Predicting and Testing Helix-Mimetic Inhibitors of the p53-Mdm2 Interaction

Joel Peter Dockray

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

March 2015

School of Molecular and Cellular Biology

University of Leeds

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

0.1. Acknowledgements

I would like to thank all those who have been integral to the success of my PhD. Some are listed in Table 0.1. Many others, while not making a scientific contribution, have made my time in Leeds a very positive experience and, although I do not attempt to list them all here, I acknowledge their important role and am grateful for their support.

Table 0.1: People who have made a contribution to my research

Person	Role	Reason for acknowledgement
Steve Homans	Supervisor	Overall PhD oversight and advice
Tom Edwards	Supervisor	Comprehensive oversight, assistance, advice, support and administration throughout my PhD
Richard Jackson	Former supervisor	Project conception and support in my first year
Andrew Wilson	Head of collaborating group	Grant management, collaboration administration and direct oversight of compound development
Gary Thompson	NMR researcher and technician	NMR technical support, pastoral support, technical advice
Arnout Kalverda	Head of NMR facility	NMR technical and pastoral support, technical advice
Jeff Plante	Predecessor	Production of GFP-Mdm2 and p53-YFP
Kerya Long	PhD chemist	Production and testing of N-linked compounds
George Preston	PhD chemist	Synthesis and purification of p53-eosin
Jon Fuller	Post-doctoral computational chemist	Initial simulation studies, technical advice
Panchami Prabhakaran	Post-doctoral chemist	Production of O-linked compounds
Valeria Azzarito	PhD chemist	Soluble compound production for NMR
Tasha Murphy	PhD chemist	Production and testing of O-linked compounds
Anna Barnard	Post-doctoral chemist	N-linked compound development
Alice Bartlett	Post-doctoral biochemist	Fluorescence polarisation assay development
Sergei Krivov	Research fellow	Technical advice
Sue Matthews	Technician	Technical support

0.2. Abstract

Aberrant protein-protein interactions (PPIs) are found in many disease states. Consequently, there is a need for PPI inhibitors for use as research tools and pharmaceutical lead compounds. Computational methods could greatly assist with the search for new PPIs.

Oligobenzamides are novel PPI inhibitors which can theoretically be produced to display any sequence of side chains. Understanding the nature of oligobenzamide binding is important for identification of the most efficient strategy of predicting oligobenzamide PPIs.

The prediction of oligobenzamide affinities using thermodynamic integration and implicit solvent methods is described. Affinities of oligobenzamides for Mdm2 predicted using implicit solvent methods bore a moderate correlation with measured affinities. Examination of MM-PBSA results using analysis of variance revealed that it is not necessary to run simulations with every member of a large combinatorial library in order to predict their relative affinities because within a particular binding site, the degree of interaction between the side chains is small. However, it could be useful to separate molecules based on their predicted binding pose because oligobenzamides can bind to Mdm2 in many different ways, depending on the choice of side chains. This insight will be valuable for future attempts to predict oligobenzamide affinities.

The ^{1}H-^{15}N HSQC NMR spectrum peaks of ^{15}N-labelled Mdm2 L33E were assigned to facilitate the future validation of binding poses. An oligoamide was shown using NMR to bind in the correct place. However, NMR testing revealed that oligobenzamides can aggregate in aqueous solution despite being soluble.

A novel FRET-based method was also developed which can be used to test potential inhibitors with a low solubility and high absorbance during their development. It was adapted for a microwell plate to facilitate future high throughput screening and an assay involving Cherry-labelled Mdm2 was tested which could be developed into an *in vivo* assay in the future.

0.3. Contents

0.4. Abbreviations

µM	µmol dm^{-3}
2xYT	two times yeast extract and tryptone
3D	3-dimensional
AA	4-(4-aminobenzamide)benzoic acid
Ac	acetyl
Aib	aminoisobutyric acid
AMBER	Assisted Model Building with Energy Refinement
ANOVA	analysis of variance
APS	ammonium persulfate
Arg	arginine
Asp	aspartate
ATM	ataxia telangiectasia mutated
ATP	adenosine triphosphate
ATR	ataxia telangiectasia and Rad3-related protein
BA	benzanilide
$B_{AC}2$	bimolecular, base-catalysed acyl-oxygen cleavage
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BFP	blue fluorescent protein
BME	basal medium Eagle
BMRB	Biological Magnetic Resonance Data Bank
BRET	bioluminescence resonance energy transfer
BSA	bovine serum albumin
C_{ar}	aromatic carbon
cdk4	cyclin-dependent kinase 4
cdk6	cyclin-dependent kinase 6
cDNA	copy DNA
CFP	cyan fluorescent protein
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHARMM	Chemistry at Harvard Molecular Mechanics
COX-2	cyclooxygenase 2
C_p	peptide carbon
CPD	composite pulse decoupling
Cys	Cysteine
D-amino acid	dextrorotatory amino acid
DFT	density functional theory
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
D-peptide	dextrorotatory peptide
DPI	dual polarisation interferometry
DsRed	*Discosoma sp.* red fluorescent protein
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced GFP
ESP	electrostatic potential
eYFP	enhanced YFP
FA	fluorescence anisotropy
Fmoc	fluorenylmethyloxycarbonyl
FRET	Förster resonance energy transfer
G1 phase	gap 1 phase
G2 phase	gap 2 phase
GAFF	general AMBER force field
GB	generalised Born
GBSA	generalised Born surface area
GFP	green fluorescent protein
Gln	glutamine
Glu	glutamate
G-protein	guanine nucleotide-binding protein
HF	Hartree–Fock
HIPK2	homeodomain-interacting protein kinase 2
His	histidine
HMQC	heteronuclear multiple-quantum correlation
HOESY	heteronuclear Overhauser effect spectroscopy
HP	high performance
HSQC	heteronuclear spin quantum correlation
Hz	Hertz
IC_{50}	half maximal inhibitory concentration
IDIVF	(An AMBER force field torsion parameter.)
Ile	isoleucine
IPTG	isopropyl β-D-1-thiogalactopyranoside
IQR	interquartile range
IS	ionic strength
ITC	isothermal titration calorimetry
jCE	Java combinatorial extension
Ka	association constant
Kcal	kilocalorie
K_d	dissociation constant
kDa	kilodalton
K_i	inhibition constant
LB	lysogeny broth
Leu	leucine
Lys	lysine

M	mol dm^{-3}
M phase	mitosis phase
mAu	milli-absorbance units
Mdm2	murine double minute 2 protein
Mdm4	murine double minute 4 protein
Met	methionine
MHz	megahertz
ml	millilitre
MM	molecular mechanics
MMFF94s	Merck molecular force field 94 static
MM-GBSA	(See MM and GBSA.)
MM-PBSA	(See MM and PBSA.)
MS	mass spectrometry
MUP	major urinary protein
MW	molecular weight
MWCO	molecular weight cut-off
N/A	not applicable
N_{am}	amide nitrogen
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NPT	constant particle number, pressure and temperature
NVT	constant particle number, volume and temperature
OPLS	Optimized Potentials for Liquid Simulations
p53	protein 53
PAGE	polyacrylamide gel electrophoresis
PB	Poisson-Boltzmann
PBS	phosphate buffered saline
PBSA	Poisson-Boltzmann surface area
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
pET	plasmid expression T7
PhD	Doctor of Philosophy
Phe	phenylalanine
pHs	pH values
Pip-1	piperidinone 1
PIPES	piperazine-N,N'-bis(2-ethanesulfonic acid)
PK	(An AMBER force field torsion parameter.)
pKa	-log10 Ka
PME	particle mesh Ewald
Pmp	pyridoxamine phosphate
PPI	protein-protein interaction
ppm	parts per million
Pro	proline
PSA	polar surface area
QM	quantum mechanics
QSAR	quantitative structure-activity relationship
R5C3	Row 5, Column 3
RanBP2	RAs-related nuclear protein-binding protein 2
RCSB	Research Collaboratory for Structural Bioinformatics
RDC	residual dipolar coupling
RED	RESP ESP Charge Derive
RESP	restricted electrostatic potential
RING	really interesting new gene
RMS	root mean square
RMSD	root mean square deviation
S phase	synthesis phase
SA	surface area
SDF	structure data file
SDS	sodium dodecyl sulphate
SE	standard error
Ser	serine
SMILES	Simplified Molecular Input Line Entry System
SN2	bimolecular nucleophilic substitution
SPR	surface plasmon resonance
SPRi	surface plasmon resonance imaging
SS	sum of squares
SWI/SNF	switch/sucrose non-fermenting
SWIB	SWI/SNF complex B
T4	bacteriophage T4
Taq	*Thermus aquaticus*
TBE	tris-borate EDTA
tBu	tert-Butyl, C(CH3)3
TE	tris-EDTA
TEMED	tetramethylethylenediamine
TEV	Tobacco Etch Virus
Thr	threonine
TI	thermodynamic integration
TOCSY	total correlation spectroscopy
Trp	tryptophan
Tyr	tyrosine
UCSF	University of California, San Francisco
UV	ultraviolet
Val	valine
VBA	Visual Basic for Applications
vdW	van der Waals
VMD	Visual Molecular Dynamics
YFP	yellow fluorescent protein

0.5. List of Figures

0.6. List of Tables

1. Introduction

1.1. Motivation

The aim of this project was to investigate whether computational methods could be used to predict the affinity of oligobenzamide inhibitors for the p53 binding site of Mdm2 and to develop practical methods which could be used to evaluate computational predictions. This section outlines some of the key reasons for studying protein-protein interactions, in particular the p53-Mdm2 interaction and the motivation for doing so using computational methods.

1.1.1. PROTEIN-PROTEIN INTERACTIONS AND THEIR INHIBITION

Protein-protein interactions (PPIs) are ubiquitous in nature at the intercellular, intracellular and subcellular levels. Control systems and signalling pathways frequently depend on these interactions and aberrant PPIs have been implicated in many diseases. Consequently, the search for molecules which can influence these interactions has in recent years become a key focus of research. Inhibitors of protein-protein interactions are in particular demand due to their potential to be developed into new research tools and pharmaceuticals.

1.1.2. HELIX MIMETICS

McMillan *et al.*[1] report the allosteric inhibition of nitric oxide synthase dimerisation by a 480 Da pyrimidineimidazole derivative. It binds to the haem group at the active site and causes a conformational change which is propagated through the enzyme to the part of the protein responsible for dimerisation. However, allosteric inhibition is unusual; PPI inhibitors are usually competitive inhibitors which bind to the interacting interface of one of the proteins.

Leung *et al.*[2] describe the selective inhibition of lactate dehydrogenase dimerisation and citrate synthase dimerisation by cyclodextrin dimers, which bind at multiple sites on the interface. However, typically inhibitors bind in a single place to one of the proteins, mimicking an interacting element of secondary structure in the other.

Smith *et al.*[3], produced a beta-sheet proteomimetic to block the active site of HIV-1 protease and Burgess[4] describes the development of beta-turn mimetics which bind to transmembrane receptor TrkA, inhibiting its interaction with dimeric nerve growth factor. However, there is most interest in inhibitors of PPIs involving an alpha helix. In 2009, Jochim and Arora[5] analysed PPIs in the PDB, identifying ones which involved an alpha helix. In their analysis, of the 8678 PDB structures containing more than one molecule, most (7066) featured an alpha-helical protein-protein interface. Furthermore, Keskin *et al.*[6] report how alpha helices appear to be particularly prevalent at shared binding sites, something they attribute to the ability of alpha helices to form a wide variety of different interactions. Consequently, many PPI interfaces of considerable scientific interest involving hubs where information is consolidated from many different sources might involve alpha-helix-mediated binding.

In addition to being an important, ubiquitous target, PPIs involving a helix are also tractable because it often not necessary to mimic the entire helix to block the interaction; a small molecule can be effective if it mimics just the key side chains, as will now be discussed.

1.1.3. SMALL MOLECULE INHIBITORS

Most of the PPI inhibitors licensed for clinical use are monoclonal antibodies which have been "humanized" to prevent them from triggering immune responses[7]. However, the use of antibodies has several drawbacks. Firstly, in therapeutic use, antibodies must be injected, to avoid proteolysis following oral administration. Secondly, because they cannot enter cells, they can only be used against extracellular targets[8]. Finally, monoclonal antibodies are expensive to produce.

A PPI involving an alpha helix can sometimes be inhibited using a peptide comprising an interacting part of one of the proteins involved[8,9]; however, *in vivo*, many peptides are susceptible to proteolysis making them unsuitable for medical use[10]. Furthermore, peptides with fewer than 15 residues do not generally have a stable secondary structure when free in solution[11]. However, there are stable peptide analogues referred to as foldamers which cannot be broken down by proteases[9]. These include:

- "Retro-inverso peptides", synthetic peptides composed of D-amino acids in which the sequence is the reverse of that to be mimicked, to account for the changed handedness. Their dextrorotatory configuration makes these compounds less susceptible to proteolysis.
- Peptoids, in which side chains are attached to the backbone nitrogen atoms.
- β-peptides, where each backbone amino group is bonded to the β carbon, rather than the α carbon, as in naturally occurring amino acids.

- Stapled peptides, helical peptides in which a covalent linkage connects two non-natural amino acids, holding the peptide in a conformation where the amide linkages are inaccessible to proteases[12].
- Other natural structures. For example, the p53-Mdm2 interaction inhibitors of Fasan *et al.*[13,14] are stable beta-hairpins (cyclic peptides) which act as alpha-helix mimetics.

In addition to foldamers, small molecule inhibitors are also amongst the compounds under investigation as possible PPI inhibitors.

Although protein-protein interfaces typically have a surface area of 15 to 30 nm^2, a few key residues, usually at the centre of the interface, are responsible for most of the free energy released during binding[15]. While peptide analogues mimic the whole of a peptide, small molecule inhibitors imitate only these important side chains using mimicking groups held in the correct position relative to each other by a rigid scaffold. The inflexibility of the scaffold increases the specificity of the molecule for the target binding site and the prearrangement of the groups potentially makes binding more entropically favourable.

Small molecule peptidomimetics have several key advantages for development as pharmaceutical agents. Their small size makes them potentially cheaper to manufacture, increases their bioavailability[16], meaning that they can be taken orally, and makes it easier to modify them to facilitate their entry into the target cells[7,8]. Consequently, the focus of this project was the development of small molecule PPI inhibitors.

1.1.4. THE SIGNIFICANCE OF P53

p53 is one of a family of tumour suppressor proteins, so called due to their ability to suppress uncontrolled cell division, instead allowing cells division only when appropriate. In response to different cellular signals, it can trigger either temporary arrest of the cell cycle (quiescence), permanent arrest (senescence) or controlled cell death (apoptosis). Its involvement in the control of cell division and apoptosis makes it the focus of a significant amount of research into potential new cancer treatments.

p53 controls the G1/S checkpoint, one of the two main points of regulation in the eukaryotic cell cycle and the more important of the two in higher organisms such as humans. The four main stages of the cell cycle are the gap 1 (G1) phase, the DNA synthesis (S) phase, the gap 2 (G2) phase and the mitotic (M) phase[17]. The G1/S checkpoint controls progression from the G1 to the S phase. Passing the checkpoint requires cyclin-dependent kinase complexes to be active, cyclin D-cdk4, cyclin D-cdk6 and cyclin E-cdk2[18]. p53 stops the cycle primarily by inducing expression of p21 (cyclin-dependent kinase inhibitor 1A). p21 binds to and inactivates cyclin-dependent kinases[19] and also binds to proliferating cell nuclear antigen (PCNA), which is required for DNA polymerisation at the start of the S phase[20].

The regulatory activities of p53 depend upon its ability to form many protein-protein interactions, all of which have the potential for study. Some parts of the molecule, the first transactivation domain for instance, bind to key protein kinases, which phosphorylate p53, activating it, when DNA damage is detected. These include ATM (ataxia telangiectasia mutated) protein and ATR (ataxia telangiectasia and Rad3-related protein) which both phosphorylate p53 on Ser15[21]. Another protein, homeodomain-interacting protein kinase 2 (HIPK2), phosphorylates p53 on Ser46[22]. Phosphorylation on Ser37 and Ser46 of p53 is thought to affect its affinity for DNA,[21,23] preventing it from inducing p21 expression. Phosphorylation of Ser15 has a less direct effect; it reduces the affinity of p53 for murine double minute protein 2 (Mdm2)[24], an E3 ubiquitin-protein ligase which controls p53 activity. One reason for investigating p53 is the density of overlapping binding sites which mean that an inhibitor of one of these interactions could potentially be adapted to block another interaction involving the same part of p53.

1.1.5. THE INTERACTION OF P53 WITH MDM2

In normal cells, Mdm2 binds to p53, causing it to become inactive. In cells with damaged DNA or other stresses, the p53-Mdm2 complexes dissociate and active p53 triggers quiescence, senescence or apoptosis[25]. In roughly 7% of all human tumours there is overexpression of Mdm2[9]. This inhibits p53 activity so it fails to induce apoptosis or cell cycle arrest despite there being damage. Where low p53 activity in a cancer patient's tumour is caused by a mutation in the *mdm2* gene (leading to Mdm2 overexpression) rather than a mutation in the gene for p53 itself, an effective treatment could be to inhibit the p53-Mdm2 interaction.

1.1.6. THE POTENTIAL OF COMPUTATIONAL METHODS

Using high-throughput screening to find protein-protein interaction inhibitors is expensive because it requires robotic machinery and it can be time consuming. The synthesis of compounds for screening can use large amounts of material and energy. It creates chemical waste and some methods involve potential hazards. *In silico* methods are safer, often quicker and potentially cheaper. They can narrow down the number of compounds which need to be made for testing, reducing the cost of pharmaceutical development and making screening accessible to the academic community.

Despite their advantages, computational methods involve experimentation with a model rather than the real system so, to realise these benefits, research is required to identify which methods reproduce experimental results and which do not.

1.2. The p53-Mdm2 Interaction

Mdm2 binds to an alpha helix of p53 near to its N-terminal end. Using phage display, Böttger *et al.*[26] identified three residues within this region, Phe19, Trp23 and Leu26 as key "hot spots" for Mdm2 binding, residues which contribute significantly to the binding energy. Binding of Mdm2 to p53 directly prevents p53 from acting as a transcription factor, causes p53 to be exported from the nucleus and permits polyubiquitination of p53 by Mdm2, thereby targeting p53 for degradation by the proteasome[9]. p53 induces expression of Mdm2 leading to negative feedback.

Modelling can be computationally expensive so it is useful to keep models of a system as small as possible, only including the important parts. Developing inhibitors of the p53-Mdm2 interaction therefore requires a good understanding of the structures of the two interacting molecules and how they associate.

1.2.1. THE STRUCTURE OF P53

The protein p53 (Figure 1.1) comprises an intrinsically disordered N-terminal region referred to as the transactivation domain, followed, after a proline-rich length, by a long, structured DNA-binding domain (roughly half of the molecule) and then, further down the protein, an oligomerisation domain.

The transactivation domain of p53 contains two subdomains. Proteins involved in replication such as the co-activator p300 typically bind to both of these subdomains. Both have regions that fold to facilitate binding with a particular target protein. The first subdomain binds to Mdm2 and its homologue, murine double minute protein 4 (Mdm4). Within it, residues 18 to 25 fold during Mdm2 binding[27].

When binding to DNA, p53 forms a tetramer through aggregation of a beta strand and alpha helix in the oligomerisation domain of different monomers. Dissociation of the tetramer exposes a nuclear export signal in the oligomerisation domain which may lead to export of the dissociated p53 monomers from the nucleus, preventing them from acting as transcription factors. Consequently, control of tetramerisation could play a role in the control of p53 activity[27]. The extreme C-terminal region after the oligomerisation domain, might also be involved in the regulation of p53 activity[27,28].

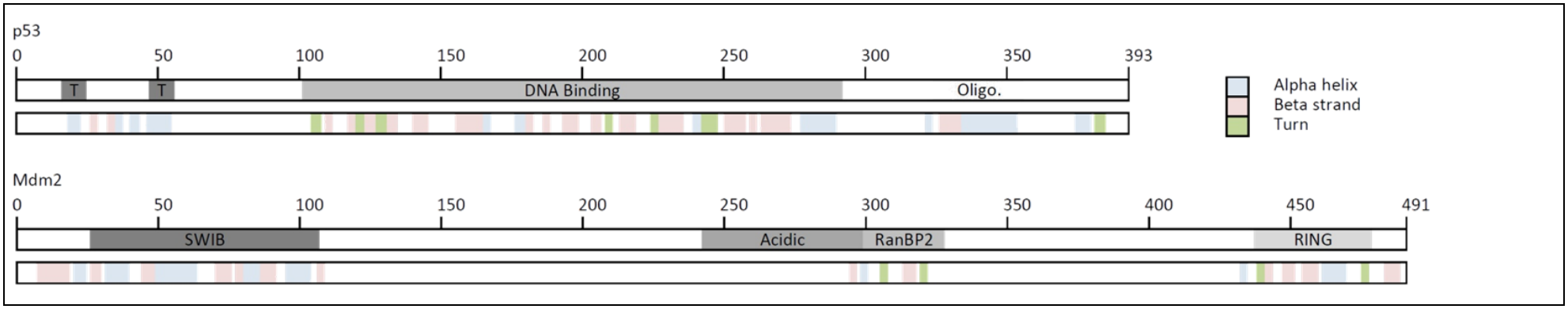


Figure 1.1: The human p53 and Mdm2 polypeptides. The numbers indicate the position of residues within the amino acid sequence. For each protein, the top band shows parts of the sequence that fold to make an important region of the protein, in p53, the two subdomains of the transactivation domain (T), the DNA binding domain and the oligomerisation domain (Oligo.): in Mdm2, the SWI/SNF (SWItch/Sucrose Non-Fermenting) complex B domain (SWIB), the acidic Asp/Glu rich domain (Acidic) and the Ran-binding protein 2 (RanBP2) and RING (Really Interesting New Gene) zinc-finger domains. The lower band shows the secondary structure: blue for alpha helix, red for beta strand and khaki for turns. This figure was constructed using Microsoft Excel with data from UniProt (www.uniprot.org) (accession codes P04637 and Q00987).

1.2.2. THE STRUCTURE OF MDM2

Mdm2 is a RING-domain family E3-ubiquitin ligase. E3-ubiquitin ligases constitute a diverse set of proteins which facilitate transfer of the protein ubiquitin from an E2 ubiquitin ligase to a specific substrate[29]. The role of an E3-ubiquitin ligase can be entirely one of recognition; however, these molecules can also play a part in catalysis.

The key parts of Mdm2 (Figure 1.1) are a p53 binding domain at the N-terminus, an acidic domain followed by a RanBP2-type zinc finger in the middle and a RING (Really Interesting New Gene) domain at the C-terminus[30].

Piperidones (derivatives of the compound piperidone) bind in the p53 binding site but at least some of their high affinity is thought to be derived from additional interactions made with residues 6 to 16 at the extreme N-terminus of Mdm2. This flexible end is thought to be able to fold over the p53 binding site[31,32].

The acidic domain is where positive and negative regulators appear to bind. It also appears to play a critical role in p53 degradation. Mutations to residues in either the 247-258 or 270-274 regions reduce p53 degradation in the cell; however, only mutations in the former have been linked to a reduction in the polyubiquitination of p53 by Mdm2, so Mdm2 may also regulate p53 activity via another mechanism[33].

In the middle of the p53 protein is a region that folds to form a Zinc finger, a supersecondary structure motif stabilised by the coordination of one of more zinc ions. In Mdm2, this region forms a 3_{10} helix followed by four beta strands[34]. It contains a binding site for the protein Ran[34], a G-protein associated with the transport of molecules between the cytoplasm and nucleus through nuclear pores[35].

Mdm2 binds to ubiquitin using the zinc ions in its RING domain. The RING domain of Mdm2 is also essential for dimerisation[30].

Figure 1.2 shows the sequence coverage of the crystal structures of Mdm2 and its homologue Mdm4 in the Protein Data Bank (PDB, www.rcsb.org)[36] at the time of writing. When predicting the binding affinity of a compound by computational modelling, the sequence coverage of the chosen structure is important; the system must be large enough to contain all of the regions that contribute to binding or are required to maintain the stability of the binding parts but too large a system makes simulations computationally expensive.

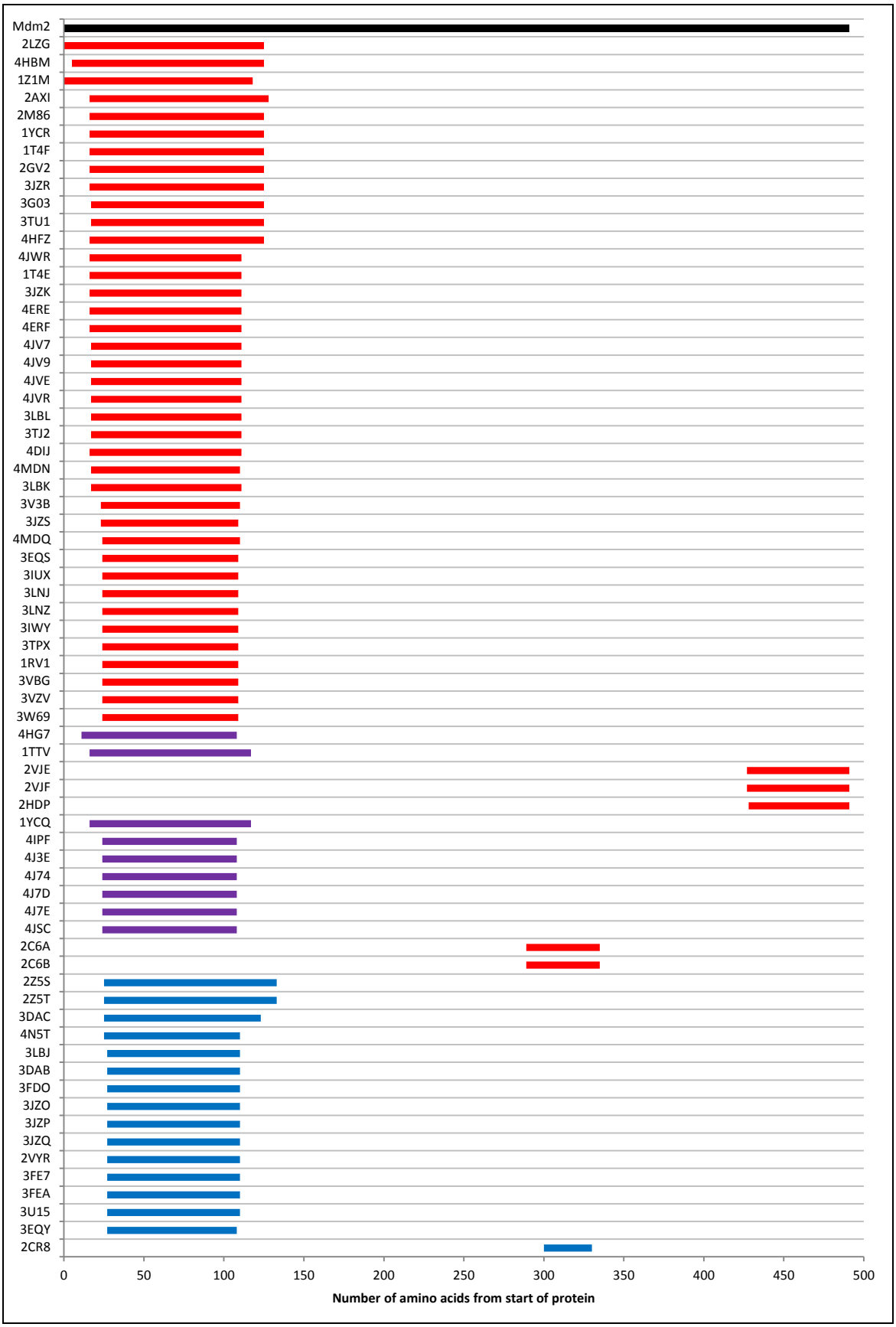


Figure 1.2: Protein Data Bank (PDB) structures containing sequences that have over 50% sequence identity with part of human p53. Red) Over 95% identity. Purple) 60 to 95% identity. Blue) 50 to 60% identity.

Structure 1T4F[37], which was used for docking and molecular dynamics simulations in this project, covers residues 17 to 125 of Mdm2. Consequently, like many crystal structures of Mdm2, 1T4F lacks the extreme N-terminal residues thought to contribute to the binding affinity of piperidones[31,32]. However, this is an assumption made in many laboratory assays where these residues are missing from the Mdm2 construct to prevent them from blocking the p53 binding site. The construct used in the wet laboratory for testing in this project covers residues 17 to 123 so computational results and experimental results should be comparable.

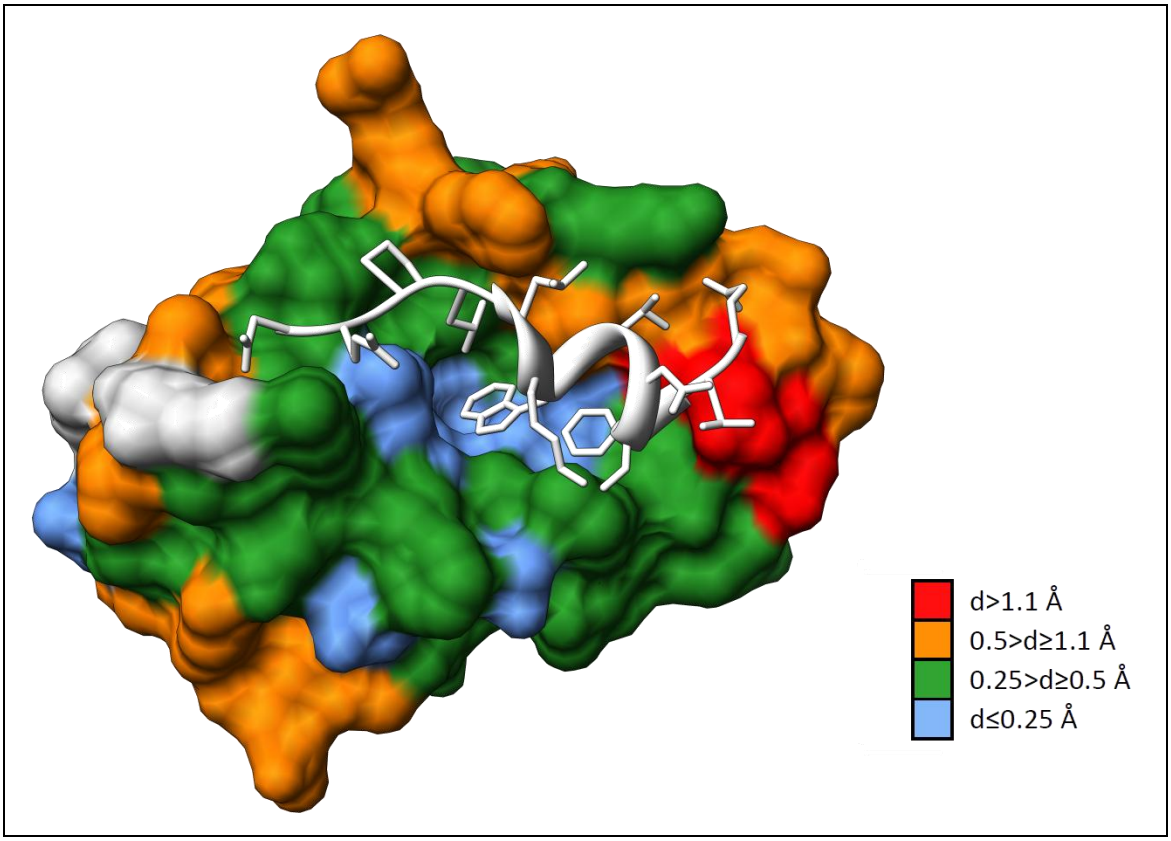


Figure 1.3: PDB structure 1YCR coloured to indicate residue movement upon ligand binding. Structure 1YCR features p53 bound to Mdm2. Each of the PDB structures 1T4E (Mdm2 with a benzodiazepine bound), 3IWY (with a dextrorotatory peptide), 3LBL (with an MI-63 analogue), 4ERE (with an N-alkylated piperidinone), 4JVR (with a morpholinone) and 4MDN (with a chloroindole derivative) was aligned with Mdm2 from the structure 1YCR (unbound Mdm2). These PDB structures were chosen to include a diverse range of bound ligands whilst ensuring 100% sequence identity and a good degree of coverage. The root mean square distance (RMSD) of each residue's alpha carbon from the average position of the alpha carbon (over all 7 aligned structures) was calculated for residues 27 to 107. The residues were coloured according to this RMSD value as indicated in the key. Notably, there is little movement around the centre of the binding site where Phe19, Trp23 and Leu26 of p53 interact; however, there is a particularly flexible region to one side of the binding site. This part is within the region covered by residues 68 to 72, a region absent from 4ERE and partially absent from 4JCR. The red colouration is thus based on results from structures 1YCR, 3IWY, 3LBL and 4MDN only. The absence of these residues from 4ERE and 4JCR could be partly responsible for the apparently high mobility in this region; however, that these residues are missing from some structures is also consistent with them having more propensity to move.

PDB structures of part of Mdm2 usually feature another ligand such as Mdm4 (2 structures at the time or writing), a peptide (16 structures) or a small molecule (31 structures). The vast majority were produced by X-ray diffraction; however, the four structures of Mdm2 parts in isolation (1Z1M, 2C6A, 2C6B and 2HDP) were produced using nuclear magnetic resonance (NMR). This highlights how the flexibility of the unbound protein makes crystallography difficult.

Superimposition of a selection of Mdm2 structures from the PDB gives an indication of which parts of the protein are most flexible (Figure 1.3). The results reveal the significant movement possible around the p53-binding site and thus highlight the need to consider protein flexibility when making binding affinity predictions.

1.2.3. INHIBITION OF THE P53-MDM2 INTERACTION

The transactivation domain of p53 is responsible for its action as a transcription factor. A transcription factor increases the activity of proteins indirectly by raising the frequency with which their genes are transcribed. Mdm2 binds to an alpha helix in the first subdomain of the p53 transactivation domain, which lies within the region comprising residues 15 to 29 of the human protein[38]. It thereby physically prevents the interaction of this region with DNA. Peptides of the residues in this region of p53 will bind to Mdm2 and compounds developed to inhibit the p53-Mdm2 interaction (Figure 1.4) are referred to as peptidomimetics, because they mimic these helical peptides.

1.2.3.1. Peptidomimetics

Small molecule, peptidomimetic inhibitors of the p53-Mdm2 interaction include β-hairpin mimetics, terphenyls, nutlins, chalcones (flavanoid precursors), aryl sulphonamides, isoindolinones, spiro-oxindoles, 1,4-benzodiazepine-2,5-diones, N^{α}-acyl-tryptophanyl-piperazides and chlorofusin[9,39].

PPI inhibitors can be identified by high-throughput screening in the laboratory[15] and this has played a major role in the identification of inhibitors of the Mdm2-p53 interaction. For example, in 2004 a group of imidazoline inhibitors of the p53-mdm2 interaction, which are referred to as Nutlins, (Figure 1.5) were discovered by the pharmaceutical company Roche using this method. The image on the left of Figure 1.4 shows for comparison how p53 and Nutlins appear to bind to Mdm2. The stereochemistry shown in Figure 1.5 is necessary for tight binding[40].

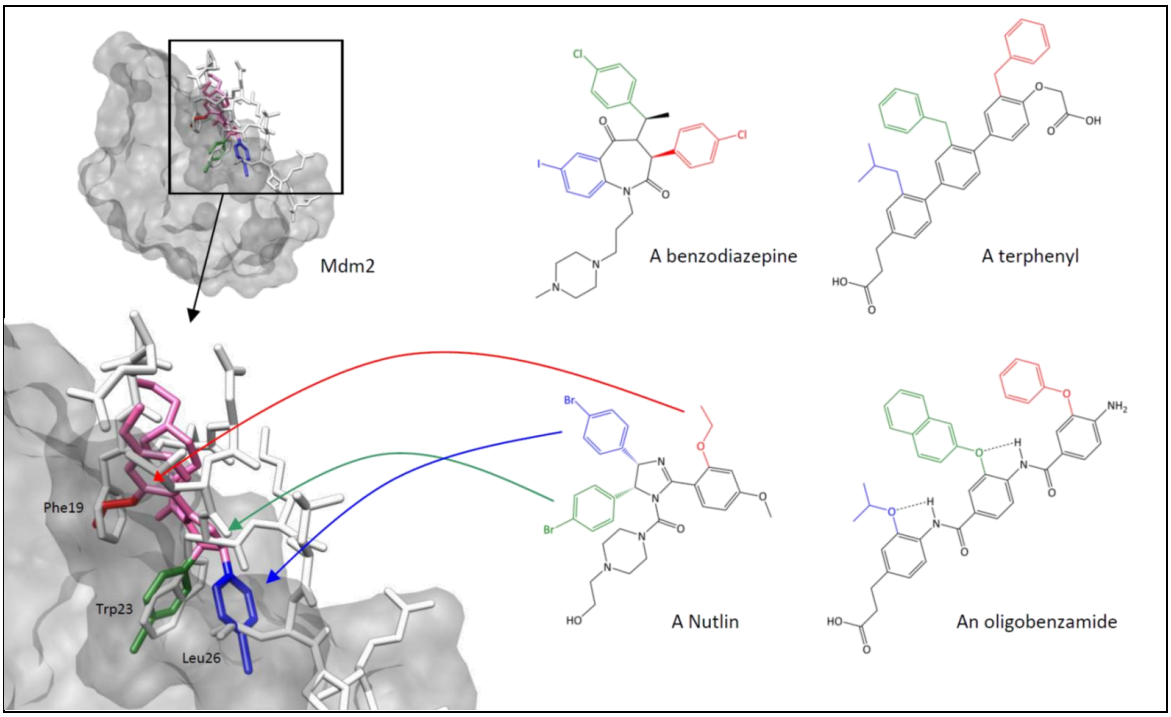
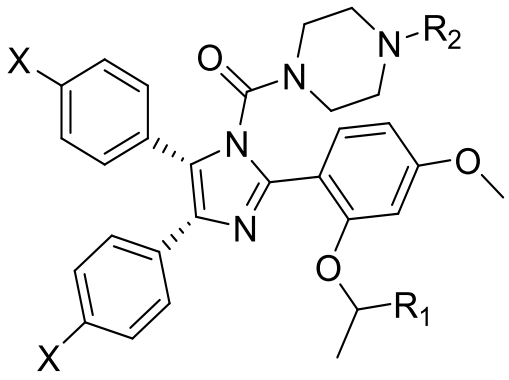


Figure 1.4: How p53-Mdm2 interaction inhibitors mimic residues Phe19, Trp23 and Leu26 of human p53. The parts of the inhibitors which mimic these residues are shown in red, green and blue respectively. PDB structure 1YCR (Mdm2 bound to the p53 transactivation domain) was aligned with structure 1RV1 (Mdm2 bound to Nutlin-2) using the Chimera (UCSF) MatchMaker tool which aligns residues based on the position of their alpha carbon atoms. On the left is shown the p53 helix (white) bound to Mdm2 (grey) from 1YCR and Nutlin-2 from 1RV1 (in colour). On the right are examples of four different types of inhibitor, a benzodiazepinedione[9], a terphenyl[39], nutlin-2[41] and an oligobenzamide[42]. Evidence of which compound side chains resided in each p53 residue binding site (indicated by the colours) came from PDB structure 1T4E (for the benzodiazepine) the modelling of Fischer *et al.*[39] (for the terphenyl), structure 1RV1 (for the Nutlin) and the modelling of Plante *et al.*[42] (for the oligobenzamide). Images from Chimera (USCF) and ChemDraw (CambridgeSoft) were combined using PowerPoint (Microsoft) and Paint Shop Pro (Corel).

It has been suggested that the rigidity of Nutlins makes their binding to Mdm2 more entropically favourable; however, their lack of flexibility also means that they cannot bind to the Mdm2 homologue Mdm4[43]. While Nutlins are the most effective known inhibitors of the p53-Mdm2 interaction *in vitro*, clinical trials conducted with some Nutlin-3 isomers[44] suggest that they are not effective inhibitors of cell proliferation *in vivo*, perhaps as a direct result of their failure to bind to Mdm4[45,46]. Mdm4 has, like Mdm2, been implicated in the control of cell proliferation through negative regulation of p53 activity and it too is often mutated in tumour cells[47,48]. Inhibition of the p53-Mdm2 interaction might be insufficient to restore p53 activity *in vivo*; compounds might also need to bind to Mdm4 and inhibit the p53-Mdm4 interaction to be effective.



Nutlin-1: X=Cl, R_1=CH_3, R_2=$C(O)CH_3$

Nutlin-2: X=Br, R_1=H, R_2=CH_2CH_2OH

Nutlin-3: X=Cl, R_1=CH_3, R_2=H

Figure 1.5: The Nutlins. Each compound has two chiral centres making 4 possible 4 stereoisomers. The stereochemistry of the active isomer of Nutlin-3 is shown. The two leftmost side chains, each attached to a chiral centre, project back into the page, away from the plane of the central imidazole ring. The method originally used to synthesise Nutlin-3 yields a racemic mixture of the active isomer and its enantiomer. The latter binds 100 times less strongly than the active isomer.

Chalcones are another example of a type of p53-Mdm2 interaction inhibitor found to be ineffective *in vivo*. They appear to not only bind to Mdm2 but also other proteins. For example, they bind to p53, preventing its interaction with DNA and thus its action as a transcription factor, thereby defeating the purpose of inhibiting the p53-Mdm2 interaction: to increase p53 activity[9].

The poor *in vivo* results obtained with the Nutlins and chalcones epitomise the difficulty with which compounds identified in the laboratory are translated into lead compounds for pharmaceutical research. They highlight the need to continue the search for inhibitors despite the extent to which the p53-Mdm2 interaction has previously been studied.

1.2.3.2. Oligobenzamides

Oligobenzamide peptidomimetics (Figure 1.6, p36) have been synthesised in the Leeds Chemistry Department by the Wilson group[49]. The scaffold of these small molecules comprises alternating amide and phenyl groups.

One of the unique merits of these compounds is that it should be possible to make a compound displaying any combination of side chains with reasonable ease by utilising the hardware which already exists for solid state peptide synthesis[50] (Figure 1.7). Monomers are synthesised first to make a library of subunits from which the desired combination can be taken and assembled by Fmoc synthesis. The potential to make many compounds is useful, firstly, because the validation of computational methods is more easily performed when there is practical data for a large number of molecules and, secondly, because modification of the sequence of side chains presented could theoretically allow many different interactions to be targeted.

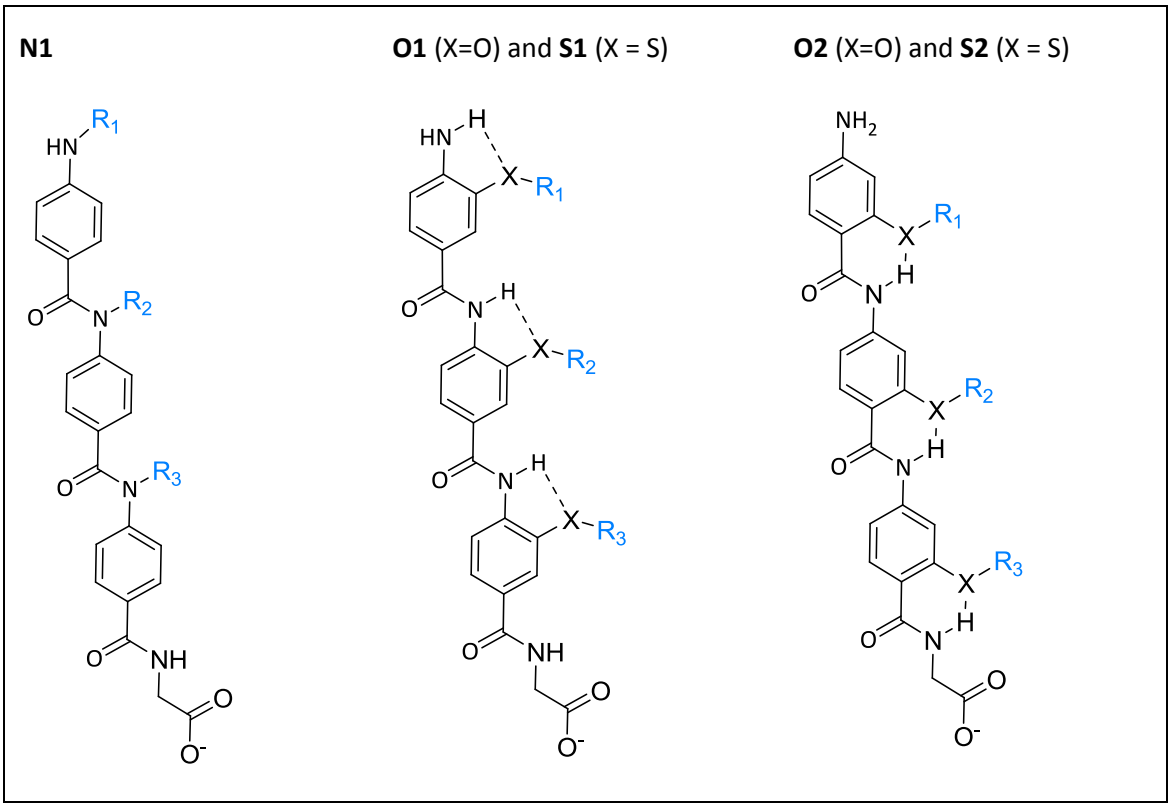


Figure 1.6: The oligoamide scaffolds used in this project. Each scaffold has three side chains, R_1, R_2 and R_3. Each side chain is held not by carbon but by an electronegative heteroatom resulting in inductive electron withdrawal from the side chain. This can affect the relative stability of side chain tautomers.

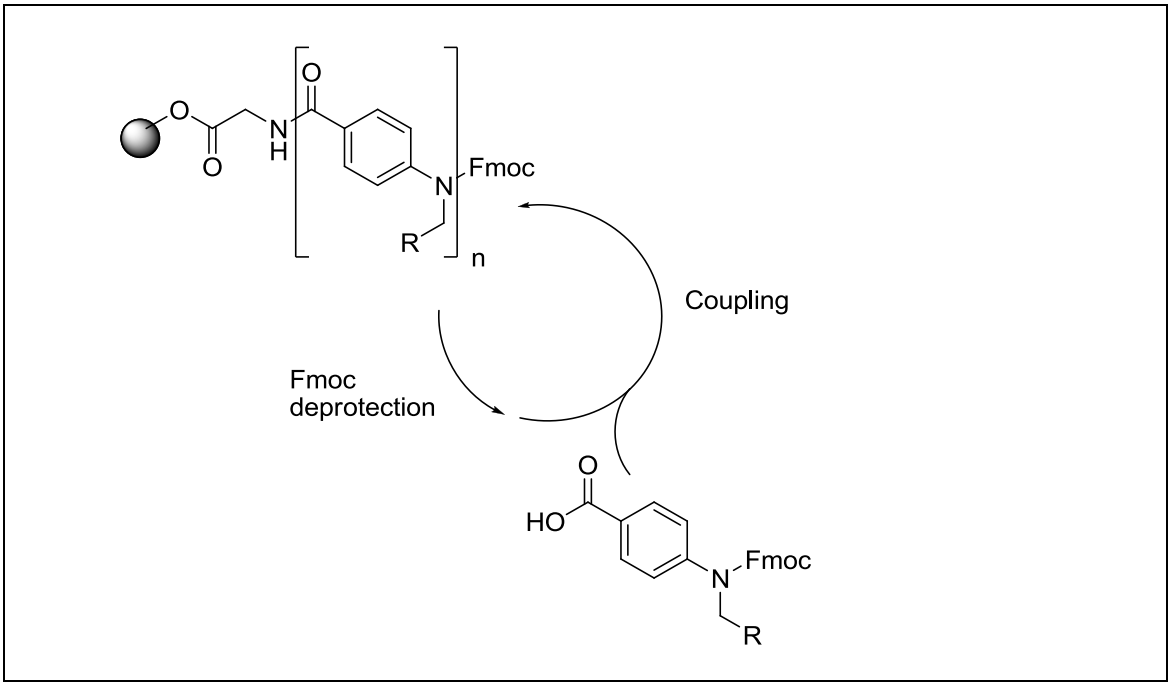


Figure 1.7: Synthesis of N-linked oligobenzamides using Fmoc protection

The combination of side chains alone is intended to control the specificity so the side chains should ideally be held prearranged so the compound cannot adapt to bind to another protein. In oligobenzamides where side chains are attached through oxygen substituents on the rings of the scaffold, intramolecular hydrogen bonding between the oxygen and amide hydrogen atoms of the scaffold restricts internal rotation. For example, attachment at the ortho ring position causes rotation about the amide nitrogen to ring carbon bonds in the core to be restricted[49]. In addition to its effect on specificity, this intramolecular hydrogen bonding can potentially favour binding because it reduces the associated decrease in entropy.

In this project, 5 different oligobenzamide scaffolds (Figure 1.6) were investigated *in silico*. Some compounds based on the N, O1 and O2 scaffolds shown have been synthesised by the collaborating chemistry group.

The limited flexibility of oligobenzamides means that the scaffold must place side chains close to their final intended positions, the locations in which they will most faithfully mimic the targeted alpha helix. Azzarito *et al.*[51] compared distances in the p53 pose of PDB structure 1YCR (p53 bound to Mdm2) with those in oligobenzamides poses. In the p53 helix, distances of 5.79 Å and 5.97 Å separate Phe19 and Trp23 and Trp23 and Leu26 respectively. Similar distances exist between the oligobenzamide side chains, specifically between the oxygen atoms connecting the side chains to the scaffold. These are, respectively, 6.18 Å and 6.17 Å in the case of O1 scaffold and 6.16 Å and 6.21 Å in the case of the O2 scaffold.

In addition to the distances between side chains, intramolecular angles are also important. In an alpha helix there are 3.6 amino acids per turn, corresponding to a 100° twist per residue. Moving from one turn to the next along one side of this right-handed helix, each amino acid, i, is positioned at an angle of 60°, anticlockwise, from residue i-3. Monte-Carlo modelling by Azzarito *et al.*[51] suggests that the corresponding angle between side chains in relaxed molecules based on the O1 scaffold with isopropyl side chains is less, 35° to 40° and for the O2 scaffold it is only 10° to 20°. However, oligobenzamides appear to have sufficient flexibility for effective alpha helix mimicry despite the intramolecular hydrogen bonding. For example, in the O1 scaffold, steric effects mean that the amide carbonyl groups are not quite in the same plane as the adjacent rings. The carbonyl groups do not form a conjugated system with the adjacent rings so rotation is permitted about the amide carbon to ring carbon bonds[49]. Molecules based on the O1 and O2 scaffolds were synthesised and tested by Azzarito *et al.*[51]. They obtained similar results with the two scaffolds despite their different connectivity, a finding which the authors attribute to similar side chain presentation enabled by this flexibility.

The flexibility of oligobenzamides can be increased by the addition of a flexible linker to the middle of the oligobenzamide chain. Kulikov *et al.*[52] discuss various methods of combining two oligobenzamide structures with a spacer to produce double alpha helix mimetics, long

molecules bent in the middle which can mimic the supersecondary structure of two adjacent alpha helices similarly arranged.

PPI inhibitors often have a low solubility because they are intended to mimic typically hydrophobic residues at PPI interfaces. Dimethylsulphoxide (DMSO) is a polar solvent which is aprotic, meaning that it cannot act as a hydrogen bond donor. It dissolves both polar and non-polar molecules and a small quantity added to an aqueous solution can increase the amount of a less polar compound, such as an oligobenzamide, which will dissolve in the solution. This enables investigations which would otherwise be precluded by precipitation to proceed. A drawback of using DMSO is that, because it increases the solubility of hydrophobic amino acid side chains, the stability of proteins can be adversely affected. This is relevant to binding assays, where denaturation can decrease specific binding or increase non-specific binding, and structural methods such as NMR and crystallography, where a protein is assumed to be in its native state.

Although oligobenzamides are straighter than the helices they are intended to mimic, oligobenzamide scaffolds do have a slight twist, a gradual bend which means that a long oligobenzamide would not be straight. The degree of curvature can be tuned by changing substituents[53] and is slightly more pronounced in scaffold O1 (the 3-O-alkylated scaffold) than in scaffold O2 (the 2-O-alkylated structure)[51]. Modelling of the scaffolds using isopropyl side chains at all three positions suggests angles of inclination of just under 160° that introduce a turn in the direction of travel of just over 20° per oligobenzamide unit. Meta-linked oligobenzamides have a much greater curvature, forming cyclic[54] or helical[55] structures if of sufficient length.

1.3. Binding Energy

1.3.1. THE THERMODYNAMICS OF INTERACTIONS

In preface to the proceeding discussion of binding energy estimation, this section introduces binding energy and how it pertains to binding affinity.

1.3.1.1. Equilibrium Constants

The binding energy is of a compound is, specifically, the Gibbs free energy of binding, in this case, within aqueous solution.

$$\Delta G = \Delta H - T\Delta S \quad \textbf{1.1}$$

Where ΔG is the Gibbs free energy change that accompanies a reaction in which there is an enthalpy change of ΔH and an entropy change of ΔS at a temperature of T.

The binding affinity of compounds can be described by the equilibrium constant of their binding.

$$K = \frac{[Complex]}{[Protein][Ligand]} \quad \textbf{1.2}$$

Where K is the equilibrium constant and the square-bracketed words represent the concentrations of the complex, free protein and free ligand.

$$\Delta G° = -RT \ln K = RT \ln K_d \quad \textbf{1.3}$$

Where ΔG is the free energy of binding and R is the universal gas constant (8.314 J K^{-1} mol^{-1}). The dissociation constant, $K_{d,}$ is the reciprocal of the association constant, K and is a measure of the affinity of two species for each other. The greater the affinity, the smaller the dissociation constant and the more negative the standard free energy of binding.

K_d is temperature dependent.

$$K_d = e^{-\Delta G°/RT} \quad \textbf{1.4}$$

1.3.1.2. Factors Affecting Binding Energy

When factors that affect binding energy are discussed, it can be useful to refer to enthalpic contributions to the free energy (due to factors which affect ΔH but not ΔS) and entropic contributions (caused by factors affecting ΔS). Table 1.1 lists some enthalpic and entropic factors that affect ligand binding. As shown by Equation 1.1 (p39), the higher the temperature, the more effect the change in entropy has on the free energy of binding. Entropic effects usually make binding less favourable so increasing the temperature tends to cause the dissociation of complexes.

Table 1.1: Enthalpic and entropic factors that can affect the standard free energy of binding for a ligand interacting with a protein

	Enthalpic contributions	Entropic contributions
Favouring binding	• Electrostatic interactions and/or hydrogen bonds, often at the binding interface, which are only made when binding has occurred or which are stronger after binding has occurred[56]	• Release of water molecules from the solvation shell around hydrophobic surfaces[57] hidden by binding
Disfavouring binding	• Lack of van der Waals forces with the solvent where there is a solvent-free void created by binding • Steric hindrance (the exchange interaction) and torsional stress due to the protein or ligand taking up an unfavourable conformation to enable binding • Electrostatic repulsion due to similarly charged atoms on the ligand and protein being bought together on binding • Breaking of hydrogen bonds and electrostatic interactions between the solvent and the binding molecules where charged atoms are brought next to a hydrophobic region on the other molecule	• Loss of translational freedom of the ligand relative to the protein. • Loss of rotational and conformational freedom of the ligand[58] • Increased rigidity within the protein upon binding for steric reasons or due to new interactions forming or interactions involving fixed ligand atoms as opposed to moving solvent. This could be at the binding site or some distance away in the protein. • A decrease in osmotic pressure upon binding

Ligand modifications that increase or decrease the enthalpy change of binding are often associated with a compensatory shift in the entropy change of binding and vice versa. This can lead to a counterintuitive change in binding energy when a chemical series is traversed[58]. Many possible causes of this phenomenon, termed enthalpy-entropy compensation, have been proposed but its origin remains unclear[59] and could be very system-specific.

As an example, Homans[60] discusses how, upon ligand binding, the entropically favourable displacement of water molecules from the Major Urinary Protein (MUP) ligand binding site is offset by the enthalpically unfavourable breaking of hydrogen bonds between the displaced water molecules. This is a result of the MUP binding site being only partially hydrated. The entropic effect would be expected to make the binding of larger ligands less favourable because this leads to the displacement of more solvent molecules; however, its compensation means that the dispersion force between the protein and ligand has the potential to predominate, favouring larger molecules with a shape complementary to that of the binding site.

Despite such observations, enthalpy-entropy compensation could, to a greater[59] or lesser[61] extent, be an artefact of binding data processing. Due to the difficulties of estimating enthalpic

and entropic components and the tendency for compensation, Chodera and Mobley[59] believe it can be most useful to focus on direct assessment of binding free energy during computational ligand development rather than estimate changes in enthalpy and entropy.

1.3.1.3. Protein-protein interactions

PPIs[6] typically have a free energy of binding that is between 25 and 80 kJ mol^{-1}. The entropically favourable burying of hydrophobic residues in the interface is thought to largely drive the association of interacting proteins[62]. Consistent with this hypothesis, proteins which are permanently part of a homodimer have hydrophobic interfaces that resemble the core of a protein[63]. Furthermore, comparison of the interacting and non-interacting surfaces of proteins that fold separately and then bind to form a heteromer reveals that the interacting surfaces are flatter and have a larger solvent accessible surface area than the non-interacting surfaces[64], features which will maximise the number of water molecules displaced (and thus the increase in entropy) upon binding.

Although the binding interfaces of heteromers have fewer charged residues[64], in some heteromers, residues forming electrostatic interactions at the interface promote binding[65–67]. The hydrophobic environment around charged and polar residues on an interacting surface is thought to increase the strength of electrostatic interactions by excluding bulk solvent, decreasing the dielectric constant[68]. Charged and polar residues might also have an important role in the stabilisation of unbound monomers, which would otherwise become denatured due to their having an excessively hydrophobic interface region[62,63]. The need to ensure monomer stability could also explain why heteromeric complexes have smaller interacting interfaces than proteins forming homomers[69].

Interacting surfaces can be divided into regions of differing residue frequency. Li *et al.* compared the centre of PPI interfaces to their periphery and observed that the centre was generally more hydrophobic, estimating a typical reduction of 190 J mol^{-1} $Å^{-2}$ to the free energy of binding due to the exclusion of solvent from the surface of hydrophobic residues in the centre[70].

Within the core, some residues are particularly important for binding[71]. These have been termed hotspots. Wells *et al.*[72] suggest that a hotspot is a residue contributing 2 kcal mol^{-1} (8.4 kJ mol^{-1}) or more to the binding energy as measured by alanine scanning. There is evidence that hotspots are found clustered together in tightly packed "hot regions"[73]. Dense packing in hot regions could partly explain why hotspots tend to be highly conserved in evolution[74].

Solvent exclusion by the tight packing of hot regions could increase the strength of electrostatic interactions involving hotspot residues[6]. The simulations of Keskin *et al.*[6] indicate that conserved residues at interfaces are less mobile so, in addition to its effect on solvation, tight

packing could make binding more entropically favourable by causing residue prearrangement. The results of molecular dynamics simulations by Rajamani *et al.*[75] are consistent with this hypothesis. In general, Smith *et al.*[76], found reduced flexibility in core interface residues as opposed to peripheral interface residues and on a larger scale, hotspots appear to occur in preformed pockets complementary to the shape of the binding partner so little conformational change is necessary for binding[77].

Despite the clear rigidity at some PPI interfaces, many proteins forming PPIs are intrinsically unstructured[78] and this project concerns the binding of Mdm2 to the intrinsically unstructured p53 transactivation domain. While the region this domain binds on Mdm2 is not unstructured, it too has considerable flexibility.

This is not the only respect in which the p53-Mdm2 interaction is atypical. While PPIs generally appear to be entropically driven by the hydrophobic effect, the binding of cytochrome c to cytochrome c peroxidase being a classic example[67], isothermal titration calorimetry reveals that binding of the p53 transactivation domain peptide to Mdm2 is entropically unfavourable with enthalpy being the driving force[79]. This enthalpy change could stem from the formation of hydrogen bonds within the p53 helix upon binding; although, analysis of simulations using the MM-PBSA method (p86) suggests that Trp23 on p53 also hydrogen bonds with Mdm2 and interacts via van der Walls forces with many nearby Mdm2 residues[80]. While the binding of wild type p53 peptide to Mdm2 might be enthalpically driven, mutations can change the balance between enthalpy and entropy, particularly if they change the orientation of Mdm2 residue Tyr100 upon binding because this has a big impact on the size of the binding site[79].

1.3.2. VALIDATION: PRACTICAL METHODS OF DETERMINING BINDING AFFINITY

In order to demonstrate that computational methods can be used to predict the binding affinity of oligobenzamides, accurate *in silico* predictions are needed; however, accurately measured affinities to test the predictions are also important.

A precursor to the successful estimation of ligand affinity is the successful identification of binding pose. In the case of oligobenzamides, this step is potentially challenging because, as discussed in the computational results chapter (p153), many binding positions are possible depending on the choice of side chains. Consequently, methods of determining where ligands bind to a protein in the laboratory are also important for validation.

This section discusses different methods in which the binding location and affinity of a ligand can be investigated experimentally.

1.3.2.1. Methods for the Assessment of Binding Affinity

The complex formed when a ligand binds to a protein is typically much larger than the ligand alone. This difference in size is exploited by many laboratory assays that detect binding. These include fluorescence anisotropy, microscale thermophoresis and surface plasmon resonance (SPR).

In fluorescence anisotropy, polarised light is used to excite a fluorescently labelled ligand. As a result of the time delay between excitation and emission of the fluorophore, the polarisation of the emitted light (the anisotropy of the fluorescence) depends on the correlation time of the particle, the average amount of time which must have elapsed for the particle to have rotated one radian in any direction[81]. Larger particles tumble more slowly increasing the fluorescence anisotropy. Consequently, the binding of fluorescently labelled ligand molecules to a protein can be detected.

A relatively new technique, microscale thermophoresis, can be used to separate particles according to their size. When placed in a temperature gradient, particles are encouraged to move down the gradient because the particles hitting them on the warmer side have more kinetic energy (and therefore exert a greater force upon collision) than the particles hitting them on the cooler side[82]. Larger particles move more slowly down the gradient and may even move towards the heat source, driven by the pressure difference caused by smaller particles moving quickly around them down the gradient.

In both of the methods above, finding the ligand concentration at which there is half maximal inhibition of the protein-protein interaction involves testing a number of different concentrations of the ligand and producing a titration curve.

The dissociation constant (K_d) of a protein-ligand interaction can theoretically be calculated using the rates of association (k_a) and dissociation (k_d)[83(p46)].

$$K_d = \frac{k_d}{k_a} \qquad 1.5$$

Surface plasmon resonance (SPR) spectroscopy can be employed to measure k_a and k_d. This technique exploits the change in refractive index of a thin metal film when species bind to its surface[84]. The ligand is bound to the metallic surface, typically gold, which coats a prism. Light is directed onto the back of a gold film and the electrons involved in the surface plasmon (electrons cyclically oscillating at the air-metal interface) become excited. This causes the absorption of light at a particular angle of incidence dependent on the refractive index at the surface and an attendant lower intensity band in the totally internally reflected light. If a solution of the protein is passed over the top of the film, any ligand protein binding changes the refractive index of the surface and this alters the angle of the lower intensity band[84]. The angle

can be measured (scanning angle SPR)[85]; however, in the last few years, a technique known as surface plasmon resonance imaging (SPRi) has been developed, in which the intensity of the reflected light is measured at a fixed angle[86]. The latter enables parallel testing of a large number of species arrayed on the chip. A problem with the use of SPR to study the binding of small molecules to a protein is that attachment of the compounds to the chip could affect their affinity for the protein.

The methods above detect a change in particle size when binding occurs. However, other methods exist which do not rely on this. One such method is dual polarisation interferometry (DPI). DPI is similar to SPR, but the refractive index of the metal surface is detected by interferometry[84]. Total internal reflection occurs multiple times at both the solvent-covered and opposite prism surfaces as the light, which is polarised, bounces down the long, thin prism waveguide. The refractive index determines the number of reflections and thus path length of the lower intensity light, which in turn determines the polarisation phase of the light relative to that of a second beam from a reference waveguide with which it is mixed at the prism end to create an interference pattern[87]. Unlike SPR, DPI can sometimes detect conformational changes in a protein[87].

Unlike the methods described above, isothermal titration calorimetry (ITC) can be used to identify the enthalpic and entropic components of binding energy[88]. The method involves measurement of the energy required to maintain the temperature of a solution of one binding component as a solution of a second binding component is titrated in to it[88]. The energy difference is quantified by way of its comparison with the energy required to maintain the temperature of a reference solution. The entropic (temperature dependent) and enthalpic (temperature independent) components of the binding energy can be determined by repetition of the titration at different temperatures[88]. ITC cannot be used for high throughput screening because, for ITC to work, a large amount of protein must be used per compound in order for the energy change to be large enough to be measured accurately[89].

In this project, binding affinity was assessed using Förster resonance energy transfer (FRET). In this method, the proximity of two fluorophore labels is detected as a result of the non-radiative transfer of energy between the molecules. This causes excitation of the donor fluorophore to cause emission from the acceptor fluorophore. FRET does not require attachment of the ligand to a surface as in SPR, nor large amounts of compound as in SPR. FRET is also better evidence of binding than a decrease in fluorescence anisotropy.

The methods of fluorescence anisotropy and FRET are now described in more detail.

1.3.2.2. Fluorescence Based Methods

When a molecule absorbs a photon of light, an electron in the molecule can be elevated to an excited state. Because only certain electronic configurations are possible in a molecule, the electronic energy is quantised; its level is restricted to a discrete number of quite broadly spaced values (quanta). Photon induced electronic transitions, because they involve a rapid change in the state of an electron, can destabilise a molecule and result in an oscillating motion. This concerted electronic and vibrational excitation (vibronic excitation) increases the number of different photon energies (wavelengths of light) which can be absorbed[90(p5)]; although, only vibrational modes which are compatible with both the ground and excited electronic states can occur (the Franck-Condon principle)[90(p7)].

The vibrational energy is transferred to adjacent molecules when collisions occur, increasing the temperature of the surroundings. The electronic energy released when the excited electron returns to its ground state can be converted to further vibrational energy or, alternatively, lost in the form of a photon, a phenomenon known as fluorescence. Because re-emission occurs after some energy has been lost, the emitted photon has less energy and consequently a longer wavelength than that of the incident light. A third mechanism of losing this electronic energy is through FRET, a phenomenon that can occur when an excited fluorophore (a fluorescent compound that can re-emit light upon light excitation) is close to another fluorophore in the ground state.

1.3.2.2.1. FRET

Following the excitation of a fluorophore, the transition dipole of the excited electron can exert a force on a nearby dipole as the molecule returns to its ground electronic state. Where two fluorophores are very close together, the resulting interaction (coupling) of the two dipoles (one in each fluorophore) can cause the concerted motion (resonance) of the dipoles resulting in the non-radiative transfer of energy from one fluorophore (the donor) to the other (the acceptor). This is known as Förster Resonance Energy Transfer (FRET).

