of this work, as I was determining structures of protein-drug complexes for proteins of which the structure had already been solved.

B4.1 Molecular replacement

The basic principle involved in molecular replacement is to use the structure of a homologous protein to provide an initial estimate of the phase for the structure under investigation. Once a homology model has been chosen, it is necessary to correct for the fact that the two structures are likely to have been solved with different space groups and unit cell dimensions. In order to transfer phase information, we therefore have to place the model onto the electron density from the novel structure in three-dimensional space and place it in the correct location within the unit cell. This is traditionally done using a search function which splits the problem into two steps: rotation and translation; so that:

$$X_2 = [C]X_1 + d$$

(where X_1 and X_2 are the homology model and the novel structure, [C] is the rotational matrix used to match the orientation, and d is the translation function used to match the position)

B4.2 Patterson function

The rotation and translation functions are carried out in Patterson space. The Patterson function is the Fourier summation of the experimentally measured intensities, where all phase angle are zero, given by:

$$P(uvw) = \frac{1}{V}\sum_{hkl}\left|\mathbf{F}(hkl)\right|^2 \cos[2\pi(hu + kv + lw)]$$

(where P(uvw) is the Patterson function at point (uvw)).

Since phase angles are set to zero, we can calculate this directly from the measured structure factor amplitude ($|\mathbf{F}(hkl)|^2$). A Patterson map, visualising this function, contains *n(n-1)* peaks where there are *n* atoms in the unit cell, and each peak corresponds to an interatomic vector. We can then use this map to locate the positions of individual atoms in the unit cell.

B4.3 Rotation and translation search

The Patterson function can also be written as an integral, such that:

$$P(uvw) = \int_v \rho(xyz) \cdot \rho(x+u, y+v, z+w) dv$$

(where the Patterson function *P(uvw)* is calculated by the infinite summation of the product of electron density ρ at point *(x, y, z)* and point *(x+u, y+v, z+w),* separated by the vector *u (uvw)*).

By calculation of the Patterson maps for both structures, the maximum overlap between the maps occurs when the two molecules are in the same orientation, calculated by the overlap function R:

$$R(\alpha\beta\gamma) = \int_v P(u) \cdot P_r(u_r) du$$

(where R is the overlap function for the rotation angles αβγ, V is the volume of the Patterson map, P(u) is the Patterson function and Pr(Ur) is the rotated Patterson function).

The rotational search is usually carried out with data in the resolution range 3-4 Å, because high resolution data are too sensitive to the difference between calculated and observed structure, and low resolution data provide too many false positive results.

Once the model is orientated correctly by rotation with regard to the novel structure, the translational position is then found by a process of trial and error that involves sampling all unit cell space. The structure factors resulting at each position are then calculated and compared with the observed structure factors from the experiment, with agreement monitored using the R factor and correlation coefficient (CC):

$$R = \frac{\sum_{hkl} \left| |F(obs)| - k|F(calc)| \right|}{\sum_{hkl} |F(obs)|}$$

$$CC = \frac{\int_V (P(x) - \overline{P}) \cdot (Q(x) - \overline{Q}) dV}{\left(\int_V (P^2(x) - P^{-2}) dV \cdot \int_V (Q^2(x) - Q^{-2}) dV \right)^{1/2}}$$

(where F(obs) and F(calc) are the observed and calculated structure factors respectively, k is a scaling constant, P(x) and Q(x) are the observed and calculated Patterson functions respectively, and $\overline{P}$ and $\overline{Q}$ are the mean values of P(x) and Q(x) respectively)

A correlation coefficient of -1 would indicate total negative correlation, whilst a correlation coefficient of +1 would indicate total positive correlation. In order to accept a solution as correct, an R factor of <0.5 and a CC of >0.5 should be reached. Once complete a packing function may also be used to avoid false solutions by monitoring clashes with symmetry-related copies in the unit cell.

B5 Macromolecular structure refinement with REFMAC5

REFMAC5 (Murshudov *et al,* 2011) was used to perform macromolecular structure refinement, in order to minimise deviation between the experimentally observed structure factors, and the calculated structure factors (from the

refinement model). In this way the electron density calculated using the model is refined to increase agreement with the experimental data.

REFMAC5 was first run using 10 rounds of rigid body refinement, in which the model is divided into rigid domains and the orientation and position of these are refined in each cycle. This provides a quick initial improvement of the agreement between observed and calculated structure factors, but uses extremely rigid constraints and therefore limits possible convergence. 10-20 rounds of restrained refinement were therefore run afterwards: in this process, REFMAC5 uses known stereochemical information about preferences for bond angles and lengths, planarity, van der Waals contact distances and chirality, to increase the ratio of observations:parameters, and refines at all atomic positions.

REFMAC5 uses the maximum likelihood method of refinement, in which a likelihood function essentially assesses whether a change makes the agreement better or worse and estimates a probability for all attempted solutions. Refinement is monitored throughout for model bias, using R_{free}, in which a set of data (usually 5%) is removed at the beginning of the process, and comparison with this unrefined set allows model bias to be detected. Success of refinement was judged by a decrease in R and R_{free}. Successive rounds of restrained refinement were completed and after each round the structure was checked and manual fixes were made in Coot.

B6 Manual rebuilding using Coot

The program Coot (Emsley *et al,* 2010) was used for manual rebuilding of all structures after each round of refinement in Refmac. This allows visualisation of the current progress of refinement by comparison of the current structure with the electron density map. Problems were fixed, such as loops poorly fitting to the data, wrong rotamers for amino acids, broken segments and mis-modelled sequences. Bond angles and lengths were checked and Coot was used to identify waters. The ligands were built using the Elbow Builder utility (eLBOW -

electronic Ligand Builder and Optimisation Workbench: http://www.phenix-online.org/documentation/elbow.htm) and added into the density seen in the binding site manually using Coot. Coot was also used to pick and check waters within the structure and to perform basic checks on geometry.