The relative rates of FRET and normal emission from a fluorophore determine what is known as the FRET efficiency, E_{FRET}, the probability of FRET occurring after a photon has been absorbed by the donor fluorophore. The rate of FRET depends on the relative orientation of the fluorescent centres, the distance between them and the extent to which the emission spectrum of the donor and the absorption spectrum of the acceptor overlap[91]. FRET is extremely sensitive to small increases in separation distance because the rate of this energy transfer is inversely proportional to the sixth power of the distance between the donor and acceptor. This occurs because the dipole-dipole interaction energy is inversely proportional to the distance cubed and the probability of energy transfer is proportional to the square of this interaction energy[92].

$$E_{FRET} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \quad \textbf{1.6}$$

Where r is the distance between the donor and acceptor and R_0 is the distance of 50% FRET efficiency, the Förster distance. In the first part of the FRET work in this project, enhanced green fluorescent protein (eGFP) was used as the donor and enhanced yellow fluorescent protein (eYFP) was used as the acceptor. The Förster distance, the distance at which there is 50% transfer efficiency, is approximately 5.6 nm for this pair[93].

$$R_0 = \sqrt[6]{\frac{9(\ln 10)\kappa^2 Q_D J}{128 N_A \pi^5 n^4}} \quad \textbf{1.7}$$

Where Q_D is the quantum yield of the donor (the proportion of absorbed light reemitted), N_A is Avogadro's number (6.022×10^{23} mol^{-1}) and κ is the dipole orientation factor, a variable describing the relative orientation of the donor and acceptor that ranges from 0 (dipoles antiparallel) to 2 (dipoles parallel)[94(p13)]. For a donor and acceptor tumbling rapidly with respect to the time period between absorption and reemission by the donor, the time-averaged orientation factor is two thirds. J is the spectral overlap.

$$J = \int_0^\infty \lambda^4 F_D(\lambda) \varepsilon_A(\lambda) \mathrm{d}\lambda \quad \textbf{1.8}$$

Where F_D is the normalised fluorescence of the donor at the wavelength λ. The overlap is scaled by the wavelength to the power 4 because the efficiency of transfer is proportional to λ^4. ε_A is the extinction coefficient of the fluorophore solution. An extinction coefficient describes the rate at which the intensity of light entering a medium is exponentially diminished.

$$-\log_{10}\frac{I}{I_0} = A = \varepsilon c l \quad \textbf{1.9}$$

Where I is the intensity of light a distance l into a solution of concentration c, extinction coefficient ε and therefore absorbance A if it enters at an intensity of I_0.

Unlike the other methods outlined in section 1.3.2.1 above, FRET can be used to carry out *in vivo* assays since if both the binding components are proteins with a fluorescent protein label, they can be synthesised in a cell. FRET is also very amenable to high-throughput screening. Different compounds can be tested in the wells of microwell plates, which can in theory be pipetted out and tested robotically[95]. Also, it is possible to perform competition assays in which a ligand competes with a labelled species (p53 for example), to bind to the second labelled molecule (Mdm2 for example). These have the advantage that the ligands do not have to be labelled, a process which might affect the binding capacity of each compound differently.

In this project, the hybrid proteins p53-YFP and p53-Cherry were used to investigate the binding of p53 to GFP-Mdm2. Figure 1.8 shows diagrammatically how fluorescent labelling of

Mdm2 and p53 can result in FRET. On binding, the YFP labelled C-terminal end of the p53 is brought into close proximity with the GFP, bringing about FRET.

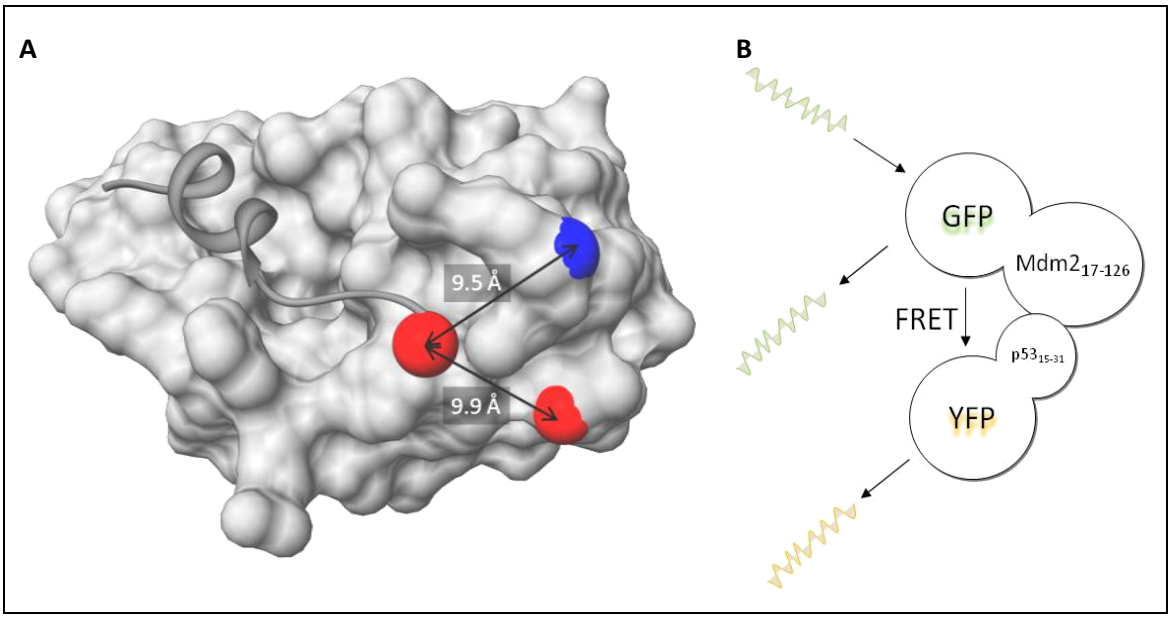


Figure 1.8: Diagrams showing how fluorescent labelling of Mdm2 and p53 can result in FRET. A) A computer generated structure of Mdm2 produced using PDB structure 1T4F showing the C-terminal end of p53 (dark grey) in red and the N (blue) and C (red) termini of Mdm2. The N-terminal end of Mdm2 was fluorescently labelled because it is closer to p53 than the C-terminal end. B) A diagram showing the principle of the FRET experiment. Binding of p53 and Mdm2 brings the GFP and YFP labels together causing FRET to occur.

Many assumptions are made when carrying out FRET measurements.

In this project, it was assumed that the YFP, GFP and Cherry labels had no effect on the binding of the peptide; however, GFP, YFP and Cherry are all considerably larger than the peptide and could get in the way of binding, reducing the binding affinity. For more details regarding these fluorescent labels see p49.

In a titration monitored using FRET, it is usually presumed that there are two FRET states, one in which the fluorophores are close and there is considerable FRET, and another in which the fluorophores are separated, resulting in little or no FRET. As the titration proceeds, the emission spectrum is assumed to transform from that one of these states to that of the other and between these points, the amount of FRET is assumed to be directly proportional to the number of donor-acceptor complexes remaining. In reality, there could be intermediate states with their own emission spectra.

A further assumption is that the FRET donor and acceptor come together as predicted, which might not be the case. For example, rather than binding with specificity, molecules might become denatured, exposing hydrophobic surfaces, leading to non-specific aggregation.

In addition, the conformational changes occurring during FRET experiments may differ from those expected. For example, differences in FRET could reflect a variation in the relative orientation of the FRET donor and acceptor rather than a change in the distance between the fluorophores. Kon *et al.*[96] describe their use of FRET to monitor conformational change within cytoplasmic dynein, specifically the distance between the tail (labelled with GFP) and the AAA domains of the head (labelled with BFP). They discuss the possibility that some of the FRET change observed might be due to conformational changes within the tail and not a change in the distance between the fluorophores. Furthermore, it is unclear which AAA domain the GFP fluorophore is closest to when the fluorophores are brought together.

In a titration, an assumption is made that the intensity of the fluorescence being measured is unaffected by factors such as the absorbance or direct fluorescence of the titrant, the concentration of which may be very high by the end of the experiment. Adding large quantities of a ligand and its solvent, for example DMSO in the case of a low solubility compound, can change the properties of the solution being observed. The resulting changes in pH or hydrophobicity could affect the fluorescence. For example, the presence of DMSO could cause partial unfolding of a protein, increasing the solvent exposure of an attached fluorophore.

The volume of solution excited by the incident light and from which fluorescence is detected also has a direct effect on the observed fluorescence. In a spectrofluorimeter cuvette, this volume is unlikely to vary. However, in a plate reader well, the volume excited by the incident light depends on the shape of the meniscus, which is affected by factors such as DMSO concentration.

Kon *et al.*[96] used FRET to investigate the effect of excess ATP on the distance between domains of dynein. To check that ATP did not alter the fluorescence of either fluorophore in isolation, they carried out controls in which ATP was added to each polypeptide-fluorophore complex in the absence of the other.

1.3.2.2.2. Fluorescence Anisotropy

Fluorescence anisotropy is another fluorescence-based method which can be used to monitor the binding of a ligand to a protein. Only the ligand is fluorescently labelled. When excited by polarised light, a rapidly tumbling ligand rotates in the time interval between absorption and re-emission, emitting light in many directions. However, if the ligand binds to a large protein, the labelled ligand, as part of the complex, tumbles much more slowly. It is only excited in some orientations and rotates little before re-emission, causing fluorescence anisotropy, polarisation of the emitted light.

The fluorescence anisotropy, r, of a sample can be calculated from the fluorescence emitted parallel, $F_{\perp}$, and perpendicular, $F_{|}$, to the plane of polarisation of the exciting light[81].

$$r = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + 2F_{\perp}} \quad \textbf{1.10}$$

There are two planes perpendicular to the plane of the exciting light and light in only one of these is detected. Consequently, the denominator of this equation is equal to the total amount of light emitted.

Sometimes the fluorescence polarisation, P, is quoted instead.

$$P = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + F_{\perp}} = \frac{3r}{r + 2} \quad \textbf{1.11}$$

Tumbling is not the only factor affecting the anisotropy. Some polarisation is lost if the emission dipole moment is at a different angle to the absorption dipole moment in the fluorophore used. Even where these two moments are parallel, the anisotropy is much less than 1; the fact that dipoles are excited at many different angles coupled with the angular dependence of the probability of fluorophore excitation limits the maximum anisotropy to 0.4[90(p358)].

An advantage of fluorescence anisotropy compared to FRET is the high signal to noise ratio. Unlike FRET, the results are insensitive to changes in the total intensity of the fluorescence. However, fluorescence anisotropy is also a less direct measure of binding because factors other than the displacement of the ligand from its receptor can affect the correlation time.

In this project, a FRET-based assay was developed to test potential inhibitors of the p53-Mdm2 interaction. A fluorescence anisotropy competition assay using p53-fluorescein[97] has previously been tested and attempts have been made to use it with oligobenzamides[42,51,98]; however, there was evidence that, as well binding to Mdm2, these compounds bound weakly to the p53 peptide, affecting its final anisotropy. This meant that the titration curves produced could not be used to calculate dissociation constants.

1.3.2.2.3. Fluorescent Proteins Used in This Project

In this project, the fluorescent proteins GFP, YFP and mCherry were used.

Green fluorescent protein (GFP) is a naturally occurring protein from the jellyfish *Aequorea victoria* which has been mutated to increase its fluorescence[99] and to make homologous proteins of different colours that include blue fluorescent protein (BFP), cyan fluorescent protein (CFP), yellow fluorescent protein (YFP)[100]. The donor and acceptor fluorophores used in a FRET experiment are known as the FRET pair. BFP and GFP are commonly used as a FRET pair, as are CFP and YFP[101].

In GFP, the fluorophore is formed by an auto-catalytic reaction beginning with nucleophilic attack of the nitrogen of Gly67 on the amide carbonyl group of Ser65 (cyclisation). Dehydration and oxidation then occur to form a 5-oxoimidazole moiety which together with the phenolic

ring of Tyr66 forms the large conjugated system responsible for the fluorescence[102,103] (Figure 1.9).

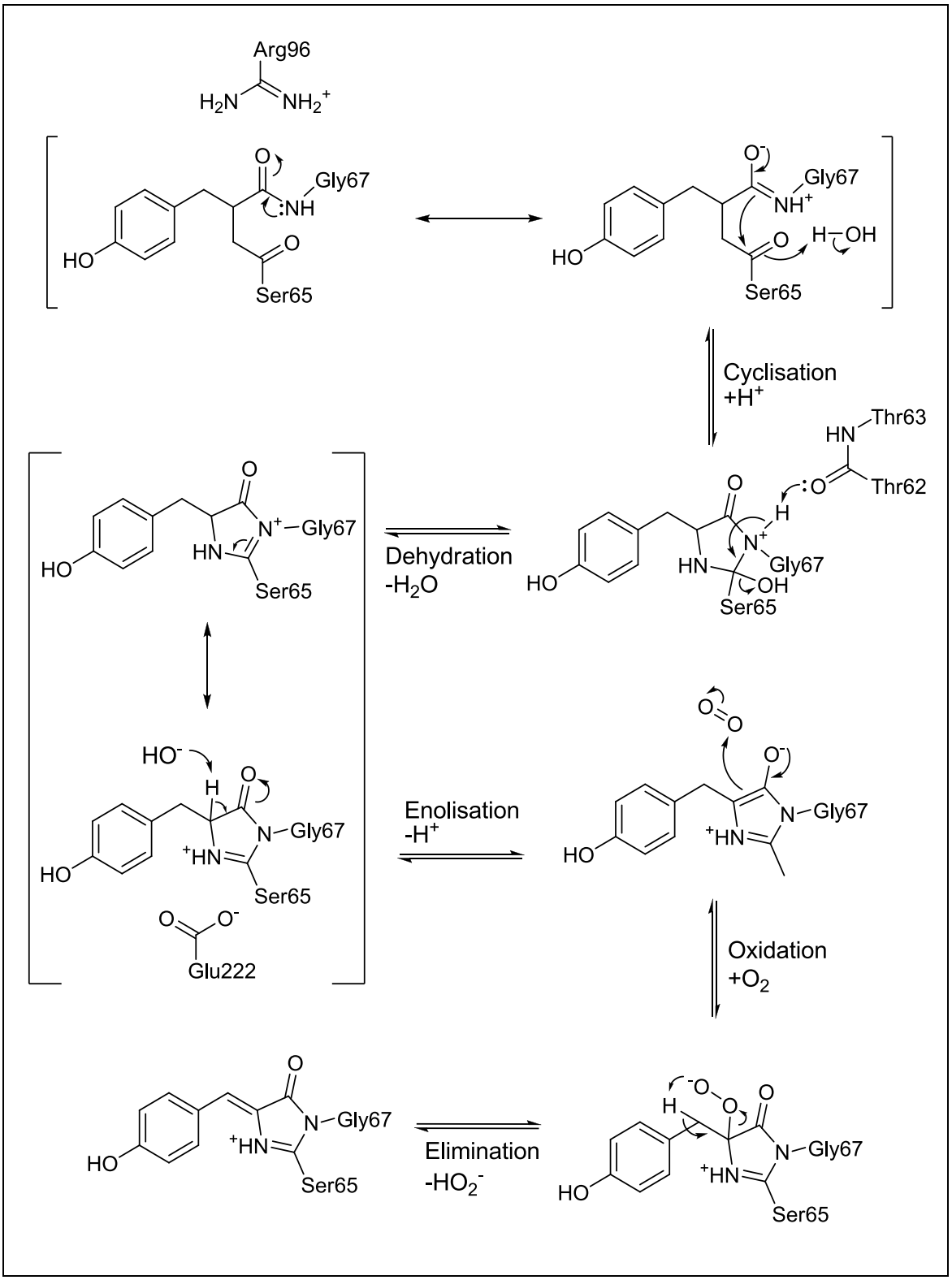


Figure 1.9: A possible mechanism for GFP fluorophore biosynthesis[102,103]. An alternative mechanism has been suggested where oxidation precedes dehydration[104].

mCherry is a mutated version of the DsRed protein, a homologue of GFP found in coral anemones of the genus *Discosoma*,[105]. The fluorophore of DsRed is similar to that of GFP and forms from residues Gln66, Tyr67 and Gly68 by a mechanism which could be analogous to that of GFP maturation[106].

1.3.2.3. Nuclear Magnetic Resonance (NMR)

This section introduces the theory of NMR, a method which enables solution state systems to be investigated at atomic resolution in a non-disruptive manner over a broad range of timescales. The dynamics of interacting proteins and the binding of PPI inhibitors to a protein can be studied by NMR so it can be used to validate predictions based on *in silico* work.

1.3.2.3.1. The principles of NMR

The wavefunction of some particles interacts with an externally applied magnetic field. These particles have a property known as spin and are said to have a magnetic moment. In NMR, a sample is placed in a strong magnetic field causing the nuclei with spin to arrange themselves relative to the field according to what is referred to as their spin state. Typically this means that a nucleus either lines up parallel (β) or antiparallel (α) to the field. The energy of the nuclei in parallel orientations is lower than that of those in antiparallel orientations. This splitting of the spin state energy levels by the magnetic field is known as Zeeman splitting[107(p23)]. The small difference in energy leads to there being slightly more nuclei antiparallel than parallel (roughly 10001 antiparallel for every to 10000 parallel for a spin ½ nucleus in a 600 MHz (14 T) spectrometer[108(p6)]. This confers a small net magnetisation on the sample which opposes the external field.

Excitation of the nuclei is possible using radio frequency electromagnetic pulses, which are applied through coils around the sample. These cause the nuclei to resonate and change state from α to β, changing in orientation (spin state). The net magnetisation of the sample is said to precess from the z-axis (that parallel to the external field) into the perpendicular xy-plane (and beyond, back on to the z-axis given a pulse of sufficient length). The net magnetisation vector of the sample precesses around the external magnetic field vector (z-axis) as a result of the rotation of the magnetic moment of each of the excited nuclei. The component of the net magnetisation in the xy plane is detected. The frequency at which this perpendicular component oscillates as the nuclei precess reflects their resonant frequency. Consequently, excitation of a sample and then detection of the frequency components in the resulting magnetisation as it decays facilitates the production of a spectrum that shows the resonant frequencies of the nuclei in the sample. As the resonant frequency of an atom depends on the difference in energy between its spin states, which in turn depends on the nature and environment of the nucleus, an

NMR spectrum contains information about the molecular structure and state of the species in a sample.

The inherent spin of a particle, a fixed property, is quantised and described by its spin quantum number, $\mathbf{I}$[107(p24)].

$$\boldsymbol{I} \in \{0, 0.5, 1, 1.5, 2, 2.5,...\} \quad \textbf{1.12}$$

The value of I for a particular nucleus depends on the numbers of protons and neutrons in the nucleus so different isotopes of a particular atom have different spin quantum numbers.

Table 1.2 shows the spin of some nuclei commonly used in NMR experiments.

Table 1.2: Nuclei often used for NMR

Nucleus	Spin	γ /10^{-7} rad $s^{-1}T^{-1}$	Natural abundance (%)	500 MHz spectrometer frequency /MHz
^{1}H	½	26.75	100	500.0
^{2}H	1	4.11	0.015	76.7
^{13}C	½	6.73	1.1	125.7
^{15}N	½	-2.71	0.37	50.7
^{19}F	½	25.17	100	470.6
^{31}P	½	10.83	100	202.6

γ, the gyromagnetic ratio indicates the size of the magnetic moment and, in the case of the ratio for nitrogen, indicates the directionality of precession. The rightmost column indicates the resonant frequency of the nucleus in the magnetic field of a 500 MHz NMR spectrometer (11.74 T)[107(p25),109(p3)].

Most nuclei chosen for NMR have a spin, **I**, of ½ because this leads to the production of a single, narrow peak on a simple NMR spectrum. Nuclei have 2**I** + 1 possible spin states (each with a different energy level), which results in 2**I** observed changes of spin state. If I is greater than ½, the increased number of peaks creates a very complicated spectrum with many peaks. Furthermore, nuclei with a spin greater than ½ also have problematically broad peaks because their anisotropic charge distribution interacts with their surrounding electrons to cause spin-state transitions (spin-lattice quadrupolar relaxation[110]).

The rightmost column of Table 1.2 gives the frequency of the radio frequency pulses required to excite each nucleus in a 500 MHz spectrometer (one with a magnetic field of exactly the strength required for hydrogen nuclei to be excited at 500 MHz). In a sample, nuclei of a particular type may be found in various functional groups. For example, hydrogen is involved in both C-H bonds and O-H bonds in ethanol. The resonant frequencies of the nuclei in different environments such as these differ slightly because the frequency is affected by the distribution of electrons near the nuclei and their associated local magnetic fields. These small differences

are usually quoted as a chemical shift, σ. The chemical shift of a nucleus is its resonant frequency relative to that of a standard expressed in parts per million (ppm).

$$\sigma = \left(1 - \frac{\nu_\sigma}{\nu_0}\right) \times 1{,}000{,}000 \qquad \textbf{1.13}$$

Where ν_σ is the resonant frequency of nuclei with a chemical shift of σ where ν_0 is the resonant frequency of the reference nuclei, which, by definition, have a chemical shift of 0. In the case of ^{1}H-NMR, the standard resonant frequency is that of the hydrogen nuclei in tetramethylsilane.

Due to the suseptibility of the chemical shift to changes in the magnetic field acting on a nucleus, it is important that all of the nuclei in a sample experience exactly the same external magnetic field. Solution state NMR is limited to small species, which rotate rapidly (have a low correlation time), because time-averaging of the magnetic field over all orientations of the species is necessary to ensure that each nucleus experiences a uniform, isotropic magnetic field and thus produces strong, easily distinguishable spectrum peaks[111].

In NMR experiments, the excited nuclei are often manipulated so that they transfer their magnetisation to others in the surroundings, so the specific nuclei excited in an experiment often differ from those which are detected at the end and contribute to the spectrum. Complex sequences of pulses targeting different nuclei allow two, three and even four dimensional spectra to be produced. These multidimensional spectra are constructed via the production of a series of one-dimensional spectra, each with a slightly different time delay between certain specific pairs of pulses.

An overview of the experiments used in this project follows.

1.3.2.3.2. The HSQC (heteronuclear single quantum coherence) Experiments

Protein NMR projects often begin with the production of a ^{1}H-^{15}N HSQC spectrum (hereinafter referred to as a ^{15}N-HSQC), which is typically obtained as a two-dimensional spectrum. An HSQC experiment detects pairs of bonded nuclei, in the case of a ^{15}N-HSQC spectrum, the primary and secondary amino groups in a sample. The spin state of a nucleus affects the spin state energies of any other nucleus with which it shares bonding electrons. This through-bond effect is known as J-coupling or scalar coupling[107(p35)].

Production of a ^{15}N-HSQC requires the generation of many 1D ^{1}H-NMR spectra, each modulated by ^{15}N nuclei with a specific resonant frequency selected by the adjustment of a specific delay length in the pulse sequence. Most of the amino acid residues in a protein bestow a single peak on its ^{15}N-HSQC spectrum because they have a single amide group. However, proline residues produce no peak because they have no amide hydrogen and tryptophan, histidine, arginine, lysine, asparagine and glutamine all have amino groups in their side chains which could potentially produce additional peaks. In practice, it is only in glutamine, asparagine

and tryptophan that the protons are not exchanged too quickly with the solution and have resonances in the range typically covered in the production of an HSQC spectrum to locate the backbone amide peaks. The N terminal amine of a polypeptide chain also exchanges its protons with water too fast for their detection.

Figure 1.10 shows the complex sequence of radio frequency pulses typically applied to the sample in an HSQC experiment. The black bars represent the pulses. Appendix B describes the exact effect of the pulses and delays on the net magnetisation of the sample and how this leads to production of the spectrum.

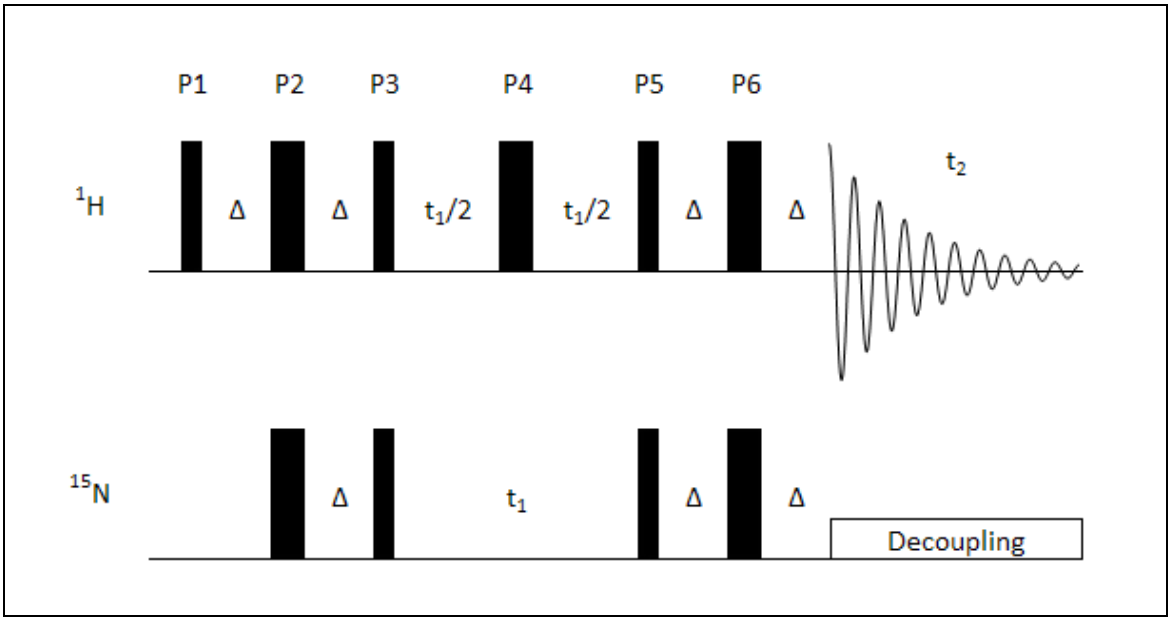


Figure 1.10: An HSQC pulse sequence[107(p242)]. The black bars represent pulses (90° for the thin bands and 180° for the thick bands). Pulses on both the hydrogen and nitrogen channels (pulses at frequencies that will excite hydrogen and nitrogen nuclei respectively) are shown. The white spaces are delays and the expression in their centre indicates the length of the delay. The delays last for milliseconds whereas the pulses are of the order of microseconds. The sequence is repeated many times with different values of t_1 to traverse the indirect dimension. The exponentially decaying wave in the proton channel shows the loss of magnetisation perpendicular to the external magnetic field. This magnetisation, which oscillates due to the precession (rotation) of the nuclei, is measured and Fourier transformed to generate a 1D spectrum along the direct dimension, t_2. A long delay lasting seconds (not shown) is required between successive executions of the sequence to allow full relaxation of the nuclei and equilibration of the NMR machine.

When a ligand binds to a protein, the magnetic field strength around the binding atoms in the protein changes. This causes some peaks on the HSQC spectrum to move, potentially facilitating identification of the binding site. When a ligand is titrated in to a solution of a protein that it binds, the peaks on the HSQC spectrum of the protein can move smoothly from their initial position to their final position or, alternatively, to disappear from their original position and appear at their new position. Which of these two phenomena is observed depends on the rate of exchange between the bound and unbound states[112]. Very fast rates of association and dissociation lead to the atoms producing each peak experiencing a time-averaged magnetic

field and the production of an intermediate peak position. In contrast, slow on and off-rates result in two distinct populations of the protein molecule, bound and unbound. As the concentration of compound is increased and the relative proportions of the two spectra change, one spectrum fades and the other increases in intensity. What is termed intermediate exchange is possible at the boundary between these two extremes, where the first order rate constants of binding and release are comparable to the difference in the atoms' Larmor frequencies (angular frequencies of precession).

1.3.2.3.3. NOESY-HSQC and TOCSY-HSQC Experiments

In this project Mdm2 L33E was used. Assignments for wild type Mdm2 had already been published but there was some variation between these and HSQC results obtained with the L33E mutant, necessitating the collection of additional spectra for assignment of the L33E peaks. Often assignment requires the production of 3D spectra, the product of triple resonance methods such as the HNCA[109(p482)] and HNcoCA[109(p491)] experiments. For these experiments, protein samples containing labelled carbon (^{13}C) are required, which are expensive to produce. In this project, the wild type assignments meant that ^{1}H-^{15}N NOESY-HSQC and ^{1}H-^{15}N TOCSY-HSQC 3D-spectra, which do not require the carbon to be labelled, were sufficient.

In a NOESY (nuclear Overhauser effect spectroscopy) pulse sequence, magnetisation is transferred through space between nearby nuclei (typically protons). The return of the magnetisation vector to the z-axis parallel to the external field following sample excitation is known as relaxation. The magnetisation vectors of nuclei with different chemical shifts precess at different angular velocities and usually relax at different rates. In NOESY, the rapid transfer of magnetisation between nearby hydrogen nuclei results in simultaneous relaxation, known as cross-relaxation, of the nuclei concerned. This results in spectrum cross-peaks.

In a TOCSY (total correlation spectroscopy) 2D spectrum, magnetisation is transferred between nuclei through bonds, as opposed to through space, by a mechanism known as scalar coupling, which is discussed further in Appendix B (p301). The transfer occurs during a delay referred to as the mixing time. Using multiple pulses in the mixing period prevents the precession of nuclear spins (chemical shift refocussing). This allows magnetisation to be transferred multiple times along the residue side chain as opposed to through a single bond as in an HSQC experiment.

Due to the large number of hydrogen atoms in a protein, both ^{1}H-^{1}H NOESY and ^{1}H-^{1}H TOCSY spectra have many peaks, which makes it difficult to distinguished each one.

The NOESY-HSQC and TOCSY-HSQC spectra (where the HSQC sequence is preceded by the NOESY and TOCSY sequence respectively, which acts as a filter) are edited versions of the 2D NOESY and TOCSY spectra, only showing peaks where there is a corresponding peak on the

HSQC spectrum. The NOESY-HSQC and TOCSY-HSQC spectra are 3D because the ^{1}H dimension of the NOESY or TOCSY and the ^{1}H dimension of the HSQC are traversed separately and plotted on two different axes[109(p289)].

A pulse sequence for the NOESY-HSQC experiment is shown in Figure 1.11 and the TOCSY-HSQC pulse sequence is shown in Figure 1.12. The effect of each pulse and delay on the sample is shown in Table 6.3 and Table 6.4 of Appendix B.

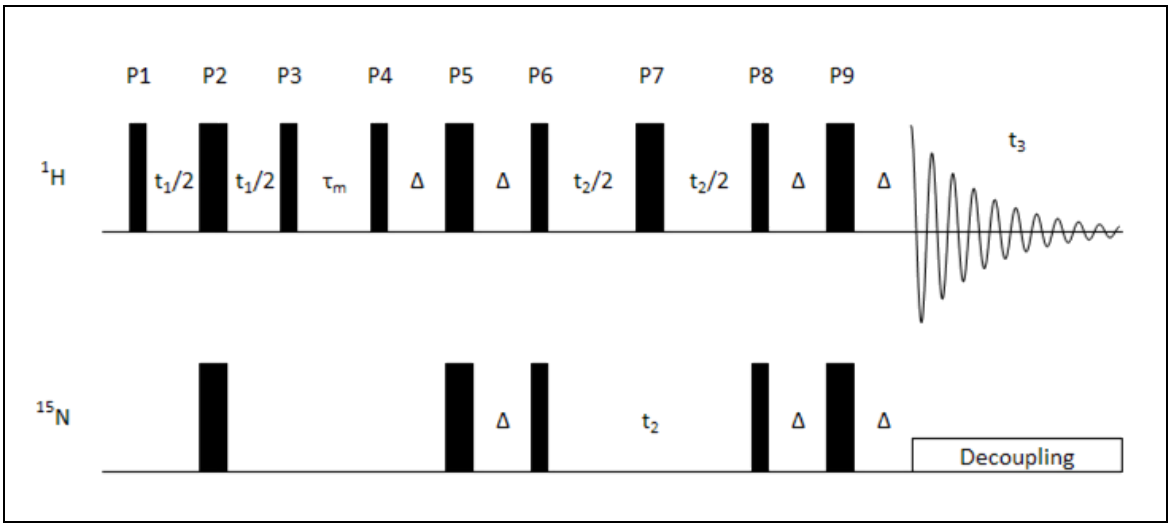


Figure 1.11: A NOESY- HSQC pulse sequence[107(p242)]. τ_m is the mixing time, the period allowed for the transfer of magnetisation between the hydrogen nuclei by the nuclear Overhauser effect. The bars, gaps and other labels have the same meaning as in Figure 1.10 and described in its caption.

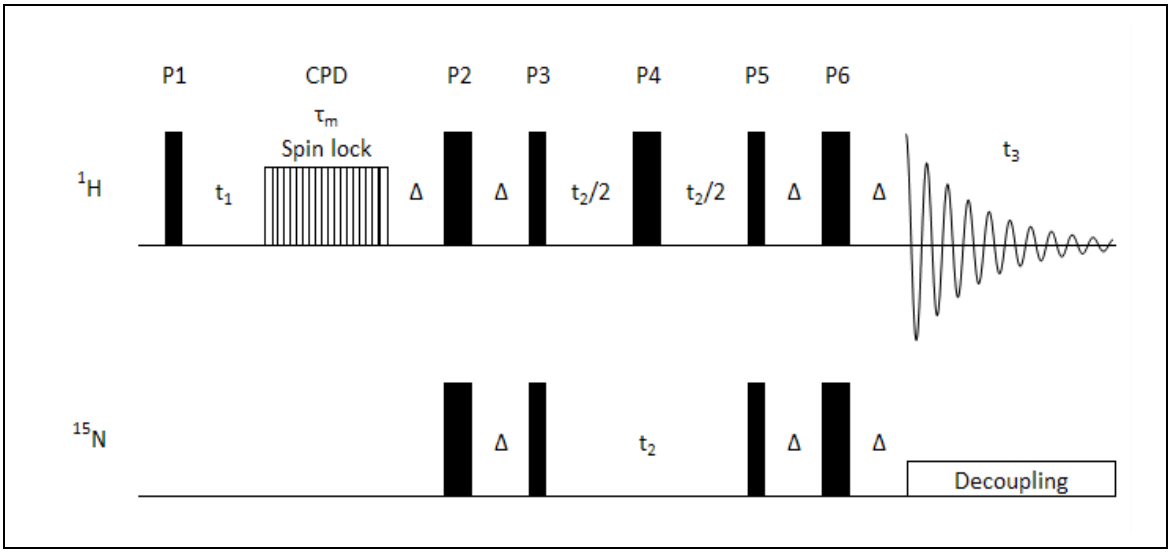


Figure 1.12: A TOCSY-HSQC pulse sequence. "CPD" indicates where composite pulse decoupling was applied. The bars, gaps and other labels have the same meaning as in Figure 1.10 and described in its caption.

Both the NOESY-HSQC and TOCSY-HSQC spectra of a protein comprise a region containing amide and aromatic hydrogen peaks in the $6 < \sigma^H < 10$ ppm part of the spectrum and a region containing peaks produced by the aliphatic hydrogen atom nuclei of the side chains in the $0 < \sigma^H < 6$ ppm part. Details of how these spectra were used to identify the HSQC peaks of Mdm2 in this project are given in the NMR results chapter (p252).

1.3.2.3.4. Expression and purification of proteins for NMR

The study of proteins often requires that they be produced in high concentrations. For example, in competition assays to test potential inhibitors of an interaction, the concentration of at least one of the two interacting molecules must exceed the dissociation constant of the complex for there to be a significant concentration of the complex to be dissociated. In NMR, signals are weak due to the small differences in nuclear energy being observed so concentrated samples are required for NMR too.

Gräslund *et al.*[113] review protein expression and purification. High concentrations are achieved by producing a large amount of the protein, purifying it efficiently and then reducing the volume of the pure solution. Usually, transgenic bacteria are transformed with DNA coding for the protein of interest as part of a plasmid vector and are then grown in a nutrient-rich medium. An antibiotic resistance marker gene in the vector enables the selection of transformed cells by addition of the corresponding antibiotic to the medium.

Cells producing large amounts of a foreign protein typically divide less frequently. Consequently, expression of the protein is usually controlled by the use of an inducible promoter, such as that of the *lac* operon, to allow expression to be delayed until the lag phase of population growth has been completed.

Within the plasmid, the protein of interest is typically expressed with one or more concatenated tags to facilitate purification. Protease sites to facilitate removal of the tags are sometimes included, especially if the tags are large protein molecules. There are several commonly used affinity purification tags[113]. His-tags, oligo-histidine sections, adhere to ions such as nickel or cobalt in a resin column. The protein tags maltose binding protein and glutathione-S-transferase bind to resin columns containing maltose and glutathione respectively.

Proteins can also be purified by methods that separate according to size (for example, size exclusion chromatography), charge (such as ion exchange chromatography) or hydrophobicity (hydrophobic interaction chromatography for instance). Size exclusion chromatography uses inert beads with holes of a particular size. Smaller proteins take a more convoluted route through the column so have a larger elution volume. To prevent the formation of disulphide bridges between protein molecules, a reducing agent such as dithiothreitol (DTT, $SH(CH_2)CH(OH)CH(OH)CH_2SH$) can be added to the buffer. After nickel affinity

chromatography, the protein will precipitate if ethylenediaminetetraacetic acid (EDTA) is not added to chelate any nickel prior to the addition of DTT[113].

Concentration is carried out using a membrane with pores large enough to let water and other molecules with a mass below the molecular weight cut-off (MWCO) pass through but small enough for the protein of interest to be retained. Centrifugation is used to accelerate filtration[113].

Separation of the proteins in samples by sodium dodecyl sulphate (Na^+ $CH_3(CH_2)_{11}OSO_3^-$) polyacrylamide gel electrophoresis (SDS-PAGE) enables the purification process to be monitored. The dodecyl sulphate ion is a detergent which denatures proteins and coats them such that the rate of electrophoretic separation depends on their mass alone.

In this project, the Mdm2 protein used in NMR experiments was ^{15}N-labelled. To produce labelled proteins, bacteria are grown in minimal media in which their only source of a particular element is a compound containing the heavy isotope[114]. Typically, $^{15}NH_4Cl$ is the source of ^{15}N and ^{13}C-glucose provides ^{13}C. If there is little bacterial growth, a solution containing vitamins and minerals such as Basal Medium of Eagle (BME) can be added to augment the minimal medium. A small amount of unlabelled rich medium can also be used to provide a trace amount of any essential compound missing from the minimal media. If growth is still inadequate, expensive, labelled rich media can be used.

Some of the isotopes used for NMR, for instance ^1H, ^{19}F and ^{31}P, are the natural, ubiquitous isotopes of the elements in question; however, others, most notably ^2H, ^{13}C and ^{15}N, are isotopes that are stable but only present in low concentrations in the natural environment. To produce labelled compounds for media containing these isotopes, isotopic enrichment is necessary. Enrichment of deuterium oxide in water can be performed via the Girdler sulphide process, which relies on the kinetic isotope effect[115]. Isotopic enrichment of ^{15}N is possible by equilibrium fractionation using the NITROX method, which exploits the greater stability of ^{14}NO and $H^{15}NO_3$ as opposed to ^{15}NO and $H^{14}NO_3$ at low temperatures[116].

1.3.3. INVESTIGATION OF THE P53-MDM2 INTERACTION USING NMR

There has been extensive investigation of the p53-Mdm2 interaction by NMR. In this project, oligoamide trimers with oxygen-linked side chains were titrated into a solution of ^{15}N-labelled Mdm2 L33E and their effect on the Mdm2 HSQC spectrum was observed. Most NMR work with aromatic oligoamides has concerned the investigation of oligomeric foldamers in isolation[55,117]; however, Lu *et al.*[118] have carried out a similar experiment previously. They titrated an oligobenzamide dimer with carbon-linked side chains into a solution of wild type Mdm2 (p268).

Other p53-Mdm2 PPI inhibitors have been more extensively researched using NMR. Wang *et al.*,[119] describe the use of HSQC experiments to measure the affinity of compounds based on an oxodihydrobenzoimidazolylacetamide scaffold for Mdm2. Isolindolinone inhibitors[120], chalcones[121] and isoquinolin-1-ones[122] have also been studied.

The adaptability of the Mdm2 binding site to different ligands was revealed by Schon *et al.*[123] who studied how the Mdm2 HSQC spectrum changed upon the binding of p53-like peptides. The dynamics of the N-terminal part of Mdm2, which is thought to occlude the p53 binding site when no ligand is present, has also been investigated[124,125]. The binding of p53-derived peptides to wild type Mdm2 is known to cause the movement of HSQC peaks created by residues some distance from the p53 binding site[126], highlighting the potential benefit of using the stable L33E mutant tested in this project in future binding studies.

The binding of Mdm2-derived peptides to p53 has also been investigated using NMR[127]. The DNA binding domain of p53 (p29) was seen to bind weakly to the acidic domain of Mdm2, a finding which is consistent with the opposing charges of these two regions.

Attempts to obtain a 1D proton NMR spectrum of oligobenzamides in this project were informative in that they suggested the aggregation of some oligobenzamides despite their being soluble. Czarna *et al.*[128] describe the use of HSQC peak movement to test multiple fragment-like inhibitors of the p53-Mdm2 interaction as part of a combined high-throughput computational and laboratory-based screening process and they highlight how NMR results can reveal essential properties of a compound such as its potensity to precipitate.

1.4. Computational Methods Used in This Project

This section introduces docking, molecular dynamics simulations and methods of estimating the free energy of binding. While the protein-protein interaction context of this project is novel, the assessment of potential inhibitors can be likened to any investigation of a set of small molecules binding to a protein of interest. Guvench and MacKerell[129] discuss the trade-off encountered by any researcher wishing to use computational methods to predict the relative binding affinity of a set of compounds, that of speed versus accuracy. They highlight the two ends of the spectrum: at one end, fast docking and scoring where the protein is considered static and the solvent is treated as a continuum: at the other end, accurate calculation of free energies based on the comparison of all atom molecule dynamics simulations generated in steps along a bound to unbound reaction coordinate, as exemplified by the method of thermodynamic integration. In this project, both docking and thermodynamic integration were carried out. Guvench and MacKerell[129] observe that some methods such as the implicit solvent MM-PBSA (molecular mechanics Poisson-Boltzmann surface area) method provide a useful compromise between speed and accuracy and, as discussed in the results, MM-PBSA proved to be most useful in this project.

1.4.1. SCREENING AND DOCKING

1.4.1.1. High-Throughput Screening

In virtual screening, a very large number of compounds are investigated with the aim of generating a rough ranking, from which can be drawn a smaller, refined set of compounds for testing by more accurate methods[130]. There are two options: ligand-based screening and structure-based screening. Shangary and Wang[44] discuss the roles that cheaper, computational methods, such as *in silico* screening and *de novo* compound design, have played in the development of p53-Mdm2 inhibitors.

Ligand-based screening seeks to predict new inhibitors through comparison with existing ligands or, in the case of enzymes, the substrate and or predicted transition state which must be stabilised. One piece of software facilitating such screening is ROCS (OpenEye)[131]. Ligand-based design is relevant to the current project in that oligobenzamides have been chosen for use as scaffolds based on their ability to place side chains in a similar relative location to those of residues in the p53 helix. However, ligand-based screening is of little relevance to the testing of potential inhibitors in this project. It is biased towards ligands which are similar to existing molecules, preventing their fair comparison. Furthermore, this bias means that ligand-based screening is not a source of interesting new chemistry.

In contrast, structure-based drug design makes direct use of the target protein's structure rather than relying on the properties of the target protein being encapsulated in those of the known binding molecules. The success of structure-based methods depends on the availability of accurate structural information, which can be difficult to obtain. Many proteins, particularly membrane proteins, are not amenable to crystallisation or NMR and a suitable structure is not available. Fortunately, there are many X-ray crystal and NMR structures of Mdm2, the focus of this project.

Structure-based screening typically involves docking, the computational fitting of compounds into the surface of a protein to predict where they will bind most strongly. In virtual screening, the docking must be rapid so the structures of the protein and ligands are often static. In such cases, a number of different conformations of the ligand are generated and docked to account for the ligand flexibility[132]. If the ligand conformation is changed during docking, this movement may be limited to a small number of degrees of freedom. For example, in ReCore (BioSolveIT)[133], the problem of finding a side chain that fits is simplified to a search for groups with the correct vectors, the vectors being the direction of the leaving bond (the bond connecting the fragment to a larger molecule) and the directionality of a pharmacophore feature such as a hydrogen bond donor. Possible side chains are rotated relative to the core structure but there is no rotation about any of the bonds within the core or the side chain. Possible side chains are ranked based on the relative geometry of the query vectors and the corresponding vectors in the side chain[133]. According to the ReCore manual, the closer the vectors and the more similar their directionality, the greater the score.

In high-throughput screening, compounds pass or fail based on whether they satisfy certain criteria, for example, based on the existence of hydrogen bond acceptors or aromatic rings in specific positions. When searching for drug-like compounds these criteria are pharmacophore constraints. Galatin and Abraham[134] used quantitative structure-activity relationships (QSAR) derived using a structure of Mdm2 bound to the p53 transactivation domain peptide to predict successfully the effect on binding affinity of small changes to the peptide sequence.

Lu *et al.*[118] used the program CAVEAT[135] to filter compounds to ensure that they all had side chains which could mimic those of Phe19, Trp23 and Leu26 in the human p53 transactivation domain, three residues identified as key "hot spots" by Böttger *et al.*[26] using phage display.

Criteria for screening can be generated using the binding positions of small, fragment-like molecules. The docking of small fragments is unreliable[136] but fragment binding sites can be determined by X-ray crystallography or NMR[137]. Fragment positions can be used to generate possible binding molecules as well as screen a database for them; positioned fragments can be grown[138] or joined[139]. The smaller chemical space of fragments as compared to that of larger

compounds has made predicting compounds by ligand-based drug design an increasingly popular alternative to high-throughput synthesis and screening[140].

1.4.1.2. Docking

In more rigorous docking, the ligand is considered flexible during the docking process; however, this comes at the expense of using more computational power. The protein is still often kept rigid in the docking process but some software can also move flexible parts of the protein. Autodock, one of the programs used in this project, can do this; although, this feature was not used in this project. If the protein has a flexible binding site but software does not permit movement of protein atoms then it is possible to dock into an ensemble of X-ray structures[141] or conformations generated by computational means. Computational methods of generating an ensemble include taking snapshots from a molecular dynamics simulation[142], using normal mode analysis[143] and docking a few compounds with a method that does allows for protein movement[144].

There are disadvantages to using multiple structures. As well as slowing down the docking process overall, using different conformations can introduce error if the conformations used do not accurately represent the possible conformations of the protein[145]. This can occur even if the ensemble members are different crystal structures[141].

When docking a molecule, a scoring method is needed to predict the relative stability of binding poses both during the docking process and to rank the resulting compound poses. Huang *et al.*[146] describe four types of scoring function: force-field based functions, empirical functions, knowledge-based methods and consensus methods which involve a combination of these.

Examples of force-field-based scoring functions are DOCK and AutoDock[146]. Force fields are discussed in detail in the context of molecular dynamics simulations on p67.

Empirical scoring functions are simpler, typically using an equation like the following.

$$\Delta G = \sum_i W_i V_i \qquad \textbf{1.14}$$

The score, ΔG, is the weighted sum of a number of chosen variables, V. These may include the numbers of electrostatic interactions, hydrogen bonds and van der Waals interactions, the number of rotatable bonds and the hydrophobic surface area at the ligand-protein interface.

The weights, W, are empirically derived using the structures of known complexes[146].

Examples of empirical scoring functions are ChemScore[147], X-Score[148] and FlexX[149].

Knowledge-based scoring functions are methods which use the experimentally determined structures of protein-ligand complexes, specifically, the frequency with which certain pairwise interactions are found in those structures.

In general,

$$\Delta G = -kT \ln\left(\prod_i \frac{\rho_i(r_i)}{\rho_i^*(r_i)}\right) \quad \textbf{1.15}$$

Where ρ_i is the probability density of finding a distance of r_i between the interacting pair of atoms, i, at the binding surface. ρ^*_i is the probability density which would be observed in a theoretical reference state in which there is no interaction[146]. k is the Boltzmann constant and T is the temperature in kelvins. Potentials of mean force are calculated using a formula of this type[150].

In this project, three docking programs were used: FlexX (BioSolveIT)[149], Autodock 4 (version 4.2)[151] and Autodock Vina[152]. In FlexX, docking is a sequential process comprising the placement of a core fragment and then incremental addition of atoms to this core until the fully docked compound has been produced. Each step is repeated hundreds of times and the highest-scoring poses (evaluated using the FlexX statistical potential scoring function[153]) are selected for the next stage[153]. Other popular docking programs utilising this incremental approach are DOCK, Glide and Hammerhead[154].

In contrast to the docking methods above, Autodock[151] uses a genetic algorithm, meaning that the state of the ligand is treated like genetic material. In the case of Autodock, the information comprises the location (three coordinates), orientation (four rotation quaternions) and conformation (one variable for each torsion angle) of the molecule[155]. A set of random poses are generated, the best poses are selected and then a new set of poses are generated based on the selected poses by combining and then mutating their information[155]. The process of pose selection, genetic recombination and pose production is repeated thousands of times to produce a population of high-scoring poses. In the event of there being many very similar poses, they are clustered by RMSD (p78) to produce distinct poses. The docking program GOLD also uses a genetic algorithm. Autodock uses a Lamarckian genetic algorithm, meaning that some fitting of each pose is performed before its genetic material (updated to take into account the effects of this local fitting) is processed to generate the next pose set[156].

In terms of the computational time required for docking, Autodock Vina is much more efficient than Autodock 4. It uses a third search algorithm for global search based on the Iterated Local Search method devised by Baxter[157]. It is a Markov chain Monte Carlo iterative method, meaning that each iteration consists of the production of a possible pose based on the current one and the acceptance or rejection of this pose based on its probability relative to the current pose. In the case of Autodock Vina, the production of the possible pose involves a mutation followed by application of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method[152]. This method uses both the value and gradient of the scoring function in the optimisation process[158(p194)].

At the beginning of this project, rapid docking and scoring were used to predict the relative binding affinity of oligobenzamides for Mdm2; however, the synthetic inaccessibility of the compounds meant that validation was not possible. Schneider[159] highlights the relatively few examples of drug discovery projects using virtual screening, citing the poor accuracy of these methods but also the need to identify synthetically accessible compounds, with docking ideally forming part of a cycle of computational and synthetic work.

Docking predicts binding poses successfully[160], but docking scores are typically poor estimates of the true binding affinity of compounds[161,162]. Oligobenzamides are large molecules and therefore have the potential to make numerous interactions when they bind. The more interactions a set of compounds can make with a binding site, the harder it is to determine their relative affinities correctly[163]. Docking aims to find the most stable pose of a compound and docking scores are based on this pose. The true affinity of a compound, in contrast, depends on all of the system's possible states, and the relative frequencies at which those states (each comprising a set of atom positions and momenta) occur as the compound and protein move around. Therefore, for accurate binding energy predictions, this phase space must be fully sampled, for example, by carrying out molecular dynamics simulations[161,162].

1.4.2. MOLECULAR DYNAMICS SIMULATIONS

A molecular dynamics simulation is a computer simulation of molecules within a virtual box. In this project, simulations of ligands bound to the protein Mdm2 were performed. Simulations require much more computation time than even flexible docking.

An X-ray crystal or NMR structure is usually chosen as the initial structure of a protein at the start of a simulation. The initial ligand atom positions may be from the crystal structure or be the result of docking. The water atoms usually come from a box of pure solvent which has previously been simulated to the point of equilibration. Ions are also added as necessary so the system has an overall neutral charge, as required by the Particle Mesh Ewald method[164] (p76).

Initially, each atom in the system is assigned a random velocity based on the system temperature. Repeatedly, the position the atoms will be in after a small time step has elapsed is calculated based on the forces acting on them. The resulting set of frames is known as a trajectory.

1.4.2.1. Necessary Assumptions

Before molecular dynamics simulations are discussed in detail, it is worth considering a couple of the many assumptions made when a molecular dynamics simulation is performed. The validity of the results depends on the validity of these.

1.4.2.1.1. Ergodicity

The probability distribution of the possible states of a system is known as its ensemble.

In a simulation, the system is usually heated up to the desired temperature using a constant volume simulation. Once equilibrated at the correct temperature, more natural, constant pressure simulation begins. (The volume of the system is allowed to vary (p82).) In the constant volume simulation, the NVT (constant particle number, volume and temperature) or canonical ensemble is being sampled. In the constant pressure simulation, the NPT (constant particle number, pressure and temperature) or isothermal–isobaric ensemble is being sampled.

The ensemble average of a property is the value of the property in each possible microstate of the system, weighted by the probability of the system being in that microstate.

$$\langle A \rangle = A_i P_i \quad \textbf{1.16}$$

Where $\langle A \rangle$ is the ensemble average, A is the property, P is the probability and i is the microstate. The probability of a compound being found in each microstate depends on the energy of that microstate and the temperature (p91).

In molecular dynamics simulations, systems are modelled over time and the averages obtained from them are time-averages.; they are only estimates of the true ensemble averages[165(p17)].

In the context of molecular dynamics simulations, the ergodic hypothesis states that, given sufficient time, a simulation will sample all of the possible states of the system and therefore that averages over the frames in a simulation reflect the ensemble average[166(p307)].

In practice, the ergodic assumption is seldom completely fulfilled[165(p17)] because the transitions between some stable states have a low probability due to the high energy of the intervening microstates. As a result, sampling may be incomplete, causing time-averages to be weighted towards the initial state of the system.

1.4.2.1.2. The Born-Oppenheimer Approximation

The mass of an atomic nucleus is thousands of times greater than the mass of each of its electrons. Consequently, when the velocity of the nucleus changes, there is a large change in momentum. This causes the nucleus to exert a large force on its surrounding electrons. Because these electrons have a small mass, the consequent change in their momentum corresponds to a large change in their velocity, so they adjust rapidly to the change in nuclear position. The Born-Oppenheimer approximation states that electrons have no momentum and adjust instantaneously to any change in nuclear position. Thus, when the movement of nuclei is considered, the electrons can be assumed to be stationary relative to the nuclei with which they

are associated. Likewise, the nucleus can be assumed to be stationary relative to electron movement, because the nucleus moves much more slowly[166(p35)].

Further consideration of this approximation depends upon an understanding of the Hamiltonian operator and the Schrödinger equation.

A fundamental particle can be treated like a spherical standing wave with a single peak at the particle's position. The function which describes this wave is the particle's wavefunction. The concept of a wavefunction can be extended to a system where the wavefunction, Ψ, describes the state of the system.

In physics and chemistry, an operator is an operation which, when applied to the state of a system, generates a value. When the Hamiltonian operator, $\mathscr{H}$, is applied to the wavefunction of a system, it returns the rate of change of the wavefunction, phase shifted by 90° (multiplied by i, the square root of -1). The shift puts it in phase with the original wavefunction.

$$\mathcal{H}\,\Psi(\mathbf{r},\mathrm{t}) = \frac{ih}{2\pi}\frac{\partial}{\partial t}\Psi(\mathbf{r},\mathrm{t}) \qquad \textbf{1.17}$$

Where h is Planck's constant, **r** is the state of the system and t is time[166(p27)].

The Planck-Einstein equation, stated in terms of angular frequency, reveals how the rate of change of a wave is also its energy.

$$\mathrm{E} = h\nu = \frac{h}{2\pi}\omega = \frac{h}{2\pi}\frac{d\theta}{dt}\theta(t) \qquad \textbf{1.18}$$

Where E is the energy, ν is the frequency (s^{-1}), ω is the angular frequency (rad s^{-1}) and θ is the angular displacement (rad). Consequently, Equation 1.17, the time-dependent Schrödinger equation can be re-written in a time-independent form[166(p27)]:

$$\mathcal{H}\,\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \qquad \textbf{1.19}$$

The result of applying the Hamiltonian operator to a system is therefore the total energy of the system and this operator is often equated to the summation of the different energies that contribute to the total energy of the system.

Let the nuclear coordinates be expressed by the vector **R** and the state of the electrons embodied in vector **r**. If the Born-Oppenheimer approximation is made, the total wavefunction, Ψ(**R**, **r**), can be modelled as the product of two separate nuclear and electronic wavefunctions (X and Ψ_e respectively), calculable using modified Hamiltonian operators, $\mathscr{H}_e$ and $\mathscr{H}_n$ respectively[166(p35)].

$$\Psi(\mathbf{r},\mathbf{R}) = \mathcal{X}(\mathbf{R})\Psi_e(\mathbf{r},\mathbf{R}) \qquad \textbf{1.20}$$

Schrödinger equations can be formulated for both the electronic (Equation 1.21) and nuclear (Equation 1.22) components[166(p27)].

$$\mathcal{H}_e\,\Psi_e(\mathbf{r}) = E_n(\mathbf{R})\Psi_e(\mathbf{r}) \quad \textbf{1.21}$$

$$\mathcal{H}_n\,\mathcal{X}(\mathbf{R}) = E_\mathcal{X}(\mathbf{R})\mathcal{X}(\mathbf{R}) \quad \textbf{1.22}$$

As shown, the coefficients E_n and E_X vary with the nuclear positions, **R**.

$$\mathcal{H}_n = -\frac{\hbar}{2m}\sum_{i=1}^{N_n}\nabla^2{}_i + \mathcal{V} \quad \textbf{1.23}$$

Where ħ is Planck's constant, h, divided by 2π, μ is the mass of each atom (assumed to be identical in this example), $\mathcal{V}$ is the interatomic potential, otherwise known as the vibronic energy and ∇ is the del-squared operator or Laplacian, which is here applied to the nuclear wavefunction[167(p12)].

In Euclidean (x, y, z) space,

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad \textbf{1.24}$$

$$\mathcal{V} = \mathrm{E}_n - \sum_{i=1}^{N_e}\sum_{a=1}^{N_n}\frac{Z_a e}{r_{ia}}. \quad \textbf{1.25}$$

Where N_e is the total number of electrons (over all atoms), N_n is the number of nuclei, Z_a is the charge of nucleus a, e is the elementary charge (the charge of a proton) and r_{ia} is the distance between nucleus a and electron i. E_n is the electronic energy as shown in Equation 1.21 and which may be calculated using $\mathscr{H}_e$.

$$\mathcal{H}_e = -\frac{\hbar}{2\mu}\sum_{i=1}^{N_e}\nabla^2{}_i(\mathbf{R}) + \sum_{i-1}^{N_e}\sum_{j=1}^{i}\frac{e^2}{r_{ij}} - \sum_{i=1}^{N_e}\sum_{a=1}^{N_n}\frac{Z_a e^2}{r_{ia}} + \sum_{a=1}^{N_n}\sum_{b=1}^{a}\frac{Z_a Z_b e^2}{r_{ab}} \quad \textbf{1.26}$$

Where μ is the mass of an electron (the reduced mass with respect to the nucleus), r_{ij} is the distance between electrons i and j and r_{ab} is the distance between nuclei a and b[167(p11)]. Note that the del-squared operator is applied to the electronic wavefunction in this equation, not the nuclear positions **R**; the value just depends upon **R.**

In a quantum mechanics simulation, $\mathcal{V}$ is estimated using the positions of the nuclei and then equation 1.22 is solved, yielding E_x(**R**) which can be used to work out how the nuclei will move and therefore what $\mathcal{V}$ will be after a small amount of time has elapsed, the time step.

1.4.2.2. <u>The Force Field</u>

Quantum mechanical calculations have limited scalability. A time complexity of $O(N^3)$ where N is the number of orbitals is typical[168] due to the need to perform matrix multiplication. Consequently, it would be too computationally challenging to perform macromolecular

molecular dynamics simulations in which all of the electronic orbitals were explicitly modelled. Quantum mechanical models are generally only used for the simulation of small molecules or a small part of a larger system (in a QM/MM simulation)[169]. Consequently, the vibronic potential energy of the system, $\mathcal{V}$ (discussed above), is usually estimated in a molecular dynamics simulation of a macromolecule using an equation containing parameters for different atoms, bonds, torsion angles and through-space interactions. This equation is known as a force field. A key assumption with most force fields is that different factors make independent contributions to the binding energy. In practice the polarisation of atoms by those in their vicinity and the limited occupancy of molecular orbitals (exchange effects) mean that forces are not entirely additive[167(p141)].

Most force fields are derived from the following standard formula for the AMBER (Assisted Model Building with Energy Refinement) force field[170]:

$$\mathcal{V}(\mathbf{R}) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi \cos[(n\phi - \delta)^2 + 1] + \sum_{nonbonded\ pairs} \left[\frac{q_i q_j}{r_{ij}} + \frac{4\varepsilon\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{4\varepsilon\sigma_{ij}^{6}}{r_{ij}^{6}}\right] \quad \textbf{1.27}$$

Where $\mathcal{V}(\mathbf{R})$ is the total potential energy associated with the bonded and non-bonded interactions between all of the particles in a simulated system where the particles are at positions described by the vector **R**. The bonded interactions are modelled by the first three of the four terms, which relate to bond stretching, bond bending and bond torsion respectively. The non-bonded interactions (electrostatic interactions and van der Waals forces) are modelled by fourth term.

b is the length of a bond being evaluated and b_0 and k_b are constants describing the equilibrium length and stiffness of bonds of that type[170]. Similarly, θ is the angle between two bonds under consideration and k_θ indicates the amount of energy stored when the angle is bent away from its equilibrium angle, θ_0. In the third term, the summed expression describes a component of the torsional energy profile for rotation about a particular bond. ϕ is the torsion angle, k_ϕ relates to the amplitude of the energy barrier, n is the number of peaks (and troughs) in the energy per 360° rotation and δ is the phase (peak position). Multiple sets of parameters for this third term allow a complex torsional energy profile to be modelled as the sum of a series of cosine waves.

The last, non-bonding interaction term accounts for electrostatic interactions and the Leonard Jones potential for each pair of non-bonded atoms (i and j), the latter comprising a highly distance-dependent, repulsive exchange force and an attractive dispersion force[170]. q_i and q_j are

the charges of the atoms. r_{ij} is the distance between them. ε is the permittivity of the environment between the nuclei and σ is the distance at which the Leonard-Jones potential is 0.

Some force fields related to the AMBER force field are the CHARMM (Chemistry at HARvard Macromolecular Mechanics) force fields and the OPLS (Optimized Potentials for Liquid Simulations) force fields but there are many others[170]. Parameters can be derived using experimental data or generated by applying *ab initio* quantum mechanics methods to small molecules[170].

1.4.2.2.1. Torsion Parameters for Oligobenzamides

Figure 1.6 (p36) shows the scaffolds at the core of the oligobenzamides investigated in this project. The blue R groups are the side chains intended to bind to Mdm2 and mimic the residue side chains of the p53 transactivation domain helix. There are five different scaffolds. In the N1 scaffold, the side chains are attached through nitrogen, in the O1 and O2 scaffolds the attachment is through oxygen and in the S1 and S2 scaffolds the attachment is through sulphur.

Vemparala *et al.*[171] discuss how the GAFF (General Amber Force Field) force field has inappropriate torsion parameters for the modelling of oligobenzamides. They calculated and validated torsion parameters for the ortho-substituted oligobenzamide bonds shown in Figure 1.5.

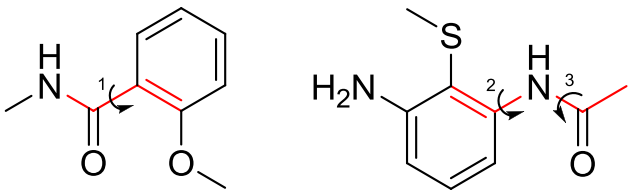


Figure 1.13: Torsion angles for which Vemparala *et al.*[171] determined parameters prior to this project. Three torsion angles within two different ortho-substituted oligobenzamide molecules are shown using arced arrows. The three bonds associated with each torsion angle are highlighted in red.

In my work, only Vemparala's amide bond parameter (torsion 3 in Figure 1.5), which is similar to the GAFF parameter, was used. However, Fuller *et al.*[172] used the aromatic carbon (C_a)-N torsion parameter calculated using the SCH_3 substituent for modelling of the O1 scaffold even though this parameter is for attachment using sulphur instead of oxygen. Fuller argues that the difference in the strength of the O-H and S-H hydrogen bonds is unimportant, a claim supported by the recent work of Liu *et al.*[173] who generated and validated torsional parameters for the C_a-peptide carbon and C_a-N bonds in both ortho and meta substituted oligobenzamides for OCH_3 and SCH_3 substituents and found them to be similar. In this current project, Liu's validated parameters were used for these torsion parameters affected by intramolecular hydrogen bonding.

Liu *et al.* note that even though ortho substituents appear to have similar effects, regardless of whether the attachment heteroatom is oxygen or sulphur (as assumed by Fuller *et al.*), where substitution is at the meta position, the choice of oxygen or sulphur does make a difference. In the meta case, sulphur significantly reduces the strength of the intramolecular hydrogen bond because steric effects are more likely to push the sulphur atom away from the hydrogen with which it is hydrogen bonded. A similar effect was observed earlier by Galan *et al.*[174] who investigated fluoro- and chloro-ortho-substituted N-methylbenzamides. As a result of their electronegativity, hydrogen bonds involving each of these halogen substituents are stronger than those involving oxygen or sulphur. However, whereas the most electronegative substituent, the fluorine, increases rigidity, the chloro group does not appear to, an observation which the authors attribute to steric effects commensurate with its greater size.

In the N1 scaffold, although there is no intramolecular hydrogen bonding, the torsion parameters could differ both for every compound tested and between the subunits within each compound. This is because, in these molecules, the substituents are on the amide nitrogen atom, in the vicinity of the rotating bonds, as opposed to some distance away, attached to the aromatic rings. This project involved both inductively electron donating and inductively electron withdrawing side chains. Consequently, the decision was made to base parameters on the torsional barrier height of the un-substituted oligobenzamide when molecules based on the N1 scaffold were simulated. The results of Liu *et al.*[173] indicated that GAFF parameters would not be appropriate but AMBER torsion parameters had not been calculated for un-substituted oligobenzamides. DFT (density functional theory) calculations had been performed on 4-(4-aminobenzamide)benzoic acid to find the torsional potential energy profiles for the C_a-C_p and C_a-N bonds[175] so the barrier heights from these profiles were used for modelling. To facilitate the fair comparison of results for different scaffolds, these N1 scaffold parameters were also used for C_a-N and C_p-C_a bonds not constrained by hydrogen bonding in the O1, S1, O2 and S2 scaffolds (for the bond on the opposite side of the ring to the substituent).

1.4.2.2.2. Calculation of Partial Charges

Gasteiger charges can be used for docking. These are estimates of the charges based solely on the connectivity of the atoms[176]. However, accurate partial charges are required for molecular dynamics simulations and unlike many molecular parameters, atomic partial charges must generally be calculated for each specific molecule because they depend on the structure of the compound beyond their immediate vicinity.