B7 Assessment of quality of the structure with MolProbity

The MolProbity facility (Chen *et al,* 2010: http://molprobity.biochem.duke.edu/) was used to assess the quality of the structures and add hydrogens to the PDB file. This program checks the protein structure against those deposited in the PDB using all-atom contact analysis, (checking for non-hydrogen-bonded atoms with significant van der Waals overlap), and geometric analysis (checking for unusual side chain or backbone conformations, to identify distorted bond angles). Structures were accepted as finished when further refinement and building did not improve the refinement statistics further and/or when the structure was placed in the top 30% by quality of structures in the PDB by MolProbity.

B8 Crystallography jargon

B8.1 Matthews coefficient

The Matthews coefficient describes the solvent content of a crystal and with the knowledge that solvent content in a protein crystal varies between 27-65%, with an average of 43%, this allows calculation of the number of protein molecules in the asymmetric unit. The Matthews coefficient Vm is given by the following equation:

$$Vm = \frac{Vc}{Mw \cdot N_{asym} \cdot N_{molasym}}$$

(where Vc is the unit cell volume in Å^3, Mw is the molecular weight of the protein in Da, N_{asym} is the number of asymmetric units in the space group, and $N_{molasym}$ is the number of molecules in the asymmetric unit)

The volume of the unit cell is calculated from the unit cell dimensions and we know the number of asymmetric units from the space group; this allows us to calculate the fractional volume occupied by proteins (V_{fp}):

$$Vfp = \frac{Vp}{Vc} = \frac{Mtot}{Dp} \cdot \frac{1}{Vc} = \frac{Mw \cdot Nasym \cdot Nmolasym \cdot Mh}{Dp \cdot Vc}$$

(where Vp is the volume occupied by protein, Mtot is the total mass of protein in one unit cell, Mh is the mass of a hydrogen atom and Dp is the density of protein at 1.35 g/cc (Matthews, 1968)

and by rearrangement:

$$Vfp = \frac{Mw \cdot Nasym \cdot Nmolasym \cdot Nmolasym}{Vc} \cdot \frac{Mh}{Dp} = \frac{1}{Vm} \cdot \frac{1}{N} \cdot \frac{1}{Dp}$$

therefore, with the mass of a hydrogen atom 1/N and N = 6.02 x 10^{23} (Avogadro's constant):

$$Vfp = \frac{1.23}{Vm}$$

Finally, the solvent content can be estimated from the Matthews coefficient by subtracting the fractional volume occupied by the protein from the total volume:

$$Vfs = 1 - Vfp = 1 - \frac{1.23}{Vm}$$

(where Vfs is the solvent content).

B8.2 Temperature factor

The temperature factor (also known as the Debye-Waller or B factor) is used to describe the thermal motion of electrons and static disorder within the protein crystal. The structure factor equation can be re-written to include this factor:

$$F(hkl) = \sum_{j=1}^{N} f_j \exp\left[2\pi i\left(hx_j + ky_j + lz_j\right)\right]\exp\left[\frac{-B_j \sin^2\theta}{\lambda^2}\right]$$

(where B is the temperature factor and with the assumption that thermal motion is isotropic)

Large B factors compromise the intensity of high resolution data. During refinement, thermal motion should be modelled as anisotropic, providing that high enough resolution data is available (<1.5Å). The B factor also includes static disorder due to defects in arrangement of protein molecules in the crystal lattice.

B8.3 R factors

R factors are the main statistics used to assess the quality of a structure during data refinement. A small R factor indicates good quality data and successful refinement. There are many different R factors used, each measuring slightly different things. These include:

- R_{sym}, which measures discrepancies between the intensity of symmetry-related reflections, the existence of which usually reflects systematic errors during the data collection process:

$$R_{sym} = \frac{\sum_{hkl}\sum_{i}\left|I_i(hkl) - \overline{I_i(hkl)}\right|}{\sum_{hkl}\sum_{i}\overline{I_i(hkl)}}$$

(where $I_i(hkl)$ is the intensity of the ith symmetry-related reflection (hkl) and $\overline{I_i(hkl)}$ is the mean intensity of all symmetry-related reflections)

- R_{merge}, which measures the discrepancy between multiple observations of an equivalent reflection:

$$R_{merge} = \frac{\sum_{hkl}\sum_{i=1}^{N}|I(hkl) - I_i(hkl)|}{\sum_{hkl} N \cdot I(hkl)}$$

(where N is the total number of the equivalent reflection *I(hkl)*)

- and R_{conv} and R_{free}, both of which monitor the refinement and model building process: R_{conv} measures the discrepancy between observed (Fobs) and calculated (Fcalc) structure factors, and is a general measure of quality of the structure. However R_{conv} can be artificially lowered if there are too many refinement parameters (model bias or over-fitting), so this is checked using R_{free}, calculated in the same way but taking into account comparison with an excluded 5% of data (the test set)

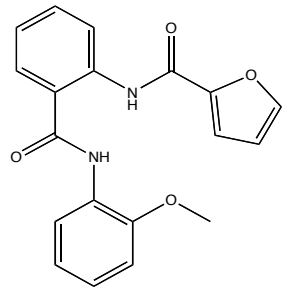
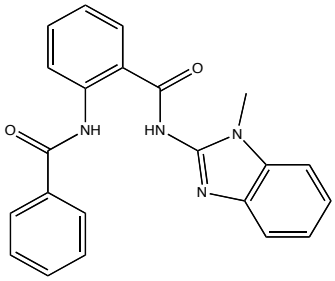
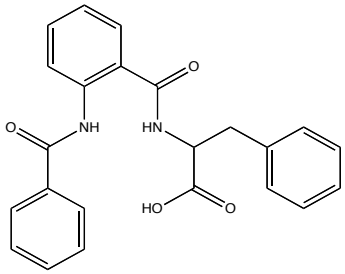
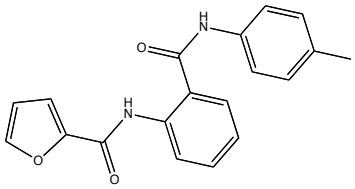
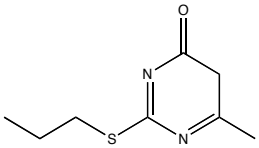
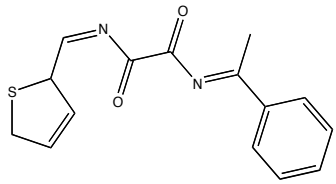
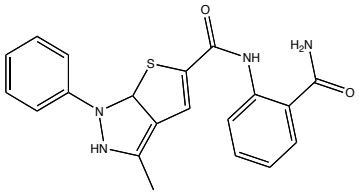
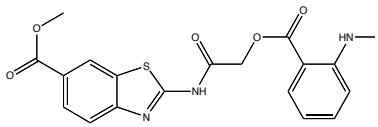
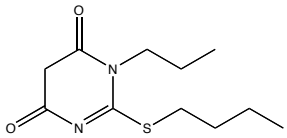
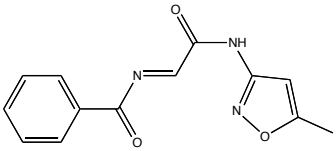
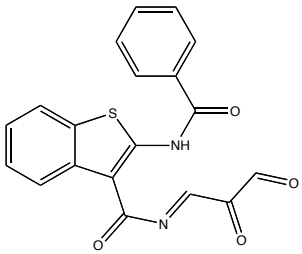
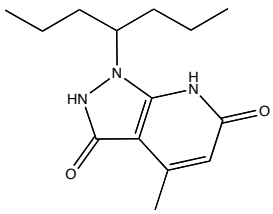
$$R_{conv} = \frac{\sum_{hkl}\left||F_{hkl}obs| - |F_{hkl}cal|\right|}{\sum_{hkl}|F_{hkl}obs|}$$

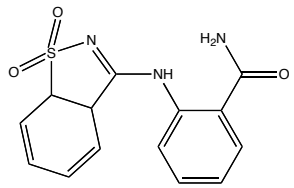
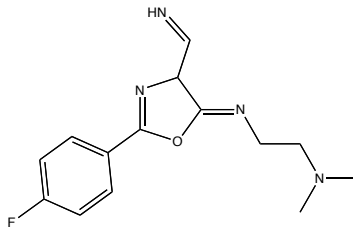
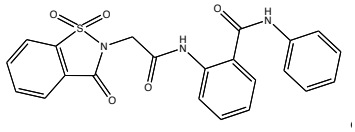
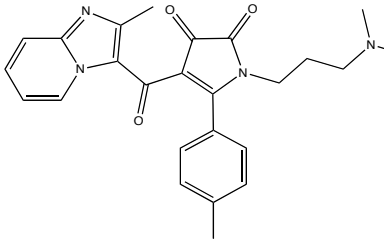
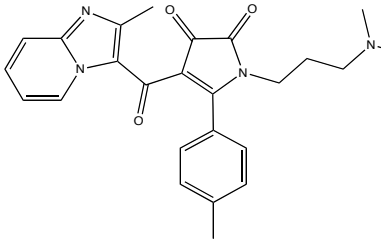
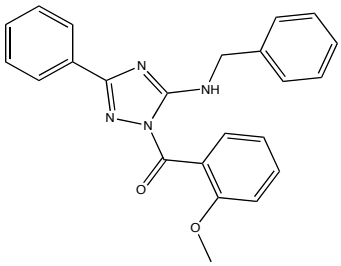
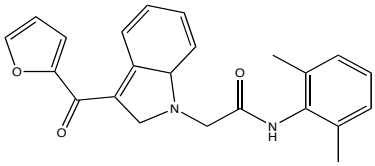
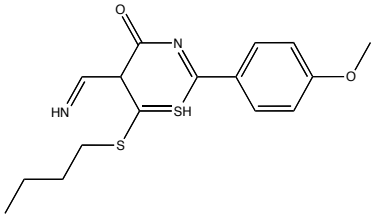
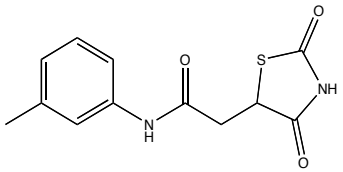
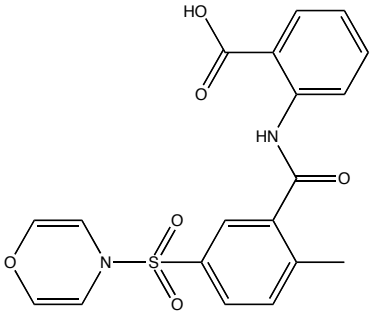
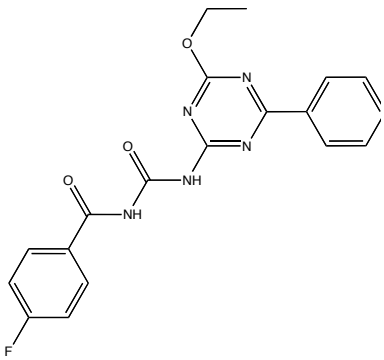
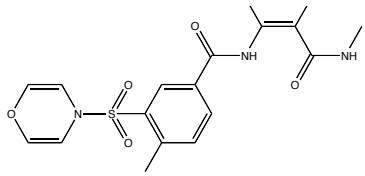
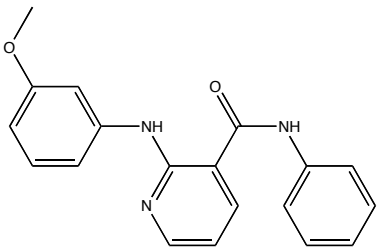
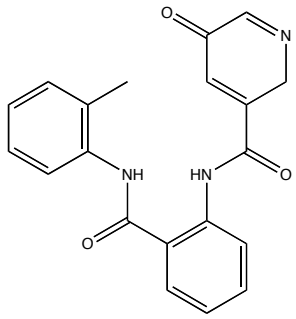
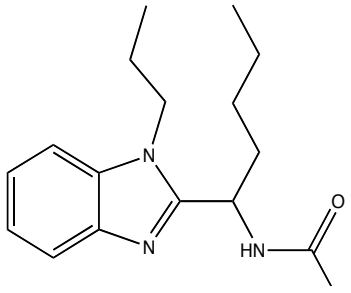
(where F_{hkl}obs and F_{hkl}calc are the observed and calculated structure factors, respectively, for reflection *hkl*)