In this project, HF (Hartree-Fock) RESP partial charges were calculated using RESP ESP Charge Derive[177]. These are derived using a Hartree-Fock molecular electrostatic potential grid calculated for the molecule, a three-dimensional array of point charges approximating the continuous Hartree-Fock approximation of the Schrodinger equation. The Hartree-Fock

approximation is based on the assumption that the wavefunction can be described as the sum of a set of functions, basis functions, which each describe an orbital of the system, and which are collectively known as a basis set.

$$\Psi = \sum_{i=1}^{N} \chi_i = \sum_{i=1}^{N} s_i \varphi_i = \sum_{i=1}^{N} \left(s_i \sum_{\mu=1}^{K} c_{\mu i} \phi_\mu \right) \quad \textbf{1.28}$$

Where Ψ is the molecular wave function, which is the sum of N molecular spin orbital functions, χ, which are each the product of a spin, s, and a molecular, spatial orbital ψ created by the summation of K atomic orbitals, φ[166(p35,p41)]. $c_{\mu i}$ is the molecular orbital coefficient indicating the contribution of atomic orbital μ to molecular orbital i.

The orbitals of atoms can be modelled using basis functions, sums of Gaussian functions, individually known as primitive Gaussian functions[166(p65,p67)]. This makes it easy to combine the orbital wavefunctions.

$$g(x,y,z) = x^a y^b z^c N_{a,b,c} \sum_{i=1}^{a+b+c} d_i e^{-\alpha_i (x^2+y^2+z^2)} \quad \textbf{1.29}$$

Where g is the function of a Gaussian orbital in dimensions x, y and z. a, b and c are coefficients describing the shape and orientation of the orbital. For example, a, b and c must sum to 0 for an s orbital, 1 for p orbitals and 2 for d orbitals. For each primitive function comprising the orbital, α controls the size of the Gaussian and d controls its contribution to the overall function. N is a normalisation factor.

In the 6-31G basis set there are 6 Gaussians used to represent the core orbitals of each atom and then 4 used to represent the valence orbitals, 3 smaller ones (referred to as contracted) and a larger one (referred to as a diffuse orbital)[166(p70)]. In this project, the 6-31G* basis set was used. The additional asterisk here signifies the use of Gaussian functions usually used to model d-orbitals when modelling heavy atoms (which do not have occupied d orbitals) to account for the polarisation of the occupied orbitals by the surrounding atoms[178(p189)].

After a molecule has been minimised, partial charges are fitted to each atom in the molecule, so that the electrostatic potential energy around the molecule is reproduced as accurately as possible.

Specifically, at each of a three-dimensional array of points around the molecule, the difference between the predicted potential derived from the partial charges using Coulomb's law and the potential estimated from the wavefunction is minimised. (The electrostatic potential energy is here defined as the amount of energy required to move a point charge with the charge of a proton from an infinitely distant point to the point in question.)

$$R = \sum_{i=1}^{N_{points}} w_i \left(\phi_i^{\Psi} - \phi_i^{\text{pred}}\right)^2 \quad \textbf{1.30}$$

R in Equation 1.30 is minimised where, for each grid point, i, of which there are N_{grid}, w is a weighting factor, φ^{Ψ} is the charge derived from the wavefunction and φ^{pred} is the potential derived from the partial charges being fitted[166, p190].

$$\phi_i^{\Psi}(\mathbf{r}) = \sum_{j=1}^{N_{nuclei}} \frac{Z_j}{|\mathbf{r} - \mathbf{r}_j|} - \int \frac{\rho(\mathbf{r})}{|\mathbf{r}_\rho - \mathbf{r}|} d\mathbf{r}_\rho \quad \textbf{1.31}$$

The potential at position **r** is the sum of the potential due to the nuclear charge (the first term in Equation 1.31) and the potential due to the electron density (the second term), which is subtracted because the electrons are negatively charged[166(p189)]. (Note that **r** is a Cartesian space position vector here as opposed to the positions of the electrons as in section 1.4.2.1.2.)

$$\phi_i^{\text{pred}}(\mathbf{r}) = \sum_{j=1}^{N_{atoms}} \frac{q_{\mathbf{r}}}{4\pi\epsilon_0 r_{ij}} \quad \textbf{1.32}$$

Where N_{atoms} is the number of atoms, q_j is the charge of atom j and r_{ij} is the distance between grid point i and atom j[166(p190)]. (ε_0 is the permittivity of free space, 8.854×10^{-12} F m^{-1}.) Charges are fitted such that the sum of the partial charges is equal to the total charge of the molecule.

The grid points used for fitting are outside the molecule so buried atoms are a long distance from any of the points, having little bearing on R in Equation 1.30. Consequently, these buried atoms can be assigned extreme charges during fitting if the size of these charges is not restrained. Charges produced using restraints are usually referred to as restrained electrostatic potential (RESP) charges[177,179].

The software used in this project to calculate partial charges (RESP ESP Charge Derive[177]) uses a hyperbolic penalty to penalise extreme values when R is calculated during the fitting process.

1.4.2.2.3. Minimisation Using the AM1 Method

Prior to the calculation of partial charges, oligobenzamides were minimised using the AM1 method. This means that the atoms in the molecule, with their orbitals, were moved and rotated so as to minimise the total energy of the molecule.

To determine the energy of a molecule, the Schrödinger equation or an approximation of it must be solved (p66). If the Hartree-Fock approximation is made, the total energy can be determined by solving the equation for each orbital separately.

$$\mathcal{F}\chi_i = E_i\chi_i \quad \textbf{1.33}$$

Where E_i is the energy of molecular spin orbital χ_i.

$\mathcal{F}$ indicates application of the Fock operator, which is similar to the Hamiltonian operator except that the energy is not the true energy, but a value which takes into account the Coulomb and exchange interactions between electrons such that the sum of the resulting eigenvalues (energies) of the molecular spin orbitals accurately reflects the energy of the whole molecule[166(p57)].

The Fock operator can be applied to the known molecular orbital wavefunctions of a molecule to calculate an energy contribution for each pair of electrons. These values can be placed in a matrix, the Fock matrix, **F**.

$$\mathbf{F} = 2\mathbf{H}\mathbf{1}_{\mathbf{1,N}} + 2\mathbf{J} - \mathbf{K} \qquad \textbf{1.34}$$

Where N orbitals in a molecule, $\mathbf{1}_{\mathbf{1,N}}$ is the row vector of ones of length N, **H** is a column vector containing the energy obtained by application of the Hamiltonian operator to each core orbital, **J** is an N x N matrix containing the energy of the Coulomb interaction between each orbital and **K** is a matrix similar to **J** which accounts for the exchange interaction[166(p61)].

In the AM1 method, three empirical functions are used to calculate the Fock matrix elements, one for the diagonal elements and two for the off-diagonal elements, one for orbitals on the same atom and one for orbitals on different atoms.

The Roothan-Hall matrix equations allow the energy of the molecule to be found from the Fock matrix. These can be expressed as a matrix equation:

$$\mathbf{FC} = \mathbf{SCE} \qquad \textbf{1.35}$$

Where **F** is the Fock matrix, **C** is the matrix of molecular orbital coefficients (containing the $c_{\mu i}$ values in Equation 1.28 (p71)), **S** is an overlap matrix to account for the deviation from orthonormality of the molecular orbitals and **E** a diagonal matrix of the orbital energies[166(p59)]. **F** and **S** are known: **C** and **E** are found by solving the equation[166(p61)].

In the AM1 method, as in many termed semi-empirical methods, the orbitals are assumed to be non-overlapping (orthonormal) making **S** of Equation 1.35 an identity matrix. Consequently[166(p60)],

$$\mathbf{FC} = \mathbf{CE}. \qquad \textbf{1.36}$$

This greatly simplifies calculation of **E** from the Fock matrix **F**.

The diagonal elements of **E**, $\boldsymbol{\varepsilon}$, include the Coulomb and exchange interaction energies twice, once for each electron involved in each interaction so when the total potential energy of the electrons (E_{elec}) is calculated from matrix **E**, the Coulomb and exchange interactions must be subtracted, as shown in the equation below[166(p62)].

$$\mathrm{E}_{Elec} = \sum_{i=1}^{N}\left[\varepsilon_i - \sum_{j=1}^{N}\left(2J_{ij} - K_{ij}\right)\right] \quad \textbf{1.37}$$

The total energy is the sum of the electronic energy (E_{Elec}, calculated using matrix **E**) and the nuclear repulsion, E_{AB}, for each pair of atoms (A and B)[166(p85)].

$$E_{Total} = E_{elec} + \sum_{A=1}^{M}\sum_{B=A+1}^{M} E_{AB} \quad \textbf{1.38}$$

In the AM1 method, for most bonds (all except for the OH and NH bonds), the following formula is used.

$$E_{AB} = Z_A Z_B (s_A s_A | s_B s_B)[1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}] + \frac{Z_A Z_B}{R_{AB}}\left[\sum_i K_{A_i} e^{-L_{A_i}\left(R_{AB} - M_{A_i}\right)^2} + \sum_j K_{B_j} e^{-L_{B_j}\left(R_{AB} - M_{B_j}\right)^2}\right] \quad \textbf{1.39}$$

Where the A and B subscripts denote the nuclei either side of a bond. R_{AB} is the distance between the nuclei, Z_A and Z_B are the nuclear charges, K, L and M are fitted constants controlling the shape of the Gaussian functions, i for atom A and j for atom B. α is also a fitted parameter. $(s_A s_A | s_B s_B)$ denotes the overlap integral of the core orbitals of atom A with the core orbitals of atom B, written assuming each atom only has an s orbital in its core so this integral is the integral for the product of the functions for the two electrons in the s orbital of atom A ($s_A s_A$) overlapping the product of the functions of the two electrons in the s orbital of atom B ($s_B s_B$)[166(p97)].

At the limit of R_{AB}=0, the energy E_{AB} is proportional to the inverse of R_{AB}, consistent with Coulomb's law (which states that the force is proportional to the inverse of R_{AB} squared). At the limit of infinite distance, E_{AB} is constant.

1.4.2.2.4. Ewald Summation

When partial charges have been calculated for all of the atoms in a system, the electrostatic interaction between atoms can be quickly estimated during a molecular dynamics simulation using a force field like the one shown in Equation 1.27. However, calculation of the electrostatic interactions in a simulation is still the most time-consuming part of the process.

The amount of time required to calculate the through-space interactions in the manner suggested by Equation 1.27, scales with the number of atoms squared because every atom can potentially interact with every other atom. The size of the dispersion force attracting two atoms is inversely proportional to the 6th power of their distance apart so, for determination of the Leonard-Jones potential, a distance cut-off can be applied to reduce the number of atom pairs which need be evaluated. In contrast, electrostatic interactions are long range so a small cut-off cannot be used without introducing a considerable degree of inaccuracy into the simulation[180].

Simulations are usually performed with what are termed periodic boundary conditions where particles leaving on one side of the simulation appear on the opposite side, as if they had entered a translated copy of the simulation box flush with the original. This maintains the total number of particles and reduces the number of particles which must be simulated for a system in bulk solvent. Atoms must be simulated as if they interact with the atoms in every image of the infinitely tessellated periodic box so there are potentially an infinite number of interactions to sum to calculate the total electrostatic potential energy.

$$\mathcal{V} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sum_{|\mathbf{n}|=0}{}' \frac{q_i q_j}{4\pi\varepsilon_0 |\mathbf{r}_{ij} + \mathbf{n}|} \qquad \textbf{1.40}$$

Where $\mathcal{V}$ here is the potential energy of the electrostatic interactions between N atoms in a periodic box and between those atoms and every atom in the images of the periodic box which would be produced by its infinite tessellation in all directions. q_i and q_j are the charges of atoms i and j respectively. $\mathbf{r}_{ij}$ is the vector from i and j. The third sum is over all **n** except for the $|\mathbf{n}| = 0$ case when i = j (because an atom does not interact with itself).

In Equation 1.40, $\mathcal{V}$ converges slowly as interactions with images of increasing distance from the origin are added. Ewald summation solves the problem of this slow convergence. It involves splitting the series in to four series which each converge quickly.

Equation 1.40 can be rewritten so every interaction is counted twice and the total is halved:

$$\mathcal{V} = \frac{1}{8\pi\varepsilon_0} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{|\mathbf{n}|=0}{}' \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \qquad \textbf{1.41}$$

This can be further rewritten to include any function f(i, j) if this is both added and subtracted so that it does not affect the final result.

$$\mathcal{V} = \frac{1}{8\pi\varepsilon_0} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\left[\sum_{|\mathbf{n}|=0}{}' \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} - f(i,j) \right] + [f(i,j)] \right) \qquad \textbf{1.42}$$

The periodic boundary conditions make it possible to calculate part of the electrostatic energy in reciprocal space using a Fourier-transformed version of the real-space formula.

Let

$$f(i,j) = \sum_{|\mathbf{k}|\neq 0} \frac{4 q_i q_j \pi}{L^3 |\mathbf{k}|^2} \exp\left(-\frac{|\mathbf{k}|^2}{4\alpha^2} \right) \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) + \frac{2\pi}{3L^2} \sum_{\mathbf{k}} q_k |r_k| \qquad \textbf{1.43}$$

Where L is the side length of the period box (assumed to be cubic in this example), α is an arbitrarily chosen constant (often set to 5/L) and **k** is the equivalent of **n** in reciprocal space[166(p337)].

$$\mathbf{K} = \frac{2\pi}{L}\mathbf{n} \qquad 1.44$$

Substitution of f(i, j) into equation 1.42 yields

$$\mathcal{V} = \frac{1}{8\pi\varepsilon_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\left[\sum_{|\mathbf{n}|}{}' \frac{q_i q_j \,\mathrm{erfc}\left(\alpha|\boldsymbol{r}_{ij}+\boldsymbol{n}|\right)}{4\pi\varepsilon_0|\boldsymbol{r}_{ij}+\boldsymbol{n}|} - \frac{\alpha}{\sqrt{\pi}}\sum_{\mathbf{k}} {q_k}^2\right] + \left[\sum_{|\mathbf{k}|\neq 0}\frac{4q_i q_j \pi}{L^3|\mathbf{k}|^2}\exp\left(-\frac{|\mathbf{k}|^2}{4\alpha^2}\right)\cos\left(\mathbf{k}\cdot\mathbf{r}_{ij}\right) + \frac{2\pi}{3L^2}\sum_{\mathbf{k}} q_k|\mathbf{r}_k|\right]\right) \qquad 1.45$$

Where erfc(x) is the error function of x[166(p338)].

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}\,\mathrm{dt} \qquad 1.46$$

Equation 1.45 shows the four rapidly converging series used in Ewald summation, two indicated in each of the sets of square brackets. The first term of the first bracket is a summation over real space. Although it is not clear from Equation 1.45, this first term alone is the sum of 3 components, the electrostatic energies of interest (component A) minus the same energies convoluted by a Gaussian function (component B) plus a third component, the interaction between each i, j interaction and its Gaussian convolution (component C). In order to obtain the energies of interest (component A) it is necessary to remove components B and C. The second term of the first square bracket, a summation in reciprocal space, cancels out component C. The first term of the second square bracket is also a summation in reciprocal space and it almost cancels out component B. It is not quite equal to component B because the situation where $|\mathbf{n}|=|\mathbf{k}|=0$ does not exist in the reciprocal space. This missing component is added by the fourth and final term (square bracket two, term two), completing the cancelation of component B and leaving component A, the electrostatic energy of the interactions.

The Ewald summation method has time complexity $O(N^{3/2})$. A faster variant of the Ewald summation method is the Particle Mesh Ewald (PME) method which has time complexity O(NlogN). In the PME method, charges are assigned to the points of a 3D grid prior to Ewald summation. The grid is treated as an array of particles and the resulting forces on each charge are then mapped back onto the actual particles[165(p312)]. Because the particles are regularly spaced, alternative expressions for the reciprocal space components can be used which allow them to be calculated faster[165(p310)].

1.4.2.3. The Time Step

The time step is the interval in a simulation trajectory between successive force recalculations. The amount of time required to carry out a simulation is proportional to the number of steps in the trajectory so it is desirable to increase the length of the time steps to reduce their total number. However, if the time steps are too long, the movement of molecules, which is continuous in reality, is not reproduced accurately in a simulation.

1.4.2.3.1. The Size of the Time Step

A time step must be chosen which is long enough for the highest frequency oscillation in the system to be modelled. The frequency at which a bond vibrates depends on the stiffness of the bond and the mass of the linked atoms, with smaller atoms resulting in higher frequencies. The fastest motion in a real system is the vibration of the bonds between heavy atoms and hydrogen due to the small mass of the hydrogen atom.

$$\upsilon = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad \text{where} \qquad \mu = \frac{m_1 m_2}{(m_1 + m_2)} \tag{1.47}$$

Where ν is the frequency at which a bond oscillates with atoms of mass m_1 and m_2 at either end if it has a stiffness of k. μ is the reduced mass of the lighter atom with respect to the heavier one.

ν can also be determined experimentally. For example, at 298 K, compounds typically have an infra-red spectrum line at 3000 cm^{-1} which is caused by the C-H bond absorbing the light and stretching[181(p423)]. Multiplying by the speed of light gives the oscillation frequency for this bond as 9×10^{13} Hz, making the period of oscillation 11 fs.

In a simulation, a method known as SHAKE[182] can be used to constrain all bonds to hydrogen. If SHAKE is not used then the time step must be small enough for this fast vibration to be modelled correctly. Using time steps of 1 fs results in 11 steps per C-H bond oscillation, which is usually sufficient for a simulation to be stable. If the size of the time step is increased to, for example, 2 fs, this can lead to extreme bond lengths occurring, the relaxation of which can destabilise the system. When the SHAKE method is used, a constraining force is applied to all bonds to hydrogen prevent these high energy states occurring so a time step of up to 2 fs can be used. SHAKE not only increases simulation stability but it also improves accuracy when comparing simulations because high energy states occur rarely so might be sampled in one simulation but not another.

The potential energy associated with different intra- and inter-molecular interactions can vary at different rates. Slow energy changes can therefore be evaluated at alternate steps, or less frequently still. The force due to these interactions can be assumed to be unchanged in the intervening steps or, alternatively, extrapolated[166(p363)] based on the rate of change of the force (dF/dt) and d^2F/dt^2.

1.4.2.3.2. The Verlet Algorithm

Once the net force on a particle is known, its acceleration can be calculated by dividing the force by the mass of the particle.

$$\mathbf{v}(t+\delta t) \quad = \mathbf{v}(t) + \mathbf{a}\delta t \tag{1.48}$$

Where **v** and **a** are vectors containing, respectively, the velocities and accelerations of atoms (assumed to have constant acceleration) in a system before (t) and after (t + δt) a time step of length δt[166(p355)].

Where the vector **r** describes the position of the atoms,

$$\begin{aligned}\mathbf{r}(t+\delta t) &= \mathbf{r}(t) + \int_t^{t+\delta t} \mathbf{v}(t)\,\mathrm{d}t \\ &= \mathbf{r}(t) + \left[t\mathbf{v}(t) + \frac{1}{2}t^2\mathbf{a}\right]_t^{t+\delta t} \\ &= \mathbf{r}(t) + (t+\delta t)\mathbf{v}(t+\delta t) + \frac{1}{2}(t+\delta t)^2\mathbf{a} - t\mathbf{v}(t) - \frac{1}{2}t^2\mathbf{a}.\end{aligned} \tag{1.49}$$

$$\lim_{\delta t\to 0} \mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a} \tag{1.50}$$

The Verlet algorithm (Equation 1.52) describes how the atomic positions after the next time step can be calculated from the coordinates in the current time step, **r**(t), the coordinates for the previous time step, **r**(t – δt), and the acceleration of the atoms, **a**[166(p356)].

$$\lim_{\delta t\to 0}[\boldsymbol{r}(t+\delta t) + \boldsymbol{r}(t-\delta t)] = 2\mathbf{r}(t) + \delta t^2\mathbf{a} \tag{1.51}$$

$$\lim_{\delta t\to 0}[\boldsymbol{r}(t+\delta t)] = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \delta t^2\mathbf{a} \tag{1.52}$$

1.4.2.3.3. Equilibration and Sampling

The root mean square deviation (RMSD) can be used to quantify the difference between two sets of paired values. If each set is considered to be a point in N-dimensional space, the RMSD is the distance between the two points.

The two sets of values can be the Cartesian coordinates (x, y, z) of the atoms in two structures.

In this case

$$\text{RMSD} = \sqrt{\sum_{atoms}\left[\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2 + \left(z_i - z_j\right)^2\right]}. \tag{1.53}$$

Using the RMSD assumes that the scales of the x, y and z axes are the same. When comparing, for example, the deviations of points on a ^{15}N-HSQC spectrum, the axes need to be scaled for the distance to still be meaningful.

One of the uses of RMSDs is to assess the equilibration of a molecular dynamics simulation, the extent to which the system has converged on a steady state where the final, fixed phase space is

being sampled and the system is thus capable of producing statistically independent as opposed to trending observations.

The RMSD can be used in a variety of ways, the simplest method being to plot the RMSD of each of the frames in a molecular dynamics simulation trajectory relative to the first frame and see when the RMSD plateaus. This was the method used in this project due to its simplicity. However, an all-to-all RMSD analysis can also be performed in which every frame is compared to every other. This reveals the extent to which populations are revisited.

Grossfield and Zuckerman[183] identify some limitations of the RMSD. The RMSD, for example, is a single variable so cannot encapsulate the entire state of the system. It could theoretically level off before the system has equilibrated if the remaining equilibration will not affect the RMSD[183].

If a simulation is performed to monitor a particular variable such as a binding affinity, then this can be monitored instead of the RMSD. While the variable might not be a good indicator of equilibration, levelling off of a particular observation suggests that any remaining equilibration will not affect its value. In the thermodynamic integration (p83) investigations performed in this project, equilibration of the derivative of the vibronic potential energy with respect to the reaction coordinate, $\mathrm{d}\mathcal{V}/\mathrm{d}\lambda$, was checked for.

Principal component analysis can be used to identify variables which better encapsulate the major dynamic modes of the system. Like the RMSD, these can be plotted against time, individually or together to see when the values of the components settle on particular set of values or appear to be moving between a limited number of sets of values, each reflecting a local minimum in the energy landscape[183]. Rather than attempt to map the phase space of a simulation on to scalar variables and then identify such low energy states, clusters of conformations can be identified by cluster analysis of the trajectory frames. A histogram can then be produced which reflects the number of trajectory frames mapping to each cluster identified. Histograms produced at intervals along a simulation trajectory can be tested to determine when they become samples of the same population, indicating that equilibration has occurred[184].

Once equilibration has occurred, it is useful to identify the length of time needed between samples for them to be independent, the correlation time. In order to calculate the error on estimated parameters, it is useful to make independent observations.

The central limit theorem describes how the variance of the means of samples from a population depends on the sample size.

$$Var(\bar{X}) = \frac{Var(X)}{n} \quad \textbf{1.54}$$

Where n is the number of values per sample and $\bar{X}$ is the sample mean of the random variable X.

If contiguous sample windows containing a certain number of frames are taken from different parts of a trajectory, each used to calculate a mean value and the variance of these means is compared to that of the global population, the number of statistically independent samples per window can be found. Alternatively, windows regularly spaced along a trajectory can be analysed and the interval between the windows increased until the variance of the windows and variance over the entire trajectory are consistent with independent sampling[183].

The most straightforward and rigorous method of producing independent observations of a variable is to carry out multiple simulations and see when the results converge. All confidence limits on simulation derived variables in this thesis are based on separate repeats. This method allows random error in docking due to insufficient sampling to be considered by starting each simulation from a separate docking run. Implicit solvent results in this project were started from the results of separate docking runs.

1.4.2.4. Maintaining the Simulated Temperature and Pressure

1.4.2.4.1. Constant Temperature

In the real world, a system exchanges energy with its surroundings. Consequently, the temperature of a real system typically equilibrates to the temperature of its surroundings. To reproduce this effect in simulations, a method is required to maintain the system temperature.

The temperature of a simulation is a function of the translational kinetic energy of the particles[165(p140)].

$$T = k \sum_i \frac{\langle |\mathbf{v}_i|^2 \rangle}{m_i} \quad \textbf{1.55}$$

Where T is the temperature, k is the Boltzmann constant. **v** is the velocity of particle i and m_i is its mass.

In the Berendsen thermostatic method, the velocities are scaled at intervals[166(p383)].

$$\mathbf{v}_{adj} = \mathbf{v}_{init} \sqrt{1 + \frac{\delta t}{\tau} \left(\frac{T_{bath}}{T} + 1 \right)} \quad \textbf{1.56}$$

Where $\mathbf{v}_{init}$ and $\mathbf{v}_{adj}$ are velocities before and after scaling in a system at a current temperature of T coupled to a heat bath of temperature T_{bath} where the time step is δt and τ is what is termed the coupling parameter. The Berendsen method ensures that the rate of change of the temperature is directly proportional to the difference between the temperature of the water bath and the

temperature of the system. The rate is also proportional to the time step and the coupling parameter.

The Berendsen thermostat is useful for controlling the temperature in a potentially unstable system as it is equilibrates. In this project, simulations were carried out with a large number of compounds and it was desirable for as many simulations as possible to be stable both to simplify production and statistical analysis. Furthermore, in the stable simulations, rapid equilibration was essential because it was not possible to manually inspect every trajectory to ensure that it had equilibrated; only a selection of trajectories could be examined. Consequently, the Berendsen thermostat was used for all simulations in this project with the exception of thermodynamic integration runs.

A disadvantage of the Berendsen thermostat is that it does not ensure the correct distribution of potential or total energy throughout the system; it merely ensures that the total energy is correct[185]. All of the simulations carried out in this project had explicit solvent molecules which will have assisted with the partitioning of energy between the parts of the system (thermalisation). However, in thermodynamic integration runs where the reaction coordinate λ is close to 0 or 1, the weak coupling of some atoms to the simulation can mean that there is insufficient exchange of energy with other atoms for thermalisation to occur[186(p118)].

Many thermostatic methods lead to correct sampling, including the Andersen thermostat[187], the Nosé-Hoover thermostat[188] and Langevin dynamics[189]. In the Nosé-Hoover thermostat method[166(p384),188], the system temperature is maintained by extension of the system to contain an additional component which has momentum and so can act as a store of kinetic energy. This component, the position of which is described by the single coordinate s, disconnects the kinetic energy of the un-extended (real) system from the total kinetic energy of the system, allowing it to be changed such that the real system has the desired temperature. The effect of the additional component in its one-dimensional extension of the system is to apply a varying but deterministic mass-dependent force to the particles similar to a friction force, which controls the momentum and speed of the particles in the real system and therefore the temperature of the real part of the extended system.

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial U}{\partial \mathbf{r}_i} - \frac{d\ln s}{dt}\mathbf{p}_i \qquad \textbf{1.57}$$

Where **p** is the momentum of a particle with position **r** in a system where the real part has a total internal energy of U and s is the position of the additional component at time t[165(p152)].

In this project, Langevin dynamics was used to maintain the temperature of thermodynamic integration runs. In this method a constant friction force is applied and there is a random perturbation.

$$\frac{d\mathbf{p_i}}{dt} = -\frac{\partial U}{\partial \mathbf{r_i}} - \gamma \mathbf{p}_i + X_i \quad \textbf{1.58}$$

Where γ is a coefficient of friction and X_i is a random variable distributed normally with mean 0 and variance equal to twice the product of the friction coefficient, the mass of the atom, m_i, the Boltzmann constant, k, and the temperature, T.

$$X_i \sim N(0, 2\gamma m_i kT) \quad \textbf{1.59}$$

1.4.2.4.2. Constant Pressure

The pressure in a system is described by the virial equation:

$$P = \frac{RT}{V} + \frac{1}{3V}\left\langle \sum_{\boldsymbol{i}} (\mathbf{r}_i \cdot \mathbf{F}_i) \right\rangle \quad \textbf{1.60}$$

Where P is the pressure, R is the universal gas constant, T is the temperature in kelvins, V is the volume and the particles, i, have position $\mathbf{r}_i$ and a force on them of $\mathbf{F}_i$[190]. The angle brackets indicate the pressure is the Boltzmann average over all possible microstates of the system.

For a single simulation trajectory frame:

$$\mathrm{P} = \frac{RT}{V} + \frac{1}{3V}\sum_{\boldsymbol{i}} (\mathbf{r}_i \cdot \mathbf{F}_i) \quad \textbf{1.61}$$

The part of molecular dynamics simulations in which data is collected is typically performed at constant pressure because, in the real world, pressure is exerted on systems by their surroundings causing them to equilibrate at atmospheric pressure. In a simulation, the volume of the periodic box occupied by the particles is changed to maintain the pressure.

Berendsen[166(p386),191] describes a simple system of pressure control involving a pressure bath where coordinates are adjusted to effect the necessary change in volume when the pressure changes and the AMBER software used in this project uses such a method[186(p25)].

$$\mathbf{r}_{adj} = \left(1 - \frac{\kappa \delta t}{\tau_p}(P - P_{bath})\right)^{1/3} \mathbf{r}_{init} \quad \textbf{1.62}$$

The rate of change of the volume is proportional to the difference between the pressure of the system, P, and the "pressure bath" pressure, P_{bath}. δt is the time step, κ is the isothermal compressibility of the system and τ_p is a constant describing the degree of coupling between the system and the bath. The rate of change of the volume is proportional to the difference between the current system pressure and the target pressure.

As with the Berendsen thermostat, use of the Berendsen barostat does not ensure correct sampling of the isothermal–isobaric ensemble in NPT simulations and other, better methods of pressure control exist. For example, as with the Nose-Hoover thermostat, the system can be extended to contain an additional moving component. The moving component occupies a

volume (due to its velocity) and consequently it occupies part of the extended system, pushing the real part of the system into the correct volume for the desired pressure[166(p387)].

A system is usually equilibrated under NVT conditions before a constant pressure run is started to ensure that the initial scaling when the barostat is enabled is not very large, destabilising the system[186].

1.4.3. THEORETICAL ESTIMATION OF BINDING AFFINITIES

In this project, binding affinities were initially estimated using thermodynamic integration and then later using the implicit solvent MM-PBSA and MM-GBSA methods. Gilson and Zhou[192] and Foloppe and Hubbard[193] review some methods by which the relative binding energy of ligands can be estimated including these methods.

Comparing simulations of two systems, each with a large phase space is inaccurate as their phases spaces might have been sampled differently in each simulation. High energy states are particularly problematic because they occur rarely but can make a significant contribution to the potential energy of a system[166(p314),183]. Furthermore, the high potential energy states of a protein-ligand complex include states in which the ligand is far from the protein surface. In such states it is debatable whether the ligand is still bound and therefore whether the state should be considered when calculating the energy of the protein-ligand complex[192]. Thermodynamic integration and implicit solvent techniques are methods that enable a fairer comparison of simulations involving different ligands.

In the MM-GBSA and MM-PBSA implicit solvent methods, the solvent molecules are not modelled explicitly (individually). Instead, the solvent is treated as a continuum[166(p592)]. This decreases the number of particles, significantly reducing the size of the space to be sampled. As a result, comparing simulations of different ligands is fairer because the results are less likely to differ due to differences in the configuration of the solvent.

In the thermodynamic integration method, the number of particles is not reduced; rather, particles are shared between the systems being compared to ensure that they take similar paths through the phase space.

1.4.3.1. Thermodynamic Integration

$$\Delta\Delta G = \Delta G_{ligandB} - \Delta G_{ligandA} = \left(G_{Bbound} - G_{Bfree}\right) - \left(G_{Abound} - G_{Afree}\right) \quad \textbf{1.63}$$

Where ΔΔG is the difference in the free energy of binding of two ligands, A and B, with free energies of binding $\Delta G_{ligandA}$ and $\Delta G_{ligandB}$ respectively. ΔΔG is equal to the free energy of their bound state (G_{Abound} and G_{Bbound} respectively) minus the free energy of the unbound molecules (G_{Afree} and G_{Bfree}).

As shown in Figure 1.14 and Equation 1.64, the difference in binding energy is equal to the change in free energy if ligand A were to be transformed in the binding site into ligand B ($\Delta G_{AtoBbound}$) minus the change in free energy if the same transformation were to be performed with the ligand unbound, free in solution ($\Delta G_{AtoBfree}$).

$$\Delta\Delta G = (G_{Bbound} - G_{Abound}) - (G_{Bfree} - G_{Afree}) = \Delta G_{AtoBbound} - \Delta G_{AtoBfree} \quad \textbf{1.64}$$

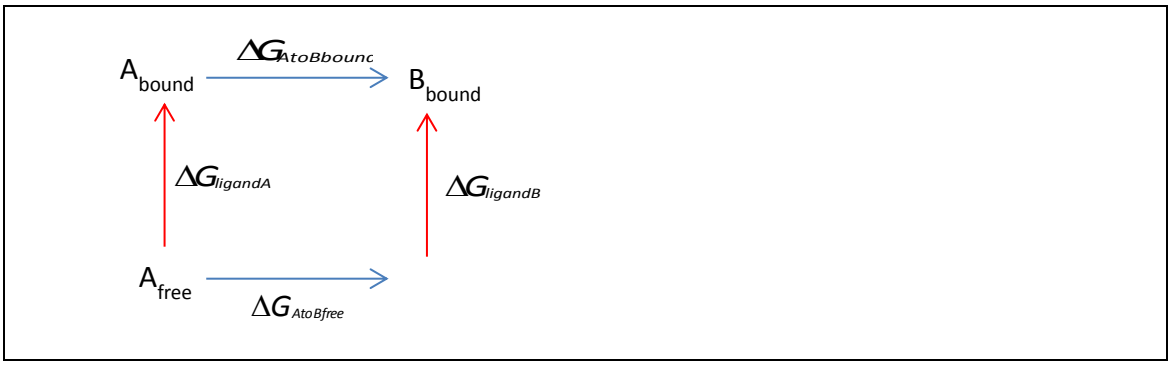


Figure 1.14: A Bordwell thermodynamic cycle to show how thermodynamic integration allows a free energy of binding to be estimated using thermodynamic integration. Two ligands are shown, A and B, both free in solution (A_{free} and B_{free}) and bound to the protein (A_{bound} and B_{bound}). The difference between the free energy change when ligand B is transformed into ligand A while bound to the protein and the free energy change when ligand B is similarly converted to ligand A when free in solution is equal to the difference in binding energy between the compounds.

In reality, there may be no mechanism by which ligand A could be transformed into ligand B, but such modifications can be studied *in silico*. In thermodynamic integration, simulations are performed at various points along the reaction coordinate, λ, of the alchemical transformation both for the bound ligand and the free ligand. In each of the simulation, the first derivative of the vibronic potential energy with respect to the reaction coordinate, $d\mathcal{V}/d\lambda$, is calculated[192]. Numerical integration of $d\mathcal{V}/d\lambda$ between λ=0 (the starting system containing molecule A) and λ=1 (the final system containing molecule B) yields an estimate of the overall change in free energy, ΔG ($\Delta G_{AtoBbound}$ or $\Delta G_{AtoBfree}$)[192].

$$\Delta G = G_1 - G_0 = \int_0^1 \left(\frac{\partial G}{\partial \lambda}\right)_{N,V,T} \partial\lambda. \quad \textbf{1.65}$$

The pressure and volume are assumed to be constant so the change in vibronic potential energy (which is a change in Helmholtz free energy) is equal to the change in Gibbs free energy (p90).

The (arithmetic) mean derivative of the potential energy from the trajectory frames analysed is assumed, making the ergodic assumption, to be an estimate of the Boltzmann average of the derivative.

$$\Delta G = \int_0^1 \left\langle {}^{\mathrm{d}\mathcal{V}}/_{\mathrm{d}\lambda} \right\rangle \, \mathrm{d}\lambda = \int_0^1 \left(\frac{\sum_i \left[e^{-\left(\frac{\mathcal{V}_i}{kT}\right)} \frac{\mathrm{d}\mathcal{V}_i}{\mathrm{d}\lambda} \right]}{\sum_j \left[e^{-\left(\frac{\mathcal{V}_j}{kT}\right)} \right]} \right) \mathrm{d}\lambda \quad \textbf{1.66}$$

Where k is the Boltzmann constant and T is the temperature.

Across the simulations, atoms which exist in compound A but not in compound B have to disappear as λ is increased and, similarly, atoms which exist in compound B only must appear. In practice, three separate TI steps are often used. Firstly, charges are removed from the disappearing atoms. Secondly, atoms appear and disappear as necessary. Thirdly, in the final step, the charges are put on to the new atoms.

The use of these three steps means that simulations are not carried out with charged atoms that are almost decoupled from the system (atoms disappearing when λ=1 or appearing when λ=0). These can potentially collide when attracted to an oppositely charged atom due to there being insufficient Pauli repulsion after the Leonard-Jones potential has been scaled by λ.

The Leonard-Jones potential (p68), is an AMBER force field component which is attractive at long distances but repulsive when atoms are close together, preventing their collision.

$$\begin{aligned} \mathcal{V}(\mathbf{R})_{\text{LJ, unchanged}} &= \sum_{nonbonded\ pairs} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right] \\ &= 4\varepsilon \sum_{nonbonded\ pairs} \left[\frac{1}{\left(\left(\frac{r_{ij}}{\sigma_{ij}} \right)^6 \right)^2} - \frac{1}{\left(\frac{r_{ij}}{\sigma_{ij}} \right)^6} \right] \end{aligned} \quad \textbf{1.67}$$

R_{ij} is the distance between atoms i and j, ε is the permittivity of the environment the atoms are in and σ_{ij} is the distance at which the potential is 0. If atoms become close together, then the Leonard-Jones potential can suddenly become very large causing the simulation to destabilise.

Even without charges, atoms may, by chance, become close in the course of their movement. In addition to performing the transformation in three stages as described, the Leonard-Jones potentials are usually replaced by what are called soft core potentials. These are Leonard-Jones potentials with an additional term, α, to ensure that the denominator of the attractive and repulsive terms is always of some magnitude[166(p579),186(p118)].

The Leonard-Jones potentials for the disappearing and appearing atoms are as follows. The total Leonard-Jones potential energy is the sum of that of the disappearing atoms, the appearing atoms and that of the unchanged atoms (calculated as normal).

$$\mathcal{V}(\mathbf{R})_{\text{LJ, disappearing}} = 4\varepsilon(1-\lambda) \sum_{nonbonded\ pairs} \left[\frac{1}{\left(\alpha\lambda + \left(\frac{r_{ij}}{\sigma_{ij}}\right)^6\right)^2} - \frac{1}{\alpha\lambda + \left(\frac{r_{ij}}{\sigma_{ij}}\right)^6} \right] \quad \textbf{1.68}$$

$$\mathcal{V}(\mathbf{R})_{\text{LJ, appearing}} = 4\varepsilon\lambda \sum_{nonbonded\ pairs} \left[\frac{1}{\left(\alpha(1-\lambda) + \left(\frac{r_{ij}}{\sigma_{ij}}\right)^6\right)^2} - \frac{1}{\alpha(1-\lambda) + \left(\frac{r_{ij}}{\sigma_{ij}}\right)^6} \right] \quad \textbf{1.69}$$

1.4.3.2. Implicit Solvent Methods

Water molecules were included in every simulation performed in the course of this project. However, later in the project, two implicit solvent methods, where the solvent is modelled as a continuum[166(p592)], were used to estimate binding affinities. The explicit water was removed from the recorded trajectories prior to analysis.

1.4.3.2.1. The MM-PBSA Method

In the molecular mechanics-Poission-Boltzmann surface area (MM-PBSA) implicit solvent method (figure 3)[194], a thermodynamic cycle is employed; the free energy of solvation for the ligand, protein and complex are used to convert a binding energy predicted *in vacuo* to one for in solution.

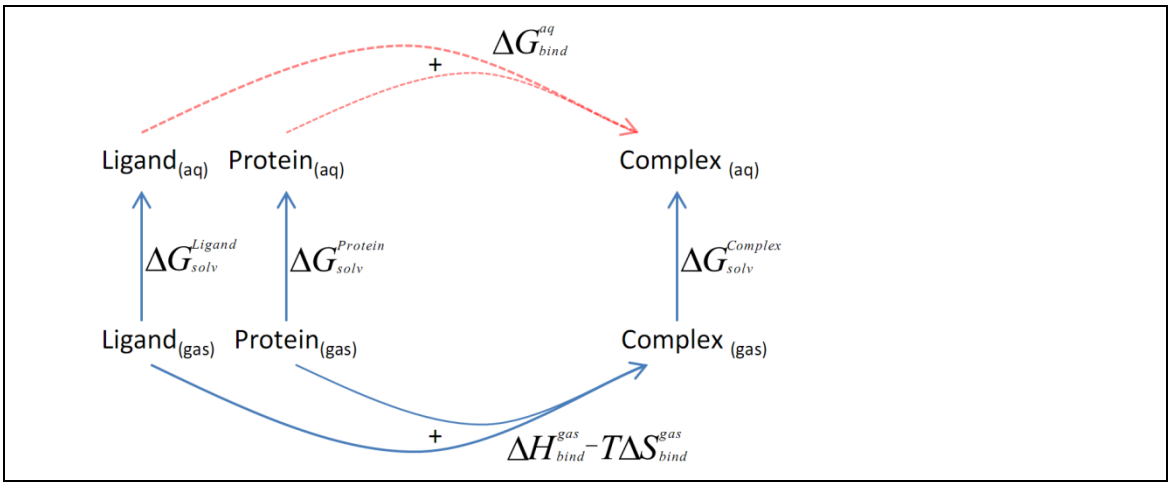


Figure 1.15: The Bordwell thermodynamic cycle used in the MM-PBSA method. The free energy of binding in solution (ΔG^{aq}_{bind}) is determined from more easily calculable energies, namely, the free energy of binding in the gas phase ($\Delta H^{gas}_{bind} - T\Delta S^{gas}_{bind}$), and the free energies of solvation of the ligand, the receptor and the complex, ΔG^{Ligand}_{solv},, $\Delta G^{Protein}_{solv}$, and ΔG^{cpx}_{solv}, respectively. This diagram has been adapted from a published figure[195].

The total free energy change of solvation can be modelled as the sum of two components, an electrostatic component, ΔG_{elec}, and a non-polar component, $\Delta G_{nonpolar}$.

$$\Delta G_{bind} = \Delta G_{elec} + \Delta G_{nonpolar} \quad \textbf{1.70}$$

When a species becomes solvated, the dielectric permittivity of the region surrounding the atoms changes. In the MM-PBSA method, the Poisson-Boltzmann equation is used to calculate the resulting change in electrostatic potential around the species. This is then used to determine the change in electrostatic energy, ΔG_{elec}.

$$\Delta G_{elec} = \frac{1}{2}\sum_{\mathbf{r}} q_i(\mathbf{r}_i)\Delta\phi(\mathbf{r}_i) \quad \textbf{1.71}$$

Where q_i is the charge of atom i at position $\mathbf{r}_i$ experiencing a change in electrostatic potential of $\Delta\varphi(\mathbf{r}_i)$[165(p292,p607)]. Calculation of the change in electrostatic potential is discussed in more detail below.

The solvent accessible surface area of the species is used to estimate the non-polar component of solvation, $\Delta G_{nonpolar}$.

$$\Delta G_{nonpolar} = \gamma A + b \quad \textbf{1.72}$$

Where A is the solvent accessible surface area and γ and b are empirically derived constants (which can be used with any molecule).

The solvent accessible surface area indicates the size of this non-polar component because it is roughly proportional to the number of water molecules in the first solvation shell around the species[166(p609)]. Two factors contribute to the non-polar component of the free energy of solvation, an energetically unfavourable cavitation component, ΔG_{cav} and an energetically favourable, van der Waals component, ΔG_{vdW}[196], both of which are proportional to the size of this first solvation shell.

$$\Delta G_{nonpolar} = \Delta G_{cav} + \Delta G_{vdW} \quad \textbf{1.73}$$

ΔG_{cav} accounts for the decrease in entropy of solvent molecules as they are forced out of the volume to be occupied by the solute and into a more dense and regular arrangement in solvent shells around the solute[197]. The volume of a complex is roughly equal to the sum of the unbound protein and unbound ligand volumes so any dependence of $\Delta G_{nonpolar}$ on species volume should be largely cancelled out when $\Delta G_{nonpolar}$ is used to calculate the binding energy of a ligand for a protein.

ΔG_{vdW} is the free energy change due to van der Waals forces (instantaneous dipole-induced dipole attractions and Pauli exclusion principle-related repulsions) the form between the surface of the solute and the first solvation shell of solvent molecules.

Returning to the electrostatic component from Equation 1.70, ΔG_{elec}, calculation of the electrostatic potential, $\varphi(\mathbf{r}_i)$, requires a partial differential equation to be solved. If the dielectric constant of the medium were the same throughout the system (independent of **r**), then the following form of the Poisson differential equation[166(p603)] could be used.

$$\nabla^2\phi(\mathbf{r}) \equiv \nabla \cdot \nabla\phi(\mathbf{r}) = -\frac{4\pi\rho(\mathbf{r})}{\varepsilon}, \quad 1.74$$

Where ρ(**r**) is the charge density and ε is the permittivity. ∇^2 indicates application of the del-squared operator (Equation 1.24) which is equivalent to finding the gradient in the x, y and z directions (∇ operator) and then finding the gradient of each of these gradients and adding these second derivative results (∇·).

It is assumed when using Equation 1.74 that the change in dielectric constant across the system is uniform when solvation occurs. In practice, the dielectric constant is position dependent due to the motion of mobile charges down the electrostatic potential gradient when the permittivity changes. A term based on the Boltzmann distribution must be introduced to account for this effect.

$$\nabla \cdot \varepsilon(\mathbf{r})\nabla\phi(\mathbf{r}) - \sqrt{\varepsilon}\kappa \sinh\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \quad 1.75$$

where κ is known as the Debye-Hückel inverse length[166(p604)].

$$\kappa = \sqrt{\left(\frac{\pi N_A e^2}{125k}\right)\frac{I}{T}} \quad 1.76$$

Where I is the ionic strength, T is the temperature and e is a constant, the charge of a proton.

Equation 1.75 can be simplified further, when the electrostatic potentials are very small ($e|\phi(\mathbf{r})| \ll kT$) because

$$\sinh x = \frac{1}{2}(e^x - e^{-x}) = \sum_{n=0}^{\infty}\frac{x^{2n+1}}{(2n+1)!} = \left[x + \frac{x^3}{6} + \frac{x^5}{120} + \frac{x^7}{5040} + \cdots\right] \approx x \quad 1.77$$

Where e is the base of natural logarithms.

The result is the linearized Poisson-Boltzmann equation.

$$\nabla \cdot \varepsilon(\mathbf{r})\nabla\phi(\mathbf{r}) - \sqrt{\varepsilon}\kappa\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \quad 1.78$$

This linearized equation often gives good results even when $e|\phi(\mathbf{r})|$ is of a similar size to kT[198] and it is the form of the Poisson-Boltzmann equation used to predict MM-PBSA binding energies by the mmpbsa.py program of the Amber suite, which was used in this project.

1.4.3.2.2. The MM-GBSA Method

The Molecular Mechanics-Generalised Born Surface Area (MM-GBSA) method of predicting the free energy of binding in a solvent is similar to that of the MM-PBSA method with the difference that the free energy of solvation is calculated using the Generalised Born equation rather than the Poisson-Boltzmann equation.

In Born's model of solvation[199], charges are treated as ions surrounded by a spherical solvent cavity of equal and opposite charge.

For a single ion,

$$G_{elec} = -\frac{q^2}{2\varepsilon R} \quad \textbf{1.79}$$

Where G_{elec} is its electronic free energy, q is the charge of the atom, ε is the relative permittivity of its environment and R is the radius of the cavity in the model, which corresponds to the van der Waals radius of the charged particle[166(p594)].

For a system containing many particles (represented using i and j below), the electrostatic energy comprises this Born solvation component (from Equation 1.79) plus the energy of the Coulomb interaction between the particles.

$$G_{elec} = \frac{1}{\varepsilon}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{q_i q_j}{r_{ij}} - \frac{1}{2\varepsilon}\sum_{i=1}^{N}\frac{q_i^2}{R_i} \quad \textbf{1.80}$$

Where r_{ij} is the distance between particles i and j and R_i is the effective Born radius of the particle.

In the model, solvation of an ion is likened to bringing the particle from a vacuum (ε=1) into the solvent sphere, which has relative dielectric ε.

$$\Delta G_{elec} = -\frac{q^2}{2R}\left(1 - \frac{1}{\varepsilon}\right) \quad \textbf{1.81}$$

When multiple particles are considered[166(p598)],

$$\Delta G_{elec} = -\left(1 - \frac{1}{\varepsilon}\right)\left(\sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{q_i q_j}{r_{ij}} + \frac{1}{2}\sum_{i=1}^{N}\frac{q_i^2}{R_i}\right). \quad \textbf{1.82}$$

The Born radius must be continually recalculated for particles in close proximity to others.

$$\frac{1}{R_i} \approx \frac{1}{a_i} - I \quad \textbf{1.83}$$

Where I is estimated using the function

$$I = \frac{1}{4\pi}\int_{|r|=a_i}^{\infty}\frac{1}{|\mathbf{r}|^4}d\mathbf{r}. \quad \textbf{1.84}$$

Where the integral is limited to the solute-occupied volume and the atom i is at position $|\boldsymbol{r}| = 0$[200].

Equation 1.83 leads to underestimation of the Born radii of atoms buried in a macromolecule. To account for this, other functions have been proposed. In this project the following function of Onufriev *et al.*[200] was tested (p169) to see if it improved results.

$$I = \frac{1}{a_i}\tanh(\alpha\Psi - \beta\Psi^2 + \gamma\Psi^3) \quad \textbf{1.85}$$

Where α, β and γ are empirically derived constants and

$$\Psi = I(a - 0.09) \quad \textbf{1.86}$$

The 0.09 was empirically derived by Onufriev *et al.*.

1.4.3.3. Entropy and Normal Mode Analysis

The change in the internal entropy of the ligand and protein upon binding is not normally included in an MM-PBSA or MM-GBSA calculation. However, this entropy change can be estimated using Normal mode analysis and then multiplied by the temperature to determine its effect on the free energy of binding. Consequently, Normal mode analysis will now be introduced, starting with an explanation of the partition function.

The entropy of a system, S, is, by definition, the logarithm of the number of ways (Ω) in which the elements of the system can potentially be arranged[165(p11)].

$$S = k \ln \Omega \equiv \log_{\exp(1/k)} \Omega \qquad \textbf{1.87}$$

Where k is the Boltzmann constant.

By definition,

$$H = G + TS = pV + F + TS = pV + U \qquad \textbf{1.88}$$

where H is the enthalpy of a system, G is the Gibbs free energy, T is the temperature, S is the entropy, p is the pressure, V is the volume, F is the Helmholtz free energy and U is the internal energy[165(p168)].

The entropy of the Universe as a whole, a closed system, is always increasing. Over time, the free energy (energy which can do work), G is converted to thermal energy which cannot do work, TS in formula 1.88.

$$\begin{aligned} U &= TS + F \\ &= T(k \ln \Omega) + (-kT \ln \mathrm{Q}) \\ &= kT \ln \frac{\Omega}{Q} \end{aligned} \qquad \textbf{1.89}$$

Where Q is what is termed the canonical partition function[165(p169)]. It is evident from Equation 1.89 that, for a system of constant pressure and volume, the partition function (Q) is the reciprocal of the factor by which Ω would be increased if all of the free energy of the system were to be used to do work (so F became 0 and Q became 1).

The canonical partition function, Q, can also be described as the sum of the Boltzmann factor for all of the microstates, i, of the system (the different quanta of Helmholtz free energy, ε_i, the system can have)[165(p12)]:

$$Q = \sum_i e^{-\left(\frac{\varepsilon_i}{kT}\right)} \qquad \textbf{1.90}$$

$$P_i = \frac{e^{-\left(\frac{\varepsilon_i}{kT}\right)}}{\sum_j \left[e^{-\left(\frac{\varepsilon_j}{kT}\right)}\right]} = \frac{e^{-\left(\frac{\varepsilon_i}{kT}\right)}}{Q} \quad \textbf{1.91}$$

Where P_i is the probability of the system being in microstate i (and have Helmholtz free energy ε_i)[165(p12)].

In quantum mechanics, some states have a limited occupancy. For example, the electron orbitals of an atom cannot contain more than two electrons. This restriction affects the distribution of particles between the possible energy states of the system and its inclusion in the partition function $Q_{quantum}$ below, is achieved using an additional variable, the degeneracy (maximum occupancy) of microstate i, g_i.

$$Q_{quantum} = \sum_i g_i e^{-\left(\frac{\varepsilon_i}{kT}\right)} \quad \textbf{1.92}$$

To understand how the entropy of a system can be estimated by Normal mode analysis, it is useful to express the total entropy of the system (S) in terms of the molecular partition function, hereinafter denoted by Z. This is the partition function for one of the molecules in the system (as opposed to the system as a whole). (The molecules are assumed to all be identical in this example.)

From Equation 1.89,

$$S = \frac{U}{T} + k\ln Q \quad \textbf{1.93}$$

For a system where the particles are distinguishable,

$$S_{dist} = \frac{U}{T} + k\ln Z^N \quad \textbf{1.94}$$

Where S_{dist} is the entropy and N is the number of molecules[201(p53)].

In reality, identical molecules are indistinguishable so the number of unique ways they can be rearranged is reduced by a factor of N![201(p53)].

$$\begin{aligned} S &= \frac{U}{T} + k\ln\frac{Z^N}{N!} \\ &= \frac{U}{T} + Nk\ln Z - k\ln(N!) \end{aligned} \quad \textbf{1.95}$$

$$\begin{aligned} \lim_{N\to\infty} S &= \frac{U}{T} + Nk\ln Z - Nk\ln N + kN \\ &= \frac{U}{T} + Nk\left(\ln\frac{Z}{N} + 1\right) \end{aligned} \quad \textbf{1.96}$$

The total internal energy of a system, U, can be expressed in terms of the molecular partition function. To do this it is necessary to express the partial derivative of the molecular partition function at constant volume in terms of the energy of each microstate (i), ε_i, and the microstate degeneracy, g_i.

From Equation 1.90[201(p184)],

$$Z = \sum_i g_i e^{-\varepsilon_i/kT}. \quad \textbf{1.97}$$

$$\left(\frac{\partial Z}{\partial T}\right)_V = \frac{1}{kT^2}\sum_i g_i \varepsilon_i e^{-\varepsilon_i/kT} \quad \textbf{1.98}$$

$$\begin{aligned} U &= N\sum_i \varepsilon_i \\ &= \frac{N}{Z}\sum_i Z\varepsilon_i \\ &= \frac{N}{Z}\sum_i g_i \varepsilon_i e^{-\varepsilon_i/kT} \\ &= \frac{NkT^2}{Z}\left(\frac{\partial Z}{\partial T}\right)_V \\ &= NkT^2\left(\frac{\partial \ln Z}{\partial T}\right)_V \end{aligned} \quad \textbf{1.99}$$

Using equation 1.99[201(p52)] to substitute U in Equation 1.96 yields[201(p52)]

$$\begin{aligned} S &= NkT\left(\frac{\partial \ln Z}{\partial T}\right)_V + kN\left(ln\frac{Z}{N} + 1\right) \\ &= Nk\left[T\left(\frac{\partial \ln Z}{\partial T}\right)_V + \ln Z - \ln N + 1\right]. \end{aligned} \quad \textbf{1.100}$$

In this project, Normal mode analysis was used to estimate the relative vibrational and translational entropy of the ligand and protein when free and when bound together. Normal mode analysis is the identification of the harmonic modes that a molecule can have, concerted atomic motions in which all of the nuclei are moving in phase[201(p196)]. Only these oscillations are stable. There is only one harmonic mode that corresponds to translation, the mode in which all of the nuclei are moving in the same direction. The others are internal vibrations, vibrational modes.

Where S_{vib} is the entropy due to the vibrational modes,

$$S_{vib} = Nk\left[T\left(\frac{\partial \ln Z_{vib}}{\partial T}\right)_V + \ln Z_{vib}\right] \quad \textbf{1.101}$$

Where Z_{vib} is the vibrational molecular partition function. The entropy can now be expressed in terms of the partition function, Z_i, of each vibrational mode, i, [166(p162)].

$$Z_{vib} = \prod_i Z_i \quad \textbf{1.102}$$

$$S_{vib} = Nk\sum_i \left[T\left(\frac{\partial \ln Z_i}{\partial T}\right)_V + \ln Z_i\right] \quad \textbf{1.103}$$

Analysis of a structure such as a molecular dynamics simulation snapshot allows the frequencies at which a molecule will vibrate in each of its harmonic modes to be determined. This allows Z_i to be calculated for each mode and thus enables calculation of the entropy.

In this project, Nmode in the AMBER package used to perform the Normal mode analysis. It first uses the Cerjan–Miller algorithm[202] to identify the second derivative of the potential energy (with respect to position), for each of the atoms in terms of the positions of all of the other atoms. The result of this calculation is a 3N×3N matrix (where N is the number of nuclei), $\boldsymbol{\mathcal{V}''}$, where each matrix element is the force an atom exerts on another multiplied by the geometric mean of the masses of the two atoms[166(p275)].

To find the Normal modes, an equation is defined which enforces harmonic motion[166(pp274-275)].

$$\mathbf{M^{-1/2}\mathcal{V}''M^{-1/2}x} = \boldsymbol{\lambda}\mathbf{x}. \qquad \textbf{1.104}$$

Where **M** is a diagonal matrix containing the masses of the atoms (so $\boldsymbol{M^{-1/2}\mathcal{V}''M^{-1/2}}$ is a matrix containing the forces each atom exerts on each other), **λ** is a column matrix containing the energy associated with each normal mode, ε_i, (the eigenvalues) and **x** is a matrix (the eigenvectors) which attributes each force component to one or more of the normal modes.

The energies associated with each harmonic of a mode, ε_i, are multiples of the energy associated with the fundamental frequency of that mode, ε_{if}, (if the vibrational zero point energy is regarded as 0)[166(p275),201(p107,p184)] and Z_i, can be calculated from ε_i[201(p185)].

$$\begin{aligned} Z_i &= \sum_{n=0}^{\infty} e^{-\varepsilon_{if} n/kT} \\ &= \frac{1}{1 - e^{-\varepsilon_i/kT}} \end{aligned} \qquad \textbf{1.105}$$

Equation 1.105 can be used to substitute for Z_i in Equation 1.103[201(p197,p198)].

$$\begin{aligned} S_{vib} &= Nk \sum_i \left[T \left(\frac{\partial \ln \frac{1}{1 - e^{-\varepsilon_i/kT}}}{\partial T} \right)_V + \ln \frac{1}{1 - e^{-\varepsilon_i/kT}} \right] \\ &= Nk \sum_i \left[\frac{\varepsilon_i/kT}{e^{\varepsilon_i/kT} - 1} - \ln\left(1 - e^{-\varepsilon_i/kT}\right) \right] \end{aligned} \qquad \textbf{1.106}$$

1.5. Molecular Dynamics Simulations and The p53-Mdm2 Interaction

1.5.1. SIMULATIONS WITH OLIGOBENZAMIDES

Performing what appear to have been the first binding affinity predictions for Mdm2 made using all-atom, explicit solvent molecular dynamics simulations, Fuller *et al.*[203] demonstrated the successful application of thermodynamic integration (TI) to oligobenzamide inhibitors of the p53-Mdm2 interaction. Their study focussed on six molecules with side chains linked through oxygen. This project builds on their work, testing a much greater number of compounds to determine if such simulations are of practical use with the larger chemical space now synthetically accessible to our collaborating chemists. Included are 31 already synthesised and tested molecules where side chains are attached via the amide group nitrogen.

Fuller *et al.*[172,203] report how an oligobenzamide appears to be stable when bound to Mdm2 in an antiparallel orientation in addition to when parallel to the alignment of the bound p53-helix. Their results are based on simulations starting from docking poses generated using Autodock where a small docking grid was used, forcing compounds to bind close to the binding site. In my work, the greater speed of Autodock Vina permitted a far greater number of compounds to be investigated and for them to be docked quickly without restriction of the searched volume to the binding site. This enabled key assumptions to be tested, such as the hypothesis that oligobenzamides bind in the p53 binding site and the assumption that, if and when they do, they bind with each side chain positioned so as to mimic a p53 residue. Fuller's suggestion that the specific combination of side chains could dictate whether parallel or antiparallel binding is most efficient, is also investigated.

Fuller *et al.* used torsion parameters intended for use with sulphur to model the intramolecular hydrogen bonding between the side chain-linking oxygen and amide group hydrogen (p69). In this project, more recently published parameters for use with oxygen were tested to see if they also yielded stable simulations.

The results of Fuller *et al.* suggested that the effect of multiple side chain modifications could be predicted in a single TI analysis in which all of the side chains were simultaneously transformed[203]. However, predicting the relative affinity of a large number of molecules by TI is still impractical. The paradigmatic model of an oligobenzamide inhibitor in which each side chain mimics a single amino acid suggests that side chains could make independent contributions to the binding energy which could be summed to predict the relative binding energy of a compound with any side chain combination. In this project, I investigate if this independence is observed or whether the concerted transformation of all side chains is essential if accurate affinities are to be obtained.

1.5.2. OTHER SIMULATIONS OF THE P53-MDM2 INTERACTION

Molecular dynamics simulations have been used extensively to study the p53-Mdm2 interaction.

The p53 residues Phe19, Trp23 and Leu26 are known to contribute significantly to the binding energy of the p53-Mdm2 interaction[26]. Espinoza-Fonseca *et al.*[204] studied the duration of interactions between aromatic residues of p53 and Mdm2 and concluded that interactions between Phe19 and Trp23 of p53 and Tyr51 and Tyr63 of Mdm2 played a key role in recognition. Verma *et al.*[205] have investigated how the polyphenols quercetin and taxifolin bind to Mdm2 and how they might displace p53 from the binding site by disrupting its interaction with the side chain of Tyr51.

Alanine scanning has been performed on p53 *in silico* using the MM-PBSA method to estimate the affinity of mutated peptides[195]. In such experiments, a simulation of the protein is performed for each residue with the residue side chain replaced by a methyl group. Applied to p53 in the human p53-Mdm2 interface, the method correctly identifies Phe19, Trp23 and Leu26 as residues critical for binding[195]. Zhong and Carlson[206] performed alanine scanning on Mdm2 and identified Leu54, Ile61, Met62, Tyr67, Glu72, Val93, His96 and Tyr100 as important for p53 binding.

Ding *et al.*[207] used quantum mechanical simulations to study the key interacting residues of p53-Mdm2. They identified p53 residue Leu22 as also important; although, they concluded that overall the p53-Mdm2 interaction was driven by van der Waals forces rather than specific residue-residue interactions.

Espinoza-Fonseca *et al.*[208] used molecular dynamics to compare the dynamics of Mdm2 with and without bound p53. Their simulations reveal that p53 stabilises the flexible p53 binding site of Mdm2 when it binds. Joseph *et al.*[43] investigated how the Mdm2 binding site is more flexible than the Mdm4 binding site, a feature which appears to allow Nutlin-3 to bind much more strongly to the former than the latter.

Shan *et al.*[209] used simulations, as well as NMR, to investigate whether the second of the two p53 transactivation domain subdomains interacts with Mdm2 (in addition to the first subdomain, which contains Phe19, Trp23 and Leu26). They found that it did, but did not undergo binding-induced folding like the first subdomain does.

Dastidar *et al.*[210] have discovered that the N-terminal part of Mdm2 can influence p53 binding through its effect on the conformation of Tyr100. Dastidar *et al.*[211] have also used simulations to study the dynamics of the first 24 residues and how these dynamics are influenced by Ser17

phosphorylation and mutation of Ser17 to aspartate. Because this N-terminal region can block access to the p53-binding site it is absent from the constructs used in this project (p32).

Using Brownian dynamics[212], ElSawy *et al.*[213] have investigated the binding of p53 and a Nutlin to both Mdm2 and Mdm4 over a longer timescale than can typically be studied in a molecular dynamics simulation. Specifically, they studied the transient interactions which p53 forms with each protein prior to formation of the stable bound state. p53 appears to interact with the N and C termini of Mdm2 before it reaches its binding site.

In this project, implicit solvent methods including the MM-GBSA method (p88) were used to predict the relative affinities of oligobenzamides for Mdm2. Chen *et al.*[214] describe their use of the MM-GBSA method to predict the relative binding energies of four small molecules (a benzodiazepinedione, a chromenotriazolopyrimidine, a Nutlin and an imidazolylmethylindole derivative) for Mdm2. The results correlated well with previously published, experimentally determined binding affinities.

The aim of this chapter has been to introduce the motivations for, background of and theory behind this project, in preparation for the subsequent results chapters. This chapter does not introduce the statistical methods employed in the project; these can be found in Appendix A (p290).

2. *In Silico* Predictions Suggest Oligobenzamide Side Chains Do Not Make Independent Contributions to the Binding Energy

2.1. Summary

This chapter concerns the docking of oligobenzamides into the crystal structure of Mdm2, the simulation of these oligobenzamides bound to Mdm2 and analysis of the results to discern the complex relationship between the choice of scaffold and side chains and the binding pose and predicted affinity of an oligobenzamide.

The paradigmatic model of oligobenzamide binding places the chains of the peptidomimetic in separate binding pockets, those usually occupied by the side chains of p53 residues Phe19, Trp23 and Leu26. Initially, such an arrangement was assumed and the properties of a side chain within each binding pocket likely to increase the affinity of a bound oligobenzamide were identified. Specifically, the ZINC database, a database of commercially available small molecules, was used to generate side chains which could potentially be added to an oligobenzamide scaffold and then, *in silico*, these side chains were systematically attached to a docked oligobenzamide and docking scores were obtained. The docking scores of oligobenzamides which differed in their side chain at a single side chain attachment position were then compared to attribute the differences in score to side chain properties.

It was not possible to test the predictions made using the ZINC database due to the large number and synthetic intractability of the compounds. Consequently, subsequent work focussed on the analysis of a smaller number of molecules and the production of more accurate binding affinity predictions using molecular dynamics simulations.

All-atom molecular dynamics simulations of oligobenzamides demonstrated that stable stimulations could be started from the poses generated by docking oligobenzamides into Mdm2 from PDB (www.rcsb.org[36]) structure 1T4F. Thermodynamic integration (TI) was used to

modify the side chains of an N-alkylated oligobenzamide[98] (the N1 scaffold molecule shown in Figure 2.1, p102) and the results were used to produce a set of predicted affinities for 31 compounds which had previously been synthesised and for which laboratory data was available. TI was also used to investigate the effect of attachment atom modification in the O1 and S1 scaffolds (Figure 1.6, p36) and to investigate the importance of certain scaffold torsion parameters for the accurate prediction of binding affinity.

The correlation between experimental and predicted results for the 31 small molecules based on the N1 scaffold was statistically significant but small. Subsequent analysis of more extensive docking results by logistic regression suggested that side chain changes permit a wide variety of different binding poses, a possibility alluded to by Fuller[172]. The results implied that, for oligobenzamides, binding energy was a property of the combination of side chains rather than being the sum of independent contributions attributable to each side chain. This finding advocated a new approach whereby side chain combinations as opposed to individual side chains were tested.

The large number of possible side chain combinations meant that, to investigate each one, a method which enabled a binding affinity to be derived from a single simulation was required. The Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics Generalised Born Surface Area (MM-GBSA) implicit solvent methods allow this.

Predicted binding energies generated using these methods were more closely correlated with experimental results. However, the correlation was only moderate. Future work would ideally use long simulations but the necessity to carry out simulations for every combinatorial library member in different binding poses could make this impractical. A possible compromise would be to treat each combination individually when docking and then, once the likely binding site of each possible ligand had been identified, investigate each side chain individually within each binding site. To investigate the interaction of factors which could affect affinity and test whether side chains made independent contributions to the binding energy both overall and within a particular binding pose (as assumed in the aforementioned proposition), predictions were made for a large number of compounds based on different scaffolds and analysis of variance (ANOVA)[215] was performed.

ANOVA tests are more reliable with complete data (when side chain interactions are being investigated, results for every possible side chain combination); however, the production of poses within a specific site for every member of a combinatorial library is a challenge. Initial testing with O-alkylated oligobenzamides used docking pose optimisation (Autodock “local” docking) to ensure that a pose was generated for every library member close to the paradigmatic binding pose. However, trajectory analysis with the MM-PBSA method yielded positive free energies of binding, which suggests that the results were unreliable. Consequently, subsequent

work used standard docking results from which poses within four specific and common binding poses were identified and the problem of missing data (having no binding pose for some molecules) was solved using impution.

The ANOVA results suggest that there is a statistical interaction between the side chain choices when treated as ANOVA factors. If binding poses are segregated by binding position prior to analysis, a small interaction is still observed, indicating that the aforementioned interaction between side chains is not solely due to the influence of side chains on where binding occurs. However, the extent of the interaction is much reduced.

At the end of this chapter, which describes the entirety of the computational work performed in this project, it is hypothesised that the lack of a high correlation between computational and experimental results could be partly attributable to error in the laboratory assay used for prediction validation. The need for an orthogonal method to assess binding affinity is highlighted. The development of such an assay is the subject of Chapter 3. The validation of predicted binding poses is also proposed due to the variety of poses predicted by docking. Small molecule binding positions can be probed using NMR and this is the subject of Chapter 4.

Key research outcomes described in this chapter:

- Identification using FlexX of the properties of side chains in the Phe19, Trp23 and Leu26 binding sites associated with increased binding affinity, most notably the need for a large side chain in the Trp23 site and a flexible one in the Leu26 site.
- Discovery that oligobenzamides can bind to Mdm2 in many different places depending on the choice of side chains.
- Discovery of a robust, moderate correlation between the predicted and experimentally determined oligobenzamide binding affinities for Mdm2 when using implicit solvent methods.
- Discovery that there is a statistical interaction between the effects of side chains at different positions on the affinity of an oligobenzamide within a particular binding site but that this interaction is relatively small in magnitude.
- Identification of oligobenzamides where side chains are attached through sulphur at the ortho ring position as warranting further investigation due to the availability of parameters for this scaffold and the stronger binding of this scaffold relative to the corresponding one with oxygen-linked side chains suggested by thermodynamic integration results.

2.2. Methods

2.2.1. PDB STRUCTURE PREPARATION

Compounds were docked into a static structure derived from PDB (www.rcsb.org[36]) structure 1T4F[37]. Structure 1T4F features Mdm2 bound to a peptide with a sequence resembling that of the p53 transactivation domain. The p53 peptide was removed by manual editing and the remaining structure was protonated using the H++ server (http://biophysics.cs.vt.edu/) as appropriate for pH 7. The server was also used to predict the orientation of side chain amide groups and the imidazolyl group orientation and δ/ε tautomerisation of each histidine. pK_a values were calculated using the default internal and external dielectric constants of 6 and 80 respectively and the salinity was also left at its default value, 150 mM.

2.2.2. COMPOUND 3D STRUCTURE PRODUCTION

2.2.2.1. The scaffolds and side chains

All of the compounds investigated in this project have a core structure corresponding to one of five scaffolds, the N1 scaffold, the O1 scaffold, the S2 scaffold, the O2 scaffold or the S2 scaffold. These are shown in Figure 1.6 (p36).

The compounds shown in Figure 2.1 (p102), hereinafter referred to as the "starting molecules", are based on these scaffolds. All of them have the side chains benzyl, napthylmethyl and isobutyl. These correspond to the side chains of the amino acids Phe19, Trp23 (actually 3-indolylmethyl in tryptophan) and Leu26 of p53 so these molecules might be expected to bind tightly to Mdm2.

Rough 3D structures of these molecules (input as Daylight SMILES[216]) were generated using Obgen, a program of the Open Babel open source suite of tools (suite version 2.2.3)[217]. The stereochemistry of the molecules was checked where applicable. Then, these were minimised using the Obminimize program of the same suite. The mmff94s force field[218] was used. Up to 10,000,000 steps of conjugate gradient minimisation were performed with Obminimize. The required convergence for termination prior to the maximum number of cycles being reached was set to 1×10^{-8} kcal mol^{-1} $Å^{-1}$. The number of steps and convergence were increased from their default values of 2500 and 1×10^{-6} kcal mol^{-1} $Å^{-1}$ because it was thought that further minimisation of the small number of side chains might reduce the amount of time required to minimise each member of the large combinatorial library subsequently produced by the attachment of these side chains to a scaffold.

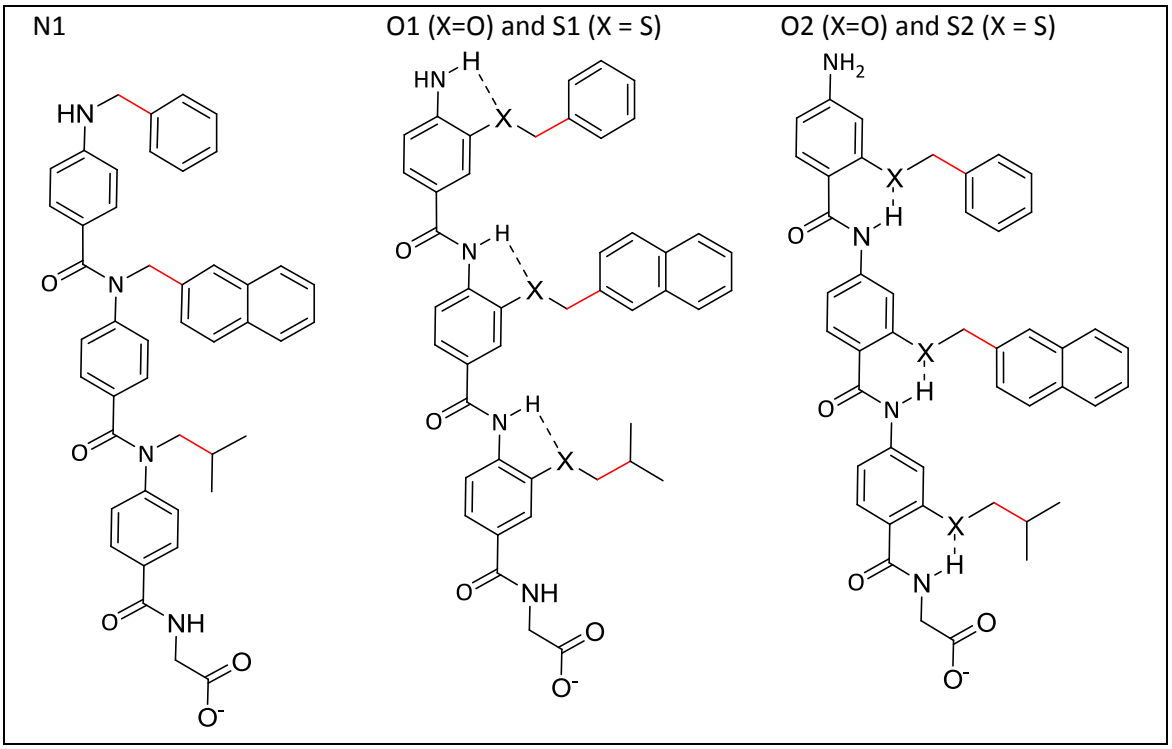
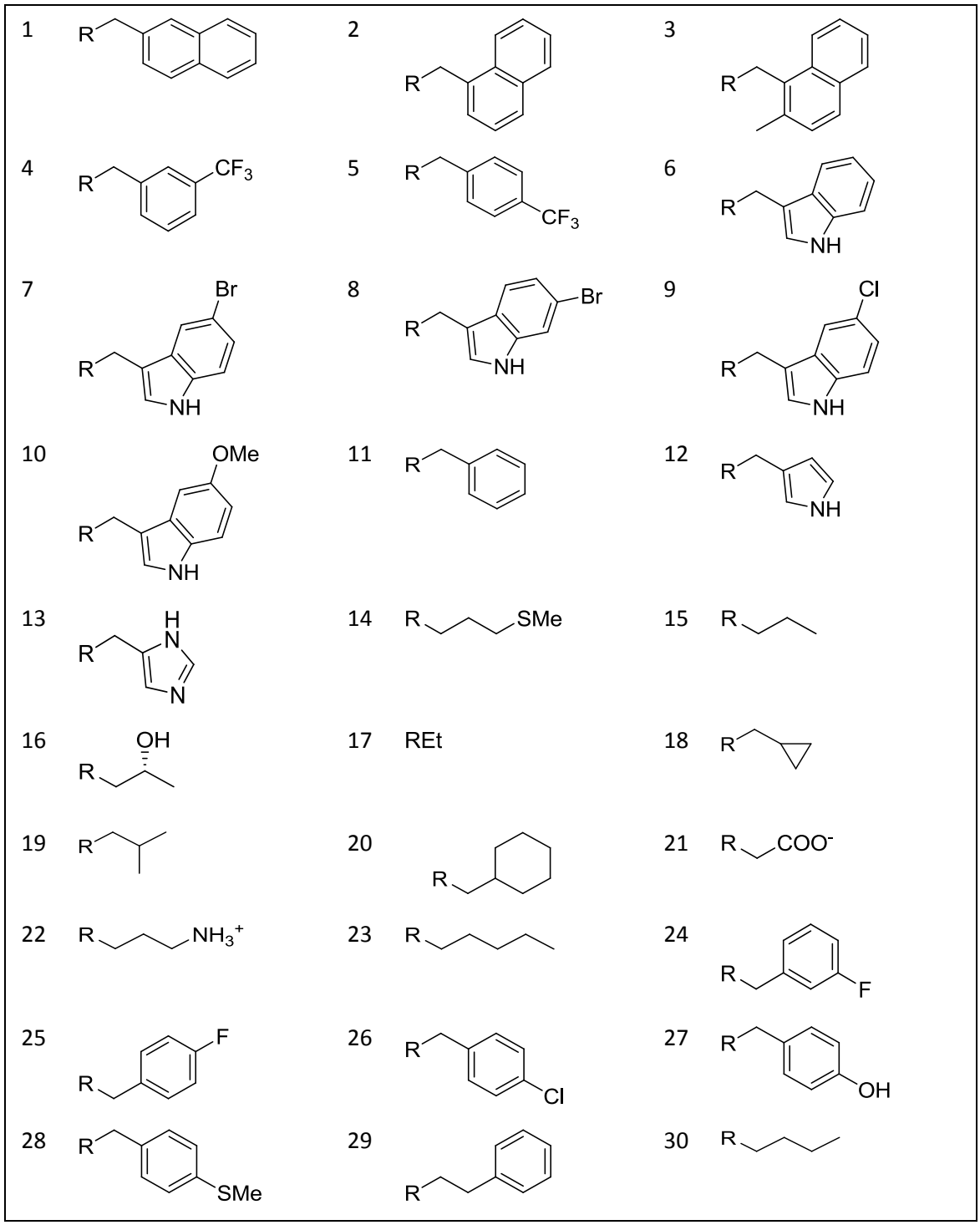


Figure 2.1: Model compounds for each oligoamide scaffold, which were used as the starting point for further modification to produce all of the compounds tested. Five compounds are indicated, the N1 starting molecule, the O1 starting molecule, the S1 starting molecule, the O2 starting molecule and the S2 starting molecule. The bonds shown in red were the bonds cut to modify the side chains unless otherwise stated. Dashed lines show the location of the intramolecular hydrogen bonds present in the scaffolds with O-linked and S-linked side chains.

The five molecules shown in Figure 2.1 were used as the starting point for generating all of the oligobenzamide structures discussed in this project. In initial work using the ZINC database[219,220], chains were modified using the high-throughput screening software ReCore and the docking software FlexX (BioSolveIT)[149]. (See p106 for details.) However, in subsequent work, which will now be described, molecules containing the side chains to replace those on the five starting molecule structures (Figure 2.1) were generated using Obgen and Obminimize (as described above) and then the side chains were spliced on to one of the five starting molecules using custom software (Side Chain Splicer, p105). The side chains are shown in Figure 2.2. They were selected on the basis of the compounds the collaborating chemistry group had made or planned to synthesise. Minimisation with Obminimize was repeated after splicing using up to 10,000 steps of conjugate gradient minimisation with a required convergence of 1×10^{-7} kcal mol^{-1} $Å^{-1}$.

The 3D structure of each side chain was generated attached to one of the units shown in Figure 2.3 (p105) before it was spliced onto one of the scaffold structures. These units were used to promote establishment of an extended side chain conformations which would not sterically clash after subsequent attachment to the scaffold.



Continued on the next page…

Figure 2.2: The side chains used in this project. Different investigations used different sets of side chains. The set used in each investigation are given in the results section. The R group is inductively electron withdrawing because the atom attaching the side chain to the scaffold is nitrogen, oxygen or sulphur. Consequently, for example, the imidazole isomer shown in 13 is the most stable protonation state in an oligobenzamide, unlike in histidine where the side chain is attached through carbon so is subject to inductive electron donation.

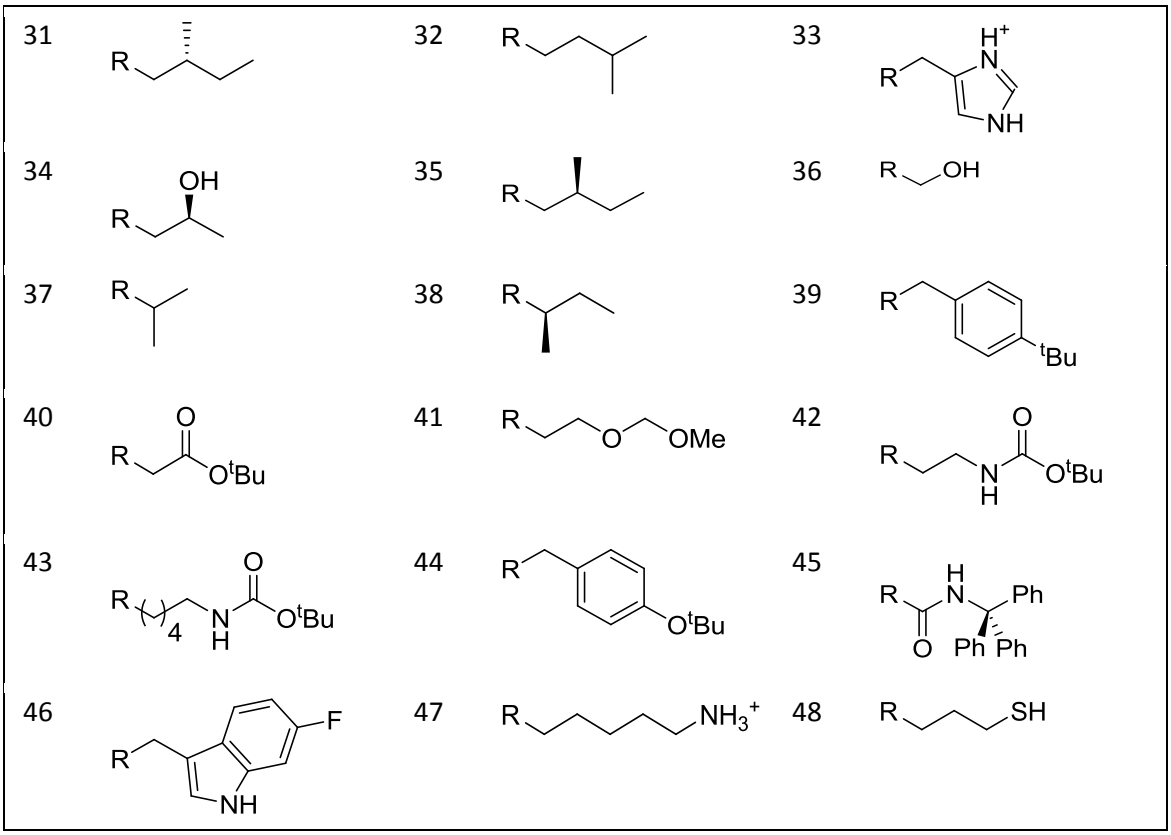


Figure 2.2 continued from the previous page.

At times in this project it was necessary to choose which side chains to investigate at each side chain attachment position. The number of side chains used for a particular investigation was decided with the total required computation time in mind. With regard to each attachment position, some consideration was given to the size of the Phe19, Trp23 and Leu26 side chains which the oligobenzamide side chains are intended to mimic. Specifically, side chains with two or more aromatic rings were only tested at the middle side chain position and there was bias towards the testing of aromatic side chains at the N-terminal position and aliphatic side chains at the C-terminal attachment position. However, there was also a desire to test a large chemical space, especially where it included structures which might soon be made in the collaborating chemistry group.

Throughout this project, conversions between Structure Data File (SDF) files (Molecular Design), Mol2 (Tripos) and PDB formats were performed using the Babel program of Open Babel[217]. This program was also used to protonate of molecules unless otherwise stated.

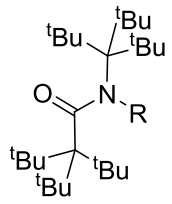
The Filter program from the Open Babel suite was used for the calculation of molecular properties. These properties included the topological polar surface area (PSA) and XlogP. The PSA of a species is the exposed van der Waals surface of the polar atoms in the molecule and their attached protons. The topological PSA is an accurate estimate of the true PSA calculated

by summing pre-calculated fragment contributions[221]. As indicated in Equation 2.1, the partition coefficient of a species is the ratio of the concentrations of its unionized form in each layer of a mixture of water and octan-1-ol, two immiscible solvents.

$$P_X = \frac{[X]^{\text{unionised}}_{\text{octan-1-ol}}}{[X]^{\text{unionised}}_{\text{water}}} \qquad \textbf{2.1}$$

In similarity to the topological PSA, the XlogP[222] is an estimate of the logP determined by the summation of predetermined logP contributions for the different parts (in this case atoms) of the molecule.

a For attachment to the N1 scaffold



b For attachment to the other scaffolds O (X=O) and S (X = S)

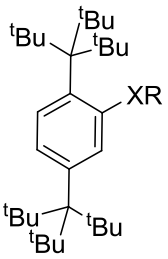


Figure 2.3: The units holding the side chain, taking the place of the scaffold, when 3D structures of each side chain were generated. Three molecules are indicated, one to generate side chains for attachment to the N-linked scaffold (a), one for attachment to the O1 and O2 scaffolds (b, X=O) and one for attachment to the S1 and S2 scaffolds (b, X=S). ^{t}Bu signifies a tert-butyl group (-C(CH_3)$_3$). These molecules were used to encourage the production of extended side chain conformations which would not sterically clash with the scaffold when attached.

3D depictions of ligands and proteins herein were, unless otherwise stated, produced using the UCSF Chimera package[223], which is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, and supported by grant NIGMS P41-GM103311.

2.2.2.2. <u>Custom Splicing Software: Side Chain Splicer</u>

Custom software, Side Chain Splicer, was used for splicing. Written in Perl, it uses trigonometric functions to calculate how atoms must be moved and then to apply these transformations to the atoms concerned.

The software accepts as input the structures of the two molecules to be joined (in Mol2 format) and the numbers of 6 atoms:

- Those either side of the bond to be cut in the first molecule.
- Those either side of the bond to be cut in the second.

- A third atom in each molecule which, together with the atoms either side of the bond, defines a plane through each molecule. These two planes (one through each molecule) are lined up following splicing by rotation of the component from the second molecule (the side chain) around the axis defined by the bond created.

Further details:

- The first molecule is not translated or rotated so it can be in a docked pose.
- The length of the bond created is the mean length of the bonds cut. Wherever possible, identical bonds were cut so the resulting bond was the correct length. (In other cases, minimisation with Amber[224–226] was applied following splicing, which will have regularised the molecular geometry.)
- In the output file, the atoms of the final molecule comprise the atoms from the first followed by the atoms from the second. The order of the atoms in each molecule is unchanged. In this project, the scaffold atoms in the starting molecules (p101) preceded those of the side chains. This enabled the production of combinatorial libraries in which the core oligoamide backbone atoms had identical numbering.

All of the custom software used in this project is of my own creation.

2.2.2.3. Preliminary work using the ZINC database and ReCore

As described in the results section (p128), oligobenzamides with a diverse range of side chains were docked using FlexX. The aim was to identify the side chain properties likely to increase the affinity of an oligobenzamide acting as a peptidomimetic with its side chains in the sites usually occupied by p53 residues Phe19, Trp23 and Leu26.

1171 side chains were tested attached to the O1 scaffold using the ReCore high-throughput screening and FlexX docking software, two parts of the LeadIT package (BioSolveIT) (version 1.0.2)[133]. The software enables a set of molecules to be cleaved to produce fragments. ReCore allows these fragments to be screened using a pharmacophore model and FlexX facilitates the subsequent combinatorial docking of a scaffold to which these side chains have been attached.

The filtering, screening and docking will now be described in more detail.

ZINC is a database of commercially available compounds (http://ZINC.docking.org/)[219,220]. The unprocessed source molecule dataset (9,959,338 molecules at the time) was downloaded from the ZINC database in Daylight SMILES[216] format. These input molecules were filtered to identify molecules which could be used in oligobenzamide synthesis to attach a side chain. In tribenzamide synthesis, each of the three monomers is synthesised separately, a process which includes attaching the side chain. Then, these three monomers are joined together to produce the tribenzamide. In the first of these two steps, side chain attachment can be through an S_N2

substitution, using a primary or secondary hydroxyl group or halide on the side chain donor molecule as a leaving group[49]. There must be only one leaving group of the same or higher leaving group ability for the correct reaction to occur. Alternatively, a $B_{Ac}2$ reaction can be carried out using an acyl chloride. This creates an ester link between the aminobenzoic acid unit and the new side chain. If an acyl chloride group is used, there can be no group of equivalent or greater reactivity in the molecule donating the side chain.

When the aminobenzoic acid monomers are joined together to generate the oligobenzamide, the amino group of each monomer must act as a nucleophile and react with the preceding activated monomer[49]. There must be no better nucleophile present in any of the side chains (for example, a sulphydryl group) because this would react in preference to the amino group. In addition, no group must be more reactive than the nucleophilic group on the oligobenzamide which the amino group must attack. The nature of this group is dependent on the coupling agent used. While it may be difficult to assess every side chain individually for inductive and mesomeric effects, which will subtly affect the reactivity of functional groups, molecules with groups which are commonly very reactive, such as those with an acyl chloride group, can be excluded as necessary.

Filtering was carried out with the aid of Filter (OpenEye Scientific Software) (version 2.0.2). Molecules with a primary or secondary halogen atom, a primary or secondary hydroxyl group or an acyl chloride group were identified. Further filters were used to ensure synthetic accessibility. Firstly, molecules were filtered to ensure that they did not have one of a number of reactive groups for which filter information was supplied with the ReCore software[227]. Secondly, molecules were only used if they would produce a small side chain ($M_r \leq 500$). Otherwise, the side chain added would be of an unsuitable size and the compound might be difficult to make or expensive to purchase. Thirdly, only molecules yielding a side chain with 40 or fewer rotatable bonds were used because compounds with small side chains are easier to dock and have a smaller conformational space to sample in any subsequent molecular dynamics simulations. Finally, only molecules containing just the atoms hydrogen, carbon, nitrogen, oxygen, sulphur, phosphorus, fluorine, chlorine, bromine and iodine were used, to remove exotic molecules which might have unusual chemistry. 1171 suitable molecules, which could be used to attach a particular side chain, were identified.

Open Babel[217] was used to generate all of the ionisation states of these molecules which would predominate at pHs between pH 5.0 and pH 9.5 (as SMILES).

The reactive hydroxyl, halogen or acyl chloride group, was removed and replaced with the group shown in Figure 2.4. The SMILES were modified using a custom Python program utilising the OEChem Python wrapper from the OEChem Toolkit (OpenEye Scientific Software) (version 1.4.2). This process created the bond shown in blue in Figure 2.4.

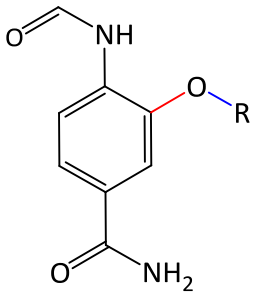


Figure 2.4: The structure joined to ZINC source molecules prior to ReCore cleavage. Attachment created the blue bond. The red bond was then cleaved by ReCore, so that every side chain had an oxygen atom which could be used later to orientate it in high-throughput screening.

3D structures were generated from the resulting SMILES using Omega (version 2.3.2) (OpenEye Scientific Software). The aromatic unit attached was intended to mimic the scaffold during structure generation and encourage the formation of an elongated side chain structure, in the same way as the units shown in Figure 2.3 (used in subsequent work). The default force field was used (a variant of the MMFF94s field[218] which lacks the electrostatic interaction terms). Up to 400 conformers with a strain energy within 10 kcal mol^{-1} of the minimum were produced per molecule, subject to a maximum run time limit of 2 minutes for each molecule. Any non-planar nitrogen atoms were inverted and where side chains contained aliphatic rings, different ring conformations were generated.

The 3D structures were cleaved using ReCore to produce fragments. The bond cleaved is shown in red in Figure 2.4. Ultra-high throughput screening of the possible side chain fragments was then carried out using ReCore. In ReCore, the problem of finding a side chain that fits between a scaffold and a protein binding pocket is simplified to a search for groups with a particular pharmacophore feature (for example, a hydrogen bond donor) in the correct orientation relative to the leaving bond vector (the direction of the bond which would connect the fragment to the rest of the molecule). In this screening, the oxygen introduced above using the OEChem Toolkit was specified as a hydrogen bond donor pharmacophore feature.

Screening with ReCore is fast because although the possible side chains are rotated relative to the core structure, there is no rotation about any of the bonds within the core or the side chain. FlexX was used to dock oligobenzamides with side chains that passed the ReCore screening process, as described in the next section.

2.2.3. DOCKING

2.2.3.1. FlexX

In initial work, fragments derived from ZINC database molecules were generated using software from BioSolveIT, ReCore[133] and FlexX[149] (version 1.0.2). In FlexX docking, a core fragment is first positioned and then this core is incrementally appended with side chains until the fully docked compound has been produced. Each addition is repeated and the highest-scoring poses are selected for the next stage. FlexX^{c}, an extension of FlexX was used to dock a library of compounds in the form of a core structure and possible side chains (generated using ReCore). FlexX uses an empirical scoring function[228]. FlexX scores mentioned herein are given in units of kJ mol^{-1}.

Initially, the three side chain positions were considered separately (one after the other). When each side chain was tested, the core docked by the software comprised the scaffold and the other two side chains. FlexX rotates this invariant, core component by a few degrees to generate an ensemble of structures on to which potential side chains can be joined. After each possible side chain has been attached to one of the core poses, the resulting structure is optimised to generate the final docking pose.

Poses were filtered using the software; only solutions where the root mean square deviation (RMSD) (p78) of the core and unmodified side chain atoms from their starting position was less than 2 Å were considered.

2.2.3.2. Autodock 4.2

Although FlexX was used initially, there is a deficiency in the FlexX scoring function, namely, its lack of terms to explicitly account for ligand and protein desolvation upon binding[229] (p130). Consequently, subsequent docking was performed using Autodock 4.2[151] and then, later, with Autodock Vina[152]. The top pose generated by Autodock 4 using the O1 scaffold starting compound (Figure 2.1) has all three side chains in precisely the positions of the Phe19, Trp23 and Leu26 side chains of p53, suggesting Autodock 4 docking is accurate (Figure 2.8, p129) and Autodock 4 does have terms to account for desolvation.

Autodock has no feature corresponding to the combinatorial docking of FlexX, which allows side chains to be attached to a core. Consequently, when using Autodock 4, side chains were attached using custom software (Side Chain Splicer, p105). The scaffold starting molecule (p101) was docked using standard docking (hereinafter referred to as "global docking"), the pose was exported, the side chains were swapped using Side Chain Splicer and then the binding position was optimised using Autodock (a process hereinafter referred to as "local docking"). The default pseudo-Solis and Wets local search method[156,230] was used for this optimisation.

Autodock 4 requires that a 3D search space be defined. Docking of the O1 starting compound using a box size greater than the size of the protein, allowing any possible binding position, did not yield poses around the p53 binding site. PDB structure 1T4F, from which the Mdm2 structure used for docking was extracted, comprises p53 bound to Mdm2. Using the cubic volume shown in Figure 2.5, which has sides of length 38.4 Å and is centred on the alpha carbon of Phe19 in the p53, did generate the expected binding pose of this molecule (Figure 2.8, p129). 128 grid points, 0.3 Å apart were used in each dimension. This is a bigger box than was used in earlier work by Fuller *et al.*[203] (22 x 22 x 18 Å).

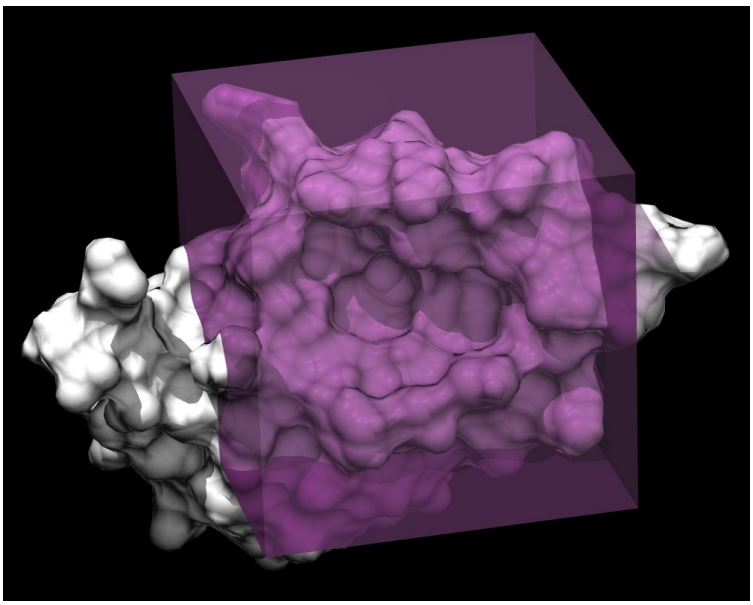


Figure 2.5: The volume of the Mdm2 structure from PDB 1T4F which was used for initial docking with Autodock 4. The cubic space shown has sides of length 38.4 Å and is centred on the alpha carbon of Trp23 in the p53 helix in structure 1T4F (not thrown).

2.2.3.3. Autodock Vina

Docking the O1 scaffold starting molecule (Figure 2.1) with Autodock Vina yielded the expected pose along with other plausible arrangements of the ligand.

25,000 compounds with different side chains were docked using Autodock Vina to investigate the effect of side chain properties on binding pose. Whereas it is impractical to dock a large number of compounds in this way using Autodock 4.2, necessitating docking of the scaffold, attachment of the side chains and then optimisation of the docking pose, the search algorithm of Autodock Vina is much faster (p63).

The results of the global docking revealed that oligobenzamides could theoretically bind to the Mdm2 structure in many different places. Consequently, because it enabled a wider search for possible binding positions, docking prior to simulations was, unless otherwise stated, performed using Autodock Vina.

Autodock Vina adjusts the amount of time spent searching for the best pose according to the number of atoms and number of degrees of freedom of the molecule being docked; however, it has a setting, the exhaustiveness, which can be used to change the balance between speed and search comprehensiveness if necessary. The makers of Autodock Vina do not reveal how this arbitrary setting is used by the program, only that the amount of time spent searching for the highest-scoring pose varies approximately linearly with the value set[231]. In this project, Autodock Vina was run with the exhaustiveness set to the default value of 8 unless otherwise stated. The appropriateness of this setting was evaluated as discussed in the results section (p186).

A search volume does not have to be defined with Autodock Vina and the program clusters poses by RMSD prior to output in an automated manner transparent to the end user[152]. (An RMSD cut-off does not have to be specified and no details of the structures prior to clustering are printed.)

2.2.3.4. Re-protonation of docking poses

To accelerate the docking process, Autodock 4 and Autodock Vina merge all non-polar hydrogen atoms (hydrogen atoms bonded to carbon) with their carbon atom, and treat each hydrocarbon unit (CH, CH_2 or CH_3) as a single, united atom. In the docking output, the united atom positions were treated as the positions of the carbon atoms and the non-polar hydrogens were added to these using Open Babel[217]. To add these protons, all hydrogen atoms were removed from the poses and then complete protonation was performed. With the exception of some renumbering of chemically identical hydrogen atoms on the same carbon, this method ensured consistent hydrogen atom numbering. This made it easier to transfer RESP charges to the resulting poses for molecular dynamics simulations as discussed in the next section.

2.2.3.5. Pose recognition

Autodock Vina automatically clusters poses prior to their output[152]. In this project, large numbers of compounds with different structures were tested, necessitating further sorting of all the docking results. Oligobenzamides were found to be able to bind in a disparate assortment of conformations so further clustering was not appropriate (p153). Instead, poses were computationally categorised based on the position of the side chains and rings of the scaffold in each molecule relative to the binding sites of p53 residues Phe19, Met20, Tyr22, Trp23 and Leu26 in the 1T4F crystal structure, the structure from which the Mdm2 structure used for docking had been extracted. Side chains were considered to be within a pocket if the centre of mass of the heavy atoms comprising the side chain (everything except for the hydrogen atoms) was within either 3 or 4 Å of the centre of mass of the heavy atoms of the p53 residue in question. (The cut-off was set on a case by case basis, as indicated in the results section.)

Hydrogen atoms were not used because 1T4F is a crystal structure so does not contain hydrogen atoms due to their small cross section.

Large cut-offs were used to reduce the chance of false negatives. In a small number of cases, a side chain centre of mass had a position such that it fell within the cut-offs of both of two adjacent binding sites. Poses for which this occurred were discarded prior to further analysis.

2.2.4. PARAMETERS

2.2.4.1. Parameters used for FlexX docking

FlexX docking was only performed with compounds based on the O1 scaffold (shown in Figure 1.6, p36). Rotation about the amide carbon to aromatic carbon bonds of the scaffold is possible because steric hindrance prevents the amide and aromatic ring atoms from lying in the same plane. Consequently, the amide and aromatic ring atoms do not form a large rigid conjugated system as is the case in benzamide[171]. In contrast, intramolecular hydrogen bonding means that rotation about the aromatic carbon to amide nitrogen bonds on the other side of the amino groups is discouraged. The FlexX torsion angle energies were modified so that they more accurately reflected these two effects (Table 2.1). No data were available for systems with an oxygen atom ortho to the amide group. Consequently, published parameters for the corresponding bond in thiomethyl acetaniline[171] (where sulphur replaces the oxygen) were used. Fuller *et al.*[172], who performed molecular dynamics simulations with these parameters, claim that this approximation is reasonable, citing the more recent work of Liu *et al.*, who have derived torsion parameters with both oxygen and sulphur and shown them to be similar.

Table 2.1: The AMBER force field torsion potential parameters for docking in FlexX

Torsion	PK /kcal mol^{-1}	Phase /°	N
ca-ca-n-c	2.617	0	3
	0.115	0	2
	4.507	0	1
n-c-ca-ca	2.075	0	2

These parameters are not entered into FlexX; they were used to calculate the torsion-associated potential energies at 10° intervals which were then used to modify the FlexX torsion_fine.dat file. In the AMBER force field, the torsion energies are modelled using cosine waves, each with a height (PK), phase and periodicity (N). The first bond (the green torsion in **Figure 2.6**A) is best described using three of these components. The second (the red torsion in **Figure 2.6**A) requires only one. IDIVF=1. In subsequent simulation work, both torsions were modelled using a single trigonometric term.

The new parameters generated by Liu *et al.*[173] were used for molecular dynamics simulations in this project, making the simulations here described the first O-ortho and O-meta-substituted trimeric oligobenzamide simulations to be performed with the correct torsion parameters for the bonds about which rotation is restricted due to intramolecular hydrogen bonding.

2.2.4.2. Parameters used for Autodock docking

Although torsion parameters were modified for FlexX docking and for molecular dynamics simulations, the Autodock 4.2 and Autodock Vina scoring parameters were not modified for the following reasons:

1. The parameters of Liu *et al.*[173] mentioned in the previous section had not been published at the time docking began.
2. Whereas FlexX was only used with O1 scaffold compounds, Autodock was used with all five scaffolds. Use of modified parameters for some scaffolds but not others could have introduced bias preventing the fair comparison of results.
3. The parameters supplied with the Autodock software had been rigorously validated for use with the Autodock software, unlike any parameters they would be replaced with.
4. Docking of the O1 scaffold starting molecule with Autodock 4.2 appeared to place it in a plausible pose with each side chain appearing to mimic the side chain of a p53 residue (Figure 2.8, p129).
5. Searching is faster in Autodock 4.2 if the torsion angles in the backbone are held rigid.
6. The end user is not expected to modify the parameter files of Autodock Vina.

The torsion angles were fixed when docking with Autodock 4 (but not Vina) in the cis conformation so the side chains were on all on one side of the molecule. This is the same procedure which was used by Fuller *et al.*[203].

2.2.4.3. Parameters used for molecular dynamics

Difficulties with synthesis of the target molecules meant that only a few compounds were available for the validation of computational results. Consequently, as the project progressed, the focus shifted from the generation a large number of estimated binding affinities to the production of a smaller number of more accurate predictions, necessitating the simulation of oligobenzamides (discussed in the next section).

Parameters for simulations came from the General AMBER Force Field (GAFF) (version 1.4)[226] with the exception of some torsion parameters. The torsion parameters for the bonds connecting the aromatic rings in the scaffold (three: ca-n, n-c and c-ca where ca represents aromatic carbon and c and n represent the sp^2-hybridised carbon and nitrogen atoms of the amide group respectively) were set to previously published AMBER force field parameters (in the case of the amide bond[171] and torsions affected by intramolecular hydrogen bonding[173]) or (in the case of the ca-n and c-ca torsions in the absence of intramolecular hydrogen bonding) were calculated from previously published torsional barrier heights[175]. The parameters for the latter barrier heights were calculated by subtraction of the total increase in AMBER 1-4 electrostatic and 1-4 van der Waals energies associated with going from the bottom of the barrier to the top of the

barrier from the torsional barrier height. The three torsion angles are highlighted on an oligobenzamide in Figure 2.6A. The barrier heights were for 4-(4-aminobenzamide)benzoic acid (Figure 2.6B). The 1-4 interaction energy (the sum of the electrostatic and van der Waals components between atoms at either end of the torsion angle) needs to be subtracted because it will be calculated and added on to the torsional energy calculated from the torsion parameters in Amber when the simulation is run, and the final barrier height in the simulation needs to equal the published barrier height.

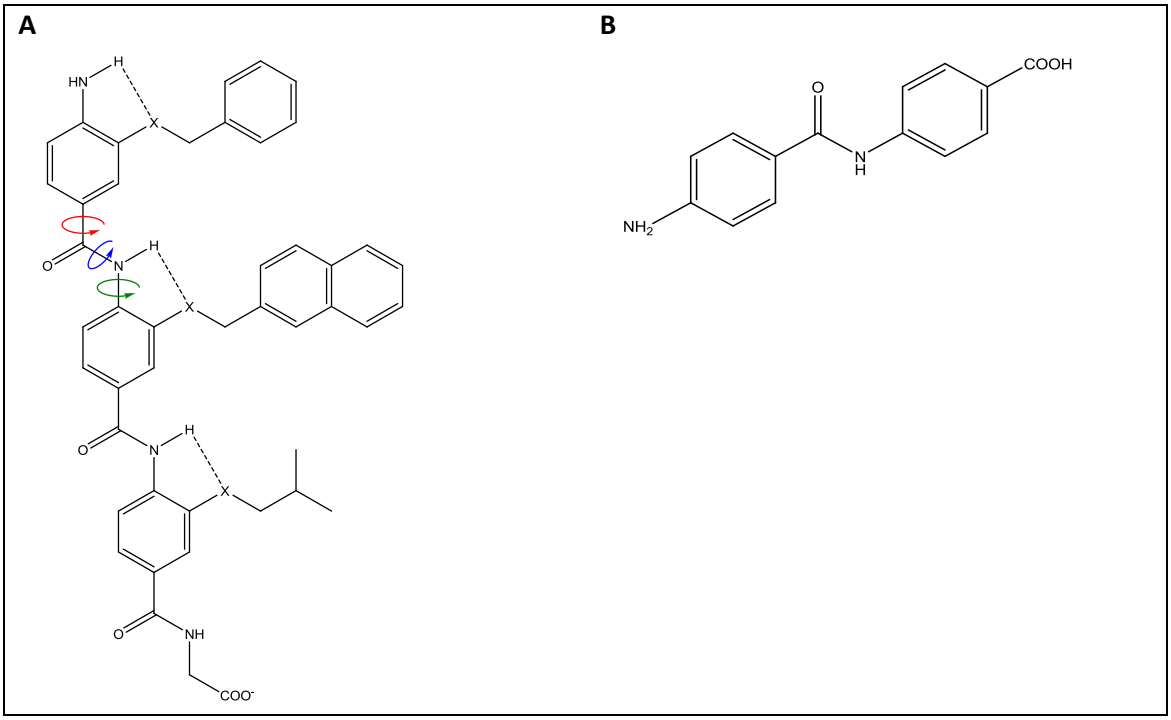


Figure 2.6: Structures relevant to the calculation of the oligobenzamide scaffold torsion parameters used in this project. A) An oligobenzamide with the bonds for which the torsion energies were modified indicated with coloured arrows. The colours correspond to those shown in the header row of Table 2.2. B) 4-(4-aminobenzamide)benzoic acid, the compound used as a model for bonds not affected by intramolecular hydrogen bonding when calculating parameters.

The 1-4 interaction energies for 4-(4-aminobenzamide)benzoic acid were calculated at 10° intervals (Figure 6.4, p310) using Amber. At each 10° point, the bond being rotated was fixed and the rest of the structure was minimised by steepest descent minimisation. Every point from 0° to 350° was actually tested for both of the ca-n and c-ca bonds because, despite the apparent partial symmetry of the molecule (Figure 2.6B), which suggests that the interaction energy should only need to be measured for torsion angles from 0° to 90°, in practice some variation was observed for the n-ca bond energy across the whole torsional space. The carboxylic acid group on the end of the molecule was left protonated, making the compound asymmetrical, and the use of steepest descent minimisation meant that the 180° flip of this terminal carboxylic acid necessary to completely minimise the energy in some cases was not possible. Where the energies for two torsion angles differed even though the symmetry of the molecule suggested

that they should be identical, the minimum energy of the two was taken, as would have been obtained using conjugate gradient minimisation (which would have found the global as opposed to the local minimum).

Table 2.2: The oligobenzamide torsional barrier heights used for molecular dynamics (in kJ mol^{-1})

Scaffold	Torsional barrier height /kJ mol^{-1}		
	ca-ca-c-n	ca-c-n-ca	c-n-ca-ca
N1	29.2[a]	70.8[b]	26.0[a]
O1	29.2[a]	70.8[b]	36.0[c]
S1	29.2[a]	70.8[b]	42.7[c]
O2	36.0[c]	70.8[b]	26.0[a]
S2	21.8[c]	70.8[b]	26.0[a]
GAFF	121.3	83.7	15.1

The barriers are shown for three key bonds found in the scaffold. The torsions are described using GAFF atom types. "ca" represents aromatic carbon and "c" and "n" represent sp^2-hybridised carbon and nitrogen respectively. These torsion angles are indicated by coloured arrows on the oligobenzamide in Figure 2.6A above, where the colour of the arrows corresponds to those used in the header of this table. Five different oligobenzamide scaffolds were tested, the N1, O1, S1, O2 and S2 scaffolds. (See Figure 1.6 (p36) for their structures.) The torsion energy parameter entered into Amber (PK), which was specified for the whole bond (IDIVF=1), is half the barrier height. There are two peaks per 360° rotation (N=2) and 0° is a minimum so the phase is 0. The 29.2 and 26.0 kJ mol^{-1} values are the result of subtraction of the 1-4 interaction energy from published barrier heights. For details of these calculations see Table 6.5 (p310). The bottom row of the table, labelled "GAFF" indicates the parameters for the bonds in the GAFF force field (version 1.4)[226]. Most notable is the difference between the ca-ca-c-n barrier height for oligobenzamides and the 121 kJ mol^{-1} GAFF barrier height for this bond, which is the result of steric hindrance in oligobenzamides rotating the amide carbonyl group out of the plane of the adjacent benzene ring[173]. References: a[175] b[171] c[173].

Table 2.2 shows the final barrier heights used. Parameters labelled with superscript "a" were produced by subtraction of the 1,4-interaction energy and the calculation is shown in Table 6.5 (p311). The PK torsion parameters (which determine the height) are half the barrier heights shown in Table 2.2 (but in kcal mol^{-1}, as required by Amber). For these torsions, there are two peaks per 360° so N is 2. As the minima are at 0° and 180°, the phase parameter is 0.

Figure 1.6 in the introduction (p36) shows the structures of the five oligobenzamide scaffolds investigated in this project. Synthesis was considerably more successful with scaffold N1 than scaffold O1. Consequently, it was important to perform molecular dynamics simulations using the N1 scaffold. In the N1 scaffold, the side chains are attached directly to the nitrogen in the core of the molecule. Consequently, N1 scaffold oligobenzamides with different side chains could potentially have different backbone torsion parameters. As discussed in the introduction (p70), because the side chains to be tested comprised both inductively electron withdrawing and

inductively electron donating side chains, it was assumed that they did not affect the backbone torsion parameters of this scaffold.

While the previously published AMBER parameters and the aforementioned barrier heights (used as a necessity due to the absence of published parameters) were originally validated with experimental data, rigorous validation of the parameters used in combination, which was considered beyond the scope of this PhD, would ideally be performed prior to their extensive use in future work. The results generated using these parameters in this project were consistent with experimental results determined by fluorescence polarisation (supplied by Kerya Long, University of Leeds) as discussed in the results section (Figure 2.22).

With regard to the side chains, the Amber[224–226] parmchk program was used to generate any undefined side chain torsion parameters not explicitly given in the GAFF force field. It does this automatically either by identifying analogous parameters from the GAFF force field and copying them or, in some cases, by empirical estimation. The results of the program were checked to verify that all of the necessary parameters had been located or calculated for every molecule.

Protein parameters came from the ff03 force field[232]. The protein was used with Gasteiger partial charges[176].

2.2.4.4. Calculation of small molecule RESP charges

RESP partial charges (p70) were calculated for oligobenzamides (as a whole molecule) using RESP ESP Charge Derive (RED) (Version II.3 or III.3)[177]. Structures were minimised using Gaussian 03 (release D.02)[233] or Gaussian 09 (release D.01)[234] using the AM1 method prior to calculation of the charges. The charges of the Hartree-Fock molecular electrostatic potential grid points were determined using the 6-31G* basis set.

RESP was used to determine atom-centric charges using the Connolly surface algorithm[235] with restraint weights of 0.0005 and 0.001. The partial charge calculation was performed such that chemically identical atoms were assigned equal charges. Identical atoms were identified using custom software. (This software uses Babel (OpenBabel) to produce canonical Daylight SMILES[216] for molecules, strings describing the connectivity of molecules, which are independent of the atom order in the starting structure and which can therefore be used to identify identical molecules.)

When side chains were swapped using Side Chain Splicer (p105), it was necessary to calculate partial charges for the new molecule. It was often desirable to retain the coordinates of the atoms but minimisation was required for charge generation. To solve this problem, a custom script was used to transfer charges from the minimised molecule, with charges assigned by

RED, to the same molecule with different coordinates, for example, the original, posed molecule prior to minimisation.

2.2.5. MOLECULAR DYNAMICS

2.2.5.1. Creation of molecular dynamics parameter files

Before molecular dynamics simulations can be carried out in Amber, a coordinate file and a parameter file have to be created for the complete system. The coordinate file contains the starting coordinates of the atoms. The parameter file contains molecule-specific parameters based on the atom types present in the molecule and their connectivity. The coordinate and parameter files were generated using the tleap program of the Amber suite.

The protocol used for production of parameter files using tleap:

1. The ff03 and GAFF force field parameters were loaded along with any additional parameters for the ligand generated by the Amber parmchk program.
2. The modified torsion parameters (discussed above) were loaded, overriding those previously set.
3. The protein was added to the system followed by the docked ligand.
4. Cubic periodic boundary conditions were defined and the resulting cube was filled with TIP3P water[236,237]. The sides of the cube were orthogonal to those of the x, y and z axes and 10 Å from the nearest atom. In the TIP3P water model, the hydrogen-oxygen bond lengths are fixed at 0.9572 Å and the bond angle is fixed at 104.52°[236].
5. If the system was negatively charged, Na^+ ions were added until it was neutral. If the system was positively charged, Cl^- ions were added until it was neutral. (The addions2 function was used, which causes the water added in the previous step to be considered when the lowest energy positions for the ions are calculated.)
6. The coordinate and parameter files were saved.

If parameters were required for the system lacking water and ions or for just the protein or ligand in isolation (for MM-PBSA or MM-GBSA calculations) then the above procedure was performed with the relevant steps omitted.

2.2.5.2. Parameter files for thermodynamic integration

The creation of parameter files for TI calculations is complicated by the requirement that all atoms except those differing between the two systems being compared have identical coordinates. As outlined in Figure 2.7 below, tleap was run 3 times per pair of compounds being compared. The initial run created a PDB file containing water and ions with atoms positioned such that they did not clash with the atoms of either of the two posed ligands. Two subsequent tleap runs were then performed as outlined by the numbered points above but with the solvent

addition step replaced by loading of the solvent and ion-containing PDB followed by manual setting of the periodic box dimensions.

All simulations were performed in Amber 10[224–226] using TIP3P explicit solvent. Solvent molecules were removed prior to MM-PBSA and MM-GBSA calculations using the Amber ptraj program. This program was also used to translate atoms into the instance of the periodic box centred on the origin.

2.2.5.3. Initial simulations

Initial simulations of a couple of oligobenzamides (the N1 and O1 starting compounds shown in Figure 2.1, p102) were carried out to see simulations started from docking poses of molecules were stable. These were carried out using the PMEMD program of Amber 10 as follows. Up to 2500 cycles of minimisation were performed, of which the first 1500 were steepest descent, the rest being conjugate gradient. The system was then heated from 100 K to 300 K over 50 ns of constant volume simulation. Then, a 5 ns constant temperature and constant pressure simulation was performed. The Berendsen weak thermostat[166(p383)] was used throughout. For details of this thermostat and the reasons it was used in all work except for thermodynamic integration see p81. A time step of 2 fs was used to speed up the simulation process. SHAKE[182] (p77) was used to control all bonds involving hydrogen. Slowly varying terms were calculated every two steps during all constant volume equilibration simulations in this project (by setting the Amber nrespa setting to 2), but on every step in constant pressure runs for more accurate simulation unless otherwise stated.

2.2.5.4. Simulations for MM-PBSA and MM-GBSA calculations

2500 cycles of minimisation were performed in Amber 10 (particle mesh Ewald version): 1500 of steepest descent followed by 1000 of conjugate gradient minimisation. Over 10 ps of constant volume simulation, the system was heated from 100 to 300 K before constant pressure simulation began at 300 K. 250 ps was allowed for equilibration of compounds as discussed in the results section (p165). When investigating changing all three side chains on the O-linked scaffold, data collection was carried out over just 250 ps following equilibration. However, 6 repeats were carried out per compound so each MM-PBSA result was based on snapshots over 1.5 ns of simulation. All other simulations analysed using implicit solvent methods were run for 500 ps following the first 250 ps of constant pressure equilibration. SHAKE[182] was used with a time step of 2 fs to control all bonds involving hydrogen.

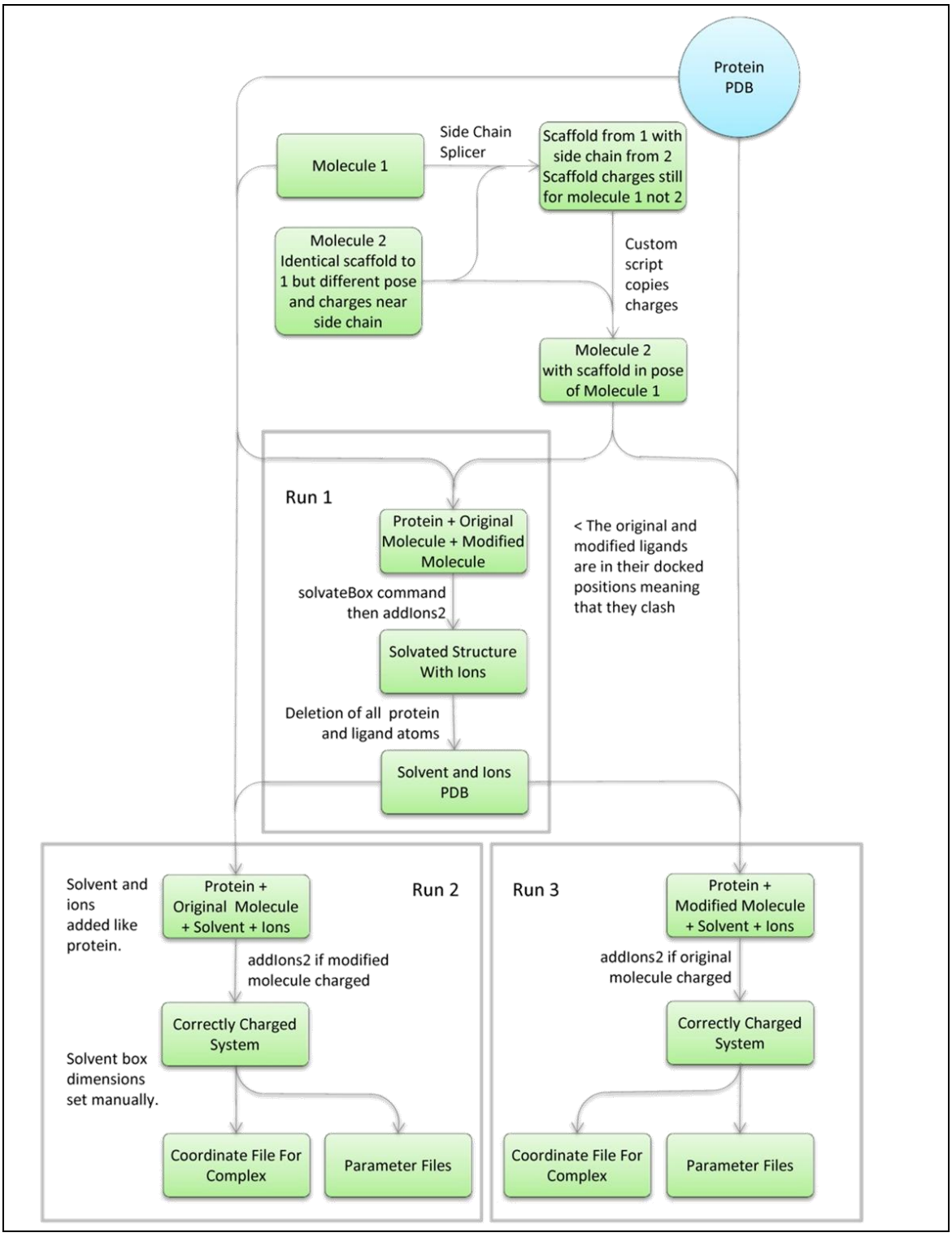


Figure 2.7: Preparation of coordinate and parameter files for TI simulations in Amber with two molecules of different structure. Two systems are required differing by only a few atoms. The original ligand (molecule 1, in its docking pose) is modified to molecule 2 by way of the side chain which differs between the molecules being swapped using the Side Chain Splicer software (p105). Charges (originally calculated with the unbound, minimised molecule) are then copied from the minimised instance of molecule 2 to the splicing-derived molecule 2 which has the scaffold in the same pose as in molecule 1. Three separate runs of the Amber tleap program are carried out (indicated by the grey regions). So that equilibration is fast, solvent and ions (K^+ and Cl^-) for neutralisation are placed in positions where they will not clash with either the original or modified ligand. A PDB file is produced containing the solvent and aforementioned ions which is then added to the protein and desired ligand when LEaP is run to generate the parameter and coordinate files.

2.2.5.5. Thermodynamic integration simulations

Thermodynamic integration (TI) was used to estimate differences in binding affinity between molecules with small differences in structure. O1 and S1 scaffold oligobenzamides with the same side chains were compared, as were O2 and S2 scaffold molecules. Compounds with the same scaffold but different side chains (for example due to the presence of an additional methyl group, halogen atom or alcohol group in one in one of the pair) were also compared. Exact details of the oligobenzamides compared can be found in the results section.

Up to 15,000 cycles of minimisation were carried out prior to TI simulations. These cycles were steepest descent, the only form of minimisation supported by Amber when TI calculations are carried out. The maximum number of minimisation cycles permitted was increased from its default value (to 15,000) because a large number of simulations were performed, increasing the chance of any one simulation failing and increased minimisation ensured that every subsequent simulation was stable. In practice, the maximum number of cycles was often not performed because (as in all Amber minimisation carried out in this project) a gradient convergence cut-off of 50 cal mol^{-1} $Å^{-1}$ was used, which is significantly less strict than the default convergence of 0.1 cal mol^{-1} $Å^{-1}$. Raising this cut-off did not appear to affect the ability to start stable simulations. The distance cut-off for direct space calculations of through-space forces between non-bonded atoms was increased from the default value of 8 Å to 12 Å during minimisation (but not in the subsequent simulations).

TI was carried out using the sander program of Amber10 as opposed to the pmemd program, which uses the Particle Mesh Ewald method of charge calculation. (pmemd does not support TI in Amber 10.) A total of 50 ps of constant volume simulation was used to heat the system from 100 K to 300 K and then 150 ps of constant pressure simulation was allowed for equilibration prior to data collection. $d\mathcal{V}/d\lambda$ was then recorded after every 500 steps. Langevin dynamics (p81) with a collision frequency of 5 $(ps)^{-1}$ was used to maintain the temperature.

Transformations were performed in three steps as described below (p121). For the removal and addition of charges (steps 1 and 3), a time step of 2 fs was used with SHAKE[182] to control all bonds involving hydrogen atoms. In the main transformation step (step 2), SHAKE was turned off and a time step of 1 fs was used. Soft core potentials (p85) were used in this middle step to increase simulation stability[238].

2.2.6. THERMODYNAMIC INTEGRATION CALCULATIONS

2.2.6.1. Integration

TI allows the difference in binding free energy between two compounds to be predicted.

As discussed in the introductory chapter (p83), this is the difference in energy between the bound systems, where the compounds are on the protein surface, minus the difference in energy between the unbound systems where the compounds are free in the solution (Equation 1.64, p84).

It is assumed that the difference in binding free energy, ΔΔG, is equal to the difference in vibronic potential energy, $\mathcal{V}$. Calculation of the difference in potential energy, $\Delta\mathcal{V}$, between the systems containing each compound (either free or bound to the protein) involves the determination of $d\mathcal{V}/d\lambda$ at points along an alchemical reaction coordinate from 0 (the starting compound) to 1 (the product). Integration under the $d\mathcal{V}/d\lambda$ curve from 0 to 1 yields the total change in potential energy (p83).

In this project, Gauss-Legendre integration was used to calculate the area under the $\mathrm{d}\mathcal{V}/\mathrm{d}\lambda$ curve as it is more accurate than using the trapezium rule[239]. This method requires the use of irregularly spaced lambda points (points corresponding to the Gauss-Legendre abscissas).

Gauss-Legendre integration is performed using points between x=-1 and x=1 but lambda values range from 0 to 1 so

$$\int_0^1 \frac{\mathrm{d}\mathcal{V}}{\mathrm{d}\lambda}\mathrm{d}\lambda = \frac{1}{2}\sum_{i=1}^{\infty}\left(w_i\left(\frac{d\mathcal{V}}{d\lambda}\right)_{\lambda_i}\right) \qquad 2.2$$

Where

$$\lambda_i = \frac{x_i + 1}{2}. \qquad 2.3$$

w_i is the weight associated with Gauss-Legendre abscissa x_i. (The number of lambda points used is discussed at the bottom of the next section.)

2.2.6.2. Transformation steps

For each O-linked and S-linked scaffolds (scaffolds O1, O2, S1 and S2), the torsional energy parameters for the dihedral about which rotation was restricted due to intramolecular hydrogen bonding were changed in TI runs from 0.5 kcal below their normal position to 0.5 kcal above. This enabled assessment of the importance of dihedral parameter accuracy in predicting the binding energy of oligobenzamides. This transformation required a single step.

TI was also used to alchemically change each sulphur atom in S-linked compounds to a methanediyil (-CH_2-) group and to do the same in O-linked compounds. These transformations were carried out in three steps. In the first, partial charges were removed from the disappearing

atoms (most notably the oxygen and sulphur). In the second step, the disappearing atoms were decoupled from the system and the new atoms (most notably those of the methanediyl group) were introduced and in the final step, partial charges were put on the new atoms. The atom type of some hydrogen atoms close to the oxygen or sulphur in the starting compound changed when the transformation to a methanediyil group was performed. The parameters of these atoms were changed by removal and replacement of these atoms along with the oxygen or sulphur in the aforementioned steps.

The effect of side chain modification was also assessed using TI investigations. Three steps were used unless no atoms were being added, in which case, no charge addition step was needed. When the side chain attachment atom or torsional barrier height was changed, 11 lambda points were used for removal and re-addition of charges and 21 lambda points were used for step 2, the main transformation step. When the side chains were modified, 8 lambda points were used for every step.

Where 21 lambda points were used, the Gauss-Legendre abscissas for n=10 and n=11 were used (with the weights halved) as opposed to the abscissas for n=21 so that the reaction coordinate (λ) was always between 0.01 and 0.99, as is necessary when soft core potentials are used in Amber[238].

2.2.6.3. Bias due to one-way conversion

When side chains were modified using TI, many conversions were only performed in one direction.

Given the large number of conversions which it was desirable to investigate and the large number of simulations which are required for each conversion (repeats for each lambda point both for the free ligand and the complex for each of the three steps outlined above), it was necessary for simulations to be run which equilibrated rapidly. This was achieved by ensuring there were no steric clashes at the start of any simulation.

Firstly, care was taken to ensure that water molecules were positioned such that they were clear of the ligand in both the original system and the modified system. (For further details see p117.)

Secondly, the size of the ligand was never increased by the addition of heteroatoms because this could create ligand atoms in space already occupied by the atoms of the protein. Although hydrogen atoms were added, compounds were usually modified to a side chain of a similar size (for example by converting one halogen atom to another) or modified by removal of atoms, a reduction in side chain size. This meant that where the molecules being compared were different sizes, the conversion was performed in one direction only. This introduces an intrinsic bias into the assessment of which compound is the more stable of two because the compound used to

generate the binding pose for both compounds (as described above) will be in its strongest binding pose, unlike the compound produced by modification. As a consequence of this bias, the side chain conversion TI results in this chapter should ideally be used for the comparison of similarly sized ligands with a common core. For example, the results for removal of a methyl group from the ortho position of a benzene ring could be compared to those of a methyl group being removed from the para position, but each individual result might not provide an accurate indication of whether the methyl group in each case improves the affinity or not; the original molecule (with the methyl group) is likely to be favoured in both cases, because this was used for the docking. Fortunately, there was no obvious evidence of this bias, which would have led to underestimation of the affinity of compounds with smaller side chains relative to those with larger side chains, in the final predictions when compared to fluorescence anisotropy results (p152).

2.2.7. MM-GBSA AND MM-PBSA

MM-GBSA and MM-PBSA predictions were made using from Amber 10 trajectories using Amber 11 because MM-PBSA processing in Amber 10 is limited to the single trajectory approach described below. Trajectory frames at intervals of 10 ps were extracted for analysis.

2.2.7.1. Single, double and triple trajectory methods for calculation of binding free energies

To predict the free energy of binding for a ligand, it is necessary to compare the energy of the free protein and ligand with the energy of their bound state. While simulations of the free ligand, free protein and complex can each be performed (a multi-trajectory approach), more accurate results can often be obtained using a single-trajectory approach, where the trajectories of the free protein and free ligand are extracted from the trajectory of the complex[240–243]. If the ligand and protein explore the same conformational space when free as when bound together then this method reduces noise because it ensures that the ligand does explore the same phase space in both the free and bound states.

The suitability of the single-trajectory method depends on the extent to which the ligand and protein have similar conformations when free and when bound. The p53 binding site of Mdm2 is very flexible and changes conformation upon binding. Furthermore, some oligobenzamides might assume a cis conformation (across the amide groups) when bound to Mdm2 so that all of the side chains are in contact with the protein but a trans arrangement when they are free in solution to spread out the side chains[244]. However, in the O1, S1, O2 and S2 scaffolds there is intramolecular hydrogen bonding which will help maintain the cis conformation in the free ligand and compounds based on the N1 scaffold can be, for steric reasons, more stable in the cis than the trans arrangement when unbound[245–247]. The uncertainty regarding the best procedure warranted testing of a variety of methods to test the robustness of the predictions.

In this project, single (Equation 2.4), double (2.5) and triple (2.6) trajectory methods were tested, as indicated in the results section. In the double trajectory method, a separate simulation of the ligand in isolation is performed, but the trajectory of the free protein comes from that of the complex. In the triple trajectory method, the free protein energies from the double trajectory method results were averaged over all of the ligands and the result was used as the energy of the free protein for every ligand. To reiterate, the same free-protein energy was used for every ligand so it had no bearing on the ranking of the compounds, as would have been the case had a protein-only trajectory been performed.

$$\Delta G_{ST} = \Delta G_{complex} - \Delta G_{protein\ (complex\ trajectory)} - \Delta G_{ligand\ (complex\ trajectory)} \quad \textbf{2.4}$$

$$\Delta G_{DT} = \Delta G_{complex} - \Delta G_{protein\ (complex\ trajectory)} - \Delta G_{ligand\ (ligand\ trajectory)} \quad \textbf{2.5}$$

$$\Delta G_{TT} = \Delta G_{complex} - \overline{\Delta G_{protein\ (complex\ trajectory)}} - \Delta G_{ligand\ (ligand\ trajectory)} \quad \textbf{2.6}$$

2.2.7.2. Parameters

MM-PBSA and MM-GBSA calculations were carried out using the mmpbsa.py program of Amber 11 (AmberTools 1.5)[224–226]. However, parameter files for the complex, ligand and protein in the absence of solvent and ions (necessary for the MM-PBSA and MM-GBSA analysis) were generated using Amber 10 (AmberTools 1.4). (AmberTools 1.4 can only make predictions using the single trajectory approach. AmberTools 1.4 was used to generate the parameter files to ensure that the parameters used for implicit solvent calculations were identical to those used in the simulations (performed before AmberTools 1.5 had been installed).)

Parameter files were generated using tleap (Amber 10). Prior to loading trajectories into mmpbsa.py, solvent and ions were removed from the trajectory of the complex and atom positions were moved into the periodic box centred on the origin using the Amber 10 ptraj program.