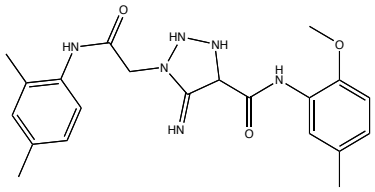
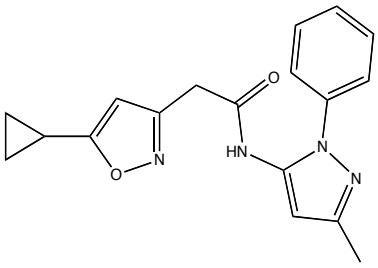
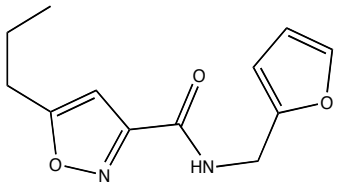
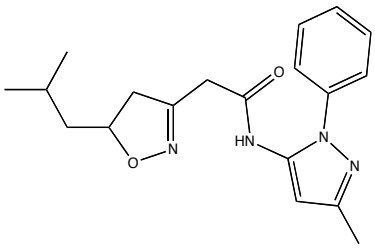
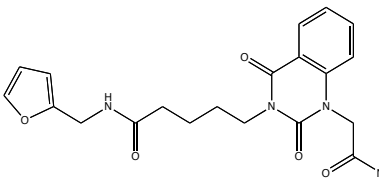
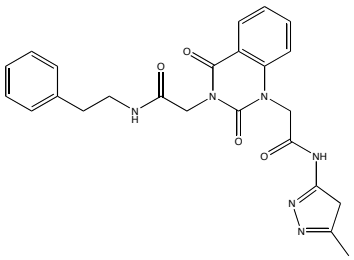
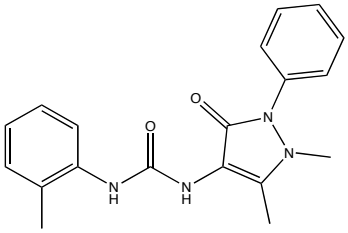
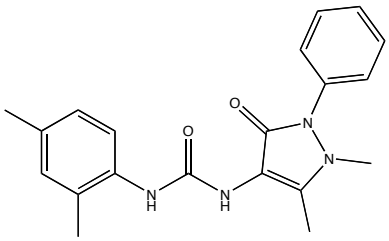
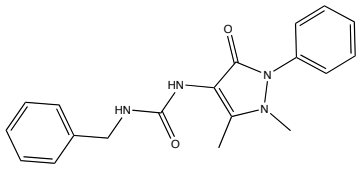
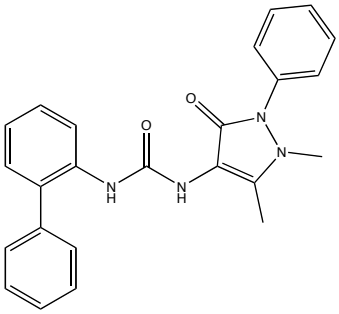
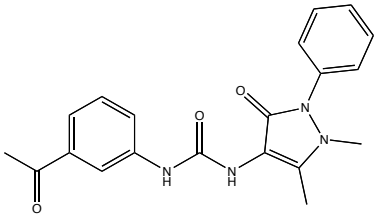
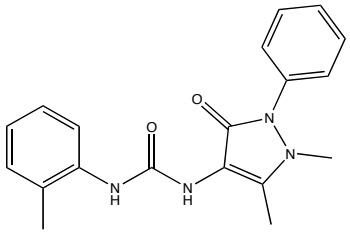
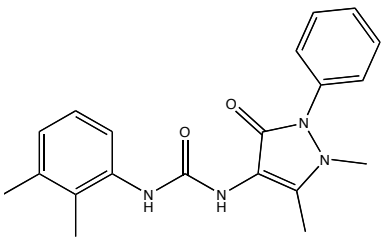
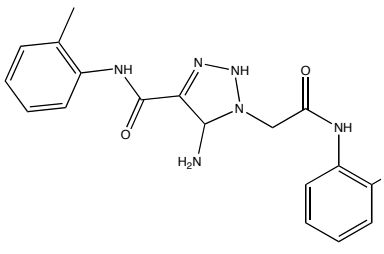
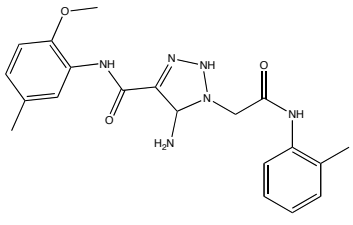
R_{sym} and R_{merge} are usually calculated for each resolution shell, since data quality naturally varies with resolution and this can help to indicate a suitable cut-off point for resolutions of data used in refinement of the model. R_{conv} (sometimes just called R) and R_{free} shoud drop during refinement, and this is the indication of a successful refinement process. If R_{conv} drops without R_{free} falling, this usually indicates model over-fitting.