Unless otherwise stated, default parameters were used for MM-PBSA and MM-GBSA analysis. One setting altered in MM-PBSA analyses was the required convergence of the energy gradient in minimisation (the drms setting) which was reduced from 0.001 to 0.0001 kcal mol^{-1} $Å^{-1}$ because sufficient computational power was available.

2.2.7.3. Buried atom radii

One of the MM-GBSA settings experimented with was the igb setting. This controls the Generalised Born model used. When set to 1, the original model described by equations 1.82 to 1.84 in the introductory chapter (p89) is used. When set to 5, a modified model is used in which the Born radii of the atoms is increased for buried atoms to account for their radius usually being underestimated within macromolecules[200]. This is described by equations 1.85 and 1.86 (p89) where the coefficients α, β and γ are 1, 0.8 and 4.85 respectively[186(p49),200]. For method 1,

the atomic radii of Tsui and Case[248] were used (PBRadii=mbondi) and for method 5, the radii of Onufriev *et al.*[200] were used (PBRadii=mbondi2).

2.2.7.4. Normal mode analysis

Normal mode analysis was used in some cases to estimate the change in entropy of the protein and an oligobenzamide ligand upon binding. Normal mode analysis involves prediction of the possible ways in which the ligand, protein and complex can resonate (p90). It was performed using the nmode program of Amber 10, which is called by mmpbsa.py when normal mode analysis is turned on. When enabled, Normal mode analysis was performed on every third frame processed in MM-PBSA and MM-GBSA analysis. Often the entropy change of the protein and compound upon binding is ignored because normal mode analysis is computationally expensive and its inclusion can introduce significant error because the calculated entropy often varies considerably between frames[249].

2.2.8. STATISTICAL METHODS

Extensive use of statistical methods was made to analyse docking and simulation results in this project. Pearson's correlation coefficient[250] and Spearman's rank correlation coefficient[251] were used to investigate the correlation between predicted binding energies and experimental results; some docking results were analysed using logistic regression[252] as detailed in the results section, and binding energies predicted using implicit solvent methods were analysed using analysis of variance (ANOVA)[215]. For full details of the statistical methods used in this project, see Appendix A (p290).

2.2.9. APPLICATION OF THE METHODS TO THE DATA

The methods sections above describe:

1. Preparation of the structure of Mdm2 from PDB structure 1T4F.
2. Production of oligobenzamides with different side chains.
3. Calculation of the properties of library members and their side chains.
4. Docking of oligobenzamides into Mdm2.
5. Modification of the side chains on docked poses *in situ* followed by re-docking (optimisation of the pose) of the new molecule.
6. Analysis of how the docking poses of library members depend on the properties of the molecules in question (regression).
7. Calculation of the atomic partial charges on selected library members.
8. Simulation of these library members bound to Mdm2.
9. Prediction of binding affinities using TI and the implicit solvent methods, MM-PBSA and MM-GBSA.

10. Comparison of TI and implicit solvent results with affinities measured in the laboratory using a fluorescence anisotropy assay.
11. Analysis of the implicit solvent results using analysis of variance (ANOVA).

Figure 6.2 (p306) shows in more detail how these methods were combined to generate the results discussed in the subsequent results section. The results section describes specifically what data sets the methods were applied to. This is also depicted in Figure 6.3 (p308) and can be summarised as follows.

Result processing in this project entailed:

1. Filtration of compounds from the ZINC database and their cleavage to produce side chains.
2. Attachment of these side chains to the O1 scaffold starting molecule (Figure 2.1, p102) (previously docked with Autodock) using ReCore and docking of the resulting molecule in each case using FlexX.
3. Analysis of the resulting docking scores, using Spearman's rank correlation coefficients to investigate how side chain properties affected FlexX docking scores.
4. Simulation of a few oligobenzamides based on the O1 and N1 scaffolds to see if the simulations were stable.
5. Generation of 3D structures for side chains suggested by the collaborating chemistry group (supplied as formulae) and the splicing of the minimised side chains on to the posed scaffolds of previously docked starting molecules (Figure 2.1, p102).
6. Global docking of 25,000 compounds with Autodock Vina and analysis of how their properties affected their predicted binding poses using logistic regression.
7. From the poses of these 25,000 compounds, extraction of
    a. The top poses of selected compounds for thermodynamic integration and their modification to produce the second of the pair of molecules to be interconverted in the TI investigation.
    b. Poses for 31 previously synthesised compounds based on the N1 scaffold for binding affinity prediction using the MM-PBSA and MM-GBSA implicit solvent methods.
    c. Poses for two libraries of molecules, one with 285 (15 × 19) combinations of side chains at positions 1 and 2 and one with 285 combinations of side chain at positions 2 and 3, for analysis using the MM-PBSA method.
8. Local docking (optimisation) of each of a combinatorial library of compounds based on the O1 scaffold comprising compounds with 12 different side chains in every combination at the three side chain positions. (These compounds are already in a

conformation close to the correct pose because they were produced by modification of the side chains on a previously docked starting molecule.)

9. Calculation of RESP charges for selected library members following minimisation and the transfer of these charges onto docked poses of the compounds ready for molecular dynamics simulations.
10. TI calculations to investigate the effect on binding affinity of changing the side chain attachment atoms and varying the backbone torsion parameters.
11. TI calculations for compounds based on the N1 scaffold and their use to predict the relative binding affinities of the 31 compounds synthesised and tested in the laboratory.
12. Prediction of binding affinities using the MM-PBSA and MM-GBSA methods for these 31 compounds, either using a trajectory of the complex (the single trajectory approach), a trajectory of the complex and a trajectory of the free ligand (the double trajectory approach) or a triple trajectory approach (p124).
13. Comparison of the predicted binding energies for these 31 compounds with experimental results from a fluorescence anisotropy assay.
14. Prediction of the binding energies of each member of the library of O1 compounds mentioned in point 8 above using the MM-PBSA method and analysis of the results using analysis of variance (ANOVA) using the side chains at each of the three side chain positions as factors.
15. Prediction of affinities for the library of compounds mentioned in point 7c above using the MM-PBSA method and analysis of the results using three separate sets of ANOVA investigations. As described in the results section, only poses corresponding to one of four binding positions (shown in Figure 2.23, p187) were selected. Each analysis involved three fixed factors: the choice of side chain at the middle position, the choice of side chain at either the first or last position and, as the third factor, either the scaffold choice or the binding position.

2.3. Results and Analysis

2.3.1. DOCKING OF SIDE CHAINS DERIVED FROM MOLECULES OF THE ZINC DATABASE ALLOWED FAVOURABLE TRAITS IN THE PHE19, TRP23 AND LEU26 BINDING POCKETS TO BE IDENTIFIED

This section describes the docking of oligobenzamides with a diverse range of side chains using FlexX and the subsequent analysis of side chain properties and docking scores. The aim was to identify the side chain properties likely to increase the affinity of an oligobenzamide binding with its side chains in the sites usually occupied by p53 residues Phe19, Trp23 and Leu26 in the p53-Mdm2 interaction.

An initial oligobenzamide binding position was required for high-throughput screening to identify potential side chains. The O1 scaffold starting compound (Figure 2.1, p102) was docked into Mdm2 from PDB structure 1T4F[37] using AutoDock 4.2 as detailed in the methods section (p109). The side chains of the compound resemble the side chains of Phe19, Trp23 and Leu26 from the p53 helix. Comparison of the highest scoring pose obtained (Figure 2.8B) with the position of the p53 helix bound to Mdm2 in structure 1T4F (Figure 2.8A), which played no part in the docking, shows that the side chains have been positioned close to their expected positions (Figure 2.8C). The heavy atom RMSDs for the first side chain, second side chain terminal benzene ring and third side chain relative to the p53 residue side chains are 2.09 Å, 1.24 Å and 1.66 Å respectively.

Molecules from the ZINC source database were cut using ReCore (BioSolveIT) as described in the methods section (p106). These were spliced on to the docked O1 starting molecule (Figure 2.8B) and screened using ReCore. 536, 505 and 508 possible side chains were identified by this high-throughput screening for the first, second and third side chain positions on the scaffold respectively. FlexX^{c} (part of FlexX (BioSolveIT)) was then used to carry out combinatorial docking. This software docks the scaffold and then attaches and tests each member of a library of possible side chains. The side chains are attached *in situ* and the resulting pose is minimised to complete the docking process.

The side chain attachment positions were treated separately; each of the possible side chains was tested at each attachment position while the other two side chains were left unchanged from those of the initial structure. (For further details see p109.) Poses with an RMSD of greater than 2 Å relative to the high-throughput screening pose generated by ReCore, in which the scaffold position is unchanged from in the Autodock pose, were discarded. Consequently all of the poses resemble that shown in Figure 2.8B where the N-terminal, central and C-terminal side chains are in the pockets usually occupied by p53 residues Phe19, Trp23 and Leu26.

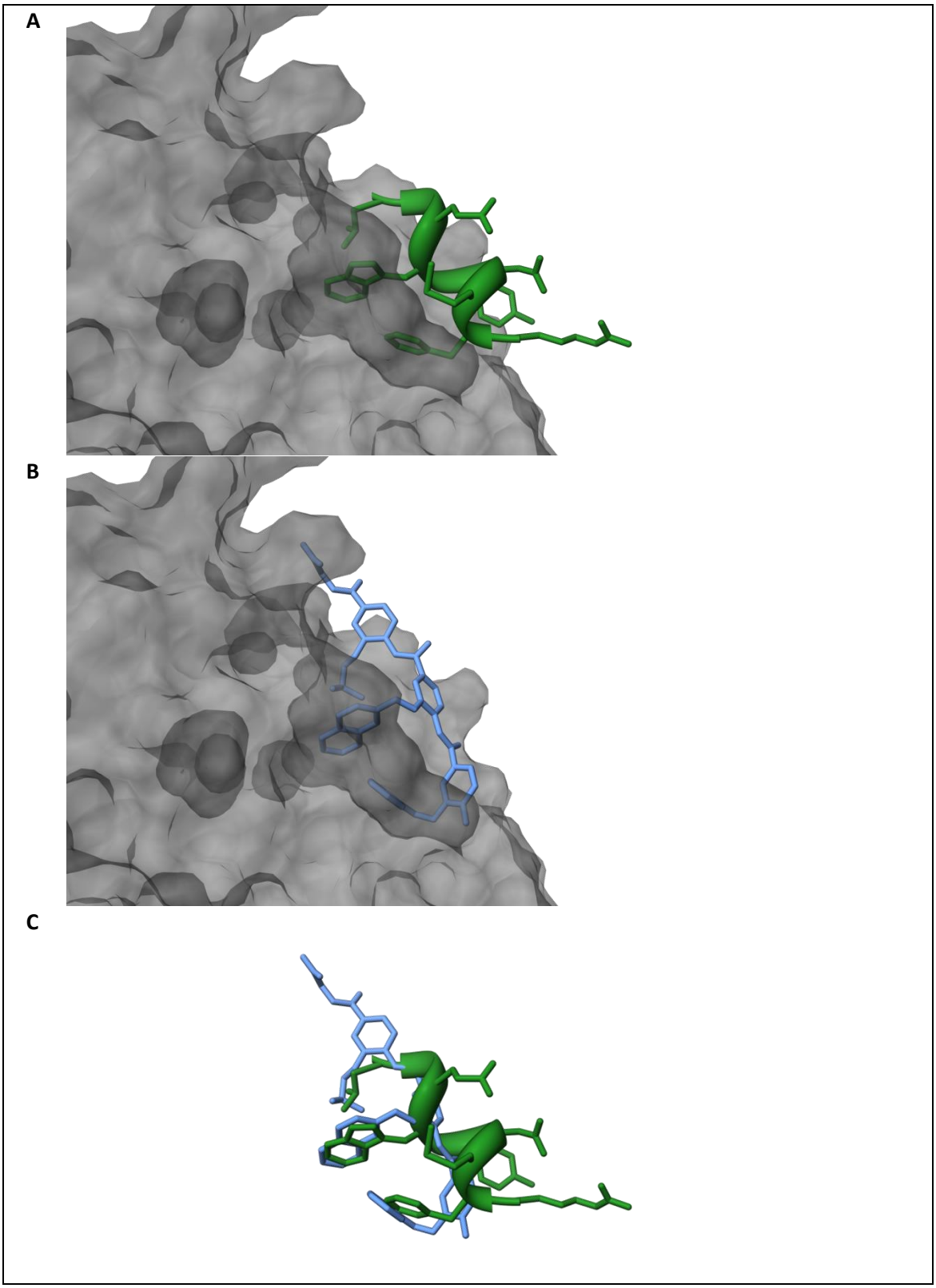


Figure 2.8: The best pose resulting from the docking of an oligobenzamide into Mdm2 from PDB structure 1T4F superimposed on the structure of p53 from the PDB structure. A) Structure 1T4F showing a modified p53 peptide (green) in the p53 binding site of Mdm2 (grey). B) The highest scoring pose generated by docking the oligobenzamide shown in Figure 2.6A (p114) (blue) into the structure of Mdm2 from 1T4F. C) The result of removal of the protein and superimposition of the original p53 structure (as in A) on the docked oligobenzamide structure (as in B). There is excellent agreement between the structures despite the p53 peptide playing no part in the oligobenzamide docking.

The properties of each side chain were calculated and Spearman's Rank correlation coefficients (r_s) were determined for each property to investigate whether its value correlated with the FlexX docking score of the resulting oligobenzamide. Positive Spearman's rank values indicate that the higher the value of the property is, the greater the docking score; however, the scores are predicted ΔG values so higher scores correspond to weaker binding. Table 2.3 (p131) shows the results. Caution is necessary when interpreting these results for several reasons. Firstly, multiple hypotheses were tested. Secondly, it was assumed that all side chains bound in the expected docking pose (side chain 1 in or near the p53 Phe19 pocket, side chain 2 in or near the Trp23 pocket and side chain 3 in or near the Leu26 pocket) and this assumption might be incorrect. Finally, docking scores are often poor measures of binding affinity[161,162].

As shown in Table 2.3, at the middle side chain position, larger side chains lead to better scores. (For the total number of atoms, number of carbons and number of heteroatoms, $r_s<-0.03$, $r_s<-0.14$ and $r_s<-0.10$ respectively based on the 95% confidence limits shown in the table.) However, at the side chain position nearest to the amino terminus, small, low molecular weight groups appear to be favoured (because larger side chains result in larger scores and higher scores signify weaker binding). According to the 95% confidence limits, for the number of atoms in, number of carbons in and molecular weight of side chain 1, $r_s>0.05$, $r_s>0.04$ and $r_s>0.08$ respectively. At the carboxyl terminus, the side chain has little room to manoeuvre so escape from the expected binding pocket requires side chain flexibility. This could explain why rigid conjugated systems are unfavourable at this position. (For the number of rotatable bonds, $r_s>0.25$ and for the number of rigid bonds, $r_s<-0.29$ based on the 95% confidence limits.)

At all three side chain positions, fragments that lead to a molecule with a high polar surface area (PSA) have good scores ($r_s<-0.27$, $r_s<-0.30$ and $r_s<-0.47$ for side chains 1, 2 and 3 respectively). More polar molecules have lower logP values and, consistent with this, side chains with a lower XlogP (increased water solubility relative to their solubility in octanol) also score favourably ($r_s>0.33$, $r_s>0.24$ and $r_s<0.29$ for side chains 1, 2 and 3 respectively). The presence of hydrogen bond donors ($r_s<-0.02$, $r_s<-0.11$ and $r_s<-0.13$) and acceptors ($r_s<-0.15$, $r_s<-0.15$ and $r_s<-0.28$) appears to increase affinity and in the central and C-terminal side chain position, side chains with a positive charge have good scores ($r_s<-0.07$ and $r_s<-0.10$ respectively).

It is difficult to reconcile the PSA, XlogP and hydrogen bonding results with the overall hydrophobic nature of the p53 binding site of Mdm2[253] and the known importance of side chain hydrophobicity for strong binding[31]. Park and Jeon[229], point out that the FlexX force field lacks terms to directly account for ligand and binding site desolvation. These unexpected correlations could be the result of this bias inherent in the FlexX scoring function. Shoichet *et al.*[254] discuss how virtual screening without considering differences in solvation energy between compounds has a tendency to return molecules that are too large or too highly charged.

Table 2.3: How the properties of side chains at the three side chain positions in the molecules docked with FlexX correlated with the docking scores of the compounds.

Property	Side chain 1					Side chain 2					Side chain 3				
	n	Med. value	Spearman's Rank			n	Med. value	Spearman's Rank			n	Med. value	Spearman's Rank		
			Low limit		High limit			Low limit		High limit			Low limit		High limit
Number of atoms	365	13	0.05	0.12	0.20	348	13	-0.17	-0.10	-0.03	359	13	-0.13	-0.06	0.01
Number of carbons	365	9	0.04	0.12	0.19	348	9	-0.21	-0.14	-0.08	359	9	-0.12	-0.05	0.02
Number of heteroatoms	365	3	-0.18	-0.11	-0.04	348	3	-0.17	-0.10	-0.03	359	3	-0.29	-0.23	-0.17
Hydrogen bond acceptors	365	1	-0.28	-0.22	-0.15	348	1	-0.28	-0.22	-0.15	359	1	-0.38	-0.33	-0.28
Hydrogen bond donors	365	1	-0.15	-0.08	-0.02	348	1	-0.24	-0.17	-0.11	359	1	-0.25	-0.19	-0.13
Molecular weight	365	190	0.08	0.16	0.23	348	186	-0.10	-0.03	0.04	359	186	-0.07	0.00	0.07
Number of formal charges	365	0	-0.10	-0.03	0.04	348	0	-0.21	-0.14	-0.07	359	0	-0.23	-0.17	-0.10
Sum of formal charges	365	0	-0.06	0.01	0.09	348	0	-0.19	-0.12	-0.05	359	0	-0.19	-0.12	-0.05
Rotatable bonds	365	3	0.10	0.18	0.26	348	3	-0.03	0.05	0.13	359	3	0.25	0.33	0.40
Rigid bonds	365	10	-0.07	0.00	0.07	348	10	0.02	0.10	0.18	359	9	-0.39	-0.34	-0.29
Predicted compound XlogP	278	1.8	0.33	0.42	0.50	262	1.7	0.24	0.33	0.41	273	1.8	0.29	0.38	0.46
Compound 2D PSA*	278	-0.72	-0.39	-0.33	-0.27	262	-0.65	-0.42	-0.36	-0.30	273	-0.67	-0.56	-0.52	-0.47

Each side chain position was investigated separately. For each side chain position, a library of compounds which differed in terms of the side chain at the position in question was docked. The side chains used for this analysis were those obtained by cleavage of molecules from the ZINC source database, filtered such that all of the oligobenzamides generated were theoretically synthetically accessible. For further details see p106. The column labelled "n" indicates the number of side chains used and therefore also the number of data points used in the subsequent regression analysis. The "Med. value" column shows the median value of the side chain property for all of the compounds in the library used to investigate the side chain position in question. Spearman's rank correlation coefficients were calculated to look for any correlation between the properties of side chains and the relative binding affinities of oligobenzamides containing them. 95% confidence limits ("high limit" and "low limit") on the values of each Spearman's rank are shown in the table on either side of the ranks themselves. For details of how these confidence limits were calculated, see p291. All of the properties shown were calculated using the Open Babel[217] Filter program. *Topological polar surface area. (See p104.) Calculation of the topological PSA and logP estimate (XlogP[222]) of a molecule requires data for the functional groups within it. The functional group composition of some side chains meant that their predicted logP and polar surface area could not be calculated, so fewer side chains were investigated when analysing the effect of these two properties on docking score.

2.3.2. LONG SIMULATIONS OF OLIGOBENZAMIDES BASED ON THE N1 AND O1 SCAFFOLDS ARE STABLE

The estimation of accurate binding affinities for ranking requires that ensembles of conformations be generated for each compound. These can be produced by running molecular dynamics simulations. This section describes some initial testing of oligobenzamides bound to Mdm2 to check that the protein and ligands were behaving as expected with the parameters used.

Docking of the N1 and O1 scaffold starting molecules (p102) using Autodock Vina generated a small number of poses of which a few (for each molecule) were consistent with binding in the p53 binding groove with the side chains occupying the pockets usually occupied by p53 side chains. 5 such poses were selected for the N-linked scaffold compound and 3 were selected for the O-linked compound (manually by visual inspection). These comprised poses both parallel and antiparallel to the p53 helix. Using the bound poses as initial structures, explicit solvent simulations were run of the complex using the new parameters (p115). These simulations were stable and the ligands remained bound to the protein. The atomic coordinate root mean square deviations (RMSDs) at the start of the trajectories are shown in Figure 2.9. See Equation 1.53 (p78) for the definition of an RMSD and a discussion of its usage in the assessment of simulation equilibration.

To investigate the rigidity of the O-linked scaffold, RMSDs for substructures of the scaffold were calculated from a 5 ns trajectory (Figure 2.10). The RMSD lines shown in red are for the 12 carbon atoms of the N-terminal and middle benzene rings of the oligobenzamide scaffold (not the atoms in the side chains). The blue lines are, similarly, for the 12 carbon atoms of the middle and C-terminal rings of the scaffold. RMSDs were calculated relative to the starting pose after alignment of the 12 atoms with those of the starting structure. Movement of the lines therefore indicates rotation of the rings relative to each other (provided that the amide bond remains in its trans arrangement). Migration of the N-terminal side chain out of the binding pocket is indicated by movement of the red line. Migration of the C-terminal side chain from its pocket is indicated by movement of the blue line and movement of the central side chain is indicated by simultaneous movement of both the red and blue lines.

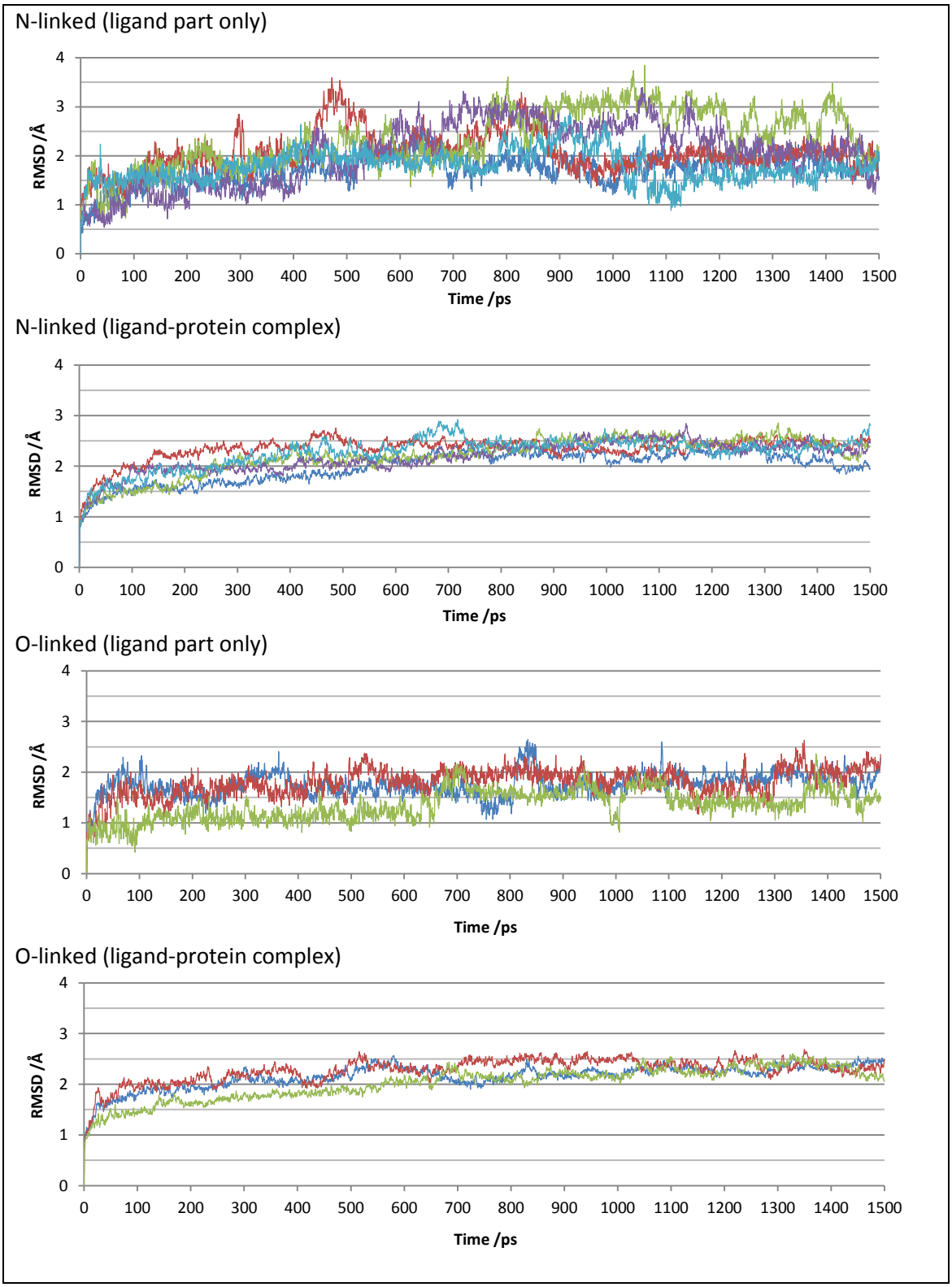


Figure 2.9: RMSDs of the ligand and ligand-protein complex in simulations of the N1 and O1 starting molecules bound to Mdm2. The compound structures are shown in Figure 2.1 (p102). These results are for the constant pressure run following a constant volume simulation and the RMSDs are relative to the structure at the start of the constant pressure part. The different lines of each graph show the results of different simulations started from different docking poses. The docking poses were manually selected prior to simulation but no selection has been applied to the trajectories shown.

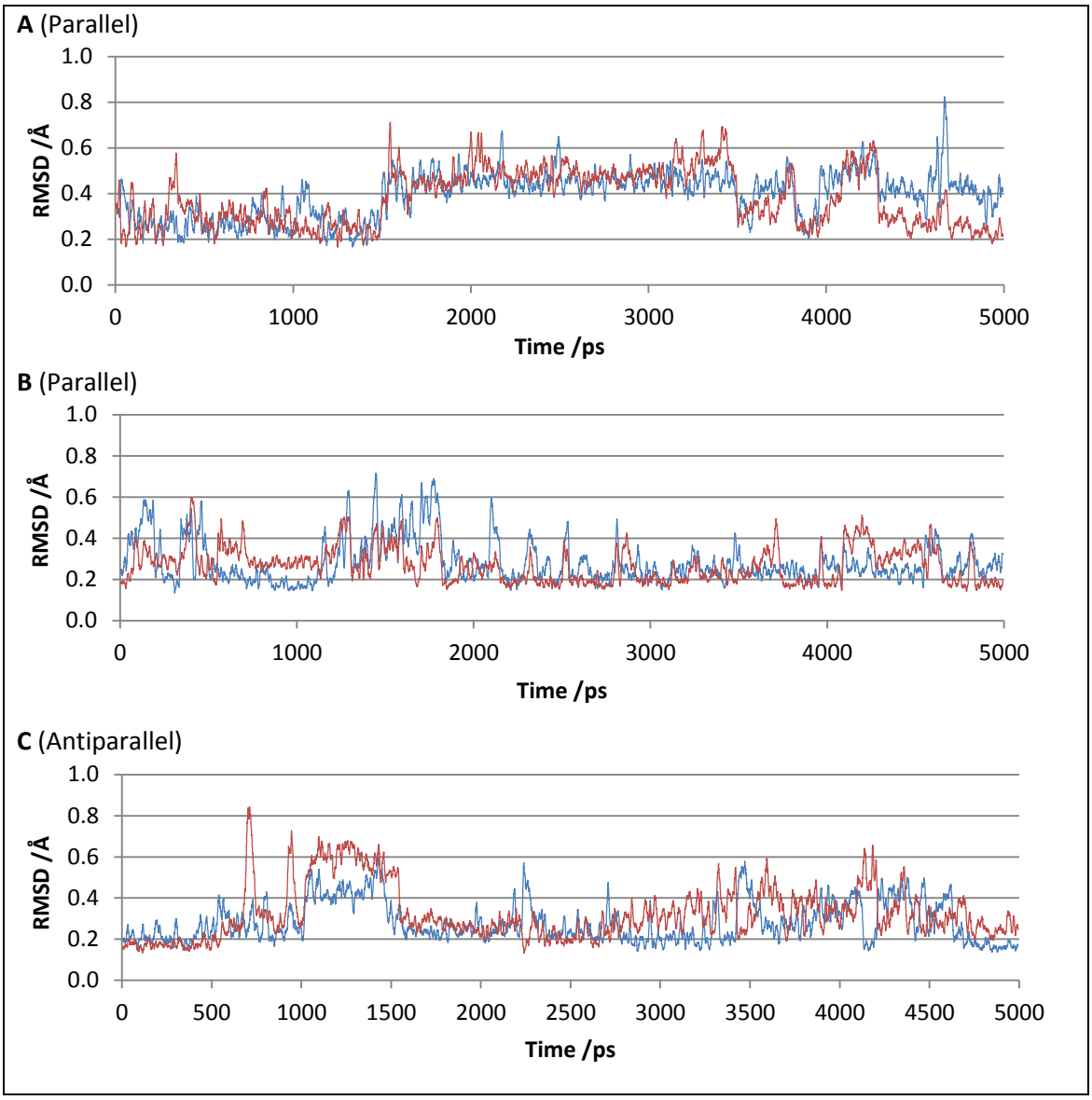


Figure 2.10: RMSD moving averages showing the relative orientation of the first and second oligobenzamide rings and the second and third oligobenzamide rings (red and blue respectively) of the O1 scaffold starting molecule starting from three different poses. The structure of the molecule is shown in Figure 2.1 (p102). The RMSDs are for the 12 carbon atoms of the two rings in question and are calculated following alignment of the atoms with those of the starting structure. A high RMSD indicates that one of the rings has rotated relative to the other so one of the side chains is projecting away from the protein (based on the assumption of that the amide bond remains trans). Simulations were started with poses parallel to the p53 in structure 1T4F and antiparallel as indicated. The RMSD moving averages shown are the mean of 21 RMSDs calculated at 500 fs intervals from 5 ps before the time point to 5 ps afterwards inclusive. Plotting an average enabled all of the data (collected at 500 fs intervals) to be used, the aim being to reduce the degree of random noise.

The results indicate some flexibility of the scaffold, despite its intramolecular hydrogen bonding; however, inspection of the trajectories (using VMD (University of Illinois)[255]) reveals that the side chains do not venture far from their binding pockets. This is evident from Figure 2.10. Using trigonometry it can be shown that rotation of both aromatic rings by 90° in opposite directions (holding the link rigid) would result in an RMSD of 1.3 Å (based on the assumption

that the ring carbon atoms are 1.4 Å apart). Consequently, the deviations shown in Figure 2.10, which are fractions of an Angstrom, even though they are affected by other factors such as bending of the scaffold, reflect only small rotations of the rings relative to each other. Most deviations are brief (for tens of picoseconds) and, where there is a longer departure (as at 1.5 ns in Figure 2.10A), the side chains do eventually return.

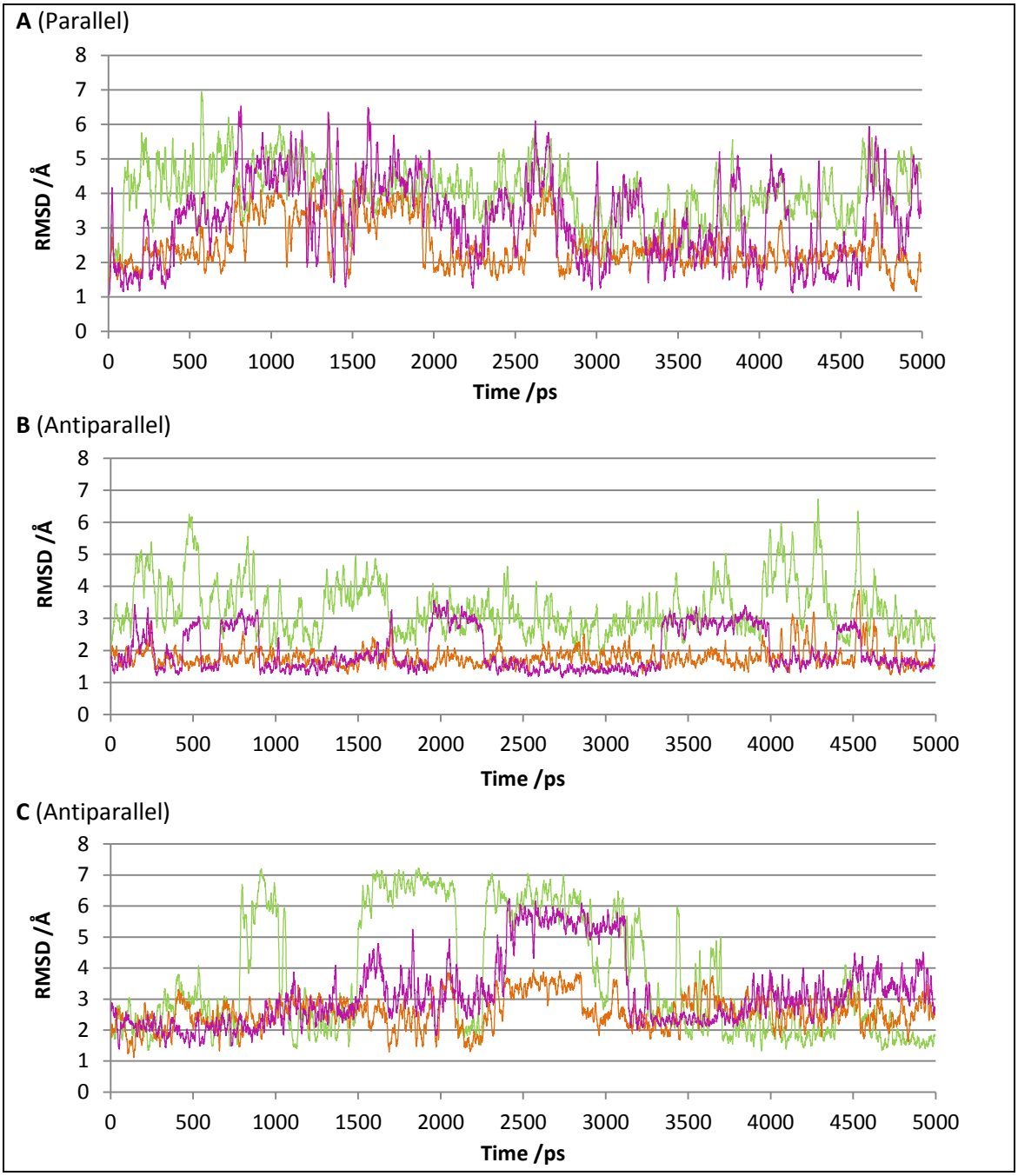


Figure 2.11: RMSD moving averages (over 21 RMSDs at 500 fs intervals) for the N1 scaffold starting molecule indicating the deviation of the side chains from their starting position relative to the centre of the scaffold. Each RMSD was generated using the carbon atoms of the central ring of the scaffold and the second carbon of the side chain. High RMSDs indicate that the side chain has moved relative to the scaffold. The green, orange and purple lines are for side chains 1, 2 and 3 respectively (the N-terminal, central and C-terminal side chains).

Figure 2.11 shows the movement of the three side chains of the N-linked scaffold relative to the central ring of the scaffold. The RMSDs are each based on 7 atoms, the 6 carbon atoms of the central scaffold ring and the first carbon of the side chain in question (the atom bonded to the nitrogen). Some movement of the side chains is visible; however, overall the side chains stay in their initial position. Figure 2.12 below shows structures from the trajectory used to produce Figure 2.11C above. The apparent deviations of side chain 1 (at the N-terminal end of the oligobenzamide) visible in Figure 2.11C (shown in green) were not caused by movement of the side chain away from its binding site but by shifting of the scaffold. This figure is discussed further on p184 because it illustrates how the motion of side chain 1 at the left-hand end of the ligand (Figure 2.12B) might, through pushing on the scaffold, cause the conformation of side chain 3 at the other end of the molecule to change (Figure 2.12C).

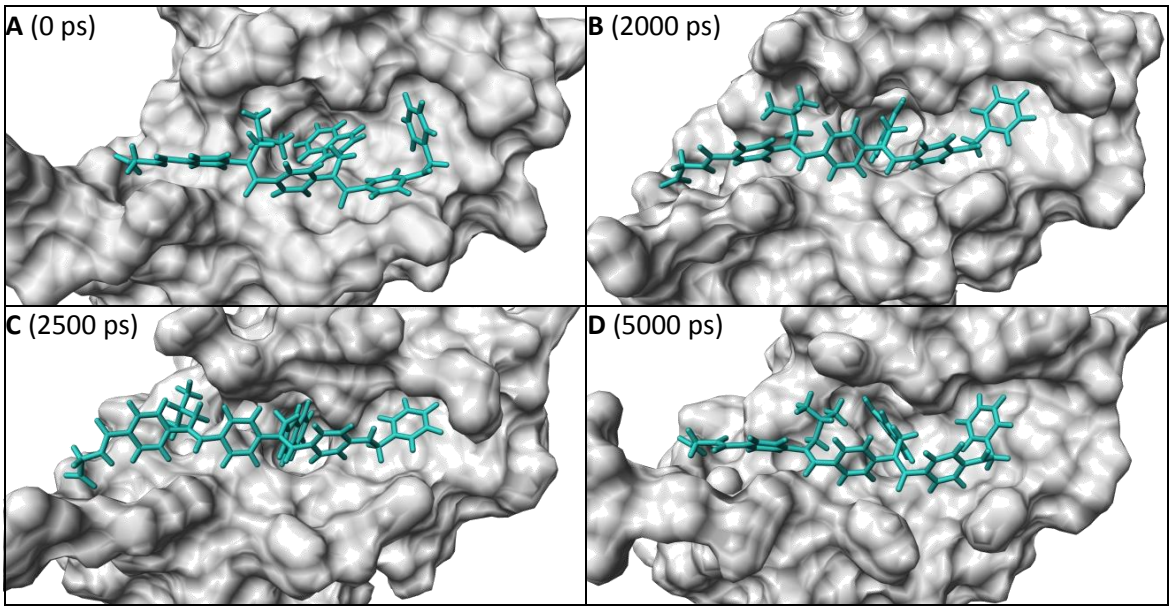


Figure 2.12: Structures from the 5 ns simulation of the N1 scaffold molecule (Figure 2.1, p102) used to produce Figure 2.11C. The oligobenzamide is shown in teal and the surface of Mdm2 is shown in grey. The time points shown have been selected such that they aid interpretation of Figure 2.11C. At the beginning (top left) and end (bottom right) of the simulation, the three side chains are in the expected binding position. At the 2000 and 2500 ps points, the scaffold is to the left of its initial position and the N-terminal side chain of the oligobenzamide (on the right in these images) has shifted relative to the scaffold to stay in its binding site. By the 2500 ps point, the C-terminal side chain has also shifted to point down, away from the scaffold. Changes in the conformation of the ligand appear to be accompanied by significant movement of the protein. Water and ions were removed from the trajectory using the ptraj program from the Amber 10 suite[224–226]. The frames of interest were extracted as PDB files using VMD 1.9.1[255]. They were aligned (to the 0 ps positions using both ligand and protein atoms) with the Protein Structure Comparison Tool V 3.0.8 (RCSB)[256,257] using the jCE algorithm (version 1.1)[258] with default parameters. The images were rendered in Chimera (UCSF).

The mobility of Mdm2 and oligobenzamides highlights the need to run simulations in order to sample different molecular conformations because the dynamics could have a significant impact upon affinity.

A single simulation would need to be of a substantial length of time to capture all of the binding states, given that the transient states visible in these simulations appear to persist for hundreds of picoseconds. Consequently, running multiple simulations from different docking poses, where possible, is appropriate.

2.3.3. THERMODYNAMIC INTEGRATION

2.3.3.1. Modification of the side chain attachment atoms revealed that sulphur in the S1 scaffold might contribute more to binding than oxygen in the O1 scaffold

Synthetic work in the collaborating Chemistry group had focussed on oligobenzamides with N-linked and O-linked side chains, yet side chains can also be attached through sulphur. Initial testing using thermodynamic integration sought to compare the O1 and S1 scaffolds and compare the O2 and S2 scaffolds. Oligobenzamide inhibitors are proposed to bind with each side chain mimicking a particular residue side chain of the p53 transactivation domain. In this model, the role of the scaffold is solely to prearrange the side chains, not contribute directly to binding through interaction with the protein. Thermodynamic integration was used to test this hypothesis through investigation of the effect of side chain attachment atom substitution on binding affinity. The results suggest that the S1 scaffold warrants further attention.

Scaffolds O1 and S1 and scaffolds O2 and S2 (Figure 1.6, p36) differ only in the atoms used to attach the side chains to the scaffold; however, they could have significantly different degrees of flexibility. The intramolecular hydrogen bonding in the scaffolds with O-linked side chains is potentially stronger because oxygen is more electronegative than sulphur. Furthermore, as discussed in chapter 1 (p70), steric effects caused by the large sulphur atom in the S2 scaffold prevent planarity of the aromatic ring and nearby amide group, reducing the potential for p-orbital alignment and conjugation[173].

The effect of scaffold flexibility on binding affinity is hard to predict. A more rigid scaffold could potentially pre-organise an oligobenzamide, ensuring that, when it is bound, weakly binding side chains do not migrate away from the protein surface but instead remain and contribute to binding. The movement of residues in an alpha-helix is constrained by the secondary structure[259] so an effective mimic of the p53 transactivation domain helix might be expected to have a scaffold which holds the side chains firmly in place. However, oligobenzamide scaffolds may not intrinsically place side chains in the optimal position. Some natural amino acids cannot be used because they are not compatible with the oligobenzamide synthetic process. These must be mimicked using different, less reactive, non-natural side chains (for example, a napthyl group in the case of phenylalanine). Consequently, some oligobenzamide flexibility might be required to allow side chains to sit where they will most effectively mimic the p53 residues.

Initially, attempts were made to perform a simple swap of oxygen in molecules based on the O1 scaffold for sulphur and *vice versa* in a set of thermodynamic integration experiments. However, it proved impossible to start stable simulations of an oxygen-linked compound produced by taking the starting pose of a sulphur-linked compound and replacing the sulphur atoms with oxygen. (In this context, an unstable simulation is one in which a very large force is created that cannot be accurately modelled so causes uncontrolled atomic motion and simulation failure.) Initial simulation at 100 K was tried and initial restraints were also tested, applied to ligand atoms on the protein side of the converted atoms; however, these did not solve the problem. TI simulations using free compounds as opposed to bound molecules were, in contrast, successful. Although simulations for conversion in the other direction from the O1 to the S1 scaffold were stable, transformation in both directions is typical to eliminate bias towards the compound used for docking and to look for hysteresis[166(p577)].

TI calculations for the modification of an oxygen or sulphur link to CH_2 were stable. Therefore, to compare oxygen and sulphur-containing structures, a TI run in which the oxygen or sulphur attachment atoms were converted to CH_2 was performed for each of a set of 12 oligobenzamides. In TI simulations, equilibration of $d\mathcal{V}/d\lambda$ appeared to occur in under 50 ps. However, due to the time required for equilibration of the RMSDs in earlier work, the initial 150 ps of each trajectory was not used for analysis.

Predicted changes in the free energy of binding upon these transformations are shown in Figure 2.13 below. Three different sizes of side chain were tested at the middle position: a small one (tertbutyl, side chain 19 in Figure 2.2 (pages 103 and 104)), a medium-sized group, (fluorobenzene, side chain 25) and a large one (methylnapthyl, side chain 3).

At each step, four repeats were performed and the confidence limits in Figure 2.13 are based on the standard errors of these results. The errors associated with the results are quite large relative to the size of the calculated changes. They suggest that the phase space was insufficiently sampled for good convergence of the results from different repeats. A disadvantage of thermodynamic integration is that many simulations have to be performed for every conversion. This reduces the amount of time for which each simulation can be performed.

Inspection of the results reveals a smaller drop in the free energy of binding in the case of ortho-attached side chains (top half of Figure 2.12) compared to with meta-attached side chains (bottom half of Figure 2.12). This is as expected, given the shorter and thus stronger hydrogen bonds in the latter case.

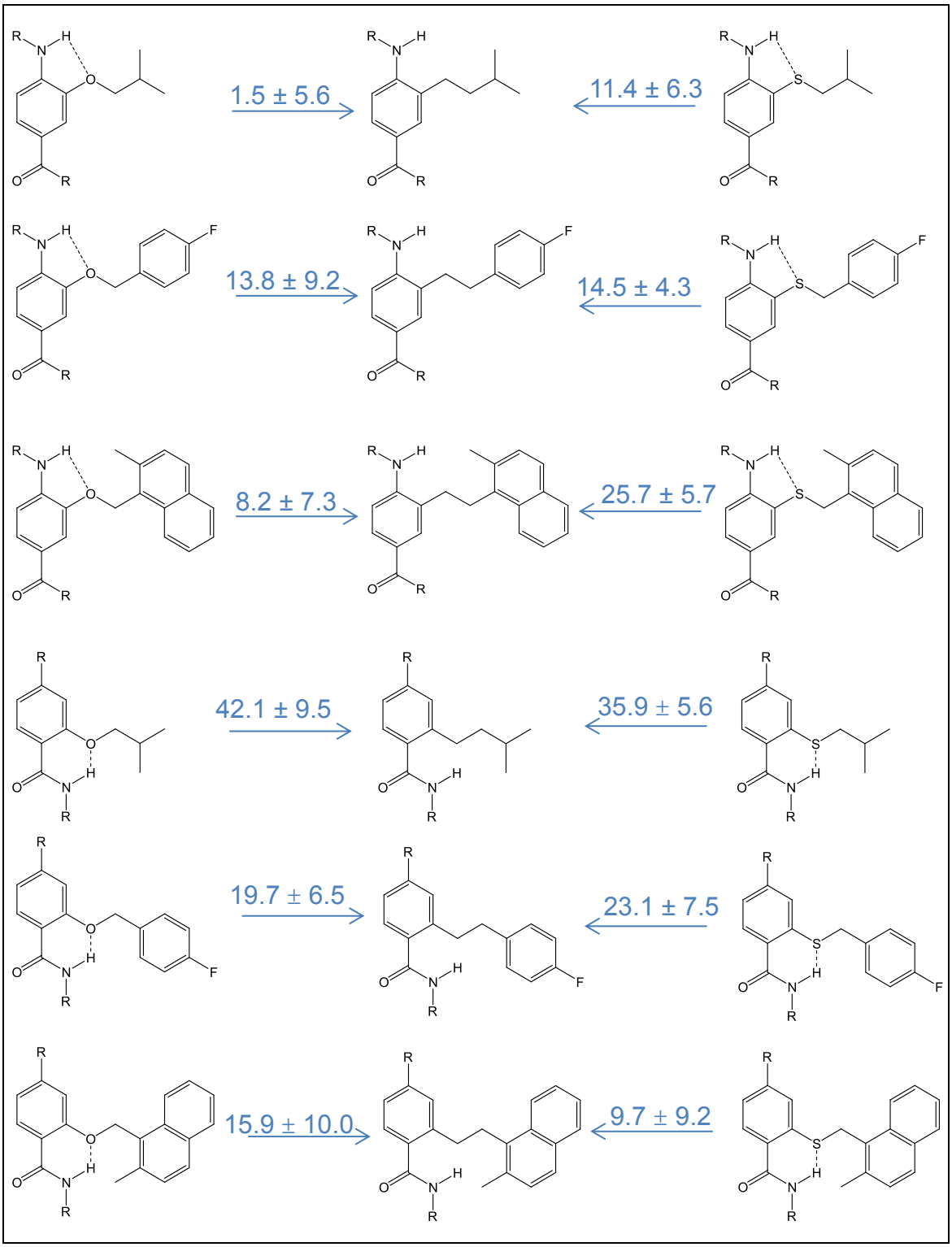


Figure 2.13: Predicted changes to the free energy of binding of 12 compounds upon transformation of the side chain attachment atoms from oxygen or sulphur to CH_2. Energies are in kJ mol^{-1}. The compounds were based on the O1 (top left), S1 (top right), O2 (bottom left) or S2 scaffold (bottom right) as indicated by the structure of the scaffold part shown, which is the central unit holding the middle side chain. Three different middle side chains were tested as indicated. Apart from being modified at the central side chain position, the compounds are otherwise identical to the starting molecules shown in Figure 2.1 (p102).

In Table 2.4, the predicted binding energy changes for the side chain-holding heteroatom conversions to CH_2 are compared to investigate if the oxygen atoms in O1 scaffold oligobenzamides or the sulphur atoms in the S1 oligobenzamides contribute more to the binding of the ligand. A t-test was used to test the significance of each difference.

The results suggest that in terms of the O1 and S1 scaffolds, the choice of attachment atom does have an impact on affinity, with sulphur leading to stronger binding. (For side chain 19, $\Delta\Delta\Delta G = 9.93$ kJ mol^{-1}, $p < 0.013$ and for side chain 3, $\Delta\Delta\Delta G = 17.50$ kJ mol^{-1}, $p = 0.018$.) Hydrogen bonds involving sulphur are weaker than the equivalent bonds involving oxygen; however, sulphur-hydrogen bond strength is more direction-dependent[260] and the increased rigidity of the S1 scaffold relative to the O1 scaffold which results from the maintenance of this bond is incorporated into the torsion parameters for this scaffold[173] (Table 2.2, p115). Prearrangement of S1 scaffold side chains could make binding more entropically favourable and also keep side chains on the surface of protein where they can contribute to the binding affinity.

No statistically significant difference in affinity was observed between O2 and S2 scaffold compounds. This could be because the S2 scaffold is more flexible than the O2 scaffold due to steric hindrance between the large sulphur atoms and the hydrogen atom with which they bond[173].

Table 2.4: Use of Student's t-test to determine if there are significant differences between the changes in free energy when oxygen is replaced with CH_2 and when sulphur is replaced with CH_2

Scaffolds	Middle side chain	Difference /kJ mol^{-1}	Standard deviation of the difference /kJ mol^{-1}	t	Degrees of freedom	P
O1/S1	19	9.94	2.64	-3.8	5.93	0.013
O2/S2	19	-6.29	3.48	1.8	4.88	0.145
O1/S1	25	0.63	3.18	-0.2	4.25	0.852
O2/S2	25	3.36	3.12	-1.1	5.88	0.331
O1/S1	3	17.50	2.92	-6.0	5.64	0.002
O2/S2	3	-6.15	4.26	1.4	5.96	0.208

In the column labelled "t" are the test statistics evaluated using student's t-distribution. In the rightmost column are the probabilities of a difference as large as the one observed occurring due to chance. Only the O1/S1 results for side chains 3 and 19 are significant at the 5% significance level. The number of degrees of freedom at which to evaluate the t-distribution was calculated using the Welch–Satterthwaite equation[261] so the assumption of equal sample variances was unnecessary. In practice, use of these degrees of freedom as opposed to the usual 2n-2 degrees of freedom (6 in this case) makes little difference to the critical values. (With 6 degrees of freedom, the 95% significance critical value of t is 2.447.)

2.3.3.2. Thermodynamic integration results suggest that predicted binding free energies are robust to small changes in the scaffold backbone torsion parameters

Different torsion parameters are used for scaffolds with oxygen-linked side chains compared to scaffolds with sulphur-linked ones as detailed in the methods section (p113). To investigate the importance of the torsional barriers used and the significance that modification of the torsional barrier height might have had in the link atom modification TI investigation above, a second TI investigation with the same 12 oligobenzamides was performed. In this, the torsion angle barrier height of the bond about which rotation is restricted due to hydrogen bonding was changed from 1 kcal mol^{-1} below to 1 kcal mol^{-1} above its normal barrier height. The results are shown in Table 2.5. A single step is required for this conversion.

Table 2.5: For oligobenzamides based on each scaffold, the TI-predicted change in binding energy associated with modification of the torsion parameter for the bond about which rotation is restricted due to intramolecular hydrogen bonding such that the barrier height goes from 1 kcal mol^{-1} below to 1 kcal mol^{-1} above the normal barrier height

Attachment position	Middle side chain	Change in potential energy predicted by the AMBER force field when increasing the torsional barrier height by 2 kcal mol^{-1} /kJ mol^{-1}					
		Attachment through oxygen			Attachment through sulphur		
		Free	Bound	Free – Bound	Free	Bound	Free - Bound
Ortho	19 (Small)	-0.576	-0.835	-0.259 ± 0.100	-0.383	-0.438	-0.055 ± 0.053
	25 (Med.)	-0.544	-0.565	-0.021 ± 0.082	-0.394	-0.411	-0.017 ± 0.039
	3 (Large)	-0.546	-1.161	-0.614 ± 0.148	-0.390	-0.454	-0.064 ± 0.072
Meta	19 (Small)	-1.127	-1.065	0.062 ± 0.132	-3.850	-3.467	0.383 ± 0.111
	25 (Med.)	-1.237	-1.210	0.027 ± 0.261	-3.781	-3.911	-0.131 ± 0.177
	3 (Large)	-1.182	-1.304	-0.122 ± 0.126	-3.787	-3.653	0.135 ± 0.078

For each scaffold, (O1 (top left), S1 (top right), O2 (bottom left) and S2 (bottom right)) scaffolds with three differently-sized side chains were tested in the middle side chain position (making 12 molecules in total). Negative values indicate where increasing the torsion energy is energetically favourable. Simulations were each performed four times. The unbiased sample variances of the free and bound values were summed to get, in the case of each molecule, the variance of the difference. 95% confidence limits based on these variances are shown in the "Free - Bound" column.

The large changes in potential energy in the bottom rows of Table 2.5 (typically -1.2 kJ mol^{-1} for the O2 scaffold and -3.7 kJ $mol^{-1}$for the S2 scaffold) could reflect the change in electrostatic energy when the atoms involved in intramolecular hydrogen bonding are brought into closer proximity by the increase in the torsional barrier height of the amide carbon-aromatic carbon bond. This stabilisation is not observed with the O1 and S1 scaffolds to the same extent, perhaps because their intramolecular hydrogen bonds are longer. This effect could be particularly large for the S2 scaffold because the torsion parameter for this bond is particularly low in order to reflect the effect of steric hindrance between the nearby sulphur atom and its hydrogen-bonded hydrogen on the extent to which the p-orbitals of the aromatic ring and amide overlap[173].

With regard to the "Free - Bound" columns of Table 2.5, the change in the free energy of binding is predicted to be negative when the torsional barrier height is increased in the O1 and S1 compounds. This implies that a reduction in scaffold flexibility increases binding affinity. However, the comparatively small changes in binding energy observed when the torsion parameters were changed by over 8 kJ mol^{-1}, most of which are not statistically significant, suggest that the difference in torsion parameters between the O1 and S1 scaffolds might not be the main cause of the any difference in the affinity of these scaffolds, as was suggested in the previous section.

S1 scaffold compounds might bind stronger than O1 scaffold compounds due to the greater size and polarisability of sulphur relative to oxygen[262(p1252)], which could potentially lead to stronger van der Waals forces between these side chain attachment atoms and the protein if they were to interact. Future work could investigate the presence of interactions between these link atoms and the protein in long simulations of O1 and S1 scaffold compounds.

In Leeds, most synthetic work has focussed on compounds with N-linked and O-linked side chains. The predicted stronger binding of some compounds based on the S1 scaffold compared to their corresponding O-linked counterparts (in the previous section) together with the predicted increase in ortho-substituted oligobenzamide affinity when the torsional barrier height is raised (discussed in this section) advocate further investigation of the S1 scaffold.

The remainder of this chapter concerns the effect side chain choice, as opposed to scaffold choice, has on predicted affinity.

2.3.3.3. Comparison of the side chains by TI allowed predictions for 31 compounds based on the N1 scaffold to be made but there was a poor correlation between these predictions and experimental results

The central aim of this project was to investigate whether relative oligobenzamide binding affinities could be predicted successfully by computational means. This section describes the ranking of 31 compounds based on the N-linked scaffold (Figure 1.6, p36) using thermodynamic integration (TI) and the comparison of the predicted ranking with single point fluorescence anisotropy measurements of their relative affinities. A poor correlation was observed and the conclusion at the end of this section is that TI is not a method by which the relative affinities of oligobenzamides for Mdm2 can be predicted accurately. TI is a method best suited to the comparison of very similar ligands with similar binding poses. The 31 compounds investigated, selected on the basis of their ease of synthesis rather than their suitability for computational comparison, had significant differences and subsequent docking discussed later in this chapter (p153) revealed that oligobenzamides can bind to Mdm2 in many different places. Compound synthesis and fluorescence anisotropy measurements were performed by Kerya Long (University of Leeds).

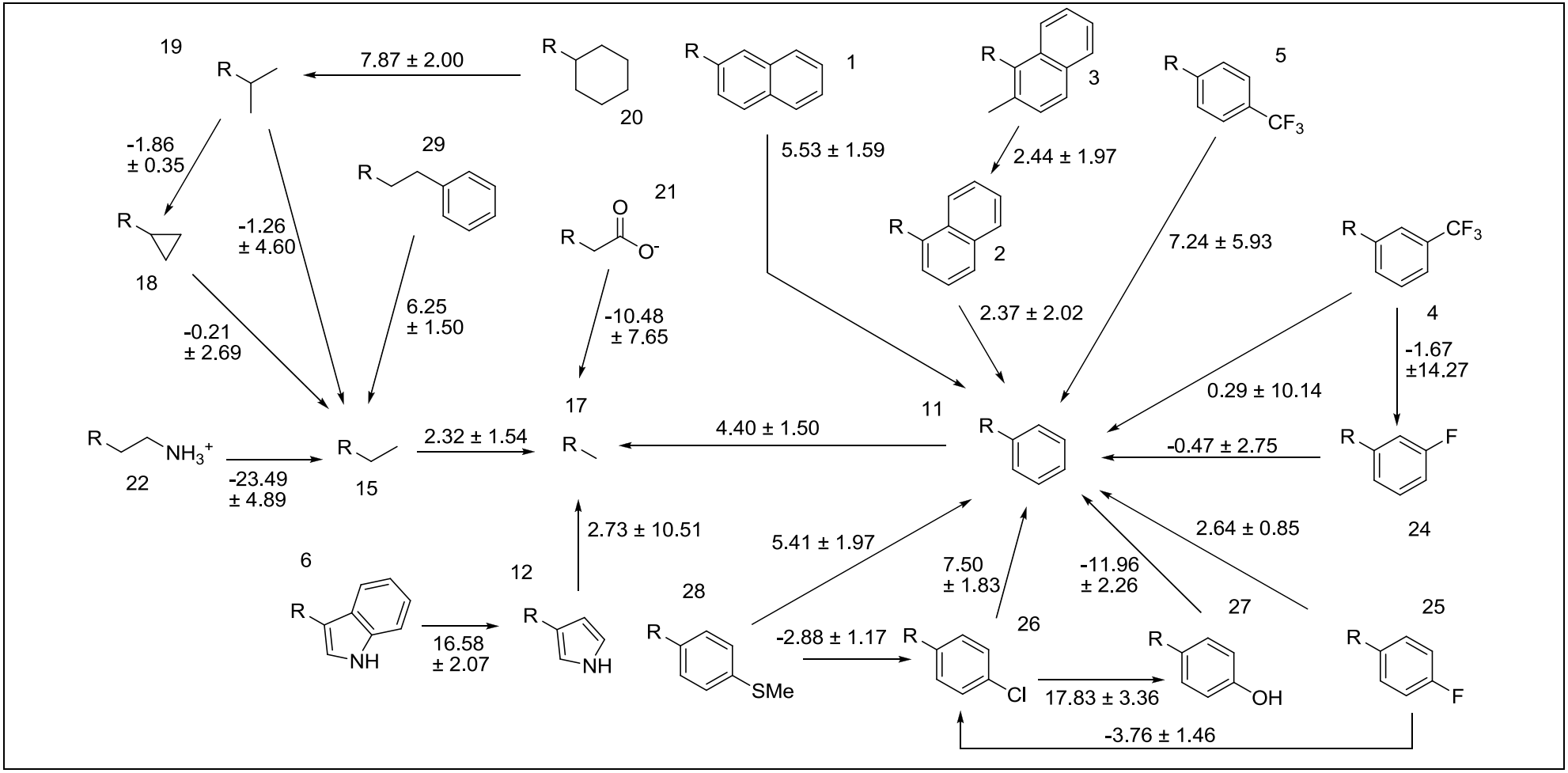


Figure 2.14: The TI transitions performed to modify the central side chain of oligobenzamides based on the N1 scaffold together with their associated calculated differences in binding energy. The numbers of the compounds correspond to the numbers shown in Table 2.6 and Figure 6.1. The calculated energies are in kJ mol^{-1} and 95% confidence limits are shown.

Figure 2.14 shows predicted ΔΔG values for conversion of the central side chain of the N1 scaffold (Figure 2.1, p102) between various possible structures. To simplify side chain comparison, it was assumed that changes were additive, so, for example, the binding ΔΔG for conversion of a chloronapthyl group to a benzyl group was the sum of the ΔΔG for conversion of a chloronapthyl group to a napthyl group and the ΔΔG for converting a napthyl group to a benzyl group.

When equilibrating simulations for TI, $d\mathcal{V}/d\lambda$ (defined on p84) equilibrated rapidly. Curves of $d\mathcal{V}/d\lambda$ against λ were typically smooth, especially for the removal of charges and addition of charges respectively, the first and last steps of the alchemical conversion. For example, see Figure 2.15 and Figure 2.16 below.

The most significant change in binding free energy shown in Figure 2.14 is that for removal of the ammoniumyl (NH_3^+) group from side chain 22, -23.5 kJ mol^{-1}. Removal of this charged group is predicted to be very favourable and much of the favourable ΔΔG (-15.4 kJ mol^{-1}) is accounted for by step 1 of the TI conversion, removal of the positive charge. Of the 31 N-linked compounds synthesised and tested, the weakest binding was found to be the one with this side chain (which was at the middle side chain position), suggesting that this side chain is genuinely very unfavourable for binding. The p53 binding site of Mdm2 is hydrophobic; p53-Mdm2 binding is driven by van der Waals forces[207]. Consequently, hydrogen bonds that the positive ammoniumyl group can make in the solvent are likely to be lost when the small molecule binds, making binding enthalpically unfavourable. Binding will also not be accompanied by the favourable entropy change which occurs when a hydrophobic group is buried and the water molecules around it are released from their regular arrangement[57]. A possible explanation considered was interaction of the charged side chain with the terminal, negatively charged carboxylate group in the free but not the bound ligand; however, simulations of the free ligand reveal that the side chain and C-terminal are too far apart and the oligobenzamide is insufficiently flexible to bring them together.

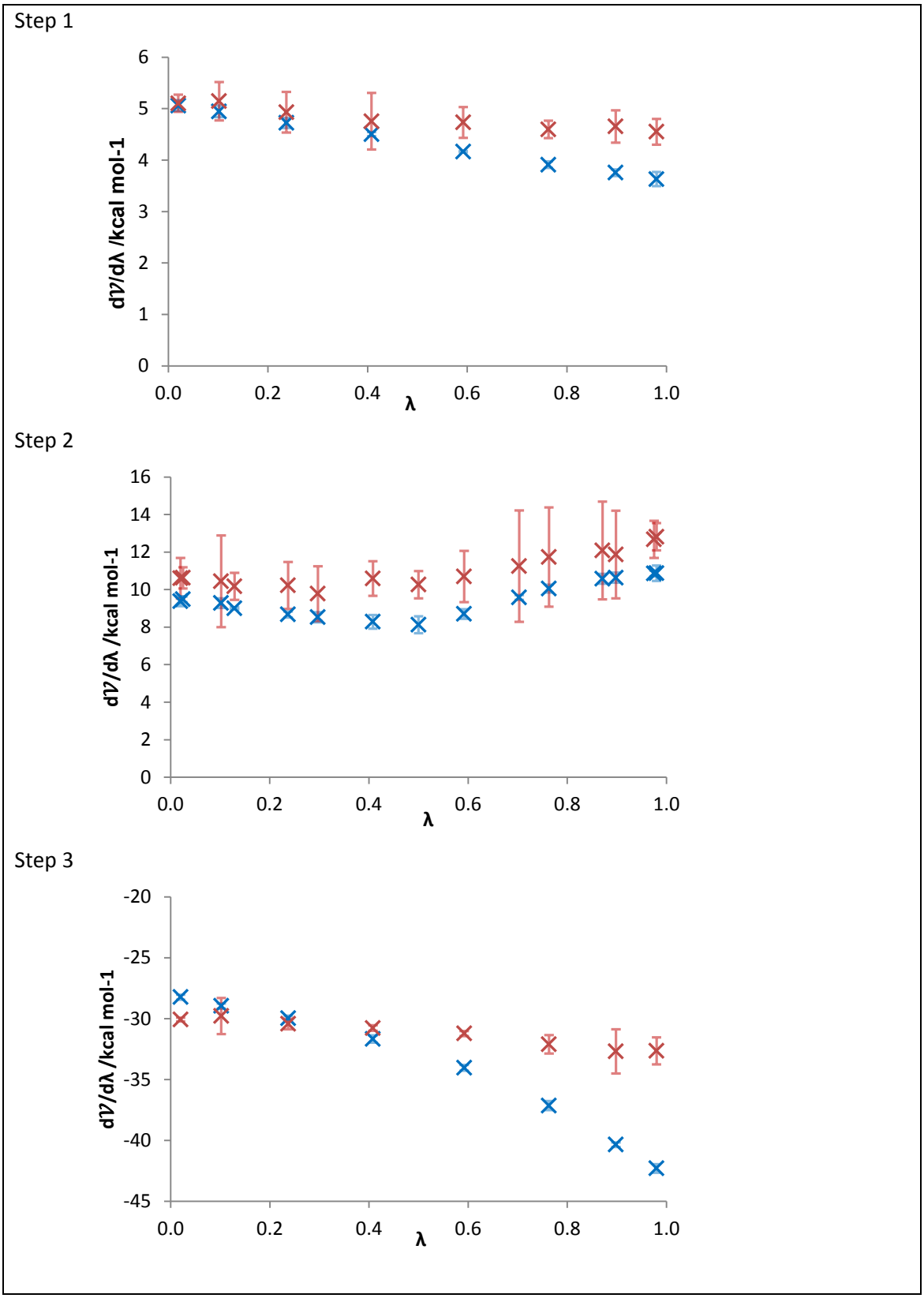


Figure 2.15: Example TI dV/dλ graphs: The modification of side chain 26 to side chain 27 at the middle position of the N1 scaffold (Figure 1.6, p36). The terminal side chains were as shown in Figure 2.1 (p102). Results for the bound ligand are shown in red. Results for the free ligand are shown in blue.