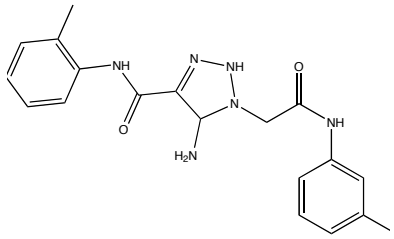
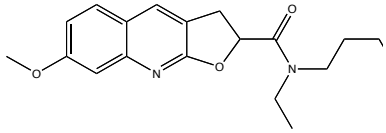
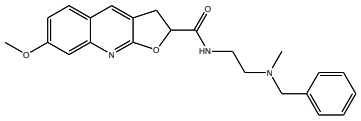
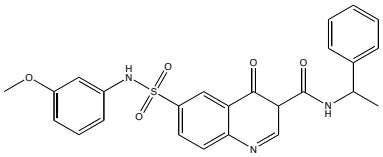
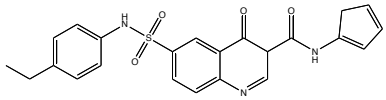
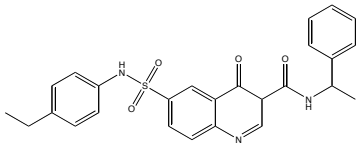
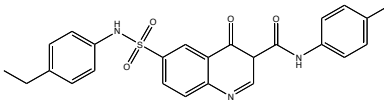
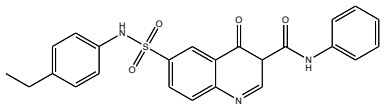
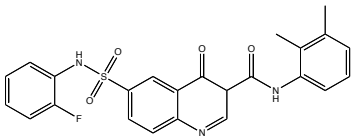
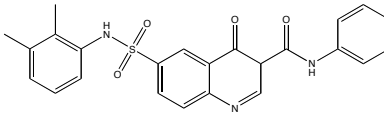
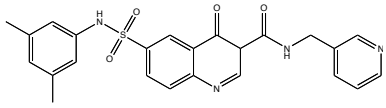
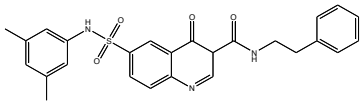
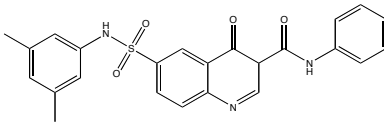
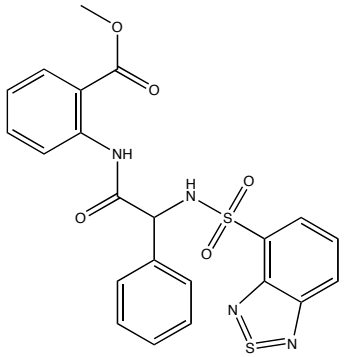
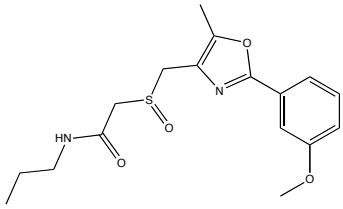
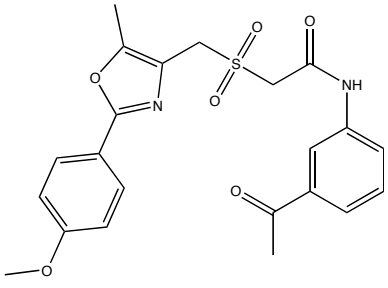
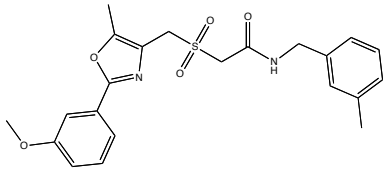
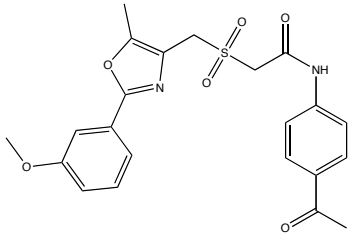
Appendix C: Structures of compounds in libraries