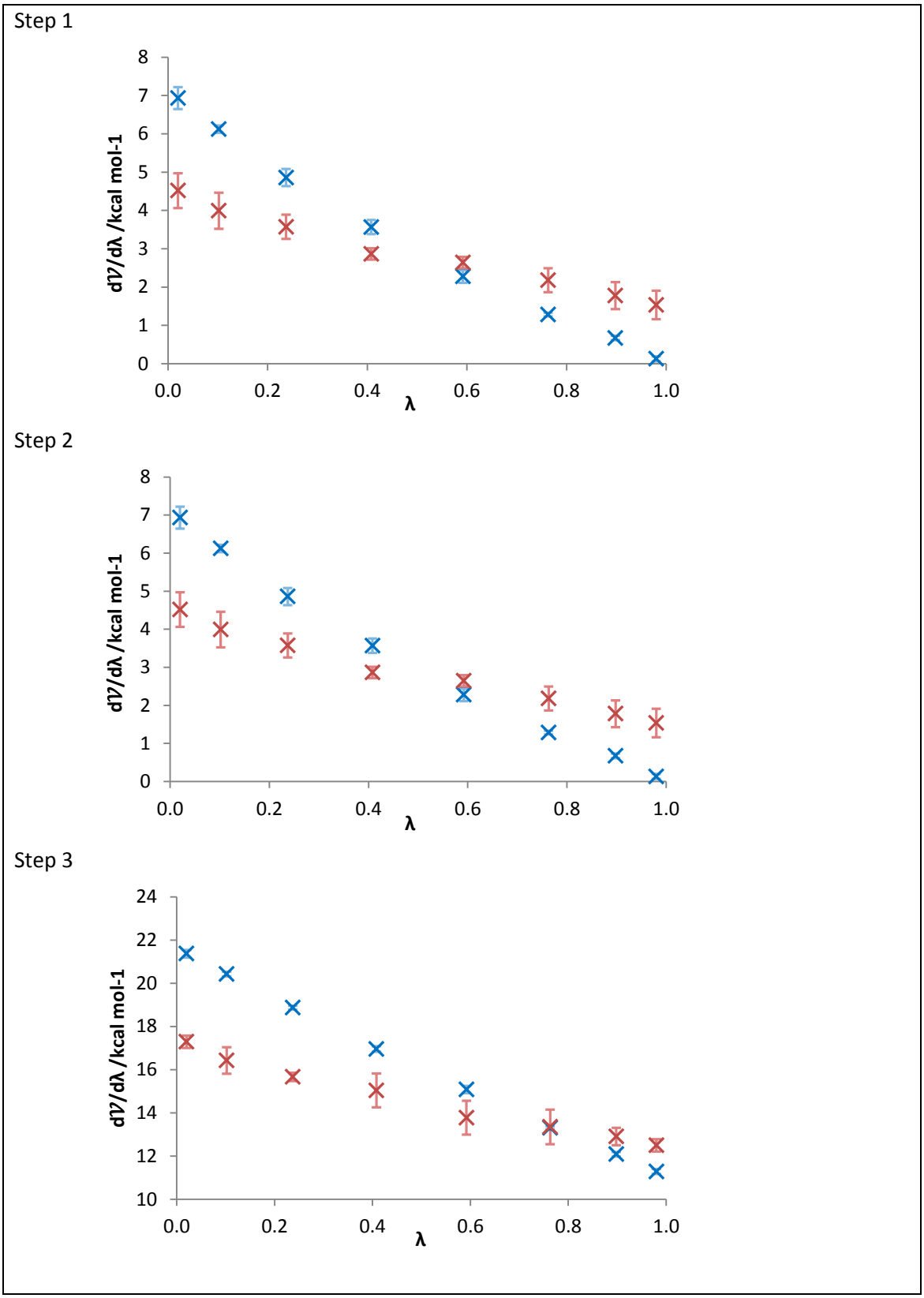


Figure 2.16: Example TI dV/dλ graphs: The modification of side chain 1 to side chain 11 at the middle position of the N1 scaffold (Figure 1.6, p36). The terminal side chains were as shown in Figure 2.1 (p102). Results for the bound ligand are shown in red. Results for the free ligand are shown in blue.

Table 2.6 shows in more detail the results summarised in Figure 2.14, TI results for modification of the side chain at the central attachment position of the N1 oligobenzamide scaffold. The result for step 3 for the conversion of side chain 21 to side chain 17 (-376 kJ mol^{-1}) appears to be unusually large in magnitude given that side chain 17 is an ethyl group. This energy change for the addition of charge to the system is due to charging of a Cl^- counter ion, which is present in the systems as a result of side chain 21 being negatively charged and the method of adding ions when simulations were set up (described on p117 and illustrated graphically in Figure 2.7 on p119). Charging of a counter ion (in this case K^+) is also responsible for the large change observed in step 3 for the conversion of side chain 22 to side chain 15 (-370 kJ mol^{-1}).

Validation of computational results would have been most straightforward using a systematic set of compounds, for example, molecules differing in their side chain at a single attachment position. However, in the event, difficulties with chemical synthesis meant that the molecules available for comparison differed in their side chains at all three side chain attachment positions, as shown in Table 2.7. To facilitate the comparison of predicted and experimental results, the side chains at each position were assumed to make independent contributions to the binding energy, such that the $\Delta\Delta G$ for any two compounds differing at all three side chain positions was the sum of three energy changes, one for the modification at each side chain attachment position. TI simulations for modification at the first and last side chain positions (full results not shown) were performed in a similar manner to those for the middle position. This enabled binding energies to be predicted for all 31 compounds synthesised by the collaborating chemistry group (Table 2.7, p150).

Table 2.7 shows, for each compound, the relative predicted free energies of binding and, for comparison, the fluorescence polarisation single point competition assay results obtained in the laboratory and supplied by Kerya Long (University of Leeds). The side chain numbers refer to the side chain structures shown in Figure 2.14 and also Figure 2.2 above (p103). The relative predicted free energies of binding ($\Delta\Delta G$) are relative to the scaffold with side chain 17 (an ethyl group) at every side chain position. The values for side chain 2, the middle side chain, correspond to the changes shown in Figure 2.14. For example, where side chain 11 is shown present at the middle position in Table 2.7, the predicted binding energy contribution of side chain 2 is -4.4 kJ mol^{-1}, consistent with the 4.4 kJ mol^{-1} $\Delta\Delta G$ for the conversion of side chain 11 to side chain 17 in Figure 2.14. Other values in this column can be calculated by summing the transitions shown in Figure 2.14. For example, in the first row of data in Table 2.7, the -9.93 kJ mol^{-1} binding free energy contribution for side chain 2 corresponds to the sum of 5.53 kJ mol^{-1} for the conversion of side chain 1 to side chain 11 in Figure 2.14 and the aforementioned 4.4 kJ mol^{-1} for the conversion of side chain 11 to side chain 17.

Table 2.6: Thermodynamic integration results for transformation of the middle side chain of the N1 scaffold starting molecule

Modification		Repeats	Step 1 ΔG /kJ mol^{-1}		Step 2 ΔG /kJ mol^{-1}		Step 3 ΔG /kJ mol^{-1}		Overall ΔΔG	
From	To		Bound	Free	Bound	Free	Bound	Free		
1	11	6	11.92 ± 0.21	13.12 ± 0.22	-77.79 ± 1.15	-91.12 ± 0.58	60.84 ± 0.72	67.44 ± 0.18	5.53 ± 1.59	
2	11	6	-8.34 ± 0.95	-5.88 ± 0.29	-73.95 ± 0.96	-81.79 ± 0.99	43.02 ± 0.90	46.04 ± 0.15	2.37 ± 2.02	
3	2	3	3.38 ± 0.03	3.58 ± 0.00	21.15 ± 0.68	18.18 ± 0.18	28.53 ± 0.24	28.87 ± 0.27	2.44 ± 1.97	
4	11	3	-31.01 ± 0.17	-31.12 ± 0.13	-38.83 ± 4.02	-41.38 ± 0.18	29.26 ± 0.63	31.63 ± 0.17	0.29 ± 10.14	?
5	11	3	-68.94 ± 0.26	-70.01 ± 0.08	-62.11 ± 2.33	-68.50 ± 0.23	34.78 ± 0.41	35.00 ± 0.05	7.24 ± 5.93	
6	12	6	22.74 ± 0.60	24.20 ± 0.20	-136.31 ± 1.65	-163.13 ± 0.17	109.72 ± 0.87	118.50 ± 0.07	16.58 ± 2.07	
21	17	3	154.70 ± 2.67	155.05 ± 0.57	163.10 ± 1.02	178.94 ± 0.72	-370.59 ± 0.67	-376.29 ± 0.06	-10.48 ± 7.65	
12	17	3	13.56 ± 3.13	14.64 ± 0.10	-35.37 ± 2.84	-39.06 ± 0.04	0.18 ± 0.01	0.07 ± 0.00	2.73 ± 10.51	?
11	17	6	0.74 ± 0.76	4.47 ± 0.14	15.96 ± 1.04	7.95 ± 0.31	0.19 ± 0.00	0.07 ± 0.00	4.40 ± 1.40	
15	17	3	6.66 ± 0.03	5.59 ± 0.01	17.12 ± 0.44	15.90 ± 0.44	0.06 ± 0.00	0.02 ± 0.00	2.32 ± 1.54	
4	24	3	8.35 ± 4.59	-13.74 ± 0.32	-51.38 ± 3.44	-27.62 ± 0.24	0.00 ± 0.00	0.00 ± 0.00	-1.67 ± 14.27	?
24	11	3	33.49 ± 0.36	30.40 ± 0.03	-52.41 ± 0.93	-51.20 ± 0.12	29.18 ± 0.43	31.52 ± 0.16	-0.47 ± 2.75	--
28	26	6	-18.91 ± 0.20	-16.04 ± 0.03	52.65 ± 1.09	52.66 ± 0.12	0.00 ± 0.00	0.00 ± 0.00	-2.88 ± 1.17	
27	26	3	133.22 ± 2.40	140.98 ± 0.27	-44.90 ± 0.93	-38.62 ± 0.08	-20.32 ± 0.12	-18.15 ± 0.02	-16.21 ± 6.44	

Continued on the next page…

Step 1 was the removal of charges from the atoms to be removed. Step 2 was the decoupling of the atoms being removed and introduction of any new atoms. Step 3 was the addition of charge to any new atoms (and this step was only necessary where atoms were being added). The first two columns show the initial and final side chains. For the structures of these side chains see Figure 2.14 (p143) or Figure 2.2 (p103). 95% confidence limits are shown on values. The symbols on the right of the table show whether the new compound binds more strongly (red) or less strongly (blue) than the original. Grey question marks and green dashes indicate where the error indicated by the 95% confidence limits is greater than the overall change, with the dash being where the error is under 3 kJ mol^{-1} and the failure to see a significant change could be because the real affinities of the original and modified side chain are very similar. Initially, the TI conversions of each step were performed three times. The results for each step were then used to calculate the 95% confidence limits on the ΔΔG for that step. The variances of the three steps were summed to obtain the variance of the overall ΔΔG. If the error on the final ΔΔG was larger than the ΔΔG itself, meaning that the overall ΔΔG was not significantly different from 0 at the 5% significance level, then a further 3 repeats were performed in an effort to reduce the error.

Table 2.6 continued: Thermodynamic integration results for transformation of the middle side chain of the N1 scaffold starting molecule

Modification		Repeats	Step 1 ΔG /kJ mol^{-1}		Step 2 ΔG /kJ mol^{-1}		Step 3 ΔG /kJ mol^{-1}		Overall ΔΔG	
From	To		Bound	Free	Bound	Free	Bound	Free		
26	27	3	20.05 ± 0.19	18.13 ± 0.02	45.50 ± 0.62	38.63 ± 0.07	-130.46 ± 0.37	-141.11 ± 0.18	19.44 ± 1.92	
26	25	3	20.43 ± 0.28	18.20 ± 0.09	-33.92 ± 0.63	-41.90 ± 0.28	-45.69 ± 0.14	-39.38 ± 0.21	3.92 ± 1.95	
25	26	3	45.20 ± 0.56	39.30 ± 0.11	34.31 ± 0.67	41.89 ± 0.00	-20.16 ± 0.04	-18.24 ± 0.03	-3.59 ± 2.18	
28	11	3	26.99 ± 0.19	28.79 ± 0.03	-22.36 ± 0.58	-29.82 ± 0.18	34.70 ± 0.46	34.96 ± 0.08	5.41 ± 1.97	
27	11	3	130.70 ± 0.80	141.13 ± 0.26	-96.11 ± 0.26	-95.20 ± 0.20	34.36 ± 0.10	34.97 ± 0.04	-11.96 ± 2.26	
26	11	3	20.18 ± 0.11	18.17 ± 0.05	-50.44 ± 0.44	-56.19 ± 0.34	34.72 ± 0.45	34.98 ± 0.10	7.50 ± 1.83	
25	11	6	45.89 ± 0.41	39.37 ± 0.06	-17.93 ± 0.52	-14.53 ± 0.17	34.48 ± 0.42	34.96 ± 0.10	2.64 ± 0.85	
29	15	6	-3.96 ± 0.49	0.71 ± 0.72	-6.84 ± 1.08	-17.12 ± 0.34	2.32 ± 0.09	1.69 ± 0.02	6.25 ± 1.50	
22	15	3	69.15 ± 0.95	84.54 ± 1.03	12.81 ± 1.31	17.26 ± 0.37	-369.58 ± 0.21	-365.93 ± 0.11	-23.49 ± 4.89	
19	15	3	47.02 ± 0.20	44.42 ± 0.08	-20.30 ± 1.79	-16.94 ± 0.39	-2.40 ± 0.04	-1.90 ± 0.02	-1.26 ± 4.60	?
20	19	3	-2.18 ± 0.06	-2.77 ± 0.01	-32.88 ± 0.68	-38.66 ± 0.17	-4.53 ± 0.40	-5.97 ± 0.02	7.82 ± 2.00	
18	15	3	-58.24 ± 0.93	-57.06 ± 0.17	-23.57 ± 0.59	-24.27 ± 0.41	-1.46 ± 0.13	-1.73 ± 0.03	-0.21 ± 2.96	--
18	19	6	20.49 ± 0.07	18.06 ± 0.03	-73.67 ± 0.26	-79.22 ± 0.27	-54.18 ± 0.15	-47.75 ± 0.05	1.55 ± 0.43	
19	18	6	53.50 ± 0.23	47.60 ± 0.07	73.61 ± 0.38	78.96 ± 0.23	-20.72 ± 0.16	-18.02 ± 0.04	-2.16 ± 0.56	

The laboratory results are in the form of a percentage inhibition relative to Nutlin-3. These values indicate the extent to which the compound inhibited the p53-Mdm2 interaction at the assay concentrations (10 μM ligand, 54.5 nM fluorescein-labelled p53 peptide and 41.5 nM Mdm2 L33E) relative to a racemic Nutlin-3 control.

Table 2.7: Comparison of fluorescence anisotropy results for 31 compounds based on the N1 scaffold with predictions of their relative binding affinity from thermodynamic integration

Side chains at each side chain position			Effect relative to Nutlin-3 in fluorescence anisotropy assay (%)	Measured rank	Predicted binding energy contribution relative to side chain 17 (ethyl) /kJ mol^{-1}				Predicted rank
1	2	3			Side chain 1	Side chain 2	Side chain 3	Total	
26	1	18	120	1	-8.18	-9.93	-3.89	-22.00	4
11	4	15	119	2	-7.29	-3.49	-3.96	-14.73	27
25	1	15	114	3	-7.48	-9.93	-3.96	-21.37	6
4	4	11	112	4	-10.93	-3.49	-7.10	-21.52	5
11	1	18	110	5	-7.29	-9.93	-3.89	-21.11	8
26	4	18	108	6	-8.18	-3.49	-3.89	-15.55	25
29	5	19	103	7	-3.28	-11.65	-5.76	-20.68	11
11	7	19	99.6	8	-7.29	-24.74	-5.76	-37.79	1
18	29	24	97.2	9	-2.21	-8.57	-6.67	-17.45	20
11	2	19	94.7	10	-7.29	-6.77	-5.76	-19.82	13
11	24	19	93.7	11	-7.29	-5.15	-5.76	-18.19	18
11	29	15	92.4	12	-7.29	-8.57	-3.96	-19.82	14
24	29	18	86.4	13	-8.88	-8.57	-3.89	-21.35	7
11	3	19	85.8	14	-7.29	-9.21	-5.76	-22.26	3
11	19	11	84.0	15	-7.29	-0.66	-7.10	-15.05	26
29	26	19	77.8	16	-3.28	-11.90	-5.76	-20.93	9
11	11	19	77.5	17	-7.29	-4.40	-5.76	-17.45	21
26	4	19	77.3	18	-8.18	-3.49	-5.76	-17.42	22
25	4	19	75.0	19	-7.48	-3.49	-5.76	-16.72	23
24	4	19	74.2	20	-8.88	-3.49	-5.76	-18.13	19
25	20	15	73.0	21	-7.48	-7.77	-3.96	-19.21	15
19	5	19	72.1	22	-1.65	-11.65	-5.76	-19.05	16
11	25	15	71.3	23	-7.29	-7.05	-3.96	-18.29	17
19	24	19	70.8	24	-1.65	-5.15	-5.76	-12.55	28
11	25	19	65.7	25	-7.29	-7.05	-5.76	-20.09	12
11	21	19	49.8	26	-7.29	10.48	-5.76	-2.57	30
11	20	19	44.9	27	-7.29	-7.77	-5.76	-20.82	10
11	18	19	38.7	28	-7.29	-2.51	-5.76	-15.56	24
11	28	19	33.4	29	-7.29	-9.41	-5.76	-22.46	2
11	27	19	31.6	30	-7.29	7.56	-5.76	-5.49	29
11	22	19	5.98	31	-7.29	21.16	-5.76	8.12	31

Fluorescence anisotropy results were supplied by Kerya long, University of Leeds. The overall change in binding energy for modification of all three the side chains was assumed to be equal to the sum of the changes associated with modification of each side chain individually. Both the practical and predicted rankings are from the strongest binding compound (1) to the most weakly binding (31).

In Figure 2.17 (p152) the predictions and experimental results are graphically compared. A high magnitude Pearson's rank correlation coefficient of -0.58 was observed. However, the results are not normally distributed which makes it more appropriate to consider the Spearman's Rank, which is only 0.40.

The observed correlation is small in magnitude. At the 5% significance level, the two-tailed critical value for Spearman's Rank with 30 degrees of freedom (n=31) is 0.35. Consequently, this rank of 0.40 would normally be considered a significant result. However, the results are not independent because of the way that they have been calculated. (As described above, the predicted relative binding energy of each oligobenzamide was calculated by way of a series of summations, specifically, addition of the ΔΔG values associated with small modifications to generate ΔΔG contributions for each side chain, followed by addition of these three side chain contributions to generate a total ΔΔG for each molecule.) Consequently, the Spearman's Rank might be less statistically significant than it appears. In conclusion, the results suggest that TI is not a useful method of ranking oligobenzamides with N-linked side chains.

There is a moderate correlation between the molecular weight of the 31 ligands and their affinity for Mdm2 as indicated by the fluorescence anisotropy assay. Spearman's rank is 0.61, suggesting that molecular weight is a much better predictor of affinity than the TI method. The same magnitude of correlation is observed using predicted logP. The lowest molecular weight compounds are, in order of increasing size, compounds 28, 26, 31 and 24. The compounds with the lowest predicted logP are, in order, 26 (negatively charged), 31 (positively charged), 28, 24 and 30. These weakly binding compounds can be found in the top left of Figure 2.17A, the top right of Figure 2.17B and, later in the chapter where use of the MM-PBSA method is described, the top left of Figure 2.22 (p174).

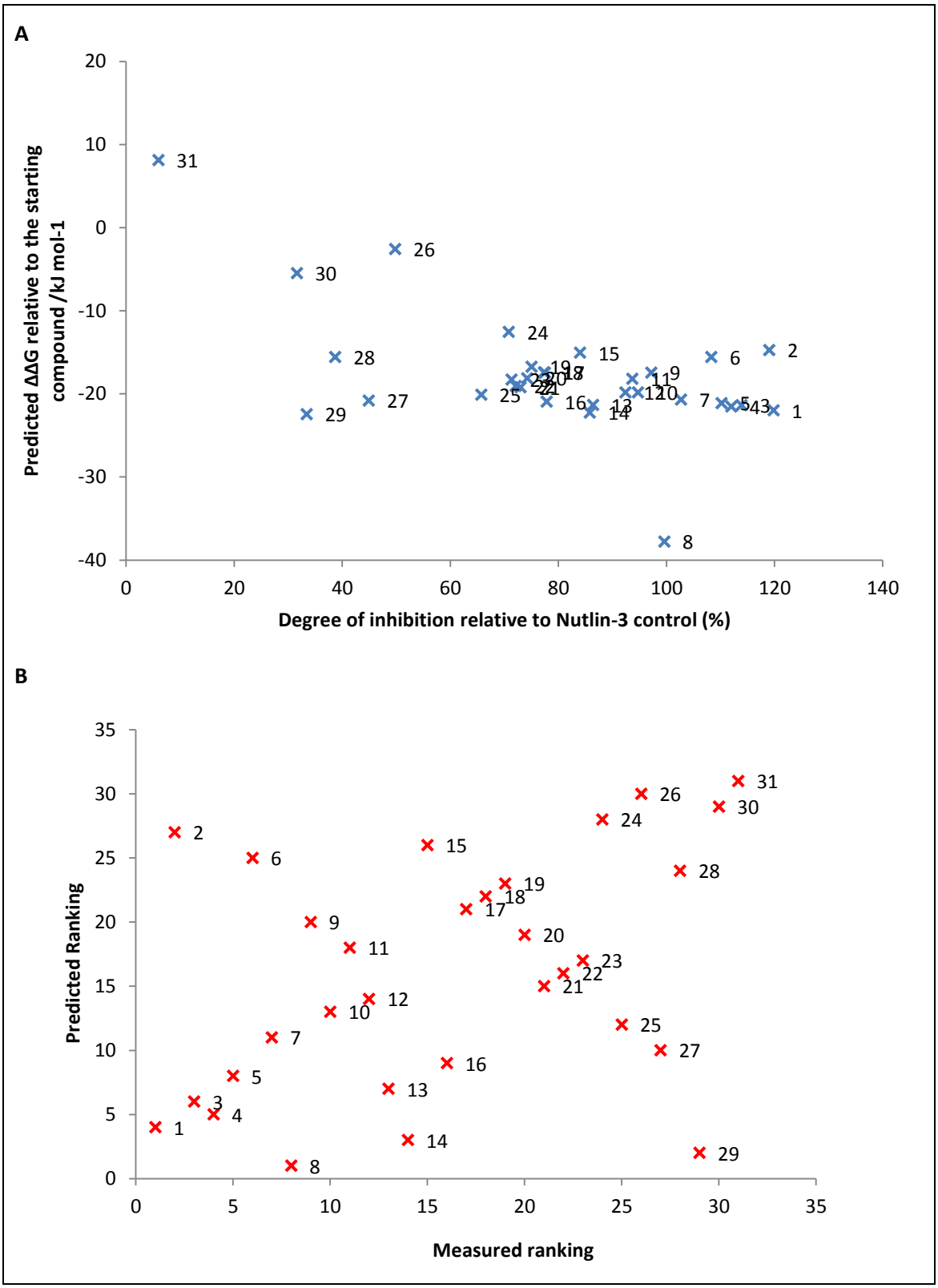


Figure 2.17: A comparison of predictions and experimental results for the 31 N-linked compounds ranked using thermodynamic integration. Each cross represents one of the 31 compounds tested. The numbers indicate the measured ranking of the compounds from 1 (the highest affinity) to 31 (the lowest). A) Comparison of the actual values. (A negative correlation is expected.) B) The same data but with the ranks plotted. It is more appropriate to consider Spearman's Rank because the data are not normally distributed. Here a positive correlation is expected because both sets of results were ranked from the most strongly binding (1) to the weakest binding (31).

2.3.4. DOCKING POSES OF 25,000 COMPOUNDS REVEAL THAT OLIGOBENZAMIDES CAN BIND IN MANY DIFFERENT POSES

2.3.4.1. Different oligobenzamides bind in different places

This section concerns the docking of oligobenzamides to investigate whether they all bind in the same, expected position, an assumption which the analysis until this point has relied upon. This assumption is necessary when using thermodynamic integration and, if invalid, could explain the poor correlation between predicted and experimental results observed above.

Gonzalez-Lopez de Turiso *et al.*[263] discuss crystal structures of two similar morpholinone inhibitors of the p53-Mdm2 interaction which feature the two compounds in significantly different binding poses. (Essentially, the groups in the Phe19 and Leu26 pockets are swapped over.) Small changes in the structure of an oligobenzamide could similarly lead to very different binding poses. Fuller *et al.*[172] reports the stable binding of an oligobenzamide based on the O1 scaffold both parallel and antiparallel to the p53-helix orientation.

To investigate where oligobenzamides with different combinations of side chains and different scaffolds bind, 25,000 compounds, roughly 5,000 for each of the N1, O1, O2, S1 and S2 scaffolds (N1: 5047, O1:4933, S1:5042, O2:5043, S2:4935), were docked with Autodock Vina and the pose positions were analysed. Table 2.8 shows the side chains used. The numbers correspond to the side chains shown in Figure 2.2 (p103).

Table 2.8: Side chains tested at each side chain position when docking 25,000 oligobenzamides to determine the effect of side chain selection on binding pose

Side chain position	Side chains tested at position	Number of side chains
1	4, 11, 13-15, 17-19, 21-30, 33, 36-40, 44	25
2	1-32, 34-46	45
3	11, 14, 15, 17-24, 30-32, 35-38, 40-43	22

See Figure 2.2 (p103) for the side chain structures. The side chains tested were selected manually on the basis of their size and similarity to the Phe19, Trp23 and Leu26 p53 side chains as described on p104.

Initially, the possibility of clustering all of the poses obtained was considered. The effect of different factors such as scaffold and side chain choice on the presence of poses in each cluster could then have been investigated. However, it was not immediately obvious how the differing connectivity of molecules should be treated in the clustering process. Furthermore, preliminary investigation revealed that the poses for different oligobenzamides, when combined, produced a broad and quite homogenous distribution of poses with little discernible clustering; while poses were focussed around the pockets of the p53 binding site and tended to be parallel or antiparallel to the p53 binding groove, they did not conform to one of a small number of clusters. Although poses of each oligobenzamide could be clustered in isolation, this did not simplify interpretation

of the disparate assortment of configurations obtained when the data for different compounds were consolidated.

To enable interpretation, the centre of mass of each side chain in each docked pose was used to evaluate the position of the side chain with regard to each of the Mdm2 binding pockets. Specifically, the distance between the side chain centre and the centre of mass of each p53 residue side chain in PDB crystal structure 1T4F ($Mdm2_{17\text{-}125}$ bound to a modified p53 peptide[37]) was calculated. If this distance was within a cut-off (3 or 4 Å as indicated next to each set of results below) then the side chain was considered to be in the pocket usually occupied by that residue in p53. (Only heavy atoms were used to calculate the centres of mass because hydrogen atoms are not present in X-ray crystal structures.) For further details see p111.

Table 2.9 below shows the most frequently occurring single side chain-pocket matches revealed by the pose analysis. For example, for the N1 scaffold, the most commonly occurring match is to find side chain 3 (the C-terminal side chain) in the pocket usually occupied by Phe19 of p53. For 99% of compounds, at least one of the poses had side chain 3 in this position and 54.5% of the time, the top pose for a compound satisfied the condition.

The results suggest that different scaffolds favour different binding poses. For example, while the side chain 3-Phe19 pocket pairing is the most common with the N1 scaffold, the side chain 2-Phe23 pairing is most common for the O1 and S1 scaffolds and for the O2 and S2 scaffolds the most common pairings are side chain 1 in the Trp23 pocket and side chain 2 in the Phe19 pocket respectively.

It appears that pockets 19 and 23 are usually occupied but it is difficult to interpret the single side chain-pocket match results below because they give little information about the pose. Table 2.10 below indicates the most commonly occurring pairs of matches. For example, it shows that 88.9% of scaffold N1 compounds have a pose with both side chain 2 in the Trp23 pocket and side chain 3 in the Phe19 pocket.

As shown, for all of the scaffolds, the most common pair of matches involves the Phe19 and Trp23 sites, a finding that suggests that these pockets are particularly important for binding, as would be expected, given their size. For all of the scaffolds, side chain 2 is one of the side chains in this top match. The second most common pair of matches always appears to involve the Leu26 pocket, consistent with the widely held view that the Phe19, Trp23 and Leu26 residues being those most important for p53 binding[26]. Adjacent side chains, particularly the middle and C-terminal side chains are seen to be able to occupy the Phe19 and Leu26 binding sites simultaneously to an extent sufficient to satisfy the 3 Å distance cut-offs.

Table 2.9: The most frequently occurring side chain to binding pocket matches for each scaffold

Scaffold	Rank	Matching	One pose (%)	Best pose (%)	Scaffold	Rank	Matching	One pose (%)	Best pose (%)
N1	1	SC3-19	99.0	54.5	S1	1	SC2-23	99.1	55.3
	2	SC2-23	98.5	57.7		2	SC2-26	91.0	19.8
	3	SC2-19	86.8	8.3		3	SC1-19	89.7	23.8
	4	SC2-26	83.6	10.6		4	SC1-23	84.0	14.3
	5	SC3-23	79.1	11.3		5	SC3-23	82.1	15.3
	6	SC3-26	73.4	4.2		6	SC2-19	79.4	5.7
	7	SC1-19	63.0	15.6		7	SC3-26	76.3	22.3
	8	SC1-22	60.2	0.3		8	SC3-19	72.5	16.7
	9	SC3-22	58.4	0.3		9	SC1-26	67.1	5.8
	10	SC2-20	38.0	2.2		10	SC3-22	58.6	1.8
O1	1	SC2-23	99.1	55.6	S2	1	SC2-19	95.6	29.8
	2	SC2-26	90.4	22.3		2	SC1-23	93.5	45.1
	3	SC1-19	90.2	26.6		3	SC2-26	91.6	35.5
	4	SC1-23	86.0	19.1		4	SC2-23	88.5	16.9
	5	SC2-19	84.6	8.1		5	SC3-23	88.2	22.2
	6	SC3-23	80.2	14.1		6	SC1-19	79.2	13.3
	7	SC3-19	76.4	14.7		7	SC3-19	79.0	5.7
	8	SC1-26	67.7	12.3		8	SC3-26	77.2	15.2
	9	SC3-26	65.4	7.9		9	SC1-26	73.3	11.3
	10	SC3-22	65.2	6.9		10	SC1-22	60.8	5.5
O2	1	SC1-23	94.7	42.6					
	2	SC3-23	92.7	32.6					
	3	SC2-26	92.4	38.0					
	4	SC2-19	91.8	16.1					
	5	SC3-19	83.0	9.9					
	6	SC2-23	81.3	15.1					
	7	SC3-26	71.4	5.2					
	8	SC1-26	68.8	15.2					
	9	SC1-19	64.8	2.9					
	10	SC1-22	49.9	2.1					

"SC1", "SC2" and "SC3" denote the 3 side chains of the ligand. The pockets in which they reside are indicated by the numbers "19", "22", "23" and "26", which refer to the residues of p53 Phe19, Leu22, Trp23 and Leu26 which, through their position in the PDB structure from which Mdm2 was extracted for docking, define Mdm2 binding pockets. For more details see p111. A cut-off of 3 Å was used for matching. The "one pose" column shows the percentage of compounds for which there was one pose with its side chains in the position shown and the "best pose" column shows the percentage of compounds for which the top pose satisfied the condition. For example, 99% of the compounds based on the N-linked scaffold, yielded, when docked, at least one pose in which the third (C-terminal) side chain occupied the Mdm2 binding pocket normally occupied by p53 residue Phe19. The structures of the five scaffolds (N1, O1, O2, S1 and S2) are shown in Figure 2.1 (p102) and the side chains tested at each side chain position are shown in Table 2.8 above, where the numbers correspond to the structures in Figure 2.2 (starting on p103).

Table 2.10: The most frequently occurring pairs of side chain to binding pocket matches for each scaffold

Scaffold	Rank	Matching		One pose (%)	Best pose (%)
N1	1	SC2-23	SC3-19	88.9	39.7
	2	SC2-26	SC3-19	62.5	6.8
	3	SC2-23	SC2-26	57.7	4.0
	4	SC2-19	SC3-23	56.4	6.5
	5	SC1-22	SC2-23	44.8	0.2
	6	SC1-19	SC2-23	43.5	6.0
	7	SC2-23	SC3-22	38.0	0.3
	8	SC2-19	SC3-26	35.0	2.5
	9	SC3-23	SC3-26	28.1	1.5
	10	SC2-26	SC3-23	25.5	1.2
O1	1	SC1-19	SC2-23	73.7	22.2
	2	SC2-23	SC2-26	67.7	11.9
	3	SC2-23	SC3-19	64.8	11.7
	4	SC1-19	SC2-26	49.4	8.5
	5	SC1-26	SC2-23	47.6	9.5
	6	SC1-26	SC3-22	44.5	5.3
	7	SC1-22	SC2-23	43.8	2.5
	8	SC2-23	SC3-22	43.0	4.7
	9	SC1-23	SC2-19	42.9	2.8
	10	SC2-19	SC3-23	38.4	3.0
O2	1	SC2-19	SC3-23	69.2	4.1
	2	SC2-26	SC3-19	67.3	7.5
	3	SC2-23	SC2-26	59.8	6.6
	4	SC1-26	SC2-19	59.7	9.2
	5	SC2-23	SC3-19	57.7	6.9
	6	SC1-23	SC2-26	57.4	15.8
	7	SC2-19	SC3-26	54.1	1.7
	8	SC1-23	SC2-19	53.5	7.6
	9	SC1-19	SC2-23	47.5	2.4
	10	SC1-23	SC1-26	46.8	9.6
S1	1	SC1-19	SC2-23	74.0	20.0
	2	SC2-23	SC2-26	65.4	6.7
	3	SC2-23	SC3-19	61.9	14.9
	4	SC2-23	SC3-26	48.8	15.1
	5	SC1-19	SC2-26	42.9	2.3
	6	SC1-19	SC3-26	41.9	11.9
	7	SC1-26	SC3-22	40.7	1.1
	8	SC1-23	SC2-19	39.2	1.0
	9	SC1-26	SC2-19	36.6	0.7
	10	SC2-26	SC3-23	36.5	5.8
S2	1	SC2-19	SC3-23	71.7	13.3
	2	SC2-23	SC2-26	64.9	8.2
	3	SC2-19	SC3-26	64.6	10.1
	4	SC1-26	SC2-19	62.7	10.6
	5	SC2-23	SC3-19	61.4	3.4
	6	SC1-19	SC2-23	61.3	7.7
	7	SC3-23	SC3-26	61.0	10.1
	8	SC1-23	SC2-19	60.1	11.7
	9	SC2-26	SC3-19	57.6	3.1
	10	SC1-23	SC2-26	55.8	18.5

"SC1", "SC2" and "SC3" denote the 3 side chains of the ligand. The pockets in which they reside are indicated by the numbers "19", "22", "23" and "26", which refer to the residues of p53 Phe19, Leu22, Trp23 and Leu26 which, through their position in the PDB structure from which Mdm2 was extracted for docking, define Mdm2 binding pockets. For more details see "Pose recognition" on p111. A cut-off of 3 Å was used for identification of the side chains in each pocket. The "one pose" column shows the percentage of compounds for which there is one pose with its side chains in the positions indicated and the "best pose" column shows the percentage of compounds for which the top pose satisfies the condition. For example, 88.9% of the compounds based on the N-linked scaffold, yielded, when docked, at least one pose in which the second (middle) side chain occupied the Mdm2 binding pocket normally occupied by p53 residue Trp23 and the third (C-terminal) side chain occupied the Mdm2 binding pocket normally occupied by p53 residue Phe19.

Table 2.11: The most frequently occurring sets of three side chain to three binding pocket matches for each scaffold

Scaffold	Rank	Side chains			One pose (%)	Best pose (%)
		1	2	3		
N	1	26	19	22	16.5	0.00
	2	23	19	22	14.6	0.00
	3	22	19	26	13.8	0.50
	4	19	23	26	13.5	0.22
	5	22	19	23	11.4	0.38
	6	19	20	26	10.4	0.42
	7	26	23	22	9.4	0.00
	8	22	23	26	7.7	0.26
	9	22	23	20	7.3	0.00
	10	23	23	22	6.9	0.00
O1	1	26	23	22	44.8	6.49
	2	26	19	22	41.8	3.14
	3	23	19	22	39.9	2.39
	4	19	23	26	22.5	2.49
	5	26	23	19	22.3	3.24
	6	23	23	22	18.1	1.70
	7	22	23	26	16.1	0.53
	8	22	19	26	12.9	0.79
	9	26	20	22	11.3	0.02
	10	23	20	22	9.6	0.02
O2	1	22	19	23	55.2	0.87
	2	22	19	26	54.8	1.03
	3	26	19	22	49.2	3.03
	4	23	19	22	37.9	1.57
	5	26	23	22	19.8	1.03
	6	22	23	26	16.8	0.65
	7	22	23	23	12.2	0.30
	8	23	23	22	8.9	0.18
	9	22	20	23	3.6	0.04
	10	22	20	26	3.5	0.04
S1	1	26	19	22	45.6	1.35
	2	19	23	26	44.1	11.66
	3	23	19	22	42.6	0.99
	4	26	23	22	35.2	1.63
	5	26	23	19	30.1	5.63
	6	22	23	26	25.9	3.27
	7	23	23	22	19.2	0.56
	8	22	19	26	17.7	1.07
	9	22	19	23	13.0	0.58
	10	26	20	22	10.7	0.00
S2	1	22	19	26	58.8	6.42
	2	22	19	23	58.2	6.14
	3	26	19	22	47.1	1.54
	4	23	19	22	39.3	1.46
	5	26	23	22	16.9	0.04
	6	22	23	26	8.8	0.67
	7	23	23	22	7.2	0.04
	8	22	20	23	7.1	0.04
	9	22	20	26	6.8	0.04
	10	22	23	23	4.5	0.18

"SC1", "SC2" and "SC3" denote the 3 side chains of the ligand. The pockets in which they reside are indicated by the numbers "19", "22", "23" and "26", which refer to the residues of p53 Phe19, Leu22, Trp23 and Leu26 which, through their position in the PDB structure from which Mdm2 was extracted for docking, define Mdm2 binding pockets. A cut-off of 4 Å was used for identification of the side chains in each pocket. For more details see p111. The "one pose" column shows the percentage of compounds for which there is one pose with its side chains in the arrangement described. For example, 16.5% of the compounds based on the N-linked scaffold, yielded, when docked, at least one pose in which the first side chain was in the pocket usually occupied by p53 residue Leu26, the second was in the pocket usually occupied by Phe19 and the third was in the pocket usually occupied by p53 residue Leu22. The "best pose" column shows the percentage of compounds for which the top pose satisfied the condition.. For further details see Table 2.9 (p155).

Table 2.11 above shows the three side chain-three pocket matches observed. While the Phe19, Trp23 and Leu26 pockets feature prominently in the side chain-pocket pairings above, the results in this last table indicate that compounds often do not occupy all three of these sites in a

particular pose. They suggest that a pose such as that shown in Figure 2.8 (p129), where the three side chains of an oligobenzamide occupy the Phe19, Trp23 and Leu26 binding sites, might be a rare, idealised conformation that reflects the similarity of the benzyl, napthyl and isopropyl side chains of the oligobenzamide shown to the phenylalanine, tryptophan and leucine p53 side chains being mimicked.

The 22nd residue of p53 is leucine. The Leu22 pocket appears frequently in Table 2.11 above; however, this might reflect its proximity to the centre of the binding site as opposed to any significant involvement in ligand binding.

2.3.4.2. Logistic regression reveals that the relationship between side chain properties and binding pose is not straightforward

The results above indicate that the choice of scaffold influences the binding pose and that oligobenzamides can potentially bind to Mdm2 in many different ways. This section describes the application of logistic regression[252] (p292) to the docking results introduced above to investigate the effect of side chain choice on binding pose.

Five Boolean questions were formulated which could be asked of a particular compound (Figure 2.18). The first, for example, was whether there was a pose where the middle side chain was in the Trp23 binding pocket of Mdm2. This question was proposed because, of the 25000 compounds tested, 52% (13090 compounds) did have such a pose and 48% (11910) did not, similar numbers, indicating that both scenarios are common, the outcome dependent on the choice of scaffold and side chains. The subsequent questions were similarly selected to provide large, roughly equal numbers of both positive and negative outcomes. While the first question, described above, was applicable to every compound, the others were not, instead being dependent on a particular orientation or conformation. However, these subsequent questions were selected to be applicable to a large proportion of the molecules.

The five questions (illustrated graphically in Figure 2.18):

1. Is there a pose where the middle side chain is in the Trp23 binding pocket of Mdm2?
2. If there are one or more poses where the middle side chain is in the Trp23 binding pocket and another side chain is in the Phe19 pocket, is this other side chain side chain 1 (the parallel conformation) or side chain 3 (the antiparallel conformation)? (Molecules with both a parallel and an antiparallel pose were excluded.)
3. If there are one or more poses where binding is parallel and involves binding in the Phe19 and Trp23 pockets, do side chains 1 and 2 bind or side chains 2 and 3? (Cases where both poses occurred were ignored.)

4. If there are one or more poses where binding is antiparallel and involves binding in the Phe19 and Leu26 pockets with adjacent side chains (straddling the Trp23 pocket), do side chains 1 and 2 or side chains 2 and 3 bind? (Cases where both poses occurred were ignored.)
5. Do any of the core parts of the compound occupy a binding site? The core was divided up into 3 sections with the amide bonds defining where each section ended and the next began. The heavy atoms were used to find the centre of mass of each section (near the middle of the benzene ring). As with the side chains, each of the core elements were assumed to be in a pocket if the distance of its centre of mass from the binding pocket centre was less than a 3 Å cut-off.

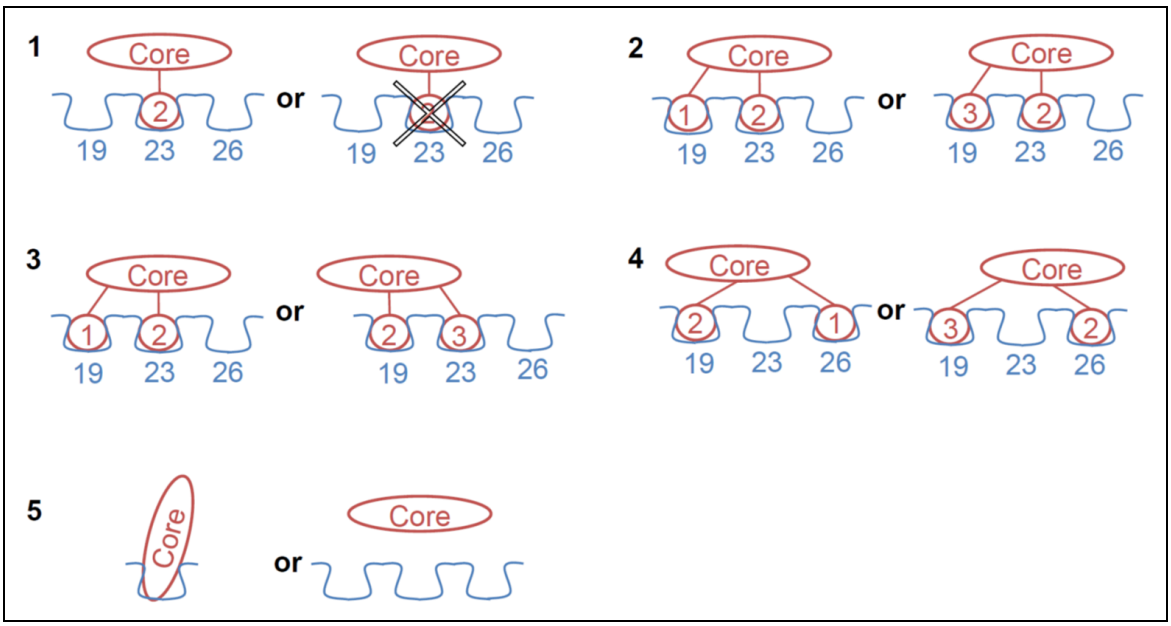


Figure 2.18: A graphical summary of the 5 questions posed when investigating how side chains affect binding pose. For the purpose of logistic regression, the outcome on the left was treated as success and the outcome on the right was treated as failure. The blue numbers "19", "23" and "26" indicate the pockets normally filled by p53 residues Phe19, Trp23 and Leu26 respectively. Q1) Is the Trp23 pocket occupied by the middle side chain? Q2) If the middle side chain is in the Trp23 binding pocket and another side chain is in the Phe19 pocket, is this other side chain side chain 1 (making the compound parallel to the p53 helix direction) or side chain 3 (antiparallel)? Q3) If binding is parallel and involves binding in the Phe19 and Trp23 pockets, do side chains 1 and 2 or side chains 2 and 3 occupy the pockets? Q4) If binding is antiparallel and involves binding in the Phe19 and Leu26 pockets with adjacent side chains (straddling the Trp23 pocket), do side chains 1 and 2 or side chains 2 and 3 bind? Q5) Is the scaffold bound in any of the binding pockets?

For analysis, the molecular weight, logP, number of heteroatoms and number of aromatic rings were calculated for each side chain and the molecules as a whole, as described in the methods section (p104). Binary logistic regression was performed to generate, for each of the five questions above, for each scaffold, a model to predict, from the properties, the answer to the question. For example, a model was fitted for each scaffold to predict whether side chain 2

would be found in the Trp23 binding pocket, based on the molecular weight (MW), logP, number of heteroatoms (H) and the number of aromatic rings (R) in side chain 2.

A binary logistic regression model predicts either "success" or "failure". For question 1 above, the "yes" outcome was treated as success and the "no" outcome was treated as failure. For the other questions, the former outcome given above was treated as success and the latter as failure. Consequently, in Figure 2.18, the success outcomes are shown on the left and the failure outcomes are shown on the right.

Table 2.12: Statistics for logistic regression models generated for each scaffold to predict whether the middle side chain bound in the Trp23 binding pocket

Scaffold	Property of side chain 2	β	SE	Wald	Significance $P(H_0)$	e^{β}
N1	MW	0.010	0.002	31.237	<0.0005	1.010
	logP	0.304	0.068	20.074	<0.0005	1.356
	Heteroatoms	0.029	0.013	4.978	0.026	1.029
	Rings	0.104	0.074	1.982	0.159	1.109
O1	MW	0.006	0.002	13.068	<0.0005	1.006
	logP	0.612	0.074	69.133	<0.0005	1.844
	Heteroatoms	0.047	0.013	14.170	<0.0005	1.048
	Rings	-0.022	0.066	0.113	0.737	0.978
S1	MW	0.004	0.002	5.609	0.018	1.004
	logP	0.581	0.070	68.969	<0.0005	1.787
	Heteroatoms	0.038	0.012	11.001	0.001	1.039
	Rings	0.058	0.062	0.863	0.353	1.059
O2	MW	0.016	0.002	60.346	<0.0005	1.017
	logP	0.695	0.126	30.323	<0.0005	2.004
	Heteroatoms	0.027	0.013	4.351	0.037	1.027
	Rings	-0.292	0.066	19.642	<0.0005	0.747
S2	MW	0.016	0.002	94.326	<0.0005	1.016
	logP	0.197	0.080	6.082	0.014	1.217
	Heteroatoms	0.044	0.011	15.818	<0.0005	1.045
	Rings	-0.214	0.059	13.336	<0.0005	0.807

β is the coefficient of the model for the property indicated in the second column from the left. The properties are as follows. "MW" is the molecular weight of the middle side chain. The "logP" was the XlogP (p104). "Heteroatoms" is the number of heteroatoms in this side chain. "Rings" is the number of aromatic rings. In the "SE" column are shown the standard errors on the β values. These standard errors are required for calculation of the Wald statistics shown in the "Wald" column. The "Significance" column gives the probability of a Wald statistic greater than the one obtained being generated due to chance if the null hypothesis (β=0) is true. e^{β} is the factor by which the odds ratio is increased when the value of the variable in question is increased by 1. For further details see p292.

Table 2.12 shows details of the fitted models for question 1. In the β column of Table 2.12 are the β parameters for the five models (one for each scaffold). Each describes the effect of a factor (for example, molecular weight) on the probability of the success outcome. Values greater than 0 indicate a higher value of the factor increases the probability of the success answer and values less than 0 indicate the opposite. Because each β value in the model has an exponential effect on

the predicted probability (as shown in Equation 6.9, p292) its effect can be more easily visualised by calculating e^{β}, the multiplier effect which decreasing the factor (such as the molecular weight) by 1 has on the odds ratio of the success outcome (1/P-1 where P is the probability of success) (p292).

It is necessary to test whether a β parameter is significantly different from 0. To check for statistical significance, the standard error on β can be calculated. A Wald statistic is similar to a t-value; it can be used to determine the statistical significance of an estimated β, given its standard error. The significance probability associated with each calculated Wald statistic is shown in the penultimate column of Table 2.12. Where the probability is less than 0.05, this indicates that, at the 5% significance level, the true β is not equal to 0 for that factor (the factor has an effect on the outcome). For full details of logistic regression and the Wald statistic see p292.

Several trends can be seen in the logistic regression results above. For all five scaffolds, increasing the molecular weight and hydrophobicity of the middle side chain significantly increases the probability of binding in the middle position. (The e^{β} values for molecular weight (MW), logP and the number of heteroatoms are greater than 1 for every scaffold.)

Models were also generated for questions 2 to 5. The e^{β} values are shown for each model in Table 2.13 below in the cases where the values were statistically significant at the 5% significance level.

Question 1 concerned the middle side chain, side chain 2. Consequently, the factors for the question 1 models (one model for each scaffold) were the molecular weight, logP, number of heteroatoms in and number of rings of side chain 2. Because question 2 concerns which way around side chains 1 and 3 are (whether the oligobenzamide binds parallel or antiparallel to the p53 orientation) the factors used in the models were the difference in molecular weight, difference in logP, difference in the number of heteroatoms and difference between the number of rings in these side chains. For questions 3, 4 and 5 the properties of the whole molecule were used as factors.

Question 2 asked whether binding was parallel or antiparallel to the p53 binding orientation. For this question, e^{β} is less than 1 in the molecular weight (MW) row for each scaffold. This indicates that a larger side chain at position 3 relative to that at position 1 favours binding in the antiparallel conformation. This conformation places the large side chain in the Phe19 binding site, which is deeper than the Leu26 site so could better accommodate a large side chain. e^{β} is significantly greater than 1 in the "rings" rows for the N1, O2 and S2 scaffolds, indicating that the more aromatic rings there are in side chain 3 relative to side chain 1, the more likely parallel

binding is (even though these increase the size of the side chain). The rigidity of aromatic side chains might prevent them from filling the Phe19 pocket in the antiparallel pose.

Table 2.13: e^{β} values from all of the logistic regression analyses where the result is significant at the 5% significance level ($P(H_0)<0.05$)

		e^{β} of property for each question				
	Question	1	2	3	4	5
Property describes / Scaffold	Property	Side chain 2	Difference SC3 –SC1	Whole molecule	Whole molecule	Whole molecule
N1	MW	↑ 1.010	↓ 0.972	↓ 0.996		↓ 0.988
	logP	↑ 1.356		↓ 0.868		
	Heteroatoms	↑ 1.029	↑ 1.088	↑ 1.207		
	Rings		↑ 1.227	↑ 1.341	↑ 1.285	↓ 0.539
O1	MW	↑ 1.006	↓ 0.985			↓ 0.996
	logP	↑ 1.844		↓ 0.753		↓ 0.866
	Heteroatoms	↑ 1.048	↑ 1.074		↑ 1.129	↓ 0.915
	Rings			↓ 0.745	↑ 1.381	↓ 0.783
S1	MW	↑ 1.004	↓ 0.983		↑ 1.008	↓ 0.997
	logP	↑ 1.787		↓ 0.689	↓ 0.823	↓ 0.878
	Heteroatoms	↑ 1.039	↑ 1.074			
	Rings			↓ 0.793	↑ 1.364	↓ 0.748
O2	MW	↑ 1.017		↓ 0.984		↓ 0.995
	logP	↑ 2.004				
	Heteroatoms	↑ 1.027	↓ 0.868	↑ 1.207		
	Rings	↓ 0.747	↑ 1.618	↓ 0.659	↑ 1.241	↓ 0.823
S2	MW	↑ 1.016	↓ 0.988	↓ 0.995		↓ 0.993
	logP	↑ 1.217		↓ 0.751		↑ 1.147
	Heteroatoms	↑ 1.045	↓ 0.821			
	Rings	↓ 0.807	↑ 2.112	↓ 0.679		↓ 0.895

e^{β} values indicate the multiplier effect an increase of 1 in the value of the property has on the predicted probability of a positive prediction. Consequently, e^{β} values greater than one indicate a positive (but not a linear) correlation between the property and the probability and values less than 1 indicate a negative correlation.

Question 3 concerned compounds binding parallel using the Phe19 and Trp23 pockets. It asked whether they bound in these pockets using side chains 1 and 2 or side chains 2 and 3. The overall properties of the molecule were used as factors. More hydrophobic molecules appear to be more likely to bind in the latter conformation because e^{β} is less than 1 in the logP row for every scaffold and for scaffolds N1 and O2, e^{β} is less than 1 in the heteroatom rows. (The presence of heteroatoms is indicative of more polar functional groups.) The Leu26 pocket is not deep. If the side chains are all hydrophobic, this pocket might not be used because it would not sufficiently bury the side chain in it, leading to entropically unfavourable ordering of the solvent molecules around the exposed part.

Larger molecules appear to be more likely to bind with side chains 1 and 2. (e^{β} is less than 1 in the MW row for scaffolds N1, O2 and S2 and e^{β} is less than 1 in the aromatic rings row for scaffolds O1 and S1.) Larger side chains were tested at side chain positions 1 and 2 than at position 3 due to the increased size of Phe19 and Trp23 compared to Leu26 in p53 (imitated by side chains 1, 2 and 3 respectively in the paradigmatic model of oligobenzamide peptidomimicry). Consequently, the results suggest that if large side chains are used at positions 1 and 2, the oligobenzamide binds with these side chains in the Phe19 and Trp23 binding pockets to take advantage of the potential to fit closely into these large pockets, maximising the van der Waals forces between the protein and small molecule.

As shown in Figure 2.18, question 4 was similar to question 3 but for parallel conformations straddling the central binding pocket. Here the effect is clear; the number of aromatic rings appears to be the only consistently important factor. The presence of rings favours binding with side chains 1 and 2 over binding with side chains 2 and 3. Aromatic side chains were primarily tested at positions 1 and 2 so these results suggest that if there are big, aromatic groups at these positions, then straddling of the Trp23 binding site is not possible, probably due to steric hindrance. Binding with side chains 2 and 3 would leave an aromatic residue at side chain position 1 free to interact with an aromatic residue of Mdm2 such as His96 or Tyr100, which would be nearby.

The Nagelkerke R^2 values[264] for a fitted model indicate the degree to which it explains the variation in the observed results. They range from 0 (no explanation) to 1 (sufficient to fully describe the observed results). The Nagelkerke R^2 values for question 4 (N1: 0.02, O1: 0.07, S1: 0.11, O2: 0.01, S2: 0.01) indicate that the models do not explain the results well. They suggest that other properties which were not considered are much more important for determining the pose. The two possible poses for this question shown in Figure 2.18 are very different so many factors could determine which one prevails.

Finally, with regard to question 5, the core appears to be less likely to be involved in binding if the side chains are particularly bulky, particularly with the N1 scaffold, where the Nagelkerke R^2 (0.298) suggests that the model explains 30% of the observed variation in the results. A possible explanation is that it is only energetically favourable for the side chains to bind instead of the core if they are large enough to fit closely into the binding pockets and not leave a void in which there are no ligand atoms to form van der Waals interactions with the protein.

Figure 2.19 below summarises all of the logistic regression results.

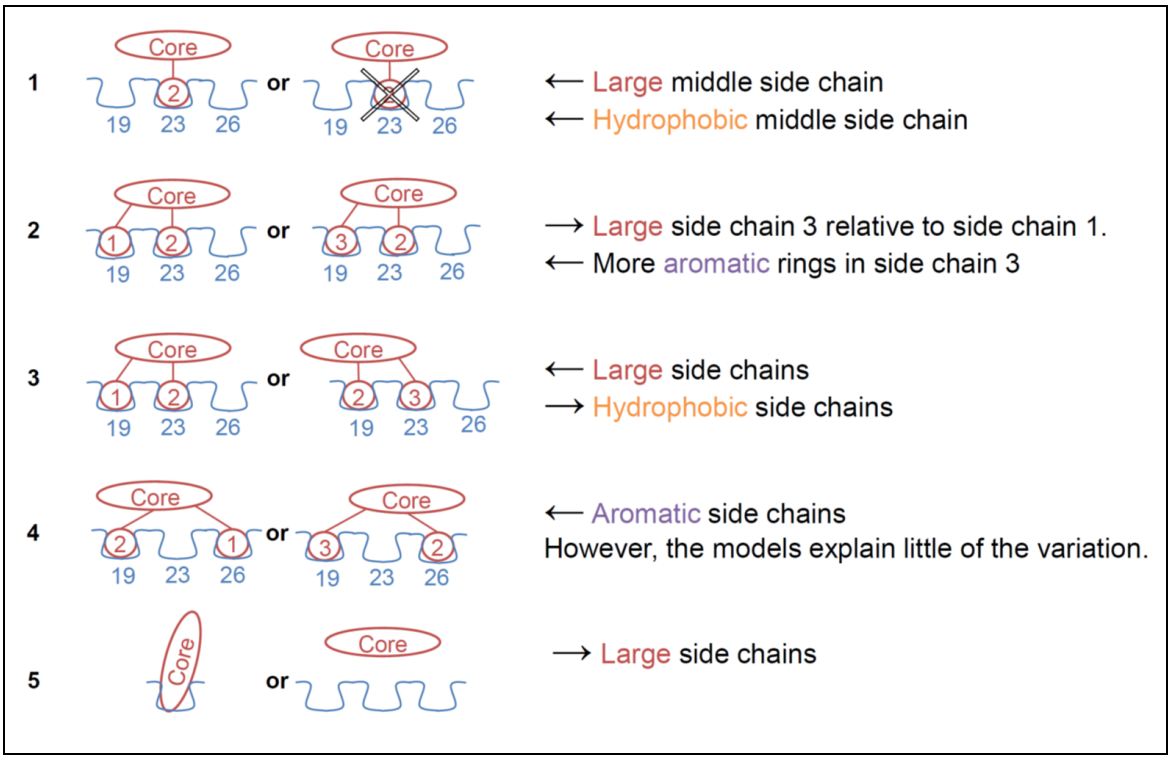


Figure 2.19: The effect of different side chain properties on which of two possible conformations are taken up by an oligobenzamide. The black arrows indicate how the relative probability of the two poses changes. For example, with regard to pair of poses 1, a large middle side chain favours the left pose relative to the pose on the right.

If oligobenzamides bind in many different ways to Mdm2, as suggested by the results of this analysis, this could be partly responsible for the poor correlation between thermodynamic integration predictions and experimental results (p151) because TI relies on the compounds being compared binding in similar positions.

It has so far been assumed that side chains each make independent contributions to the binding energy but if the choice of side chains has a large impact on binding position then this would suggest that binding affinity is a complex function of the combination of side chains rather than the sum of independent factors. TI is not suitable where side chains do not act independently because it uses significant computing power and this makes it impractical to use it to assess every member of a large combinatorial library.

TI is best suited to assessing the effect of a small change to a molecule as opposed to ranking a whole library. Furthermore, to assess the performance of a method, the results of all of the side chains should ideally be completely independent and this is difficult to achieve using TI.

2.3.5. MM-GBSA AND MM-PBSA RESULTS

2.3.5.1. Binding energies predicted by implicit solvent methods correlate better with experimental results

In this section, the challenge of predicting the relative affinities of the 31 N-linked compounds introduced on p142 is revisited. A moderate correlation (with a Spearman's rank of 0.55 to 0.60) is shown to exist between the measured ranking and the ranking predicted by both the MM-PBSA and MM-GBSA methods. The correlation is shown to be robust to changes in the number of docking poses used, the ionic strength, whether a separate trajectory is run of the free ligand and the set of Amber Born radii used (in the case of the MM-GBSA method).

As described above (p151), thermodynamic integration results did not correlate well with measured affinities. The docking results above (p153) suggest that this could be because an assumption necessary for the TI approach to be used, specifically that oligobenzamides all bind in a similar place, is invalid. The MM-GBSA and MM-PBSA methods (introduced on p86) can be used to rank a series of compounds using as little as one trajectory per compound. Consequently, unless the trajectories are very long, a variety of different poses can be tested for every member of a large library. Because each side chain combination can be investigated, there is no need to assume that the oligobenzamides being compared all have a similar pose and study each side chain in isolation.

Development of a fast method of testing a large number of compounds could be more useful than the development of a slower, more reliable method of affinity prediction because the affinity of a small number of compounds can be assessed accurately in the laboratory. Effort is underway to automate oligobenzamide production using microwave synthesis[245]. Furthermore, assay methods such as the fluorescence polarisation assay developed by the collaborating chemistry group[42] and a FRET assay described in the next chapter can potentially be performed with large numbers of compounds on a microwell plate. These methods could soon allow large scale oligobenzamide production and testing in the laboratory, making a slow computational method obsolete.

To assess the ability of the MM-GBSA and MM-PBSA methods to reproduce experimental results, they were used to predict the binding affinities of the 31 compounds for which fluorescence polarisation experimental data was available (introduced on p142). Simulations were started from the docking poses with the highest scores from each of three docking repeats.

To investigate the rate at which simulations of oligobenzamides with N-linked side chains equilibrated, the RMSDs for the first part of 50 randomly selected trajectories were studied (Figure 2.20). By visual inspection of individual trajectories from this set in isolation, it was decided that, in the majority of cases, equilibration had occurred after 250 ps (that the RMSD

appeared to be fluctuating around a constant value) and this was the equilibration time used for subsequent work. Inspection of Figure 2.20 suggests that the RMSDs could still be increasing after 250 ps but this could be because viewing the trajectories overlayed is similar to viewing an average RMSD. If the mean RMSD of multiple trajectories involving different ligands is plotted, equilibration appears to take longer than if individual trajectories from the same set are inspected. This is because the average RMSD keeps increasing until every single simulation in the set has equilibrated, which, given variation in the required equilibration time for systems containing different compounds, could be significantly longer than most simulations require.

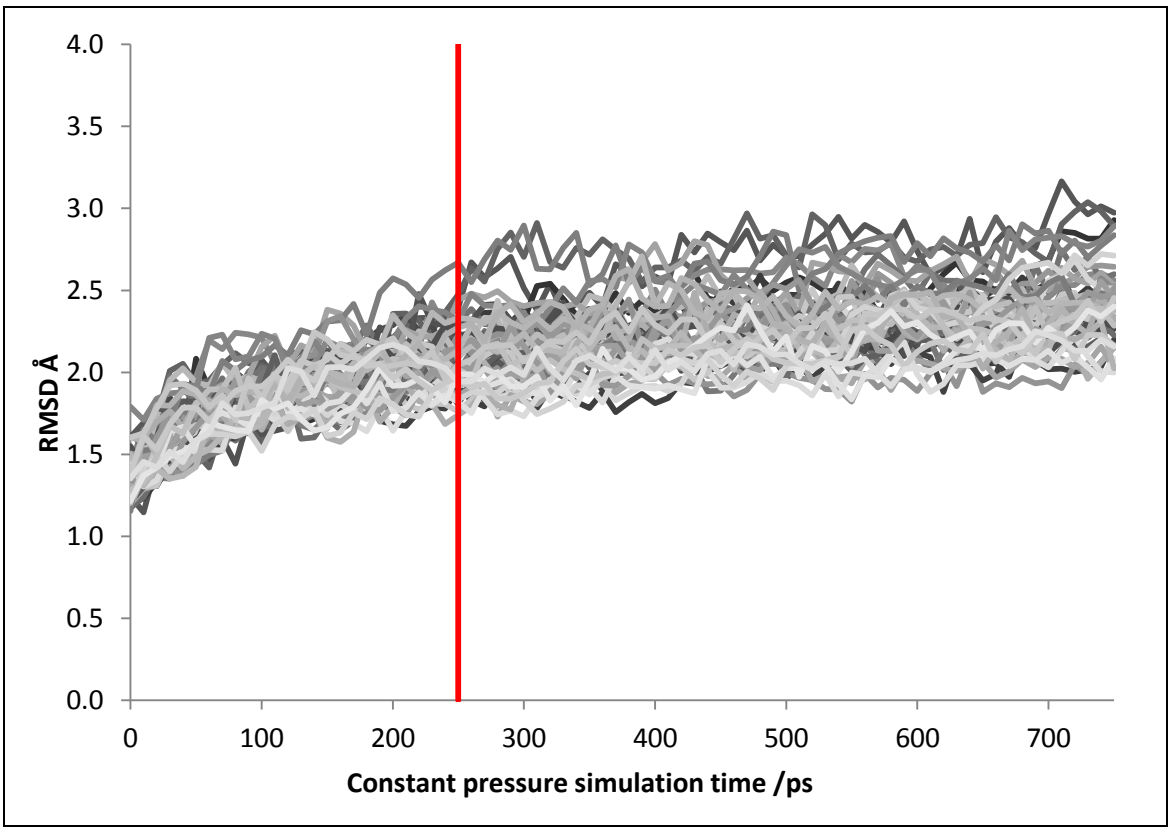


Figure 2.20: Equilibration of compounds based on the N-linked scaffold: RMSDs of the protein and ligand atoms in a random selection of 50 trajectories of 573 different molecules bound to Mdm2. The RMSDs are relative to the structure prior to any simulation. Some equilibration occurred during constant volume simulation prior to t=0. Water and ions were removed prior to analysis. Snapshots were aligned prior to the RMSD calculation (but after removal of water and ions) to account for rotation of the entire system. The red line at 250 ps indicates the point after which snapshots were used for MM-PBSA and MM-GBSA calculations.

After equilibration at constant volume, simulations were run for 750 ps at constant pressure. As discussed above, the first 250 ps were discarded leaving 500 ps, from which MM-PBSA and MM-GBSA results were derived. Longer simulations could have been carried out but the desire was to use the same simulation lengths as would be used to investigate large numbers of compounds later.