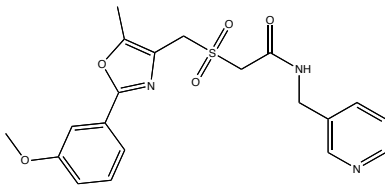
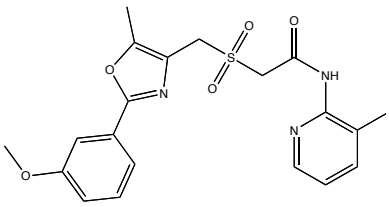
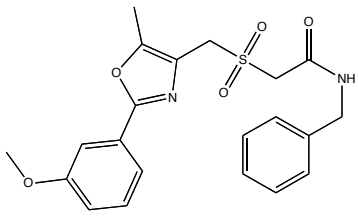
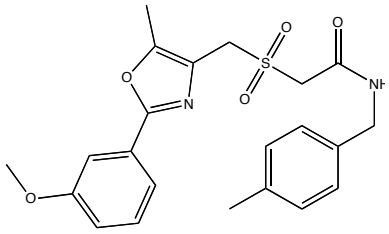
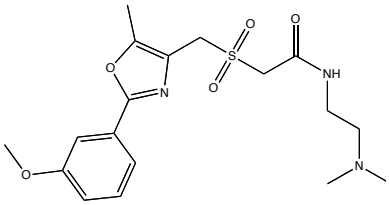
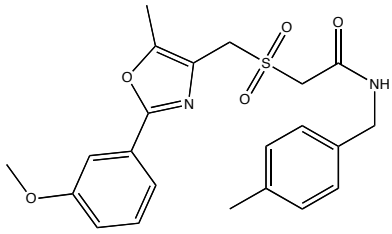
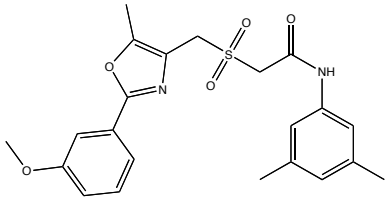
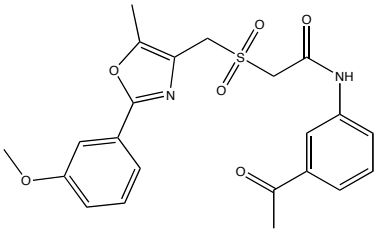
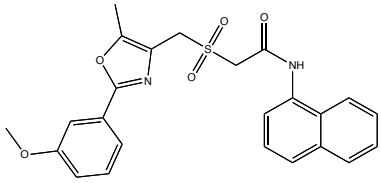
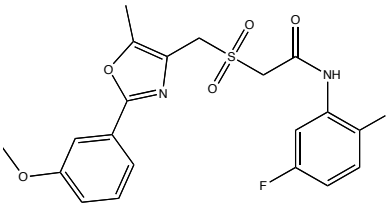
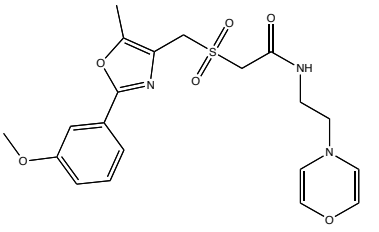
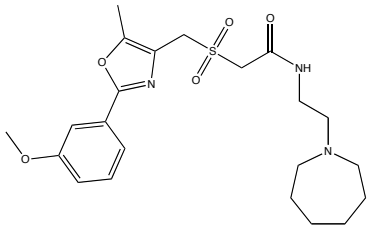
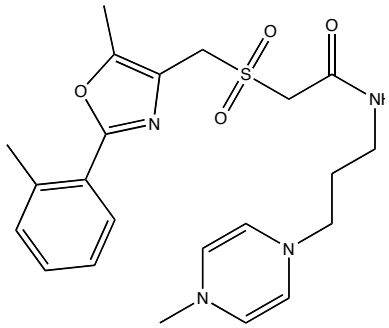
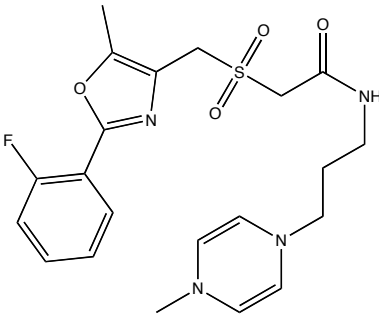
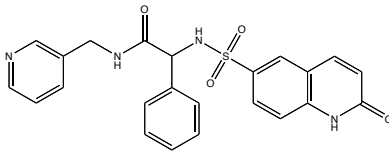
C1 ChemDiv library from virtual screening

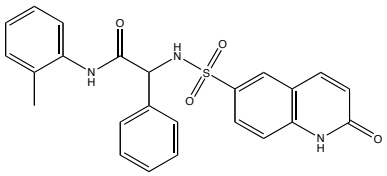
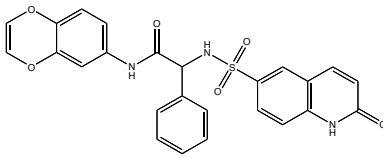
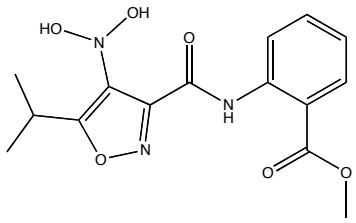
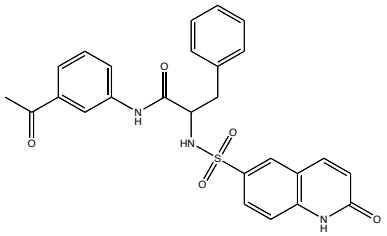
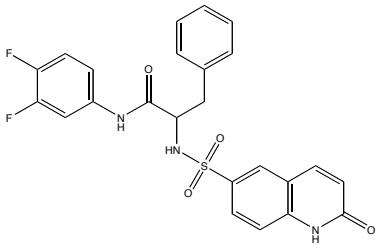
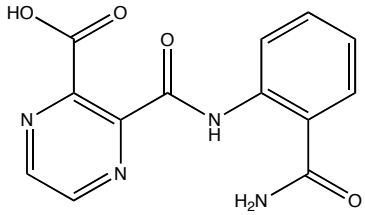
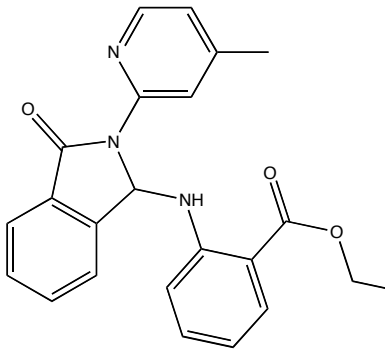
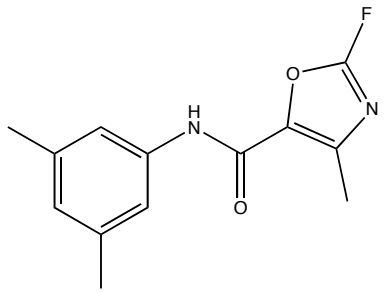
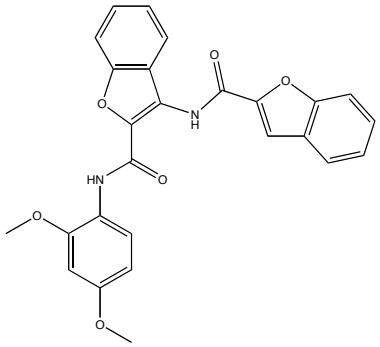
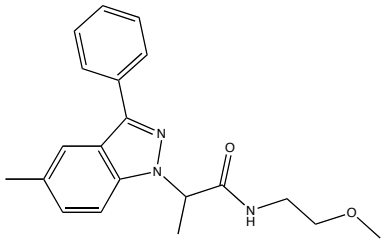
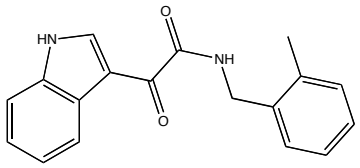
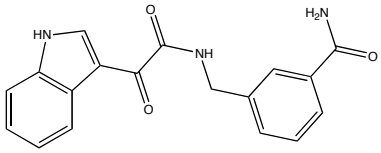
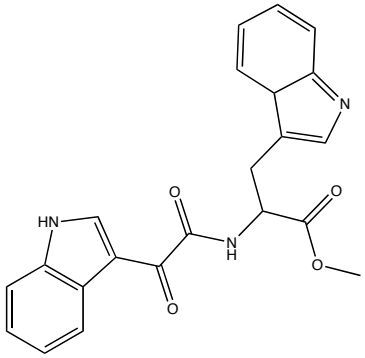
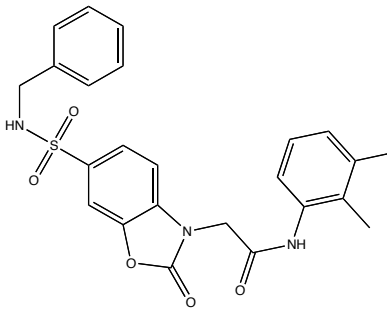
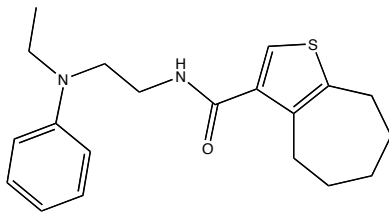
A1 	A2 	A3 
A4 	A5 	A6 
A7 	A8 	A9 
A10 	A11 	A12 

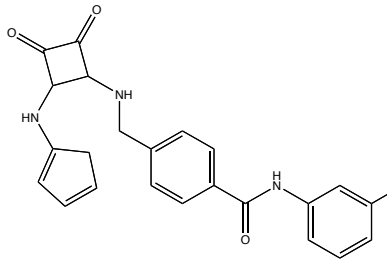
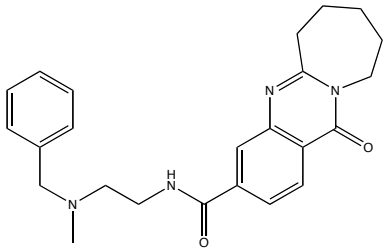
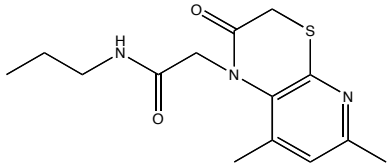
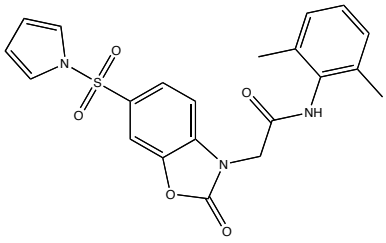
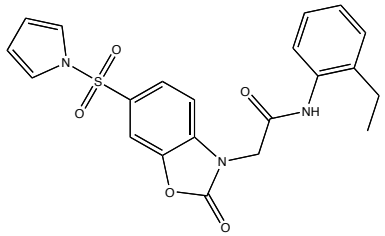
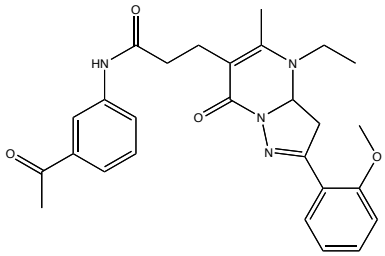
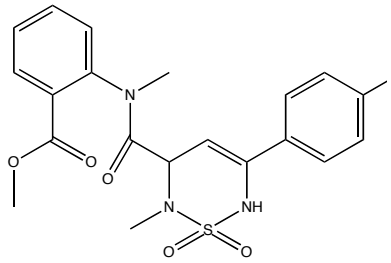
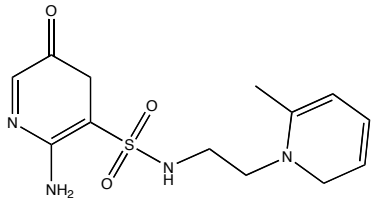
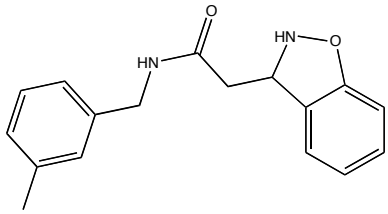
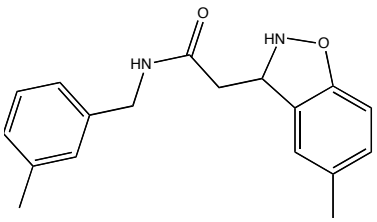
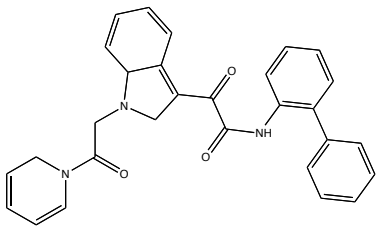
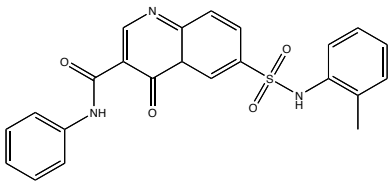
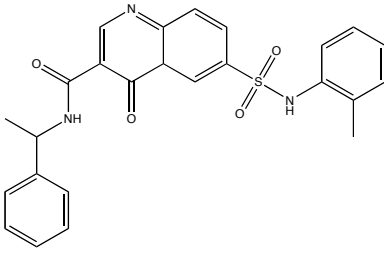
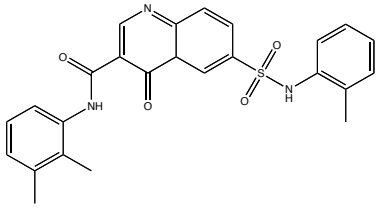
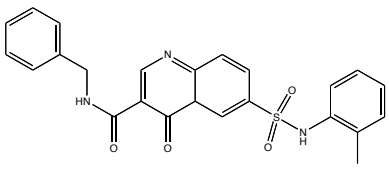
B1 	B2 	B3 
B4 	B5 	B6 
B7 	B8 	B9 
B10 	B11 	B12 
C1 	C2 	C3 