Table 2.14: The predicted free energies of binding for 31 compounds based on the N1 scaffold, generated using the MM-GBSA and MM-PBSA methods

	GBSA predictions /kJ mol^{-1}					PBSA predictions /kJ mol^{-1}				
Trajectory Method	Single	Double	Triple	Single	Single	Single	Double	Triple	Single	Single
IS	0	0	0	0.35	0	0	0	0	0.35	0
Side chains	igb=1				igb=5	indi=1				indi=4
4-4-11	-185.9	-172.3	-192.2	-182.9	-169.6	-151.7	-138.6	-160.5	-151.9	-500.1
11-1-18	-188.1	-181.1	-181.9	-185.3	-173.3	-155.8	-149.0	-148.0	-156.1	-497.0
11-2-19	-195.6	-189.4	-199.8	-192.5	-180.5	-158.6	-152.9	-162.4	-158.7	-551.6
11-3-19	-187.4	-179.1	-209.6	-184.4	-172.5	-151.7	-143.2	-175.6	-151.9	-508.5
11-4-15	-184.8	-168.1	-206.1	-181.9	-167.9	-148.1	-132.1	-169.9	-148.5	-543.3
11-7-19	-184.9	-173.1	-175.6	-182.0	-171.3	-151.4	-139.9	-143.5	-151.7	-509.0
11-11-19	-171.7	-163.5	-166.9	-168.7	-157.5	-138.6	-131.0	-122.5	-139.0	-502.4
11-18-19	-168.8	-166.8	-191.6	-165.8	-153.7	-140.2	-138.3	-166.7	-140.3	-501.7
11-19-11	-164.7	-166.2	-188.5	-161.6	-149.5	-134.2	-135.7	-152.8	-134.2	-535.5
11-20-19	-177.7	-171.9	-169.8	-174.7	-165.2	-148.5	-143.2	-143.5	-149.0	-511.2
11-21-19	-153.7	-150.2	-179.9	-147.4	-139.0	-123.3	-120.2	-146.2	-123.8	-769.2
11-22-19	-158.3	-151.7	-176.7	-158.4	-137.4	-123.7	-117.8	-143.1	-123.7	-236.7
11-24-19	-172.1	-161.5	-187.3	-169.2	-157.9	-139.5	-129.3	-150.3	-139.8	-500.4
11-25-15	-172.7	-167.5	-197.3	-169.7	-157.4	-136.1	-131.6	-157.7	-136.1	-522.2
11-25-19	-177.9	-169.5	-178.6	-174.9	-166.4	-142.5	-134.9	-140.9	-143.1	-511.3
11-27-19	-171.8	-161.4	-183.2	-168.8	-162.2	-126.7	-116.9	-133.1	-126.7	-514.0
11-28-19	-191.3	-183.7	-189.3	-188.4	-176.6	-156.6	-149.1	-149.7	-157.0	-506.9
11-29-15	-173.4	-167.0	-172.7	-170.5	-157.4	-142.2	-135.7	-149.2	-142.4	-530.8
18-29-24	-174.7	-169.6	-166.0	-171.8	-162.2	-142.5	-137.4	-137.2	-142.7	-492.8
19-5-19	-175.7	-168.9	-175.1	-172.9	-162.4	-148.8	-142.3	-141.4	-149.2	-464.1
19-24-19	-171.4	-161.4	-164.6	-168.4	-157.4	-141.4	-131.4	-133.7	-141.7	-501.2
24-4-19	-196.4	-186.0	-198.7	-193.6	-180.5	-160.5	-150.9	-166.1	-160.8	-529.8
24-29-18	-183.0	-182.3	-170.1	-179.9	-167.8	-147.3	-146.3	-132.1	-147.4	-559.7
25-1-15	-190.5	-184.4	-194.2	-187.6	-176.6	-155.8	-150.5	-151.2	-156.1	-505.8
25-4-19	-181.0	-168.9	-192.5	-178.2	-166.8	-146.6	-135.0	-157.4	-146.9	-484.2
25-20-15	-173.3	-167.1	-154.2	-170.4	-162.6	-142.1	-135.9	-122.3	-142.4	-503.6
26-1-18	-192.7	-188.8	-192.3	-189.8	-178.3	-158.7	-155.0	-159.4	-159.1	-516.7
26-4-18	-195.1	-179.5	-193.7	-192.2	-178.3	-160.4	-145.9	-158.1	-160.8	-551.9
26-4-19	-203.7	-185.2	-203.2	-200.8	-188.0	-167.3	-150.0	-170.5	-167.6	-546.0
29-5-19	-190.4	-184.1	-196.4	-187.4	-176.0	-154.1	-148.1	-159.8	-154.5	-533.5
29-26-19	-193.6	-188.9	-210.4	-190.5	-185.5	-162.7	-157.3	-180.8	-162.8	-564.8
Median	-181.0	-169.6	-188.5	-178.2	-166.8	-148.1	-138.6	-150.3	-148.5	-511.2
IQR	18.1	16.1	20.0	18.1	18.6	15.0	15.1	17.9	15.1	33.1

For both the MM-GBSA and MM-PBSA methods, 5 different sets of results are shown: results for the single, double and triple trajectory methods with default settings and then results calculated using the single trajectory method with either the ionic strength or another parameter as shown changed as indicated. In the single trajectory method, the energies of the protein and free ligand are calculated using the trajectory of the complex, ignoring the other component of the complex. In the double trajectory method, the energy of the free ligand is calculated using a second simulation of the complex in isolation. Finally, in the triple trajectory method, a separate simulation is usually run of the free protein in addition to simulations of the complex and free ligand. In this case, because the estimates were solely for the purpose of ranking, the mean protein ΔG from the single trajectory method was used as the free protein ΔG for all ligands when using the triple trajectory method. For further details see p123. The "igb" setting controls the Amber GB method used. When igb=1, the original method of calculating the Born radii is used (described by equations 1.82 to 1.84 in the introductory chapter (p89)). When igb=5, the Born radii

are scaled to account for the radii of buried atoms being underestimated by the original method[200]. For further details see p124. The "indi" parameter is the dielectric constant used for the volume inside macromolecules, expressed as a multiple of the permittivity of a vacuum (8.85×10^{-12} Fm^{-1}). See p89 for more information on relative permittivity and its use in the MM-PBSA method. IS=Ionic strength. IQR=Interquartile range.

Table 2.14 shows binding energies predicted by the MM-PBSA and MM-GBSA methods averaging over the top three poses from three docking repeats so each result is based on 9 simulations, 4.5 ns of simulation time in total.

The rationale behind this decision to average over the top three poses is worth discussion. Docking scores are typically poor measures of binding affinity[161,162]. This is to be expected, given the need for a scoring function to predict affinity from a single pose as opposed to an ensemble of the possible conformations which a compound can assume within a binding site. Because the docking poses generated by Autodock Vina might not be ranked correctly, it could be unreliable to base predicted affinities on simulations run from just the highest scoring poses of each compound.

Rather than averaging over the top three poses, an alternative would have been to run simulations from multiple poses, calculate accurate affinity predictions from each trajectory and then select the best (most negative) predicted binding energy. However, the best affinity calculated might have been the best due to chance, not because the pose used to start that simulation corresponded to the correct binding site. Therefore, using this method could have led to the results being based on the wrong trajectories. Furthermore, there might not be one predominant binding site; the top poses found by docking could all indicate valid binding sites with similar affinities, the laboratory-measured binding constant being a function of the binding constant at each of these sites.

Initially, trajectories from the three highest scoring binding poses for each compound were used because three is the lowest number of readings from which a mean with reasonably sized confidence limits can be calculated.

The large number of results shown in Table 2.14 reflects the variety of different methods which can be used to estimate binding affinities. Presenting results for just those methods that lead to good correlations with the measured affinities would give an unrepresentative picture of the true accuracy of the methods. Results for different methods are also useful to assess the robustness of the predictions to changes in the settings used.

One method choice, which is indicated in the third row of Table 2.14, is the number of trajectories to use to calculate each affinity. In the single trajectory method, the protein and ligand coordinates are extracted from the trajectory of the complex, assumed to be

representative of the unbound molecules and used to calculate the energy of the free protein and free ligand. In the double trajectory method, the energy of the free ligand is calculated using a second trajectory of the ligand in isolation. A triple trajectory method can also be performed in which a third simulation is run of the free protein. In this work, because the estimates were solely for the purpose of ranking, the "triple trajectory" results were generated by using the mean protein ΔG (from the single trajectory method results) as the free protein ΔG for every ligand. For further details see p123.

Two of the Generalised Born methods available in AMBER (switched between using the igb setting) were tested (using an ionic strength of 0). These were the original model (igb=1) described by equations 1.82 to 1.84 in the introductory chapter (p89) and a modified model (igb=5), where a function with empirically derived parameters is used to rescale the Born radii of atoms based on the degree to which the atoms are buried and in so doing account for the underestimation of Born radii within macromolecules[200]. For further details see p124. The original model was also tested using an ionic strength of 0.35 M, which is close to the ionic strength of the buffer used in the laboratory assays. MM-PBSA results were also obtained using two different ionic strengths and two different values for the internal dielectric constant.

Table 2.14 reveals how different results are obtained, depending on the trajectories used to estimate the solvation energy of the free ligand and free protein. The range of the results was higher when using the triple trajectory method as opposed to the single and double trajectory methods. This suggests that subtraction of the vacuum and solvation energies of the unbound protein calculated using the trajectory of the complex compensates for some of the variation in the potential energy of the complex. The results therefore advocate the use of a single or double trajectory approach.

While the results of the different methods shown in Table 2.14 differ, they are largely similar with the exception of the MM-PBSA results obtained with a dielectric constant of 4 in the rightmost column. The calculated binding energies are all large in magnitude (very negative) and these are especially so, with a predicted median free energy change of -512 kJ mol^{-1}. The change in entropy of the protein and ligand is not intrinsically included when predicting binding affinities using the MM-PBSA and MM-GBSA methods. To investigate whether this missing entropic effect was responsible for the extreme values, normal mode analysis was performed on the trajectories.

Normal mode analysis is computationally expensive and it has been found to introduce significant error because the results often vary significantly between snapshots. Consequently, it is often not performed when similar ligands binding to the same site are compared[249,265]. Similar entropy changes were calculated regardless of the trajectory method used and there was little variation between the results for the different compounds. For example, as shown in Table 2.15,

the interquartile range of the results for the different compounds using the single trajectory method was 21.6 J mol^{-1} K^{-1}, which is just 5% of the median entropy change (approximately).

The rightmost column of Table 2.15 shows the amount which must be added and subtracted from each value in the single trajectory entropy column to generate its upper and lower 95% confidence limits. (A mean entropy was calculated for each of the three repeats (using the top three poses in that repeat) and the confidence limits are based on these three means.) The median of the confidence limit values is 25.7 J mol^{-1} K^{-1} so the median distance between the upper and lower confidence limits of the entropy values (twice this, 51.5 J mol^{-1} K^{-1}) is more than double the aforementioned interquartile range of the entropies themselves. Although increasing the number of repeats could increase the accuracy, the large error is consistent with the findings of the research cited above[249,265], which found that including the effect of entropy as predicted by normal mode analysis often does not improve the accuracy with which the relative affinities of ligands can be calculated.

The binding energy of the effective p53-Mdm2 interaction inhibitor Nutlin-3 is -42.4 kJ mol^{-1} based on its published K_i of 26 nM (assumed to have been measured at 298 K)[266]. While including the effect of the entropy might not improve the ranking, its inclusion shifts most of the binding energies predicted by the MM-PBSA and MM-GBSA methods from around -150 kJ mol^{-1} and -175 kJ mol^{-1} respectively to more reasonable values around -20 kJ mol^{-1} and -55 kJ mol^{-1}. It does not shift the MM-PBSA results obtained with a dielectric constant of 4 in the rightmost column of Table 2.14 to reasonable values, suggesting that these results are anomalous.

Table 2.15: Normal mode analysis of simulations of the 31 N-linked compounds synthesised

Side chains	Entropies /kJ mol^{-1} K^{-1}				Overall entropy for each trajectory method /J mol^{-1} K^{-1}			95% Limits For Single Trajectory Method (±) /J mol^{-1} K^{-1}
	Complex	Protein from complex trajectory	Ligand from complex trajectory	Ligand from free ligand trajectory	Single	Double	Triple	
4-4-11	16.34	15.35	1.435	1.421	-444.3	-430.1	-442.3	7.1
11-1-18	16.21	15.37	1.276	1.262	-435.8	-422.2	-409.5	30.4
11-2-19	16.22	15.35	1.282	1.275	-418.1	-411.8	-420.0	18.6
11-3-19	16.24	15.35	1.316	1.312	-426.4	-422.0	-430.0	34.4
11-4-15	16.23	15.35	1.321	1.303	-435.2	-417.5	-432.9	24.2
11-7-19	16.27	15.37	1.324	1.303	-423.2	-402.0	-394.7	18.4
11-11-19	16.20	15.36	1.252	1.242	-407.9	-398.0	-401.2	38.6
11-18-19	16.15	15.35	1.216	1.214	-416.9	-415.4	-425.6	29.6
11-19-11	16.17	15.34	1.233	1.228	-412.4	-406.7	-424.4	11.7
11-20-19	16.22	15.37	1.254	1.250	-406.1	-402.2	-390.4	27.8
11-21-19	16.18	15.38	1.217	1.217	-415.7	-416.0	-401.1	25.7
11-22-19	16.17	15.35	1.240	1.243	-425.0	-428.4	-436.2	46.4
11-24-19	16.22	15.36	1.268	1.261	-415.3	-409.0	-407.4	19.4
11-25-15	16.18	15.35	1.245	1.241	-409.0	-405.3	-421.2	29.4
11-25-19	16.22	15.36	1.266	1.261	-405.5	-400.6	-403.4	11.0
11-27-19	16.20	15.35	1.270	1.266	-419.2	-414.6	-429.4	35.7
11-28-19	16.24	15.37	1.319	1.307	-441.2	-429.2	-424.4	19.9
11-29-15	16.23	15.38	1.249	1.253	-397.8	-401.7	-380.5	10.7
18-29-24	16.21	15.37	1.266	1.269	-433.0	-436.8	-425.3	20.3
19-5-19	16.23	15.35	1.303	1.299	-419.6	-415.4	-430.9	17.5
19-24-19	16.20	15.36	1.240	1.236	-407.1	-403.1	-400.4	35.6
24-4-19	16.29	15.37	1.360	1.341	-446.7	-427.4	-417.5	10.7
24-29-18	16.22	15.37	1.272	1.276	-425.4	-429.5	-422.6	18.8
25-1-15	16.24	15.38	1.285	1.276	-431.5	-421.8	-401.9	26.8
25-4-19	16.29	15.36	1.357	1.340	-431.2	-414.0	-413.9	42.4
25-20-15	16.21	15.36	1.251	1.245	-401.0	-395.0	-398.4	61.6
26-1-18	16.23	15.37	1.297	1.292	-435.6	-430.9	-420.2	18.4
26-4-18	16.27	15.36	1.352	1.341	-440.1	-429.5	-434.4	21.8
26-4-19	16.29	15.37	1.365	1.352	-449.8	-437.1	-424.3	47.4
29-5-19	16.26	15.35	1.350	1.351	-442.0	-443.0	-455.2	39.5
29-26-19	16.23	15.37	1.293	1.293	-434.2	-433.3	-429.7	34.9
Median	16.22	15.36	1.276	1.275	-425.0	-416.0	-421.2	25.7
IQR	0.04	0.02	0.068	0.057	21.6	23.4	26.9	N/A

Entropies for the complex, free protein and free ligand were calculated from trajectories of each Mdm2-compound complex using Amber nmode. Subsequently, another value for the entropy was calculated for each compound using simulations containing only the ligand. The overall entropy difference for single, double and triple trajectory approaches was then calculated. In the single trajectory method, the entropy change of binding was estimated using only the entropies calculated from the trajectory of the complex. In the double trajectory method, the entropy of the free ligand from the trajectory of the ligand in isolation was used. In the "triple trajectory" method, the mean protein entropy (the average of the values in the third column) was used as the free protein entropy for every ligand. (See p123.) The rightmost column indicates the error in the single trajectory method entropies. It shows the amount which must be added and subtracted from each entropy in the 6th column to generate its 95% confidence limits. The error in normal mode analysis entropies means that the entropy change of the protein and ligand is often ignored when making MM-PBSA and MM-GBSA predictions[249]. IQR=Interquartile range.

The results shown above were generated using the top three poses produced in each docking repeat. To investigate whether this was the best number of poses to use, the correlation between the predicted relative affinities and affinities measured in the laboratory using fluorescence polarisation by Kerya Long (University of Leeds) was determined using different numbers of poses from each repeat (Figure 2.21). For this analysis, the binding energies predicted with the full MM-PBSA method and the single trajectory approach were used (for reduced noise). The relative internal dielectric for buried atoms and the ionic strength were left at their default values of 1 and 0 respectively.

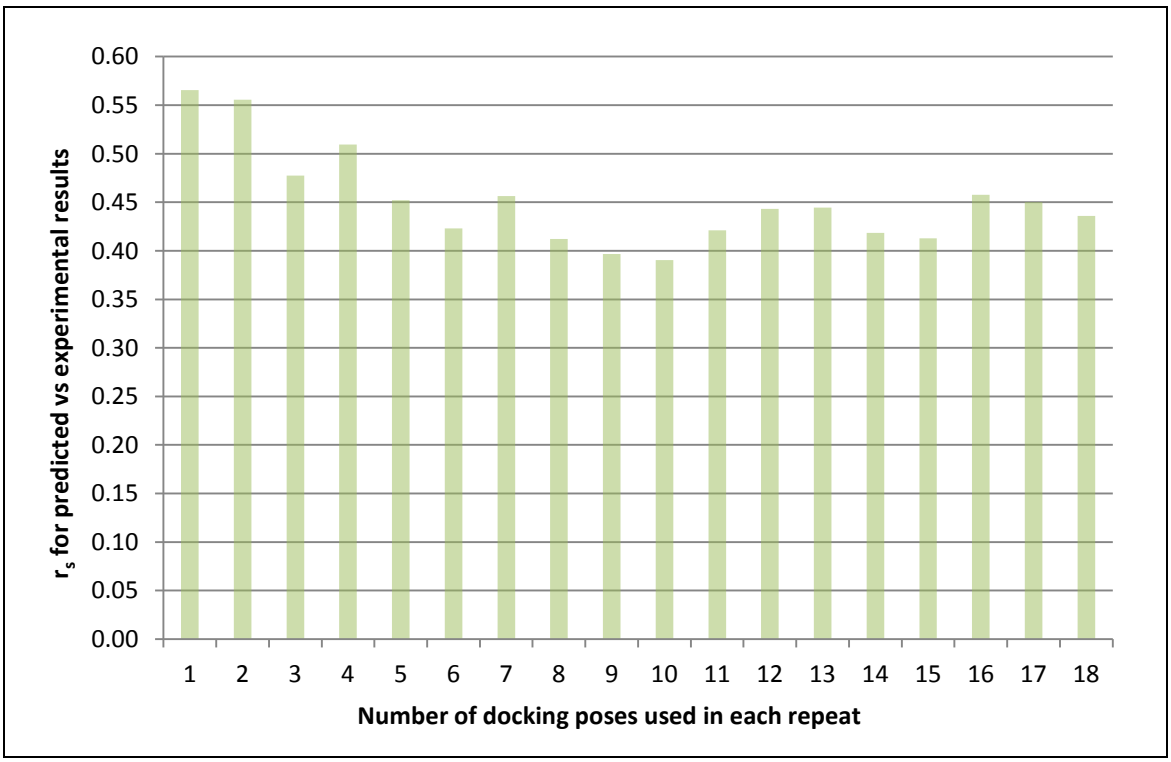


Figure 2.21: The effect of the number of docking poses of each molecule considered on the magnitude of the Spearman's Rank coefficient for the correlation between fluorescence polarisation results and predicted MM-PBSA binding energies. The single trajectory method was used with the internal dielectric constant set to 1 and the ionic strength set to 0 (the default values).

There is a drop in the correlation in going from using two poses to using three or four. Although, this result might be due to chance, it suggests that the top two poses are representative of where oligobenzamides actually bind but that the third and subsequent poses are less so. Consequently, the coefficients shown in Table 2.16 and Table 2.17 below were determined using the first two docked poses from each of the three docking runs.

Table 2.16 and Table 2.17 shows the correlation between predicted and experimental results obtained using different methods. A moderate correlation was obtained for both the single and double trajectory methods using both MM-PBSA and MM-GBSA, using either Pearson's product moment correlation coefficient or Spearman's rank and using an ionic strength of 0 or

0.35 M. However, they are consistent with the results of the triple trajectory method being less reliable than those produced by the single and double trajectory approaches.

Table 2.16: The effect of the number of trajectories used and including entropy on the correlation coefficient between predicted binding affinities and fluorescence polarisation single-point assay results

Trajectory Method	Method	No entropy		With entropy	
		Pearson	Spearman	Pearson	Spearman
Single	GB	0.563	0.551	0.580	0.573
	PB	0.613	0.556	0.614	0.498
Double	GB	0.564	0.569	0.567	0.481
	PB	0.587	0.523	0.566	0.394
Triple	GB	0.277	0.265	0.256	0.191
	PB	0.320	0.314	0.292	0.233

There were 31 compounds in total. All of the correlations are statistically significant. MM-GBSA and MM-PBSA can be performed using a 1, 2 or 3 trajectory approach. These were implemented as discussed in the methods section. Correlations were calculated with and without normal mode analysis entropies included. The colour reflects the coefficients shown in each box; the strongest correlations are highlighted in blue and the weakest ones are indicated in red.

Table 2.17: The effect of the GB method used, internal dielectric constant and ionic strength on the correlation coefficient between predicted binding affinities and fluorescence polarisation single-point assay results

Method	Ionic strength /M	No entropy		With entropy	
		Pearson	Spearman	Pearson	Spearman
GB igb=1	0	0.563	0.551	0.580	0.573
GB igb=1	0.35	0.545	0.551	0.498	0.505
GB igb=5	0	0.549	0.511	0.508	0.517
PB indi=1	0	0.613	0.556	0.614	0.498
PB indi=1	0.35	0.613	0.560	0.568	0.549
PB indi=4	0	0.343	0.376	0.350	0.368

There were 31 compounds in total. These results were calculated using the single trajectory method. The Generalised Born (GB) and Poisson Boltzmann (PB) methods were used to calculate the electrostatic component of the solvation energy as indicated. Two different GB models provided by Amber, switched between using the igb parameter, were used, the original model (igb=1) and a modified model (igb=5). In the latter, a function with empirically derived parameters is used to rescale the Born radii of atoms. For further details see p124. When using the PB method, two different relative dielectric constants were used for buried atoms, the default value of 1 and a higher value of 4, set using the indi parameter. For more information on relative permittivity and the MM-PBSA method, see p89. The colour reflects the coefficients shown in each box; the strongest correlations are highlighted in blue and the weakest ones are indicated in red.

Figure 2.22 shows graphically the correlation with the MM-PBSA method with an ionic strength of 0 and internal dielectric constant (relative permittivity) of 1. The correlation is negative because the more negative the free energy of binding for a compound, the more strongly it binds to Mdm2 and the greater the extent to which it is likely to displace the labelled p53 peptide in the competition assay which was used for laboratory testing. For more details of fluorescence anisotropy and the competition assay see p48.

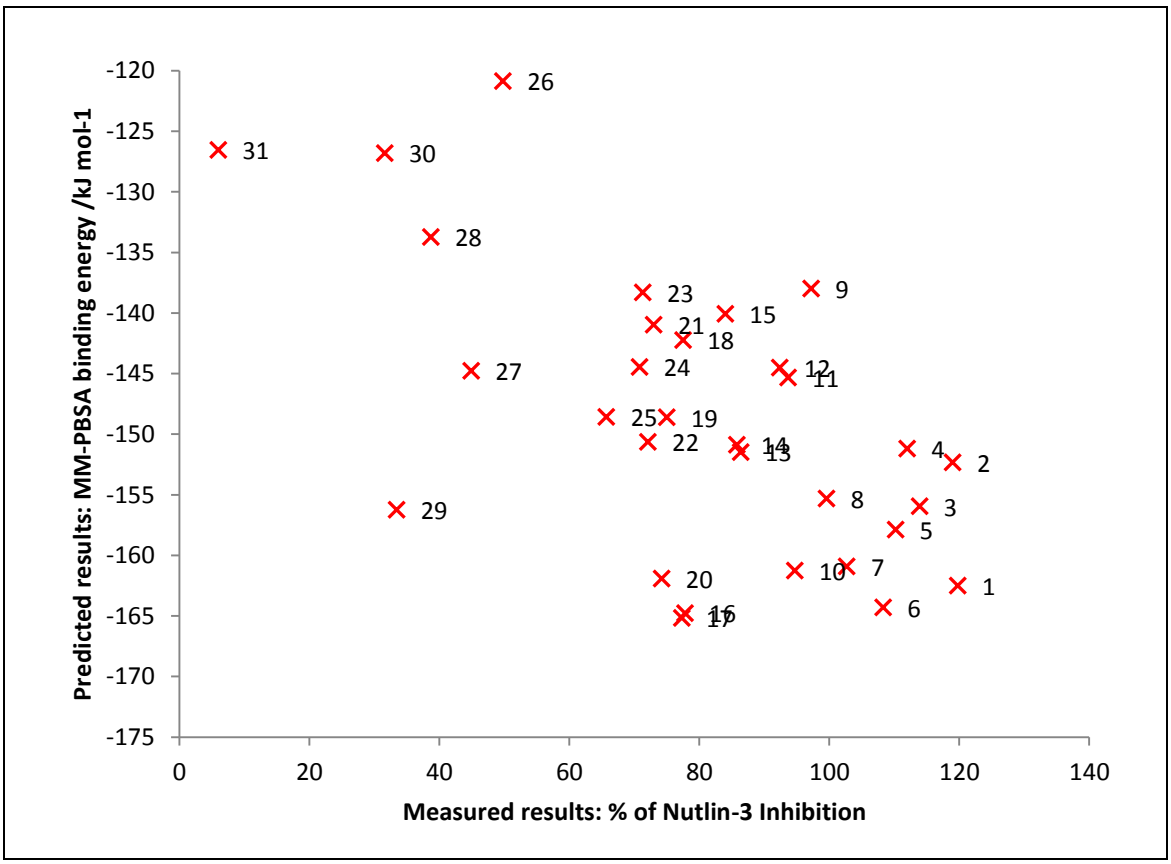


Figure 2.22: The correlation between MM-PBSA results and single point assay results. The single trajectory method was used with the internal dielectric constant set to 1 and the ionic strength set to 0 (the default values). The data are positively skewed on the binding energy axis and negatively skewed on the abscissa, supporting use of Spearman's rank as opposed to Pearson's product moment correlation coefficient. A negative correlation is apparent from this graph, as expected given that the more negative the free energy of binding, the greater the affinity of the compound for Mdm2 and thus the greater the expected inhibition at the concentrations tested in the competition assay (10 µM oligobenzamide, 54.5 nM fluorescein-labelled p53 peptide and 41.5 nM Mdm2 L33E). Although the results suggest that there could be a linear correlation between the predicted free energies of binding and the percentage inhibition, the correlation would not be expected to be linear given the exponential relationship between binding energy and association constant.

As shown in Table 2.16 and Table 2.17, there is not a clear improvement in the correlation coefficients observed when the entropy from normal mode analysis is added. This is interesting because, due to its lack of intramolecular hydrogen bonding, the N1 scaffold is the scaffold most likely to lose entropy upon binding.

In conclusion, the results in this section reveal a moderate correlation between predictions and experimental results for compounds based on the N1 scaffold. While there exist many ways in which MM-GBSA and MM-PBSA results can be calculated and predictions were made after the experimental data were made available, the results indicate that the observed correlation is visible using both these implicit solvent methods and using a variety of different settings. The best correlations were obtained with the MM-PBSA method using a single or double trajectory approach.

Although a few compounds based on scaffold O1 were made, it was only possible to compare practical and computational results for 6 compounds and there was a high degree of error in the experimental results. A strong correlation was not observed and so the small number of readings meant that no conclusion could be drawn.

2.3.5.2. ANOVA: side chains interact with each other but the interactions are small in magnitude

Testing every member of a large combinatorial library in multiple binding positions is computationally demanding. Once every library member has been docked and the likely binding sites of ligands have been identified, side chains could be investigated individually within each binding site by a method such as thermodynamic integration. This might overcome the problem of ligands binding in different places but require fewer simulations. To predict whether such methods would work, a deeper understanding of the relationship between the choice of scaffold and side chains and binding pose and predicted binding affinity is necessary. This section describes the production of MM-PBSA predictions for every member of a large combinatorial library and analysis of the results using ANOVA (described on p294) to determine the complexity of model necessary for affinity prediction.

To eliminate the effect that side chains have on each other's binding through their effect on binding pose it is necessary to compare the affinities of oligobenzamides all bound in the same position. In this project, two separate approaches were used to generate docking poses at a particular site.

The first method entailed taking the O1 scaffold starting compound pose (Figure 2.1, p102) with its side chains in the Phe19, Trp23 and Leu26 binding sites, modifying the side chains *in situ* and then optimising the pose using Autodock as described in more detail below. This method is useful in that it ensures that a pose is obtained for every combination of side chains. ANOVA analysis is more reliable when there are equal numbers of results for every cell (factor value combination) (p296)[267].

The second method involved docking poses as normal (using Autodock Vina) and then selecting poses which bound at a particular site (p186). This method results in missing data where

docking fails to generate a pose for a particular ligand within the site of interest and the results thus require more complex treatment (p296).

The inherent problem of missing binding data where compounds in a library do not bind to a protein may be one of the key reasons why ANOVA has not previously been used to investigate the interaction of functional groups in ligand binding.

2.3.5.2.1. Simulations started from poses derived by side chain replacement followed by pose optimisation within a particular binding site suggest a three-way interaction between the three side chains but the predicted binding energies are positive, which implies the docking poses are unnatural

Simulations were run with every possible combination of 12 side chains at the three side chain attachment positions on scaffold O1 (1728 compounds in total). The side chains used were numbers 1, 2, 3, 4, 5, 6, 11, 19, 22, 26, 47 and 48. See Figure 2.2 (p103) for the structures. Simulations were started from the best pose obtained in each of six separate sets of Autodock 4 local searches. As detailed in the methods (p109), a local search in Autodock involves optimising (minimising the potential energy of) a pose close to the binding pose. Here, the docked O1 scaffold starting molecule (Figure 2.1, p102) with side chains in the Phe19, Trp23 and Leu26 binding sites (Figure 2.8, p129) was taken, the side chains were changed (using Side Chain Splicer, p105) and then the local search was used to position the compound as best possible.

Six short simulations (500 ps) were performed for each compound. The second 250 ps of each trajectory was used for analysis so a total of 1.5 μs of simulation was used for calculation of each binding energy (because there were six repeats).

Although this local optimisation method was convenient because it resulted in a set number of poses for every combination of side chains and the subsequent ANOVA was balanced, the free energies of binding subsequently predicted using the MM-PBSA method were positive. This is evident from the last line of Table 2.18, which shows various statistics for the variables calculated. While the MM-PBSA results exclude the effect of the entropy change of the ligand and receptor upon binding, this is typically negative (as shown in Table 2.15, p171) so increases the free energies when included, rather than reducing them.

The almost universally positive binding energies predicted are not consistent with the known binding of many oligobenzamides to Mdm2 when present at concentrations far below 1 mol dm^{-3}. Consequently, the docked poses used to start the simulations might not have been representative of where compounds actually bind. The local docking method finds the best binding pose for compounds assuming that they bind in the same position as the starting compound. This assumption might be incorrect.

Table 2.18: Statistics for the predicted binding energies of O1 scaffold oligobenzamides

		Median	Range	Interquartile range
Gas phase ΔΔG /kJ mol^{-1}	Van der Waals	44.28	56.99	9.00
	Electrostatics	40.67	302.21	68.30
	Complex	7046.0	520.2	78.9
	Receptor	6695.5	305.5	56.7
	Ligand	277.0	151.0	26.8
	Overall	83.59	310.00	71.15
Solvation ΔΔG /kJ mol^{-1}	Poisson Boltzmann	-60.09	287.33	67.27
	Surface area	-4.240	5.212	0.758
	Complex	1403.3	584.6	96.4
	Receptor	1331.7	267.5	48.9
	Ligand	121.2	207.8	27.6
	Overall	-64.15	288.21	67.49
	Total	19.37	41.53	6.06

Predictions were made using the MM-PBSA method. The simulations were started from poses generated by *in situ* side chain substitution followed by local pose optimisation using Autodock.

Despite their unusually high values, a full factorial ANOVA was performed on the MM-PBSA predicted binding energies. A full factorial ANOVA is one in which the fitted model contains all of the possible interaction terms. The model used in this case can be described using the following formula:

$$Y_{ijkn} = m + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + (abc)_{ijk} + \epsilon_{ijkn} \quad \textbf{2.7}$$

Where Y_{ijk} is the binding energy of a ligand (actually raised to the power 0.65 here to increase residual heteroscedasticity as discussed below) with side chains i, j and k, which it is assumed can be predicted by the summation of 7 coefficients, looked up using i, j and k. The first is a constant, m. This is 0.74 for the results in this section ($F_m=3.20\times10^5$, $P(m=0)<0.0005$). The next three terms (a_i, b_j and c_k) are main effects, variables which each depend on the value of a single factor, in this case, the side chain at one of the side chain attachment positions. The fitted model has one value of a for every side chain found at side chain position 1, one value of b for every side chain found at side chain position 2 and so on. The two way interaction terms follow. The fitted model has a two-dimensional array for each of these three terms, from which a value is selected when a prediction is made based on the two side chains in question (1 and 2, 1 and 3 and 2 and 3 for $(ab)_{ij}$, $(ac)_{ik}$ and $(bc)_{jk}$ respectively in Equation 2.7). Finally, $(abc)_{ijk}$ is the three-way interaction. Fitting this model generates a three-dimensional array of values for this last term.

The model states that prediction using the coefficients is imperfect; for every ligand binding energy prediction there is a residual error, ε_{ijkn} where n is the repeat number.

Before inspection of the fitted model, it is necessary to consider the assumptions of ANOVA because if these are not met, the chance of a false positive result is increased (the probability of a null hypotheses being rejected increases above the chosen significance level). Firstly, the values analysed must be independent. Secondly, the residuals left at the end of fitting (the error values) should be homoscedastic, meaning that the variance of these residuals should be the same for every set of conditions (for every set of side chains in this case). Thirdly, the residuals should be normally distributed.

The TI predictions generated earlier in this chapter were not independent due to their method of calculation whereby components for each side chain were added. In contrast, the MM-PBSA results are independent (the first assumption).

To test for homogeneity of the variances of the residuals (the second ANOVA assumption), Levene's test[268,269] was performed in SPSS. (For this analysis, an ANOVA models were used with all of the full factorial interaction terms except for the three-way interaction, the inclusion of which would have reduced the number of energies in each cell too much.) Use of the raw, untransformed MM-PBSA binding energies for modelling yielded a Levene's test F value of 1.645, which is too great to have occurred due to chance alone ($P(H_0)<0.0005$ based on 1727 and 8640 degrees of freedom); there was not equality of variances. Plotting the residual standard deviations from each cell (side chain combination) against the mean predicted binding energy of that cell reveals that the data are heteroscedastic. The variance increases with the free energy of binding. In an effort to correct for the apparent trend, a transformation was applied; the square root of each of the values ($\Delta G^{0.5}$) was tested. However, this was too extreme a transformation. It did not just remove the positive correlation between binding energy and residual variance but went further, creating a negative correlation. Consequently, exponents between 0.5 (taking the square root) and 1 (applying no transformation) were tested and raising the energies to a power of 0.65 produced the most homoscedastic residuals.

The Levene's test F-value for the transformed data is 1.370, which is still statistically significant ($P(H_0) <0.0005$, $d_1=1727$, $d_2=8640$); however, F is small in magnitude. Because tests for homogeneity of variance are sensitive and the ANOVA test is reasonably robust to homogeneity of variances when there is a large amount of data, the use of tests for homoscedasticity before ANOVA is controversial[270(p308)]. Consequently, the remaining heterogeneity of the variances following the aforementioned transformation was ignored.

When an ANOVA is performed, it is assumed that the residuals follow a normal distribution. The residuals obtained using the untransformed data have a positive skew ($g_1=0.174$). The distribution is also slightly leptokurtic (pointed) ($g_2=0.619$). The positive skew is removed by the transformation ($g_1=-0.002$, $SE_{g1}=0.024$). The kurtosis is not but the degree of kurtosis is

small (g_2=0.517 after transformation), so this was ignored. For a discussion of what constitutes a small kurtosis and the interpretation of Levene's test results for large samples see p298.

The results of the full factorial ANOVA (Table 2.19 below) reveal that there is a statistical interaction between every side chain (a three-way interaction) (F=1.151, P(H_0)<0.0005). Instead of each side chain contributing independently to the binding energy, the affinity of each side chain depends to an extent on the side chains at each of the other positions, even where the side chains in question are at opposite ends of the scaffold. Because the simulations used to generate the predictions in this section were started from similar poses due to the docking method used, this result suggests a degree of mechanical coupling between the side chains of an oligobenzamide as it moves within a particular binding site.

Table 2.19: The results of fitting an ANOVA model to investigate the effects of side chain choice on the predicted binding energy of O1 scaffold oligobenzamides

Source of variation	Sum of Squares /kJ^2 mol^{-2}	Degrees of freedom 1	Mean Sums of Squares /kJ^2 mol^{-2}	F	Significance
SC1	6.87×10^4	11	6246.757	510.664	<0.0005
SC2	2.01×10^4	11	1823.358	149.057	<0.0005
SC3	2.27×10^4	11	2065.213	168.828	<0.0005
SC1*SC2	3982.857	121	32.916	2.691	<0.0005
SC1*SC3	1.13×10^4	121	92.996	7.602	<0.0005
SC2*SC3	3234.467	121	26.731	2.185	<0.0005
SC1*SC2*SC3	1.87×10^4	1331	14.078	1.151	<0.0005
Error	1.06×10^5	8640	12.233	N/A	N/A

Predictions were made using the MM-PBSA method. The simulations were started from poses generated by *in situ* side chain substitution followed by local pose optimisation using Autodock. SC1 is the side chain at position 1 (the amino end of the compound). SC2 is the side chain at position 2 and SC3 is the side chain at position 3. For each source of variation, the first number of degrees of freedom for the F-test to check its significance is shown. The second number of degrees of freedom for these tests was 8640. For a definition of the mean sums of squares see p295.

The significance of the three-way interaction (F=1.151, P(H_0)<0.0005) reveals that the side chain at one position can not just influence the affinity of the other two side chains but also the way in which the other two side chains influence each other. The docking results on p153 suggest that side chains could influence each other's effect on binding through their impact on the binding pose. Because the ANOVA investigation described in this section focussed only on binding poses in the expected positon with side chains occupying the Phe19, Trp23 and Leu26 binding sites, its results suggest that, in addition, side chains affect the binding of each other within a particular binding site. The interaction after binding pose has been accounted for is investigated further on p190, where more reliable, negative predicted binding energies are used.

Because this section describes a balanced ANOVA (equal numbers of results for every combination of side chains), the F-values for the two-way interactions (F = 2.691, 7.602, 2.185) and main effects (F = 510.7, 149.1, 168.8) indicate the extent to which the side chains affect the results after the three-way interaction has been accounted for. Specifically, that P<0.0005 for every term of the model indicates that, at the 0.5% significance level, every term is useful for explaining the results. The side chains have an effect which is independent of the other side chains in addition to the aforementioned interaction effects.

The proportion of the variation in a set of data which is explained by a fitted ANOVA model is indicated by its R-squared and adjusted R-squared values. Both range from 0 (no predictive ability) to 1 (perfect prediction of the results by the model) (p297). For the ANOVA described above, SS_{Error} (p297) is 1.06×10^{5} $(\text{kJ mol}^{-1})^{0.65\times2}$ and the total adjusted sums of squares (SS_{Total}), which is independent of the model, is 2.55×10^{5} $(\text{kJ mol}^{-1})^{0.65\times2}$, making R^2 equal to 0.585.

To investigate the magnitude of the interactions, three other models were tested and the (unadjusted) R-squared values were determined for each using the new SS_{Error} and the previously calculated SS_{Total} above.

- The model with no three-way interaction (SS_{Error} = 1.25 $(\text{kJ mol}^{-1})^{0.65\times2}$):

$$Y_{ijkn} = m + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + \epsilon_{ijkn} \quad \textbf{2.8}$$

- The model after the first and last side chain two-way interaction has been removed (SS_{Error} = 1.36 $(\text{kJ mol}^{-1})^{0.65\times2}$):

$$Y_{ijkn} = m + a_i + b_j + c_k + (ab)_{ij} + (bc)_{jk} + \epsilon_{ijkn} \quad \textbf{2.9}$$

- The model with no two-way interactions (SS_{Error} = 1.44 $(\text{kJ mol}^{-1})^{0.65\times2}$):

$$Y_{ijkn} = m + a_i + b_j + c_k + \epsilon_{ijkn} \quad \textbf{2.10}$$

As above, Y_{ijkn} is the (transformed) free energy of binding, a_i, b_j and c_k are the main effect terms for the N-terminal, central and C-terminal side chain attachment positions respectively, $(ab)_{ij}$, $(ac)_{ik}$ and $(bc)_{jk}$ are the two-way interaction terms for the side chain 1-side chain 2, side chain 1-side chain 3 and side chain 2-side chain 3 interactions respectively and ε_{ijkn} is the residual error where n is the repeat number.

Table 2.20: Incorporating different degrees of interaction between the side chains: R-squared and adjusted R-squared values for four ANOVA models to predict the relative affinities of O1 scaffold oligobenzamides

Model	Unadjusted R-squared	Adjusted R-squared
No interaction	0.438	0.436
No interaction of side chains 1 and 3	0.467	0.452
No three-way interaction	0.511	0.491
Complete factorial design	0.585	0.501

Unadjusted R-squared values (see p297) range from 0 to 1 and describe the fraction of the variation in the results explained by the model. Adjusted R-squared values (see p298) facilitate a fairer comparison of models with different numbers of factors. A complete factorial ANOVA design is a model where the equation describing it has a term for every possible interaction between the factors. Each increase in the adjusted R-squared value down the table is statistically significant according to the Wald statistic for the highest order interaction in the more complicated model. ($P(H_0)<0.0005$ in every case.)

Table 2.20 shows the R-squared values for the models. The complete factorial design (the model with all of the possible interaction terms) accounts for less than 60% of the variation in the results ($R^2=0.585$). This highlights the significant error in all of the calculated binding energies. This error would be reduced by increasing the simulation time. (Because the fitted model has a three-way interaction term dependent on the choice of all three side chains and these were the only factors distinguishing one oligobenzamide from another in this controlled computational investigation, the residual error reflected by this R-squared is entirely due to variation between repeats due to sampling error.)

A large drop in the unadjusted R^2 value is observed in going from the complete factorial model with all of the interactions to the model with no interaction effects at all. An ANOVA model will usually fit a set of the results better if the number of terms is increased because the additional parameters can absorb some of the random variation in the data. Adjusted R^2 values (Equation 6.24 (p298)) take this effect into account. They can provide a better indication of how well models with different numbers of fitted parameters perform relative to each other.

The adjusted R-squared values (Table 2.20 above) suggest that the three-way interaction makes only a small improvement to the fit of the model. Consequently, while the interactions are statistically significant, the interactions appear to be quite small in magnitude and thus explain little of the variation in the results.

The simple model with no interactions appears to be quite predictive relative to the other models. The small effect sizes of the interactions imply that processing time might be best spent creating an accurate model of the main effects as opposed to simulating large numbers of compounds in an effort to model the interaction terms. Linearly scaling the unadjusted R-squared values so that the R-squared of the full factorial ANOVA was 1 (as if there were no sampling error) would bring the unadjusted R-squared value for the no-interaction-term model

up to 0.75. Consequently, if longer simulations were run or more repeats were performed to reduce sampling error, it might not be necessary to perform simulations with every member of a library; up to 75% of the variation in predicted affinity between oligobenzamides with different side chain combinations could be revealed through study of each side chain attachment position in isolation.

2.3.5.2.2. The interaction of the side chains is evident in many components that contribute to the binding energy

To determine whether a particular component of the binding energy was responsible for the observed interactions (for example, the change in the electrostatic energy or the surface area-dependent component of the free energy of solvation), ANOVA was performed using some of these components as the dependent variable.

Some of the results of this analysis are shown in Table 2.21. Data for these variables were not transformed prior to analysis because the results of Levene's tests (column 2) revealed significant homoscedasticity, with the exception of those variables asterisked, for which the p-value of the complete factorial design should be ignored.

In Table 2.21, $\Delta G_{MM-PBSA}$ is the MM-PBSA predicted binding energy, which is the sum of two components (Equation 2.11 and Equation 2.12). The first is the difference in energy of the complex and the unbound receptor and ligand in the gas phase, $\Delta H_{gas,complex} - \Delta H_{gas,receptor} - \Delta H_{gas,ligand}$, which can also be written $\Delta H_{vdW} + \Delta H_{Elec}$, where ΔH_{vdW} is the change in potential energy associated with the van der Waals forces and ΔH_{Elec} is the change in the electrical potential energy. Because the dissociated protein and ligand came from the trajectory of the complex, there is no predicted change in internal energy. The second is the difference in solvation energies of the complex and the free receptor and ligand, $\Delta G_{solv,complex} - \Delta G_{solv,receptor} - \Delta G_{solv,ligand}$. This is calculated using the Poisson Boltzmann equation (ΔG_{PB}) to account for changes in the electrostatic energy and an empirical surface-area based method to account for the non-polar element of the solvation (ΔG_{Cavity}). Consequently, it can also be written $\Delta G_{PB} + \Delta G_{Cavity}$.

$$\Delta G_{MM-PBSA} = (\Delta H_{vdW} + \Delta H_{Elec}) + (\Delta G_{PB} + \Delta G_{Cavity}) \quad 2.11$$

$$= (\Delta H_{gas,complex} - \Delta H_{gas,receptor} - \Delta H_{gas,ligand}) + (\Delta G_{solv,complex} - \Delta G_{solv,receptor} - \Delta G_{solv,ligand}) \quad 2.12$$

Table 2.21: Statistics for the ANOVA models created to investigate interaction of the side chain factors in the determination of various components of the MM-PBSA-predicted binding energies

Dependent Variable	Levene's Test		p-values for complete factorial design	Unadjusted R-squared values for model				Adjusted R-squared values for model			
	F	p		Complete factorial design	No three-way interaction	No interaction of side chains and 3	No interaction	Complete factorial design	No three-way interaction	No interaction of side chains 1 and 3	No interaction
ΔH_{vdW}	1.367	<0.0005	0.038	0.447	0.356	0.296	0.264	0.337	0.330	0.277	0.261
ΔH_{Elec}	1.411	<0.0005	0.044	0.911	0.896	0.893	0.884	0.893	0.892	0.890	0.884
ΔG_{PB}	1.391	<0.0005	0.080	0.915	0.901	0.898	0.889	0.898	0.897	0.895	0.889
ΔG_{Cavity}	1.369	<0.0005	0.006	0.442	0.346	0.295	0.266	0.330	0.321	0.276	0.264
$\Delta H_{gas,complex}$	1.063	0.049	0.445	0.544	0.474	0.462	0.439	0.453	0.453	0.447	0.438
$\Delta H_{gas,ligand}$	2.667*	<0.0005	<0.0005*	0.970	0.965	0.932	0.818	0.964	0.964	0.930	0.817
$\Delta G_{solv,complex}$	1.032	0.192	0.109	0.770	0.733	0.726	0.714	0.725	0.723	0.719	0.713
$\Delta G_{solv,ligand}$	4.024*	<0.0005	<0.0005*	0.987	0.985	0.963	0.890	0.985	0.984	0.962	0.890
$\Delta G_{MM-PBSA}$	1.367	<0.0005	<0.0005	0.585	0.511	0.467	0.438	0.501	0.491	0.452	0.436

The MM-PBSA predictions ($\Delta G_{MM-PBSA}$) can be divided into their constituent components in two different ways, by type or by species. In the top four data rows the separation is by type. ΔH_{vdW} is the *in vacuo* change in van der Waals potential energy associated with binding and ΔH_{elec} is the *in vacuo* change in electrical potential energy. ΔG_{PB} is the difference in solvation free energy between the complex and free protein and ligand due electrostatics. The difference between the solvation energies due to the non-polar component of the solvation energy predicted using the solvent accessible surface area is ΔG_{Cavity}. In the lower four data rows, the MM-PBSA-predicted free energy of binding is split into components relating to the complex and components relating to the free ligand. $\Delta H_{gas,complex}$ and $\Delta H_{gas,ligand}$ are the *in vacuo* energies given by the AMBER force field of the complex and free ligand respectively. $\Delta G_{solv,complex}$ and $\Delta G_{solv,ligand}$ are the free energies of solvation of complex and ligand respectively. The values in the top four data rows are not independent to the results in the bottom four data rows as shown in Equation 2.11 and Equation 2.12 (p182). The p-values for the three-way interaction in the complete factorial design are shown. For all eight possible dependent variables indicated, the p-values for the highest order interaction in the three smaller designs lacking some interaction terms are all less than 0.0006, meaning that the differences are significant at the 0.1% significance level, with the exception of the $\Delta H_{gas,ligand}$ and $\Delta G_{solv,ligand}$ results. *The gas phase binding energies and solvation energies for the ligand in isolation have a broad range of variances so the F-test results are unreliable.

Table 2.21 shows R-squared values for models with different numbers of factors. Unadjusted R-squared values (p297) range from 0 to 1 and describe the fraction of the variation in the results explained by a model. Adjusted R-squared values are similar but account for the number of factors in a model (p298) so facilitate the fairer comparison of models with different numbers of factors.

The unadjusted R-squared values for the full factorial models give an indication of the extent to which there was sufficient sampling of the phase space to determine accurately each component of the MM-PBSA predicted binding energy. For example, the unadjusted R-squared values for the gas-phase enthalpy of the ligand ($\Delta H_{gas,ligand}$) and the free energy of solvation of the ligand ($\Delta G_{solv,ligand}$) have very high unadjusted R-squared values (0.970 and 0.987 respectively) because the phase space of the ligand is much smaller than that of the complex so is sampled more comprehensively within the simulation time used.

The significances of the three-way interactions in each ANOVA analysis are shown in the table. The p-values in the fourth column indicate that the three-way interaction discussed earlier does not stem from a single component of the binding energy; an interaction is observed with many of its contributing parts. Where the existence of the three-way interaction is borderline for some components, there are significant pairwise interactions (not shown) between all of the factors.

Data rows 1 to 4 of Table 2.21 show the results of ANOVA with the dependent variable (the variable which the model seeks to explain) changed to each of the different physical factors that contribute to the binding energy (shown in Equation 2.11) (ΔH_{vdW}, ΔH_{Elec}, ΔG_{PB} and ΔG_{Cavity}). Interactions effects appear to have the greatest impact on ΔH_{vdW} and ΔG_{Cavity} because, with these, the adjusted R-squared values drop significantly upon removal of the interaction terms.

ΔH_{vdW} and ΔG_{Cavity} both depend on short range interactions. In the case of ΔH_{vdW}, the attractive dispersion force between two atoms is proportional to the 6^{th} power of their distance apart. ΔG_{Cavity} is calculated using the change in solvent accessible surface area and this only changes when parts of the protein and ligand touch. Given the short range of these effects, it is noteworthy that the greatest drop in adjusted R-squared occurs upon removal of the two-way interaction between the terminal side chains. A possible explanation is that side chains influence the binding of each other through their effect on the position of the scaffold within the binding site. In order for certain side chains to be accommodated at the ends of the molecule, the scaffold might need to move and this could cause the nature of other side chains' interaction with the protein to change. Figure 2.12 (p136) illustrates this activity. It shows structures from one of the 5 ns simulations of the N1 scaffold starting molecule (Figure 2.1, p102) performed to verify simulation stability. Concerted movement of side chain 1 (shown on the right-hand side of each image) and the scaffold core is evident upon comparison of Figure 2.12A with Figure 2.12B. It is possible that side chain 1 pushes against the protein and moves the scaffold. Within

roughly 200 ns, side chain 3 adapts to the new scaffold position (Figure 2.11C (p135), Figure 2.12C). Eventually the scaffold and both side chains return to a conformation similar to that of their initial pose (Figure 2.12D).

Data rows 5 to 8 of Table 2.21 show the results of ANOVA with the dependent variable set to each of the ligand-dependent components shown in Equation 2.12, namely the force field-derived energy changes for the gas phase complex and free ligand ($\Delta H_{gas,complex}$ and $\Delta H_{gas,ligand}$) and the solvation free energies of the complex and ligand generated using the Poisson-Boltzmann equation and solvent accessible surface areas ($\Delta G_{solv,complex}$ and $\Delta G_{solv,ligand}$). As shown in the table, the three-way interaction is not statistically significant in terms of $\Delta H_{gas,complex}$ and $\Delta G_{solv,complex,}$ the variables relating to the complex (P=0.445 and P=0.109 respectively). These results imply that the interaction between side chains manifest in the total energy change $\Delta G_{MM\text{-}PBSA}$ might stem from side chain interaction in the free ligands as opposed to their complexes. While the p-values for the three-way interaction significance in relation to the free ligand energies, $\Delta H_{gas,ligand}$ and $\Delta G_{solv,ligand}$, must be treated with caution (due to heteroscedasticity), these results are consistent with this hypothesis. Furthermore, removal of all interaction terms from the $\Delta H_{gas,complex}$ and $\Delta G_{solv,complex}$ models results in the adjusted R-squared values dropping by 0.015 and 0.012 respectively whereas removal of the interaction terms from the $\Delta H_{gas,ligand}$ and $\Delta G_{solv,ligand}$ models results in the adjusted R-squared values dropping by considerably more, 0.147 and 0.095 respectively, consistent with side chain interactions occurring in the free ligand rather than in the ligand-protein complex. However, these results are misleading because they reflect the variation caused by interaction effects as a proportion of the total variation (in the affinity of different oligobenzamides) as opposed to in absolute terms. Between different ligands, $\Delta H_{gas,complex}$ and $\Delta G_{solv,complex}$ vary more than $\Delta H_{gas,ligand}$ and $\Delta G_{solv,ligand}$ due to the presence of the protein in simulations of the complex, which can interact with the ligand in many ways and assume different conformations as it does so. Consequently, the interaction effects may just be a lower proportion of the total variation in the case of the complex because there is more variation in total. Furthermore, the statistical interactions apparent in the unbound ligand components of the binding energy could involve the protein because the trajectory of the free ligand used for this analysis was extracted from the trajectory of the complex.

Although it is unlikely that statistical interactions between the side chain factors exist solely in the unbound ligand, the extent to which the pose of a ligand affects the interaction of the side chains is worth investigation because it could have an impact on the best approach for affinity prediction in future. In the next section, the effect of the pose on the side chain interaction effects is confirmed.

The movement of the scaffold from side to side illustrated in Figure 2.11 and discussed in relation to the ΔH_{vdW} and ΔG_{Cavity} components above suggests that side chains do not just

influence each other through their effect on binding position but also through scaffold-mediated mechanical coupling within a particular binding site. In the subsequent section (p191) this is tested for.

2.3.5.2.3. The binding pose has a statistically significant effect on the interaction between the side chains

The previous section described the use of ANOVA to analyse the MM-PBSA-predicted binding free energies of oligobenzamides based on the O1 scaffold. The O1 scaffold starting compound (Figure 2.1, p102) was docked using Autodock to generate an initial pose. Subsequently, for each side chain combination of interest, the side chains of the posed starting compound were modified *in situ* and then local optimisation was performed to generate a pose with which a simulation could be initialised. Most of the binding energies calculated from these simulations were positive, which suggests that oligobenzamides do not actually bind in the poses predicted by this method. The docking results on page 153 consistent with this hypothesis, suggesting that oligobenzamides often fail to bind with a side chain in each of the three p53 hotspot binding sites as assumed in the previous section.

In subsequent results now described, standard, global docking as opposed to just a local search was performed for all compounds prior to simulations, with the aim of simulating compounds in their true binding position.

Three docking runs were performed using Autodock Vina with the exhaustiveness setting (p111) set to the default value of 8. To verify that the docking process was sufficiently exhaustive to obtain the best binding poses, a fourth docking run was performed with the exhaustiveness set to 16 to see if this led to higher scoring docking poses in the results. (According to the documentation, the time required for docking is roughly proportional to the value of the exhaustiveness setting.) The mean top pose scores for the three original docking runs were -7.41, -7.40 and -7.40 and the mean top pose score in the fourth was -7.49 kcal mol^{-1}. The higher docking score for the last run with the higher exhaustiveness is statistically significant. However, this difference is small compared to the range of docking scores obtained for the top poses. Consequently, an exhaustiveness of 8, the default recommended value for this setting, appears to be sufficient. Box and whisker plots for the docking scores of the top pose for each compound showing the range of top pose scores are shown in Figure 6.5 (p311).

To generate starting poses for simulations, the results of the three main docking runs described above were searched to identify compounds in each of the four poses indicated in Figure 2.23. These unsystematic poses were chosen because the 25,000 compound docking analysis above (p153) revealed that they were poses in which oligobenzamides were very commonly found. Missing values (p296) are problematic in ANOVA analysis. The use of these common poses minimised the number of combinatorial library members which could not be obtained in any

one of the poses (resulting in a missing MM-PBSA prediction) and so increased the reliability of hypothesis testing. It also means that the results are more applicabile to oligobenzamide binding in general. If there was, in a particular repeat, more than one pose for a compound which satisfied one of the conditions (A, B, C or D) then only the pose with the highest docking score was used. A compound can satisfy the requirements of poses B and D at the same time; however, data for poses B and D were never analysed simultaneously.

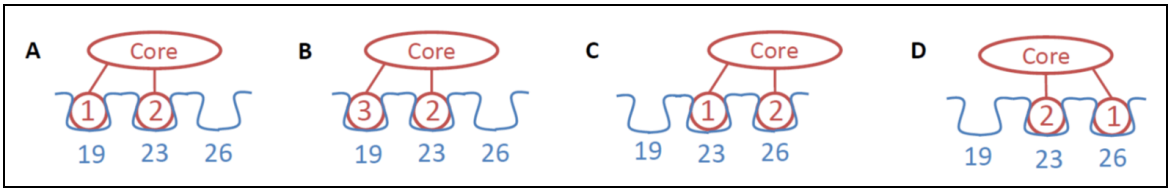


Figure 2.23: Four poses, A, B, C and D referred to in the ANOVA results. The scaffold is shown in red: Mdm2 is shown in blue. The labels 19, 23 and 26 indicate the pockets normally occupied by p53 residues Phe19, Trp23 and Leu26. For further details of pose recognition see the section entitled "Pose recognition" starting on p111.

Simulations were run starting from all of the selected poses. Figure 2.20 above shows RMSD values for 50 N1 compounds when simulations are started from global docking poses generated using Autodock Vina. A similar rapid rate of equilibration was observed with oligobenzamides based on the O1, S1, O2 and S2 scaffolds (results not shown). The first 250 ps of constant pressure simulation was removed from all trajectories prior to their analysis using the MM-PBSA method.

Initially, five ANOVA analyses were performed, one for each scaffold. The three factors in each model were the pose (A, B, C or D) and the side chains at the first (N-terminal) and second (middle) side chain positions. The aim was to investigate how the pose factor affected the interaction between the side chains when using different scaffolds. The sparsity of the results was too great to carry out an ANOVA with all three side chains as factors in addition to the pose. A large amount of impution (p296) would have been required, which could have biased the results.

Table 2.22 shows the side chains used. Every combination of side chains was tested. The numbers refer to the side chain structures shown in Figure 2.2 (p103).

Table 2.23 gives some summary information about the data. Preliminary analysis revealed that while docking of a compound derived from any of the five scaffolds commonly yielded a pose in position A, for some scaffolds there were few compounds with a pose satisfying conditions B, C or D.

Table 2.22: Side chains tested at each side chain position when using the side chains at positions 1 and 2 as factors in ANOVA models

Side chain position	Side chains tested at position	Number of side chains
1	4, 11, 13, 15, 19, 21, 22, 24, 25, 27, 29, 33, 37, 39, 44	15
2	1, 3, 5, 6, 9-13, 15, 21, 22, 25, 26, 30, 32, 37, 39, 44	19
3	19 (N1 scaffold) / 37 (O1, S1, O2, S2 scaffolds)	1

The side chain numbers refer to the side chain structures in Figure 2.2 (p103). 285 (15 × 19) side chain combinations were tested. In each ANOVA investigation two poses were investigated so there were 570 cells (possible sets of conditions) in total. The side chains tested were selected manually on the basis of their size and similarity to the Phe19, Trp23 and Leu26 p53 side chains, as described on p104.

Table 2.23: Summary statistics for the MM-PBSA data following separation by scaffold type and impution ready for ANOVA analysis

Scaffold	Poses Studied	Number of Cells	Number of Results		ΔG Values /kJ mol^{-1}		Levene's Test	
			Total	Imputed	Median	Q_3 - Q_1	F	P
N1	A,B	570	1308	74	-33.34	6.57	1.417	<0.0005
O1	A,D	570	1255	98	-32.45	7.08	1.291	0.001
O2	A,C	570	1126	134	-34.06	6.40	1.143	0.057
S1	A,D	570	1205	136	-33.49	6.70	1.307	0.001
S2	A,C	570	1349	50	-32.84	6.58	1.172	0.020

The pose letters refer to the arrangements depicted in Figure 2.23 (p187). The number of cells is the number of possible combinations of factor conditions, in this case 2 × 15 × 19, the number of poses multiplied by the product of the numbers of side chains at each position. For details of the impution carried out see the "Missing values" section beginning on p296. Non-parametric statics for the calculated free energies of binding follow the impution statistics, including the interquartile range (Q_3-Q_1). Finally, on the far right are the results of Levene's test for the homoscedasticity of the residuals. F is the test statistic. In the P column is the probability of a Levene's test test statistic value equal to or greater than the one shown in the penultimate column being observed if the residuals are homoscedastic.

Table 2.23 shows how, to further reduce the sparsity of the results for this first investigation, only poses which satisfied two of the four pose conditions were analysed in the ANOVA for each scaffold: pose A and another pose condition which was often fulfilled when using the scaffold in question. The table indicates the number of poses investigated for each scaffold, the number of values imputed for each analysis, nonparametric statistics for the predicted ΔG values and the F and p values of Levene's test for residual homoscedasticity.

The ΔG predictions are negative (as shown in the median ΔG column of Table 2.23), in contrast to the results obtained after the local search docking above, which were positive (Table 2.17, p177). This is consistent with only the global docking finding the correct binding positions.

The negative ΔG values are large in magnitude. The small size of the interquartile ranges (Q_3-Q_1) in Table 2.23 relative to the median ΔG values suggests that there is a similarly sized component absent from each prediction. No normal mode analysis was carried out to account

for the decrease in entropy of the ligand and protein upon binding. When the size of this entropy change is determined for similar ligands, there is typically little variation in the size of the values obtained[249] so the systematic error in the results could be this unaccounted for entropic component.

The F values for Levene's test (p298) are less than 1.5 so the heterogeneity of the residual variances is low in magnitude in all of the models, even though, in most cases (where the p-values are less than 0.05 in the rightmost column of Table 2.23) the heterogeneity is significantly different from 0. For a discussion concerning the interpretation of Levene's test results for large samples see p298.

Table 2.24 gives the significance of each of each of the sources of variation when fitting the complete factorial design (Equation 2.7, p177) and the same model lacking the three-way interaction (Equation 2.8, p180). The p-value for the three-way interaction in the former model is less than 0.05 for every scaffold except the O1 scaffold, for which it is 0.16. The geometric mean of the three-way significances averaged over every scaffold is 0.018. This suggests that there is, in general, a significant three-way interaction. It confirms that the pose affects the affinity of the side chains and, furthermore, that it affects the nature of the statistical interaction between the side chains. This is consistent with one of the factors being the pose position because in different poses different parts of the protein will mediate the interaction of the side chains so their interaction could be different.

Table 2.24: For five different scaffolds, the significance of the ANOVA sources of variation for models fitted to investigate the effect of pose, N-terminal side chain and middle side chain choice on the MM-PBSA-predicted free energy of binding

Source of Variation	Statistical significance of source of variation (Results significant where P<0.05)									
	Full factorial model for each scaffold					Model with no three-way interaction				
	N	O1	O2	S1	S2	N	O1	O2	S1	S2
P	0.000	0.012	0.000	0.000	0.008	0.000	0.005	0.000	0.000	0.003
SC1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SC2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
P*SC1	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.003	0.000
P*SC2	0.000	0.052	0.008	0.003	0.060	0.000	0.030	0.008	0.004	0.053
SC1*SC2	0.020	0.003	0.081	0.005	0.021	0.690	0.006	0.210	0.010	0.179
P*SC1*SC2	0.013	0.160	0.020	0.040	0.001	N/A	N/A	N/A	N/A	N/A

The factors of pose, N-terminal side chain and middle side chain are denoted by P, SC1 and SC2 respectively. Where the significance is shown as 0.000, this means that the probability is less than 0.0005.

For each of the N1, O2 and S2 scaffolds, the model lacking the three-way interaction term has an insignificant p-value for the side chain 1-side chain 2 interaction (N1: 0.69, O2: 0.21, S2: 0.18). This suggests that the side chain interaction is heavily pose-dependent because it indicates that if the pose is not considered, the ANOVA analysis detects no systematic pattern

attributable to side chain interaction; the pose information is needed to make sense of the interaction between the side chains.

In the absence of the three-way interaction, a statistically significant interaction is still observed between the pose and each side chain, with the exception of the pose and side chain 2 interaction for the S2 scaffold (p-value 0.053) which is almost significant and could be a false negative result (a type II error).

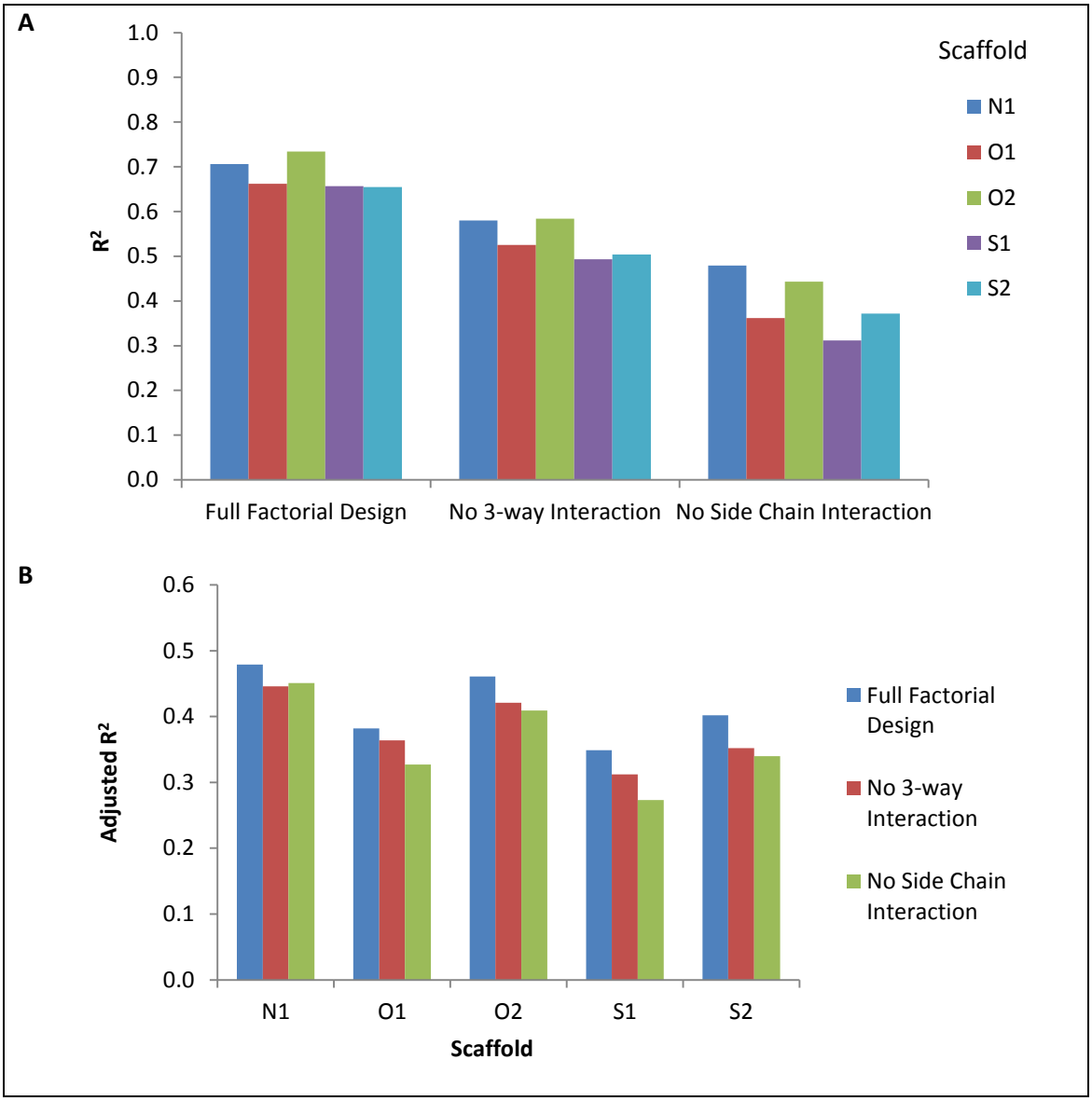


Figure 2.24: R-squared values for ANOVA models fitted to investigate the effects of pose, N-terminal side chain and middle side chain choice on the MM-PBSA-predicted free energy of binding. As shown, oligobenzamides based on all five scaffolds were investigated. A) Raw R-squared values. B) Adjusted R-squared values.