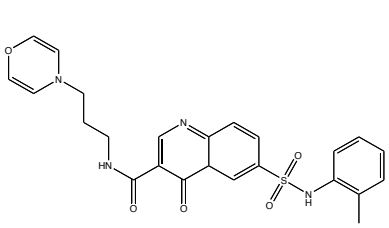
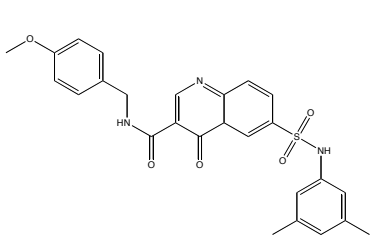
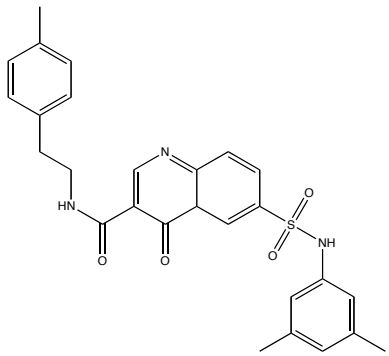
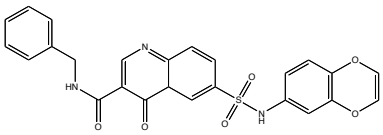
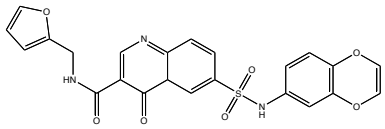
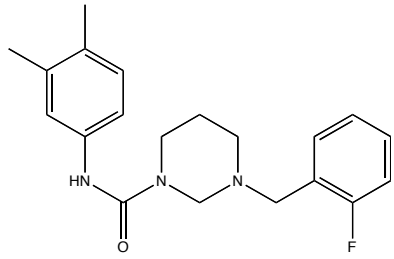
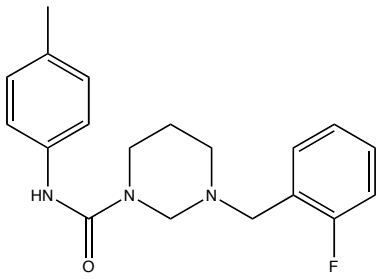
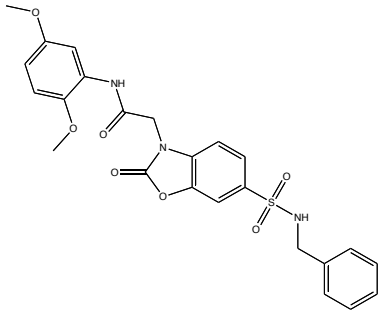
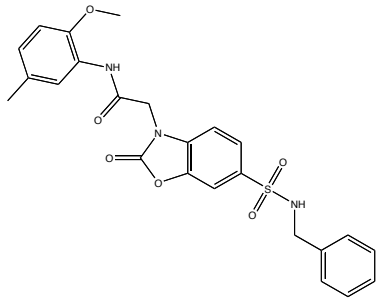
C4 	C5 	C6 
C7 	C8 	C9 
C10 	C11 	C12 
D1 	D2 	D3 
D4 	D5 	D6 

D7 	D8 	D9 
D10 	D11 	D12 
E1 	E2 	E3 
E4 	E5 	E6 
E7 	E8 	E9 
E10 	E11 	E12 

F1 	F2 	F3 
F4 	F5 	F6 
F7 	F8 	F9 
F10 	F11 	F12 
G1 	G2 	G3 

G4 	G5 	G6 
G7 	G8 	G9 
G10 	G11 	G12 
H1 	H2 	H3 
H4 	H5 	H6 

H7 	H8 	H9 
H10 	H11 	H12 
A1.2 	B1.2 	C1.2 
D1.2 	E1.2 	F1.2 
G1.2 	H1.2 	A2.2 

B2.2 	C2.2 	D2.2 
E2.2 	F2.2 	G2.2 
H2.2 	A3.2 	B3.2 
C3.2 