Figure 2.24 above gives R-squared and adjusted R-squared values for the full factorial models, models with no three-way interaction and models with no interaction terms (Equation 2.10, p180). The drop in the adjusted R-squared values upon removal of the (pose-side chain 1- side

chain 2) three-way interaction term (0.02 to 0.05 depending on the scaffold) is noticeable in Figure 2.24B. It is larger, for example, than the drop of just 0.01 in the adjusted R-squared value seen when removing the three-way (side chain 1-side chain 2- side chain 3) term from the O1 scaffold model where the three side chains were factors (Table 2.20, p181). This further supports a role for the pose in determining the nature of the interaction between the side chains.

2.3.5.2.4. After changes in binding pose have been accounted for, an interaction between the side chains still exists

In the analysis described above, the first side chain choice, second side chain choice and pose were treated as fixed factors. The ANOVA analysis revealed that the pose affected how the side chains interacted. This section describes the results of the data being split again, this time by pose and treatment of the scaffold as a factor.

Separate ANOVA analyses were performed for each pose with three factors: the scaffold and the side chain choice at each of the two attachment positions. First, the analysis was performed using the first, N-terminal side chain and the middle side chain. Then the analysis was repeated using the middle side chain and the last, C-terminal side chain. Table 2.22 (p188) shows the side chains used in the former case (with attachment positions 1 and 2) and Table 2.25 below shows the side chains used in the second case (with attachment positions 2 and 3).

Table 2.25: Side chains tested at each side chain position when using the side chains at positions 2 and 3 as factors in ANOVA models

Side chain position	Side chains tested at position	Number of side chains
1	11	1
2	1, 3, 5, 6, 9-13, 15, 21, 22, 25, 26, 30, 32, 37, 39, 44	19
3	11, 14, 15, 17-19, 21-23, 30, 32, 36-38, 40	15

The numbers refer to the side chain structures in Figure 2.2 (p103). The side chains tested were selected manually on the basis of their size and similarity to the Phe19, Trp23 and Leu26 p53 side chains as described on p104.

Table 2.26 shows summary statistics for the results and including the degree of impution required and the results of testing for residual homoscedasticity. Notice that a significant degree of impution was required, which increases the chance of a type II error (observing no interaction where there is one in reality). The F values for Levene's test are low, as in Table 2.23, so conclusions can reliably be drawn from the results.

Table 2.26: Summary statistics for the MM-PBSA data following separation by pose and impution ready for ANOVA analysis

Side Chains	Pose	Scaffolds Considered	Number of Cells	Number of Results		ΔG Values /kJ mol^{-1}		Levene's test	
				Total	Imputed	Q_2	$Q_3 - Q_1$	F	P
1 and 2	A	N1,O1,O2,S1,S2	1425	2874	335	-32.90	7.01	0.928	0.928
	B	N1,O1,O2,S2	1140	2126	324	-32.74	6.74	1.123	0.030
	C	O2,S2	570	1417	30	-33.88	6.89	1.340	<0.0005
	D	O1,S1	570	1133	136	-32.56	5.95	1.141	0.059
2 and 3	A	N1,O1,O2,S1,S2	1425	3234	180	-34.50	7.07	1.118	0.013
	B	N1,O1,S1	855	1908	144	-34.73	6.46	1.295	<0.0005
	C	O2,S2	570	1622	6	-35.81	6.33	1.306	<0.0005
	D	O1,S1	570	1387	68	-33.01	5.99	1.229	0.004

The number of cells is the number of combinations of conditions. 1425, 1140 and 570 are the result of multiplying 285, the product of 15 and 19 (the numbers of side chains used at each side chain position) by the number of scaffolds investigated in each case. Q_3-Q_1 is the interquartile range; 50% of the values lie within this distance of each other. Non-parametric statistics (the median and interquartile range) are appropriate because the energies are not normally distributed. Levene's test is a test for homoscedasticity of the residuals after fitting. F is the test statistic and P is the probability of an equal or greater test statistic being produced due to chance if the residuals are actually homoscedastic (the null hypothesis).

Table 2.27: The statistical significance of sources of variation in eight ANOVA models fitted to investigate the influence of scaffold and side chain choice on affinity

Side chains changed	Pose	Full factorial model three-way significance	Significance of the two-way interactions for the model with no three-way interaction		
			Sc*SA	S*SB	SA*SB
1 and 2	A	0.082	<0.0005	<0.0005	0.018
	B	0.034	<0.0005	<0.0005	0.528
	C	0.064	0.028	0.581	0.007
	D	0.026	0.009	0.228	0.180
2 and 3	A	0.084	<0.0005	0.001	0.505
	B	0.087	<0.0005	0.009	0.019
	C	0.515	0.002	0.009	0.019
	D	0.241	<0.0005	0.072	0.001

The models are vary in terms of the pair of adjacent side chains used as factors (1 and 2 or 2 and 3) and the pose that the oligobenzamides are in (A, B, C or D). The factors Sc, SA and SB are the scaffold, the first side chain of the pair and the second side chain of the pair respectively. The significance of the three-way interaction for the full factorial ANOVA is shown and then the significance of the two-way interactions in the model with no three-way component. Type III sums of squares were used throughout so the two-way significances refer to their respective two-way interaction after the other two-way interactions have been accounted for. (The significance of an interaction is the probability of the observed results occurring by chance if the interaction does not exist.)

For each pose, Table 2.27 shows the significances of the three-way interaction in the full factorial model and of the two-way interactions in the model lacking the three-way term. When only compounds with similar poses are compared, an interaction between the side chains is still

observed. In every case except one, either the three-way interaction or the two-way side chain interaction in the model with no three-way interaction is significant at the 5% significance level.

2.3.5.2.5. The choice of scaffold could affect not just the effect of each side chain on affinity but also the interaction between the side chains

The oligobenzamide scaffolds have different degrees of flexibility. Figure 2.9 (p133) showed RMSD values for two different ligands, the N1 and O1 scaffold starting molecules (Figure 2.1, p102). After equilibration, the RMSDs of the two ligands' atoms relative to their starting pose are different, over 2 Å for the N1 scaffold and under 2 Å in the case of the latter. The N1 scaffold could be more flexible than the O1 scaffold because it lacks intramolecular hydrogen bonds.

Interconversion of the O1 and S1 scaffolds by TI (p137) revealed that the side chain attachment atom in these scaffolds has an impact on affinity, an effect which could reflect differences in flexibility between these scaffolds as well, given the increased strength of hydrogen bonds involving oxygen relative to those involving sulphur and the increased directionality of the latter relative to the former. A steric effect unique to the S2 scaffold (the large size of the sulphur atom and its close proximity to the hydrogen atom with which it interacts) is thought to increase its flexibility[173].

The differences in flexibility between the scaffolds could result in differing degrees of mechanical coupling between the side chains. The results in Table 2.27 above support this hypothesis because they indicate that, at the 5% significance level, the interaction of side chains depends on the choice of scaffold. Specifically, the geometric mean of the three-way interaction significances when side chains 1 and 2 are modified (0.082, 0.034, 0.064 and 0.026) is 0.046, which is less than 0.05. For the modification of side chains 2 and 3 (0.084, 0.087, 0.515 and 0.241) the geometric mean is 0.174, which is not significant. This latter result could differ because the poses A, C and D are defined using the positions of side chains 1 and 2 so poses within the A, C and D groups may have a variety of conformations around side chain 3, reducing the overall fit of the model, including the effect of the scaffold. Consistent with this, the adjusted R-squared values for the models involving modification of side chains 2 and 3 are generally lower than those involving modification of side chains 1 and 2 (Figure 2.25).

There is typically a 10% difference between the adjusted R-squared values of the full factorial models (the models with the scaffold-side chain-side chain three-way interaction) and the corresponding models lacking no interaction terms (Figure 2.25) and much of the drop in R-squared is realised when the three-way term is removed. Therefore, the effect of the scaffold on the interaction between the side chains could be of some magnitude and the effect of the scaffold on the communication of the side chains should be investigated further.

Azzarito *et al.*[51] discuss how different scaffolds can adapt to arrange side chains in the same positions and thus have very similar affinities. The Mdm2 binding site is also flexible and has been shown to adapt to bind different peptides with similar affinity[79]. The results herein are not consistent with these ideas. Instead, they suggest a prominent role for the scaffold in not just determining the affinity of each side chain but also in affecting how those side chains communicate.

The docking results discussed above (question 5, p164) indicate that the scaffold can often bind directly to the protein. Although the significance values of the main effects in the ANOVA here are biased due to the impution, they are consistent with this conclusion, in that they show a large scaffold main effect ($p<0.05$) in all cases except for modification of side chains 2 and 3 in pose D where the p value is 0.85. If the scaffold itself interacts with the Mdm2 protein and shifts the positon of the side chains so as to optimise its own binding, this could explain its influence on the side chain effects.

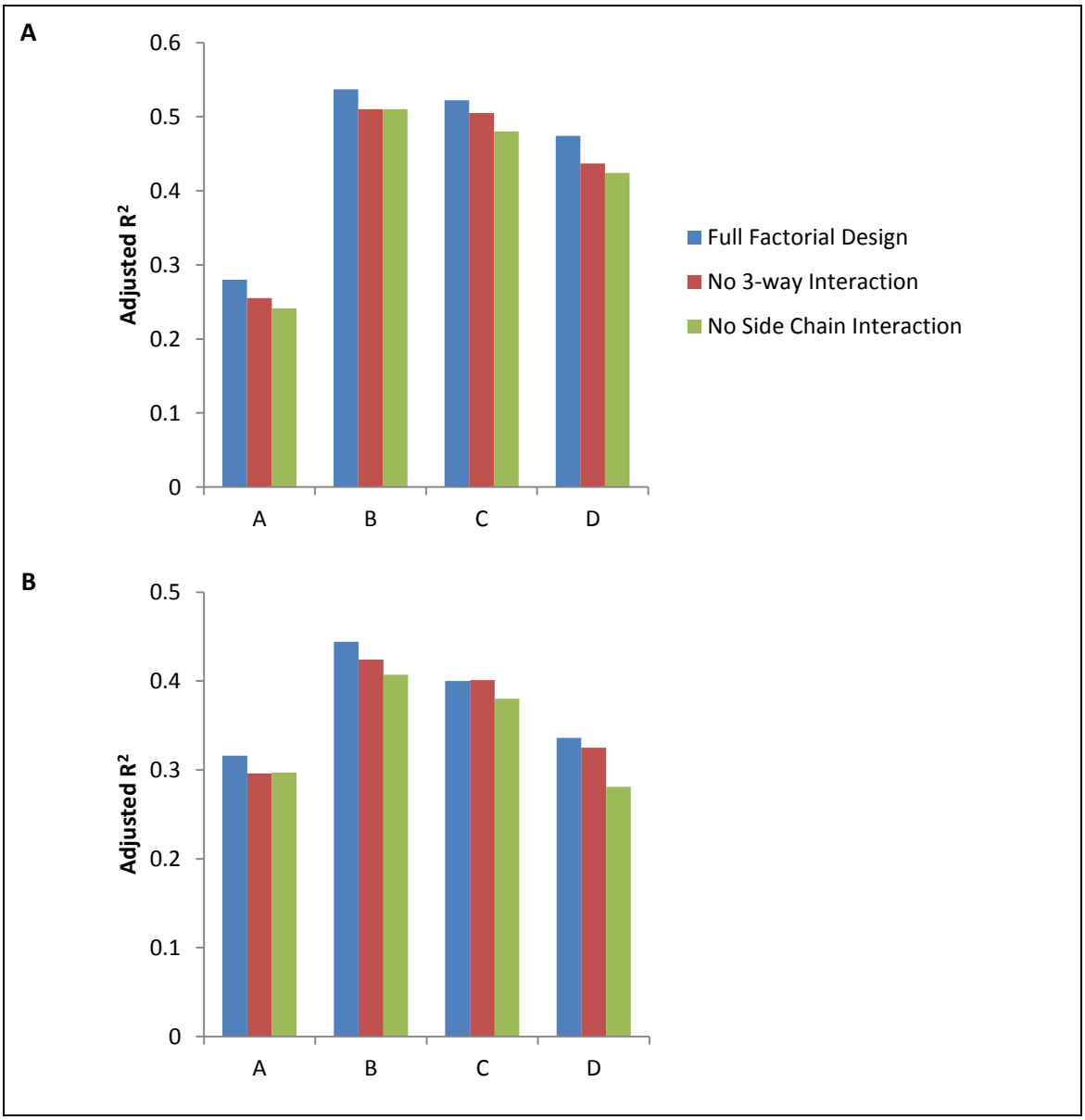


Figure 2.25: Adjusted R-squared values for the eight ANOVA models referred to by Table 2.26 and Table 2.27. A) Using the N-terminal side chain and the middle side chain as factors. B) Using the middle and C-terminal side chains as factors.

2.4. Conclusions and Further Work

The initial aim of this project was to demonstrate that the relative binding affinities of oligobenzamides can be predicted computationally. The prediction of accurate binding affinities requires that binding poses be generated for possible ligands, that molecular dynamics simulations be run from the docked poses and the trajectories then analysed to estimate the free energy of binding for each compound.

One of the challenges of simulating oligobenzamides is correctly calculating the energy associated with the torsion angles of the backbone. Modification (using TI) of the barrier heights for the bonds about which rotation is restricted in the O and S-linked scaffolds had little effect on their binding energy so any effects side chains have on the scaffold torsion energies in these scaffolds is unlikely to prevent the accurate ranking of oligobenzamides. Because in the N1 scaffold the nitrogen atoms holding each side chain are within the backbone, the torsion parameters could vary depending on the choice of side chains, making this scaffold more challenging.

In an attempt to predict the affinities of 31 compounds based on the N1 scaffold, two approaches were tried, namely, using TI and using implicit solvent methods. In the former, the effect of small changes to the side chain at each of the three side chain positions was evaluated and it was assumed that side chains made independent contributions to the binding energy. The correlation between the predicted free energies of binding and the experimentally determined affinity ranking of the compounds was found to be poor. Specifically, Spearman's rank, the absolute value of which ranges from 0 for no correlation to 1 for a perfect correlation, was found to be just 0.4. In contrast, when the implicit solvent MM-PBSA and MM-GBSA methods were used to make predictions, a much greater correlation with experimental results was observed. While the exact correlation depended on the settings used, Spearman's rank values of roughly 0.55 were typical, provided that the trajectory of the free protein was extracted from the trajectory of the complex and, when using the MM-PBSA method, the relative internal dielectric was not increased from its default value of 1. The greater correlation obtained with the implicit solvent methods could reflect the fact that it was unnecessary to assume compounds all bound in the same pose; all of the compounds could be docked and simulated.

To investigate further the assumption that all oligobenzamides bind to Mdm2 in the paradigmatic pose of the O1 scaffold starting molecule (where the side chains occupy the binding sites of p53 residues Phe19, Trp23 and Leu26) regardless of their side chain structure, 25,000 compounds were docked using Autodock Vina. Oligobenzamides were predicted to bind in many different poses, indicating that this prerequisite for TI may be violated.

To investigate the extent to which the side chains of an oligobenzamide influence the binding of each other, both through their effect on binding pose and within a particular binding pose, the MM-PBSA method was used to predict the affinities of a large combinatorial library of oligobenzamides in different poses and the affinities were studied using ANOVA.

When the choice of side chain at each position was treated as a factor, a statistical interaction between the side chains was evident; however, inclusion of interaction terms in the models did not significantly increase their adjusted Nagelkerke R^2 values, indicating that the interactions are small in magnitude.

Because the ANOVA results show a statistical interaction between binding pose and the choice of side chains, grouping compounds based on their likely binding position (from docking) would be appropriate when predicting the effect different side chains have at each side chain position. However, due to the small magnitude of the interactions between side chains, after accounting for the effect of side chain choice on the binding position, it might not be appropriate to test large combinatorial libraries in each binding site.

The models used do not explain a significant amount of variation in the results. This suggests that making predictions using longer simulations might be useful, something which would only be practical if independent contributions from each side chain were assumed, reducing the number of compounds that require testing. Within each possible oligobenzamide binding position, each side chain position could be investigated separately to determine the contribution different functional groups at that position are likely to make to the binding affinity. To predict the relative binding affinity for a new molecule, the compound could be docked to identify where it was likely to bind and this would indicate which set of side chain contribution predictions to sum to get the total predicted binding energy of the compound.

To predict the side chain properties likely to increase affinity at a particular binding site, it is useful to generate oligobenzamide poses within that site. Local optimisation of a docked oligobenzamide following subsequent *in situ* modification of its side chains using Autodock led to MM-PBSA values which were positive. This suggests that the optimised poses were not realistic.

FlexX is a docking program designed to place ligands in a pose close to a conformation previously generated by high throughput screening. A fragment library was docked using FlexX and the docking scores were analysed to identify the desirable properties of side chains when an oligobenzamide binds in the expected, idealised conformation occupying the Phe19, Trp23 and Leu26 binding sites. A smaller side chain is favourable at the first side chain attachment site, a larger side chain is favoured at the second position and aliphatic side chains as opposed to aromatic ones are favourable at the last side chain position. These findings are consistent with the properties of the p53 side chains mimicked in this binding pose. However, the results also

suggested that more hydrophilic side chains would bind more strongly, a finding which is not consistent with earlier work[31,253] and could be explained by the lack of terms to account for ligand and protein desolvation in the FlexX scoring function[229,254]. These findings highlight the need to run simulations and use more accurate methods of affinity prediction than scoring.

Sampling of all the possible torsion angles of oligobenzamides may require tens of nanoseconds worth of simulation time[172]. Fuller *et al.* demonstrates how Hamiltonian replica exchange can be used to increase the efficiency of sampling[203]. Simulations (replicas) are started at various lambda values and the lambda values are then swapped between the simulations at intervals (equivalent to the exchange of coordinates and momenta between simulations running at different lambda values). This reduces the chance of each simulation becoming stuck in an energy minimum and ceasing to explore the phase space[271]. Replica exchange could be used in future work to reduce sampling error.

While the TI results presented in this chapter did not correlate strongly with experimental results this could reflect the way in which TI was applied; TI might be a useful method for comparing ligands which docking suggests have very similar binding poses. One of the problems with TI is the requirement to perform many simulations in order to determine the effect of the reaction coordinate (λ) on the potential energy of the system. In this respect, a method known as lambda dynamics[272] could be a useful alternative. In this method, lambda is a variable which can change freely during the simulation so its effect on the potential can be ascertained in a single simulation. While a lambda dynamics simulation might have to be long, it would not necessary to simulate multiple systems to equilibration as when using the TI method. Multiple lambda variables can be used. This could allow multiple functional groups to be tested at a particular side chain position in a single simulation or even simultaneous modification at multiple side chain attachment positions.

In order for methods such as TI and lambda dynamics to be effective, the poses of ligands must be determined accurately. When docking was performed in this project, the flexibility of the protein was neglected. In future work, flexible docking could be performed in which the entire protein or just the protein side chains can move[152] or, alternatively, compounds could be docked into an ensemble of structures generated by a molecular dynamics simulation. Flexibility can lead to poorer docking because information in the in initial crystal structure is lost[145]. However, the Mdm2 binding site is flexible so conformational change might be essential to facilitate accurate determination of binding poses. (See Figure 1.3 on p32.) After fully flexible docking, it might be difficult to define where the p53 binding pockets are for the purpose of pose interpretation. If an ensemble of structures were produced by molecular dynamics simulations, the p53 peptide could be simulated in the binding site. This would not only allow the pocket positions to be tracked but also force them to stay open.

Synthetic work on oligobenzamides has focussed on compounds with N-linked[245] and O-linked[49] side chains as opposed to ones where attachment is through sulphur. Modification of side chain attachment atoms using TI suggested that the sulphur atoms in oligobenzamides based on the S1 scaffold contribute more to binding energy than the corresponding oxygen atoms in the O1 structure. Consequently, some sulphur-linked oligobenzamides might bind more strongly than their equivalent oxygen-linked molecules and the sulphur linked scaffolds should be investigated further.

It is necessary to demonstrate a high correlation between predicted and experimental results to show the practical use of implicit solvent methods in the prediction of the oligobenzamide affinities. While the computational methods used in this chapter could be developed further the experimental results used for their evaluation must be accurate too for a high correlation to be obtained. The experimental results used in this chapter were generated using a fluorescence polarisation competition assay. In this assay, an inhibitor binding to Mdm2 displaces a fluorescently-labelled p53 peptide. The release decreases the anisotropy of the label fluorescence and this is detected. Unfortunately, the collaborating chemistry group had found evidence of compounds binding directly to the fluorescently labelled peptide[42]. Furthermore, some variation in the ranking used in this chapter was observed when the assay was repeated. An orthogonal assay was thus needed to assist with validation of the computational findings. The next chapter details the development of a FRET-based assay which could be used instead of or in addition to the fluorescence polarisation assay in future.

The importance of correct binding pose prediction on the accuracy of predicted affinities warrants validation of the oligobenzamide docking poses generated by Autodock Vina. This is challenging because there is no crystal structure of an oligobenzamide bound to Mdm2. Chapter 4 describes the production of ^{15}N-labelled Mdm2 L33E and its use in NMR studies to investigate the binding of small molecules to Mdm2.

3. Novel FRET Assays for Monitoring Inhibition of the p53-Mdm2 Interaction

3.1. Summary

This chapter describes the development of FRET-based assays to determine the relative affinity of potential inhibitors for the p53 binding site of Mdm2. These assays all use GFP-labelled Mdm2 but the FRET acceptor fluorophore differs. Initial trials using Yellow fluorescent protein (YFP) were of limited success due to the significant overlap of the GFP and YFP emission spectra. Subsequent trials using eosin were much more successful and it was possible to develop both a cuvette-based assay and a plate reader-based assay. Each of these assays was validated with some inhibitors. The development of a plate reader assay is significant because it indicates the potential for robotic, high-throughput screening.

Despite the success of this *in vitro* assay, it could not be developed for *in vivo* work because p53-labelled eosin cannot be expressed within a cell. Consequently, a second protein FRET acceptor was trialled, mCherry. Two p53-mCherry proteins were made, one using a wild-type peptide and the other using a mutated peptide known to bind more strongly than the wild-type[273]. The FRET assay results reflected this difference of affinity. This observation, coupled with the fact that inhibitors known to bind to the p53 binding pocket of Mdm2 displace the p53-eosin and p53-mCherry proteins in their respective assays, suggests that, despite the labelling, the p53 peptide in each case still binds in the expected position, validating the use of the FRET assays for inhibitor testing.

In the context of FRET analysis, oligobenzamides exhibit some inconvenient features; they do not bind strongly to Mdm2, they are coloured, they bind to some aromatic small molecule fluorophores and they have a low solubility. The assays described in this chapter use various methods to overcome these significant problems. Firstly, the eosin and mCherry assays are performed with 10% dimethylsulphoxide (DMSO) to increase the solubility of a low-solubility

compound, such as an oligobenzamide. Secondly, when testing inhibitors, the drop in GFP-Mdm2 fluorescence, rather than the fluorescence of the acceptor or the ratio of GFP to acceptor fluorescence, is used to monitor FRET, so any change in acceptor fluorophore emission due to oligobenzamide binding has no effect on the measurements. Finally, the high absorbance and low solubility of oligobenzamide compounds makes testing high inhibitor concentrations impossible, something which would normally make it impractical to find the affinity of weakly binding compounds. Thus, in the eosin assay here described, a no-eosin control is used to eliminate mathematically the compound absorbance and make the endpoint of the titration (equal fluorescence of the mix and the control) known. The latter makes it possible to fit a titration curve to the assay results and determine the IC_{50} of an inhibitor without having to add inhibitor until the level of FRET reaches a plateau.

Key research outcomes described in this chapter:

- Development of FRET-based assays for testing oligobenzamides and other inhibitors with low solubility, low affinity and significant absorbance.
- Adaption of assays using p53-eosin for microwell plates in preparation for high-throughput screening.
- Testing of two p53-mCherry proteins which could in future be used in an *in vivo* assay.

3.2. Introduction

The previous chapter describes the computational prediction of binding affinities, specifically, prediction of the affinities of a diverse library of oligobenzamides for Mdm2. The compounds investigated included 31 oligobenzamides based on the N-linked scaffold which had been synthesised and for which fluorescence anisotropy (FA) single-point assay data were available, allowing *in silico* and *in vitro* results to be compared. There was a moderately strong correlation between Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) predictions and FA laboratory results. The differences between measured and predicted binding affinities could indicate that the computational predictions are not very accurate but will also reflect error in the laboratory measurements.

In the FA assay used to generate this laboratory data, a compound is titrated into a solution of unlabelled Mdm2 and fluorescein-labelled p53. The p53-fluorescein is excited with polarised light and the polarisation of the light emitted by the fluorescein is monitored[42]. When an inhibitor of the p53-Mdm2 interaction is added, the p53-fluorescein which was bound to the Mdm2 is released. Free in solution, the peptide can rotate faster resulting in a decrease in the anisotropy of its fluorescence. This is a useful assay due to its high signal to noise ratio and the small amount of peptide required; however, there is some evidence that compounds bind to the fluorescein label[42]. If they do, the overall anisotropy is not directly indicative of the proportion of peptide which is bound, making any calculated K_d inaccurate. Consequently, there is a need for a method that directly measures the binding of Mdm2 to p53 peptide rather than the mobility of the p53 peptide. The results of such a method might correlate better with *in silico* predictions.

3.3. Methods

3.3.1. FLUORESCENCE MEASUREMENTS

3.3.1.1. Equipment

All measurements were carried out in Dulbecco A phosphate buffered saline (PBS) (Oxoid) (160 mM NaCl, 3 mM KCl, 8 mM Na_2HPO_4, 1 mM KH_2PO_4) with 10% DMSO.

Fluorescence was measured using either a quartz cuvette in an RF-5301PC (Shimadzu) spectrofluorimeter kept, unless otherwise stated, at 10°C using a refrigerated water bath or using microwell plates, at 30°C using a POLARstar Optima (BMG Labtech) plate reader.

The cuvette volume was always between 500 and 700 µl of solution. The path lengths through the cuvette solution were 10 mm and 4 mm in the case of the excitation and emitted light respectively.

In the plate reader, black 96 well plates (Falcon) were used with a well volume of 100 µl. Curves were fitted to results using the nls (non-linear least squares) function of the programming language R[274].

3.3.1.2. Materials

FRET experiments were performed using GFP as the donor fluorophore and three different acceptor fluorophores: eosin and two proteins, YFP and mCherry. The concentration of the fluorescently labelled peptide and each protein was determined from their absorbance. The extinction coefficient was assumed to be unaffected by the labelled species. Absorbances were measured at 25°C using a Cary 300 Bio spectrophotometer (Varian).

L33E GFP-$Mdm2_{17-126}$ was produced fully purified by Geoffrey Plante (University of Leeds) at a concentration of 225 µM in pH 7.3 buffer comprising 25 mM Tris, 150 mM NaCl, 5% glycerol, 1 mM EDTA and 1 mM DTT. Aliquots were stored at -80°C. It was made prior to the start of the FRET work and the amount was sufficient to cover all of the laboratory work undertaken.

p53-YFP was supplied in pH 7.9 buffer comprising 20 mM Tris, 500 mM NaCl and 500 mM imidazole by Geoffrey Plante (University of Leeds). The protein was stored aliquoted at -80°C. An attempt was made to purify the p53-YFP by gel-filtration and remove the imidazole; however, the yield of gel filtration to remove the imidazole was prohibitively low. Consequently, the results shown in this thesis were produced with the imidazole-containing stock. The imidazole was significantly diluted in the experiments performed.

An oligopeptide comprising residues 13 to 33 of p53 (Sigma Aldrich) was supplied labelled with eosin at a concentration of 600 or 800 μM in DMSO by George Preston (University of Leeds). It was stored at -80°C.

Two p53-mCherry proteins, the wild-type and a mutant, were expressed and purified (p210).

Oligobenzamides PPY-2-75 and PPY-2-75 were supplied by Panchami Prabhakaran (University of Leeds) dissolved in DMSO at a concentration of 10 mM. They were stored at -20°C.

The known inhibitors Nutlin-3[40] (racemic mixture) and R5C3[275] were obtained as freeze dried solids from Sigma Aldrich. They were dissolved in DMSO upon arrival and stored at 5°C.

The peptide comprising residues 13 to 33 of p53 can be considered to be an inhibitor of the p53-Mdm2 interaction. It was obtained unlabelled from Sigma Aldrich.

Unlabelled Mdm2 was used in a fluorescence anisotropy experiment. It was supplied by Alice Bartlett (University of Leeds) in pH 7.3 buffer comprising 25 mM Tris, 150 mM NaCl, 5% glycerol, 1 mM EDTA and 1 mM DTT.

3.3.1.3. FRET measurements

This chapter describes the development of FRET-based assays in which the affinity of potential inhibitors for the p53 binding site of Mdm2 is assessed based on their ability to displace a labelled p53 peptide from GFP-labelled Mdm2. Three different FRET acceptors were tested. In each case, experiments were performed to validate the assay. Firstly, the effect of acceptor-labelled p53 on the spectrum of GFP-Mdm2 was determined in the presence and absence of the known p53-Mdm2 interaction inhibitor Nutlin-3. Secondly, the labelled p53 was titrated into GFP-Mdm2 to determine the dissociation (K_d) of the interaction. The K_d can be compared to that of the unlabelled proteins and is necessary for the conversion of inhibitor IC_{50} values into K_i values as described below. Thirdly, compounds were titrated into a mixture of GFP-Mdm2 and the labelled p53 to see if the relative affinities of known inhibitors were consistent with previously published work and, where possible, whether oligobenzamides generated a titration curve.

Experiments were performed in a spectrofluorimeter cuvette, with the exception of titrations involving the addition of compound to a mixture of GFP-Mdm2 and p53-eosin; these were carried out both into a cuvette and across a microwell plate, the latter in preparation for future high-throughput screening.

3.3.1.3.1. Titration of labelled p53 into GFP-Mdm2

3.3.1.3.1.1. With p53-YFP as the FRET acceptor

To determine the dissociation constant, a 500 µl solution was made up comprising 5 µM GFP-Mdm2 and 10 µM p53-YFP in PBS with 10% DMSO. The GFP fluorescence was measured at 505 nm, exciting at 470 nm. The sample was then serially diluted with a solution lacking the p53-YFP component but otherwise identical and the fluorescence measurement was repeated after each dilution.

There is some excitation of YFP in 470 nm light and some emission by YFP at 505 nm. To determine how much direct excitation and fluorescence of the p53-YFP contributed to the measured fluorescence, the experiment was repeated without the GFP-Mdm2 component. (A solution of 10 µM p53-YFP was diluted with buffer.) The results of this control experiment were subtracted from the fluorescences of the main GFP-Mdm2-containing experiment to obtain the GFP fluorescence.

Serial dilution entailed the removal of solution from the cuvette and then its subsequent replacement with diluent. The addition of cold solution was used to keep the cuvette cold. This dilution method allowed the starting main run and the control p53-YFP-only run solutions to be set up with identical p53-YFP concentrations, reducing the error when the direct YFP fluorescence was subtracted. (If the high p53-YFP concentration fluorescences had been measured at the end of a long titration involving many small additions, this error could have been large.) After each addition, the solution was mixed by pipetting.

Equation 3.1 was fitted to the fluorescences (after subtraction of the direct p53-YFP fluorescence)[276]. It was assumed that the GFP-Mdm2 fluorescence was proportional to the Mdm2-p53 concentration.

$$F = F_0 - A\Big([\mathrm{Mdm2}] + [\mathrm{p53}] + K_d - \sqrt{([\mathrm{Mdm2}] + [\mathrm{p53}] + K_d)^2 - 4[\mathrm{Mdm2}][\mathrm{p53}]}\Big) \quad \textbf{3.1}$$

Where F is the fluorescence, F_0 is the fluorescence in the absence of p53-YFP, K_d is the dissociation constant, [Mdm2] and [p53] are the concentrations of GFP-Mdm2 and p53-YFP respectively and A is another constant, specifically the change in fluorescence when 1 mol of Mdm2 binds to 1 mol of p53 in a volume of 1 litre. For each dissociation constant to be found, the mean logarithm of the K_d values from different repeats was determined and confidence limits on the K_d value were calculated using the standard error of the log K_d values[277].

$$SE_{10^x} = SE_x \frac{d}{dx} 10^x = (\ln 10) 10^x SE_x \quad \textbf{3.2}$$

3.3.1.3.1.2. With p53-eosin as the FRET acceptor

500 µl of a solution of GFP-Mdm2 at a concentration of 1 µM in PBS with 10% DMSO was placed in a cuvette and p53-eosin was titrated in such that the p53-eosin concentration in the cuvette increased exponentially with successive additions. The solution added contained 1 µM GFP-Mdm2 and, like the analyte, had 10% DMSO to ensure that only the concentration of p53-eosin (and the total volume) changed with each addition. To cover a large concentration range, two p53-eosin solutions were used, a 0.23 µM solution and a 30 µM solution. The GFP fluorescence at 505 nm was read during the titration, exciting at 450 nm. Excitation was also carried out at 450 nm when taking emission spectra. To mix the solution after each addition, a magnetic stir bar was moved up and down in the cuvette using magnets held outside the cuvette. The mean logarithm of the K_d and the standard error of this value were determined. These were used to calculate the K_d and its confidence limits.

3.3.1.3.1.3. With p53-mCherry as the FRET acceptor

The K_d was determined for each p53-Cherry protein by way of its titration into 1 µM GFP-Mdm2 in PBS with 10% DMSO. Equation 3.1 was fitted to the results of each titration to determine a value for the p53-mCherry dissociation constant. p53-mCherry was added from solutions of two different concentrations, as with the titration of p53-eosin into GFP-Mdm2, in order to cover a wide concentration range. The volume of each addition was progressively increased such that the concentration of p53-mCherry in the cuvette increased exponentially by a factor of 2.4 until a limit (5 µM) was reached. The concentration was then increased in a linear manner so that many observations were made at the high concentration end of the titration curve. This greatly increased the accuracy with which the dissociation constant (K_d) values could be calculated because the K_d has a big impact on the gradient of the titration curve as it begins to plateau.

3.3.1.3.2. Titration of compounds into mixtures of GFP-Mdm2 and labelled p53

3.3.1.3.2.1. With p53-YFP as the FRET acceptor

The known p53-Mdm2 interaction inhibitor Nutlin-3[40] was tested to see if the FRET assay enabled its K_i to be determined correctly. A solution of 5 µM Nutlin-3, 2 µM GFP-Mdm2 and 2 µM p53-YFP in PBS with 10% DMSO was titrated into a solution lacking the Nutlin-3 but otherwise identical. A similar titrant but with 350 µM Nutlin-3 was then added to increase the concentration further. The initial volume in the cuvette was 500 µl. Excitation was at 450 nm and the ratio of YFP (530 nm) to GFP (495 nm) fluorescence was monitored. Mixing was carried out using a very small magnetic stir bar, which was moved vertically through the solution. To determine the IC_{50} value (the inhibitor concentration at which there was 50% inhibition), the ratios from each repeat were fitted to equation 3.3, which may be derived from the Hill equation[278] by setting the Hill coefficient to 1. The mean of the log IC_{50} values was

calculated together with its standard error. The IC_{50} value and 95% confidence limits were found from these.

$$R = R_0 - A\frac{[I]}{[I] + IC_{50}} \quad \textbf{3.3}$$

Where R is the ratio, R_0 is the ratio in the absence of inhibitor, A is the range of the fluorescence change and [I] is the concentration of inhibitor.

3.3.1.3.2.2. With p53-eosin as the FRET acceptor

Titrations were carried out into a cuvette in the spectrofluorimeter or, alternatively, across a microwell plate, where the fluorescence was measured using a plate reader. In the spectrofluorimeter, excitation was at 450 nm and emission was measured at 495 nm. In the pate reader, a 450 nm filter was used for excitation and emission was detected using a 492 nm filter. The emission wavelengths, which are significantly lower than the maximum emission wavelength of GFP, were chosen to avoid the detection of p53-eosin, which emits very little light of wavelengths shorter than 495 nm. The filters used in the plate reader had a bandwidth of 5 to 7 nm.

Potential inhibitors were titrated into 1 μM GFP-Mdm2 and 2 μM p53-eosin in PBS with 10% DMSO. In the cuvette assay, the initial volume of the analyte was 300 μl. The inhibitor solution (titrant) was added so as to increase the concentration of inhibitor in the cuvette by a factor of 1.5 with every addition. To enable a large range of concentrations to be covered, inhibitor solutions of two different concentrations, x and 38x were used consecutively in each titration. Depending on the inhibitor in use x equalled 1.8 or 2.7 μM. These titrant solutions also contained 1 μM GFP-Mdm2, 2 μM p53-eosin and 10% DMSO to ensure that only the concentration of inhibitor changed. The increasing volume of solution in the cuvette was factored in when calculating the size of each addition. Mixing was performed using a miniature magnetic stir bar as described above. Each main titration was followed by a second control titration, identical in every way except that the p53-eosin was replaced with buffer in both the titrant and the analyte.

When the assay was performed using microwell plates, 16 exponentially increasing inhibitor concentrations from 0.11 to 230 μM were used for both the main and control titrations.

The main and control titrations were repeated at least 3 times. For each repeat, the ratio of the GFP fluorescence with p53-eosin to that of the control was found at each inhibitor concentration. A curve based on Equation 3.4 was fitted to these ratios to determine an IC_{50} value for each repeat. Calculation of the mean log IC_{50} value and its standard error enabled the IC_{50} value and its 95% confidence limits to be determined for each inhibitor.

$$R = R_0 + (R_\infty - R_0)\frac{[I]^n}{[I]^n + IC_{50}{}^n} \quad \textbf{3.4}$$

Where R is the ratio of the GFP fluorescences, the no-p53-eosin GFP fluorescence divided by the with-p53-eosin GFP fluorescence. R_0 is the ratio when there is no inhibitor, but with results from the cuvette experiments, the value of R_0 was assumed to be equal to the ratio for the lowest concentration point used. In the plate reader, R_0 was found to be 0.442 ± 0.014 for the majority of runs. Where this was not the case (for example, R_0 was 0.55), a run was considered anomalous and was repeated.

In Equation 3.4, n is the Hill coefficient. It was normally fixed at 1 but the plate reader results obtained with the oligobenzamide PPY-2-75 could not be fitted using a Hill coefficient of 1. They were consistent with a Hill coefficient of 2 so fitting was carried out with n fixed at 2 instead. This decision is discussed further in the results section on p235.

At high inhibitor concentrations there should theoretically be no FRET in either the main or control titrations. Their fluorescences should thus be identical, making R_∞, the ratio of these two fluorescences (at a very high inhibitor concentration) equal to 1. Consequently this value (1) was used when fitting Equation 3.4 to the ratios determined in the plate reader. Contrastingly, in the spectrofluorimeter experiment, a maximum ratio of 0.946 ± 0.015 was observed with Nutlin-3 and all of the results were fitted using this result. It is possible that the DMSO denatured some GFP-Mdm2 and p53-eosin in the long p53-eosin-containing experiment in the spectrofluorimeter, causing some aggregation which no amount of inhibitor could dissociate.

The efficacy of PPI inhibitors was quantified by determination of their dissociation constant (K_i). Where possible, the K_i was evaluated in preference to the IC_{50} value (the concentration at which the PPI disrupts 50% of the complexes) because it is not dependent on the concentrations of the proteins.

In the absence of inhibitor:

$$[\mathrm{Mdm2}]_0 = \frac{1}{2}\Big([\mathrm{Mdm2}]_\mathrm{T} - [\mathrm{p53}]_\mathrm{T} - \mathrm{K_d} + \sqrt{([\mathrm{p53}]_\mathrm{T} - [\mathrm{Mdm2}]_\mathrm{T} + \mathrm{K_d})^2 + 4[\mathrm{Mdm2}]_\mathrm{T}\mathrm{K_d}}\Big) \quad \textbf{3.5}$$

Where K_d is the p53-Mdm2 dissociation constant. $[p53]_T$ and $[Mdm2]_T$ are the total concentrations of p53 and Mdm2 respectively and $[Mdm2]_0$ is what the concentration of free Mdm2 would be if there were no inhibitor present.

$$[\mathrm{p53.Mdm2}]_{50} = \frac{1}{2}[\mathrm{p53.Mdm2}]_0 = \frac{1}{2}([p53]_T - [\mathrm{Mdm2}]_0) = \frac{1}{2}\big([\mathrm{p53}]_\mathrm{T} - [\mathrm{p53}]_0\big) \quad \textbf{3.6}$$

Where $[p53.Mdm2]_0$ is the free p53-Mdm2 concentration when there is no inhibitor and $[p53.Mdm2]_{50}$ is the same concentration but at the point of 50% inhibition.

According to Nikolovska-Coleska *et al.*[276],

$$K_i = \frac{[I]_{50}}{\frac{[p53]_{50}}{K_d} + \frac{[\mathrm{Mdm2}]_0}{K_d} + 1} \quad \textbf{3.7}$$

Where

$$[\mathrm{I}]_{50} = \mathrm{IC}_{50} - [\mathrm{Mdm2}]_{\mathrm{T}} + [\mathrm{p53.Mdm2}]_{50} + \frac{\mathrm{K_d}[\mathrm{p53.Mdm2}]_{50}}{[\mathrm{p53}]_{50}} \quad \textbf{3.8}$$

$[p53]_{50}$ and $[I]_{50}$ are the concentrations of free labelled p53 and inhibitor respectively at the point of 50% inhibition.

3.3.1.3.2.3. With p53-mCherry as the FRET acceptor

Inhibitors of the p53-Mdm2 interaction were titrated into 2 μM GFP-Mdm2 and 5 μM wild-type p53-mCherry in PBS with 10% DMSO up to a concentration of 60 μM. GFP was excited at 440 nm and GFP fluorescence was measured at 505 nm. No no-p53-mCherry control was used; Equation 3.4 was fitted as before to find each IC_{50} value, except that in this case R represented the raw GFP fluorescence. The minimum fluorescence, maximum fluorescence and Hill coefficient were allowed to vary when fitting each curve.

3.3.1.4. Additional experiments to investigate the GFP-Mdm2 and p53-eosin assay

3.3.1.4.1. Titration of unlabelled Mdm2 into p53-eosin

A high K_d was measured for the binding of GFP-Mdm2 to p53-Cherry. To investigate whether this was partly due to the presence of the GFP label, a titration was carried out without the GFP label to see if the K_d was lower. Unlabelled Mdm2 was titrated into 250 nM p53-eosin in a FluoroLog (Horiba) spectrofluorimeter. The titrant and analyte contained 10% DMSO. The eosin was excited with polarised light at 525 nm and its emission at 543 nm was measured using a second polarising filter. Both filters could be positioned horizontally (H) or vertically (V) and the anisotropy at each point was calculated using the fluorescence (F) obtained with each of all four possible filter orientations, HH, HV, VH and VV.

$$A = \frac{F_{VV} - \left(\frac{F_{HV} \times F_{VH}}{V_{HH}}\right)}{F_{VV} + 2 * \left(\frac{F_{HV} \times F_{VH}}{V_{HH}}\right)} \quad \textbf{3.9}$$

3.3.1.4.2. Titration of an oligobenzamide into p53-eosin in the absence of GFP-Mdm2

Previously published fluorescence anisotropy results suggest that oligobenzamides bind to p53-fluorescein[42]. Eosin is a similar compound to fluorescein so oligobenzamides might also bind to p53-eosin, an interaction which would complicate the analysis of GFP-Mdm2 and p53-eosin competition assays.

In a control experiment, the oligobenzamide PPY-2-75 was titrated into a solution of 1.6 μM p53-eosin in PBS with 10% DMSO across a microwell plate to investigate whether PPY-2-75 has any effect on p53-eosin fluorescence in the absence of GFP. 492 nm and 590 nm filters were used for excitation and emission respectively. A dilution factor of 1.5 was used to create 28 exponentially increasing concentrations from 0.9 nM to 50 μM. The plate was read at 30°C.

3.3.1.4.3. Single point assays measuring FRET and fluorescence anisotropy

To perform single point assays, compounds were added to a solution of 1 μM GFP-Mdm2 and 2 μM p53-eosin in a cuvette in the FluoroLog (Horiba) spectrofluorimeter and the FRET and fluorescence anisotropy of the p53-eosin were measured in quick succession. The fluorescence anisotropy of the eosin can be measured in the presence of GFP-Mdm2 if it is excited at 535 nm because this wavelength is too high to excite the GFP-Mdm2. A circular dichroism cuvette was used to minimise the solution volume and therefore the amount of labelled peptide used.

3.3.2. PRODUCTION OF P53-MCHERRY

3.3.2.1. Site directed mutagenesis: Production of doubly His-tagged wild-type and singly His-tagged mutant p53-mCherry constructs

A pET-28b-TEV plasmid encoding a 279 residue fusion protein comprising a p53-like peptide (Figure 3.1) and the fluorescent protein mCherry was produced by Jillian Augustine (University of Leeds). The encoded protein had oligo-histidine tags at both the N and C termini. The mutant peptide used had previously been shown to have increased affinity for Mdm2[279] with a K_d of 5.6 nM in 10% DMSO, considerably lower than the wild-type p53 transactivation domain peptide K_d of 0.9 μM[273]. It shares the key residues Phe19, Trp23 and Leu26 with the wild-type peptide. The crystal structure of the mutant with Mdm2 (PDB structure 3G03[280]) suggests that the L22Y mutation is at least partly responsible for the increased affinity; Tyr22 fills a hydrophobic space which is solvent-accessible when the wild-type binds (with Leu22)[273]. The mutant peptide also binds to Mdm4 more strongly than the wild-type peptide. In this case, the hydrophobic part of the Thr27 side chain is thought to contribute to the affinity through its interaction with Mdm4 residue Met53. In the wild type peptide, the polar backbone carbonyl group of Pro27 takes its place close to Met53.

I designed primers for site-directed mutagenesis to transform the construct above into, firstly, one with only one His tag (Figure 3.2) and, secondly, one encoding the wild-type p53 peptide (still with two His tags) (Figure 3.3). The site-directed mutagenesis was performed by Jillian Augustine who then supplied the resulting constructs.

MGHHHHHHHHHHSSGHENLYFQGSQLTFEHYWAQLTSSKGEEDNMAIIKEFMRFKVHMEGSVNGH
EFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEG
FKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDG
ALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGR
HSTGGMDELYKLEHHHHHH

Blue – Oligo-histidine tag
Green - p53-like sequence
Yellow - TEV protease cut site
Red - p53-mCherry

Figure 3.1: The amino acid sequence of the double His-tagged mutant p53-mCherry protein (encoded by DNA initially produced by Jillian Augustine). Residues are coloured as shown in the key.

Before site-directed mutagenesis (No stop codon)	GACGAGCTGTACAAG...CTCGAGCACCACCACCACCACCACTGA D E L Y K L E H H H H H H X
After site-directed mutagenesis Stop codon present	GACGAGCTGTACAAGTAACTCGAGCACCACCACCACCACCACTGA D E L Y K X
One of the two site-directed mutagenesis primers	5′- gacgagctgtacaagtaactcgagcaccacca – 3′

Figure 3.2: Diagram to show how a new stop codon was introduced into the p53-mCherry construct. The capitalised DNA sequences show the end of the mCherry coding sequence in the construct. Underneath each is the encoded amino acid sequence. Two primers are used for mutagenesis, each the reverse complement of the other.

Mutant construct (unmodified)	CAGCTGACCTTTGAACATTATTGGGCGCAGCTGACC AGCAGC Q L T F E H Y W A Q L T S S
Corresponding part of wild-type *mdm2* DNA	CAGGAAACCTTTAGCGATCTGTGGAAACTGCTGCCGGAAAACAGC Q E T F S D L W K L L P E N S
Corresponding part of forward primer	CAGGAGACCTTTTCAGATTTATGGAAGCTGCTGCCCGAAAACAGC Q E T F S D L W K L L P E N S
Forward primer (100 bp)	5'- gcggccatgaaaacctgtacttccagggatcccaggagacctttcagat ttatggaagctgctgcccgaaaacagcaagggcgaggaggataacatggc -3'

Figure 3.3: Diagram to show how the mutant p53 peptide was mutated to the wild-type peptide in a single site-directed mutagenesis experiment. Two primers are used for mutagenesis, each the reverse complement of the other. The top sequence shows the part of the mutant construct which was mutated. The two DNA sequences below show the corresponding part in the wild-type *mdm2* gene and in the forward primer. The entire sequence of the forward primer is shown at the bottom. Where sequences differ from that of the mutant construct, this is indicated in red.

3.3.2.2. Production of the singly His-tagged wild-type p53-mCherry construct

To produce a wild-type construct with only one His-tag, the two constructs (the wild-type with two His-tags and the mutant with one His-tag) were cut with *PstI* and *ClaI* (New England Biolabs) and the larger fragment from the wild-type construct digest (with the N-terminal end of p53-Cherry containing the wild-type p53 peptide) was ligated into the smaller fragment from the mutant-construct digest (with the C-terminal end of p53-Cherry with no His-tag) (Figure 3.12). *ClaI* was chosen because it cuts within the *kan* gene of the vector so kanamycin could be used to select for transformants containing both parts of the desired plasmid correctly orientated.

100 ng μl^{-1} of plasmid DNA was incubated with both *PstI* and *ClaI* in the supplied NEBuffer 4 with bovine serum albumin (BSA) at 37°C as directed by the manufacturer. After 2 hours' incubation, solutions were run on a 1% agarose gel made with and run in tris-borate EDTA (TBE) buffer (20 mM Tris-acetate, 0.5 mM EDTA). A constant voltage of 60 V was applied. RedSafe (iNtRON Biotechnology) was added to the gel to stain bands, which were cut out on a SafeImager blue light transiluminator (Invitrogen). The desired pieces of DNA were purified from their excised gel bands using the QIAquick gel extraction kit (QIAGEN) as directed by the manufacturer. DNA was eluted in the supplied TE buffer (10 mM pH 8 Tris-Cl buffer with 1 mM EDTA).

Ligation was performed over 1 hour at 37°C using T4 DNA ligase (Invitrogen) with the supplied buffer. DNA concentrations were measured using a NanoDrop 1000 Spectrophotometer (Thermo Scientific) and a ratio of 3 large DNA fragments to every small fragment was used in the ligation. DH5α cells were transformed immediately as described below.

3.3.2.3. Transformation of DH5α (cloning competent) cells

100 µl of DH5α *Escherichia coli* cells (*fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) in transformation buffer (10 mM PIPES, 15 mM $CaCl_2$, 250 mM KCl, 55 mM $MnCl_2$ (pH 6.4)) was thawed on ice. 9 µl of purified DNA in TE buffer (10 mM pH 8 Tris-Cl buffer with 1 mM EDTA) was added. The cells were left for 30 minutes on ice, incubated at 42°C for precisely 30 seconds (heat shock) and then returned to the ice for 5 minutes.

Standard practice is to incubate cells in 1 ml of 2xYT broth for 45 minutes at 37°C and then plate 40 µl or 400 µl of media on plates with antibiotic; however, attempts to transform with the single-His-tagged wild-type p53-mCherry construct and then plate repeatedly yielded no colonies suggesting inefficient ligation or transformation. Colonies were produced successfully by transfer of the transformed cells into 1 L of 2xYT media in a 2 L baffled flask (incubated at 37°C with shaking). After 45 minutes, kanamycin was added to the flask and it was left

overnight at 37°C to give any correctly transformed cell a chance to multiply. A 1 million times serial dilution of the saturated solution which resulted was spread on kanamycin-containing plates the next day, which were then incubated at 37°C to obtain individual colonies.

3.3.2.4. Identification of transformants with the correct construct

Colony PCR suggested that all of the colonies obtained had the desired construct and this was confirmed for some colonies by sequencing.

For colony PCR, cells from each selected colony were transferred to 30 µl of purified water. The water was boiled for 10 minutes. Then, 3 µl was used to complete a 30 µl PCR reaction mixture. Taq polymerase (New England Biolabs) was used with ThermoPol buffer (New England Biolabs) and custom primers for the N and C termini of the protein:

5′-AAAGGATCCCAGGAAACCTTTAGCGATCTGTGGAAACTG-3′ (N)

5′-AAACTCGAGTTACTTGTACAGCTCGTCC-3′ (C)

The protein-coding region was cut in the middle during the splicing process so PCR between the ends of the protein was sufficient to confirm that both halves of the open reading frame were present in a construct.

Details of the PCR program used for colony PCR are shown in Table 3.1.

Table 3.1: The program used for colony PCR

Stage		Temperature /°C	Time /s
Initial denaturation		96	120
Cycles (20)	Denaturation	94	30
	Annealing	40	30
	Extension	72	90
Final extension		72	240

3.3.2.5. Expression and purification of p53-mCherry

Both of the p53-Cherry constructs were purified in the same manner.

Novagen Rosetta 2 (Merck Biosciences) cells were transformed with the p53-mCherry construct as directed by the manufacturer. After inoculation, 6 litres of 2xYT broth (16 gl^{-1} tryptone soya broth (Oxoid), 10 gl^{-1} yeast extract (Oxoid) and 2.5 gl^{-1} sodium chloride) was incubated at 37°C until the optical density at 600 nm was 0.8. 1 mM IPTG was then added to induce expression and the cells were incubated at 18°C overnight.

After 15 hours, the cells were pelleted using 15 minutes' centrifugation at 6000 × g and then re-suspended in 20 mM Tris buffer (pH 7.9) with 0.5 M sodium chloride and 20 mM imidazole. DNase was added. Suspensions were sonicated (5 s 20 times using a Soniprep 150 (MSE) with mixing halfway through) and then centrifuged at 15000 × g for 30 minutes. The supernatants were filtered using a Millex-GP 0.2 µm filter (Millipore). Next, metal affinity chromatography was carried out. A 5ml HisTrap HP Ni^{2+} column (GE Healthcare) was equilibrated with the resuspension buffer and then loaded with each lysate. After washing (with 100 ml of the resuspension buffer), fractions (30 ml) were eluted using the same buffer but with progressively increasing concentrations of imidazole (30 mM, 150 mM, 300 mM, 500 mM and 1M).

The 300 mM imidazole fractions were concentrated using a Centricon YM-10 (10 kDa molecular weight cut-off) filter (Amicon) to 5 ml, passed through a Millex-GP 0.2 µm filter and then loaded on to a Superdex 75 Sephadex gel filtration column (GE Healthcare) which was equilibrated and then run with a pH 7.3 buffer comprising 25 mM Tris, 150 mM NaCl, 5% glycerol, 1 mM EDTA and 1 mM DTT. A single gel filtration column was used to process both proteins, the mutant first and then the wild-type.

SDS-PAGE gels were run with samples from mCherry-coloured fractions as indicated in the results section. The stacking gel (pH 6.8) comprised 125 mM Tris buffer, 0.1% SDS, 0.1% APS, 50 mM 5% v/v acrylamide, 0.25% v/v TEMED. The resolving gel (pH 8.8) comprised 375 mM Tris buffer, 0.1% SDS, 0.1% APS, 15% v/v acrylamide, 0.2% TEMED.

Protein from all peaks was stored at -80°C. Selected samples were concentrated using a Centricon filter (10k MWCO) to concentrations of up to 194 µM (mutant) and 385 µM (wild-type). The concentrations were determined using the mCherry absorbance, which was assumed to be unaffected by the preceding p53 peptide.

3.4. Results and Analysis

3.4.1. INITIAL TESTING OF GFP-MDM2

GFP-Mdm2 was expressed and purified by Jeff Plante (University of Leeds). He reported that the fluorescence of the GFP-Mdm2 appeared to decrease as a solution containing the protein was repeatedly pipetted. It is important for the fluorescence of the GFP to be constant for accurate FRET measurements to be made so this claim was investigated further.

A drop in fluorescence was observed upon mixing by pipetting. To investigate the effect of more agitation of the solution, mixing was then performed magnetically, using a small magnetic follower placed at the bottom of the 3 ml cuvette. The shutter was left open and the GFP fluorescence of a stirred solution and an unstirred control were monitored over time (Figure 3.4). As shown, the switch to magnetic mixing did not solve the problem of decreasing GFP fluorescence, suggesting that this mixing method was also too vigorous. Furthermore, it was deemed that regular use of the large amount of protein required for the 3 ml cuvette would be uneconomical so the cuvette size was reduced and this prevented conventional use of the stir bar. Therefore, to mix the solution in subsequent work, the magnetic follower was agitated vertically using two magnets (large stir bars) moved up and down manually either side of the cuvette. This appeared to solve the mixing problem because the GFP fluorescence no longer dropped upon mixing.

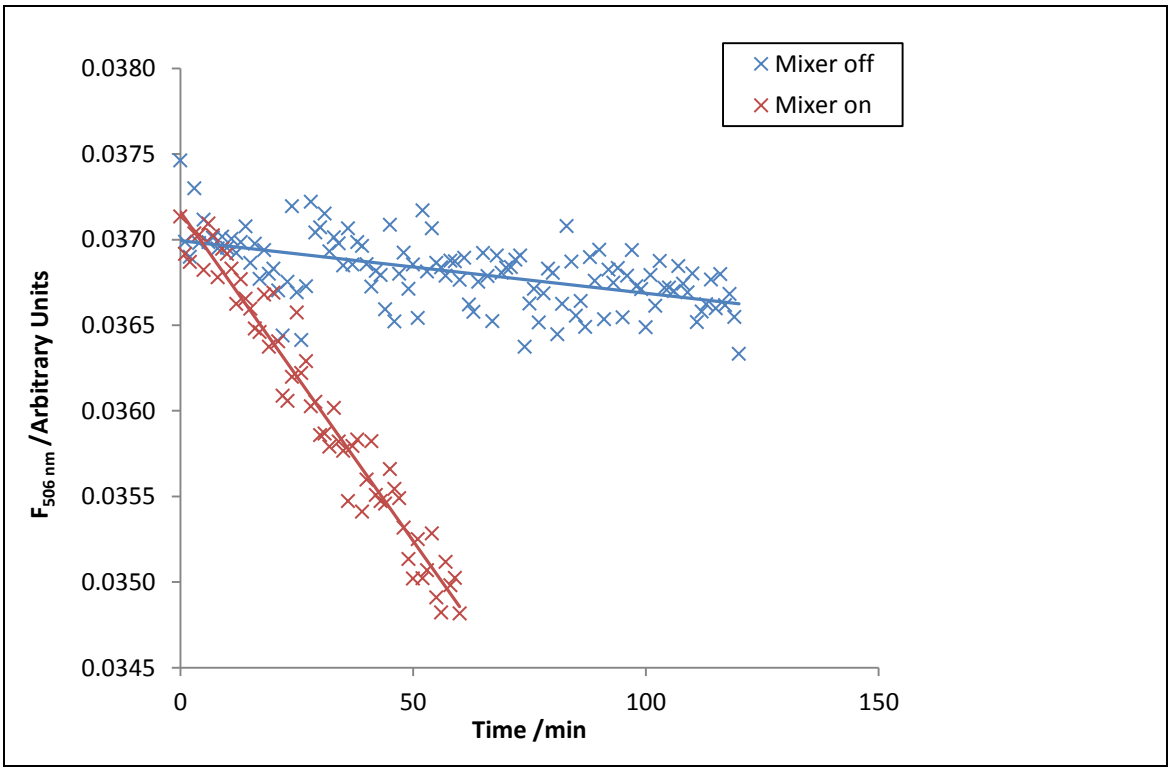


Figure 3.4: The effect of mixing on the fluorescence of a 0.56 μM GFP-Mdm2 solution. Red) With magnetic mixing. Blue) Without mixing.

3.4.2. EXPERIMENTS WITH P53-YFP

3.4.2.1. Titration of p53-YFP into GFP-Mdm2 solution

The GFP emission was reduced when p53-YFP was titrated into GFP-Mdm2 solution suggesting that FRET to the YFP was occurring (Figure 3.5). The dissociation constant (K_d) was found to be 1.3 ± 1.1 μM, which is consistent with the K_d of 0.9 μM found by Czarna *et al.*[273] for the binding to Mdm2 of an unlabelled peptide with a sequence corresponding to that of the p53 transactivation domain. The size of the 95% confidence limits on the calculated K_d reflects the difficulty of measuring a K_d accurately when using GFP and YFP. The emission spectra of these fluorophores overlap considerably so FRET has little effect on the combined spectrum. This is one of the reasons why GFP is rarely used with YFP; BFP (Blue Fluorescent Protein) with GFP and CFP (Cyan Fluorescent Protein) with YFP are more popular combinations[101].

3.4.2.2. Addition of Nutlin-3 to a solution containing p53-YFP and GFP-Mdm2

Addition of the known p53-Mdm2 interaction inhibitor Nutlin-3 (Figure 3.6A) to the p53-YFP and GFP-Mdm2 solution caused the GFP fluorescence to be almost fully restored to its initial value. Nutlin-3 did not cause the expected decrease in fluorescence at the YFP-emission peak; however, the fluorescence at this wavelength includes emission from both fluorophores. Given that the GFP emission increased, the YFP-emission must have decreased, consistent with disruption of the p53-Mdm2 interaction by the inhibitor.

Titration of Nutlin-3 into a solution of GFP-Mdm2 and p53-YFP yielded results to which a curve based on the Hill equation was fitted with the Hill coefficient fixed at 1 (Figure 3.6B). This gave a 50% inhibition midpoint (IC_{50}) at 1.23 ± 0.16 μM.

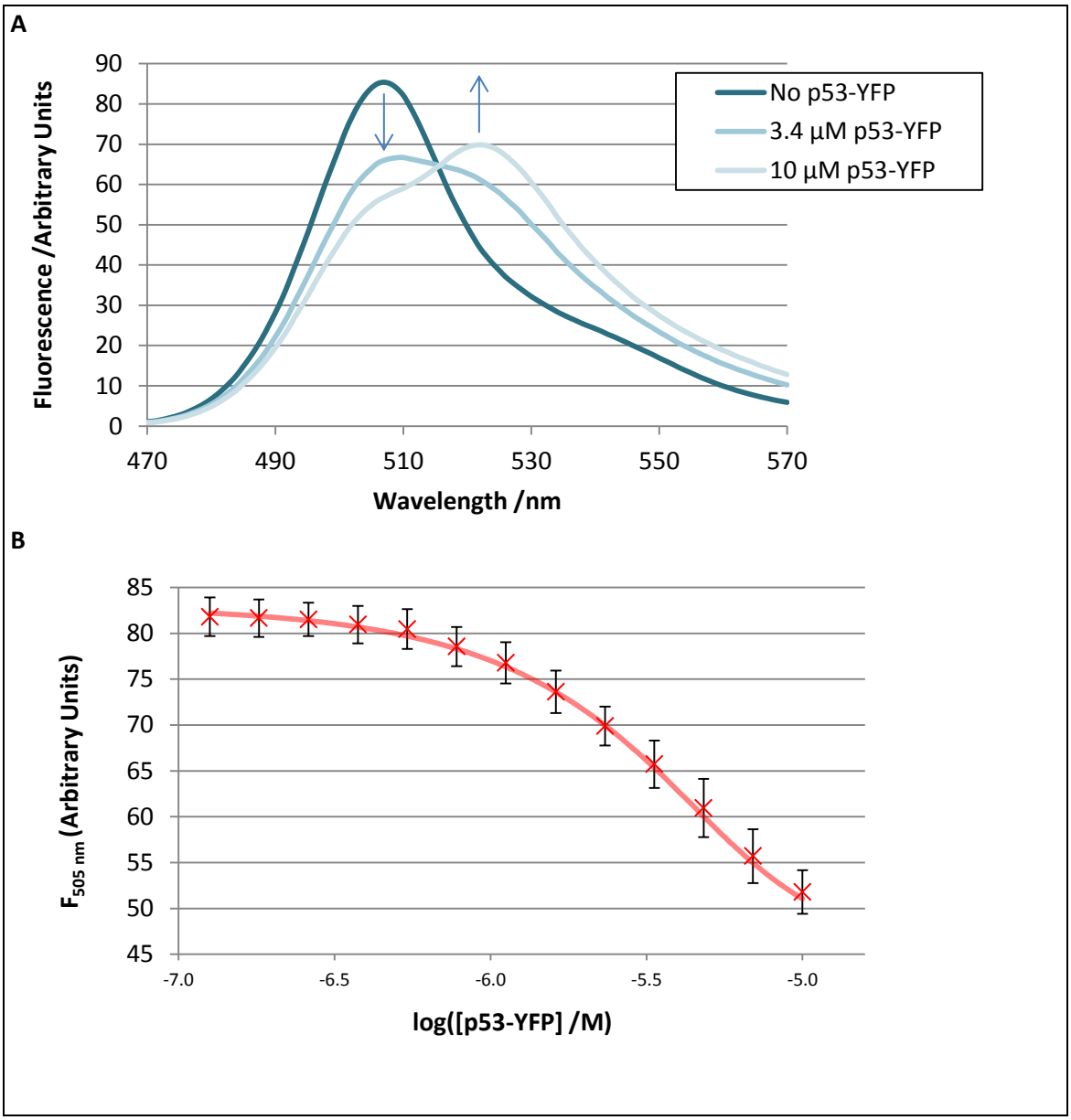


Figure 3.5: Graphs to show the consequence of titrating p53-YFP into 5 µM GFP-Mdm2. A) How changes in the concentration of p53-YFP affect the fluorescence spectrum. The GFP fluorescence falls as the concentration of p53-YFP increases. B) How addition of p53-YFP affects the fluorescence at 505 nm. The highest concentration tested was 10 µM because the YFP absorbance would be significant at higher concentrations. Excitation was at 460 nm.

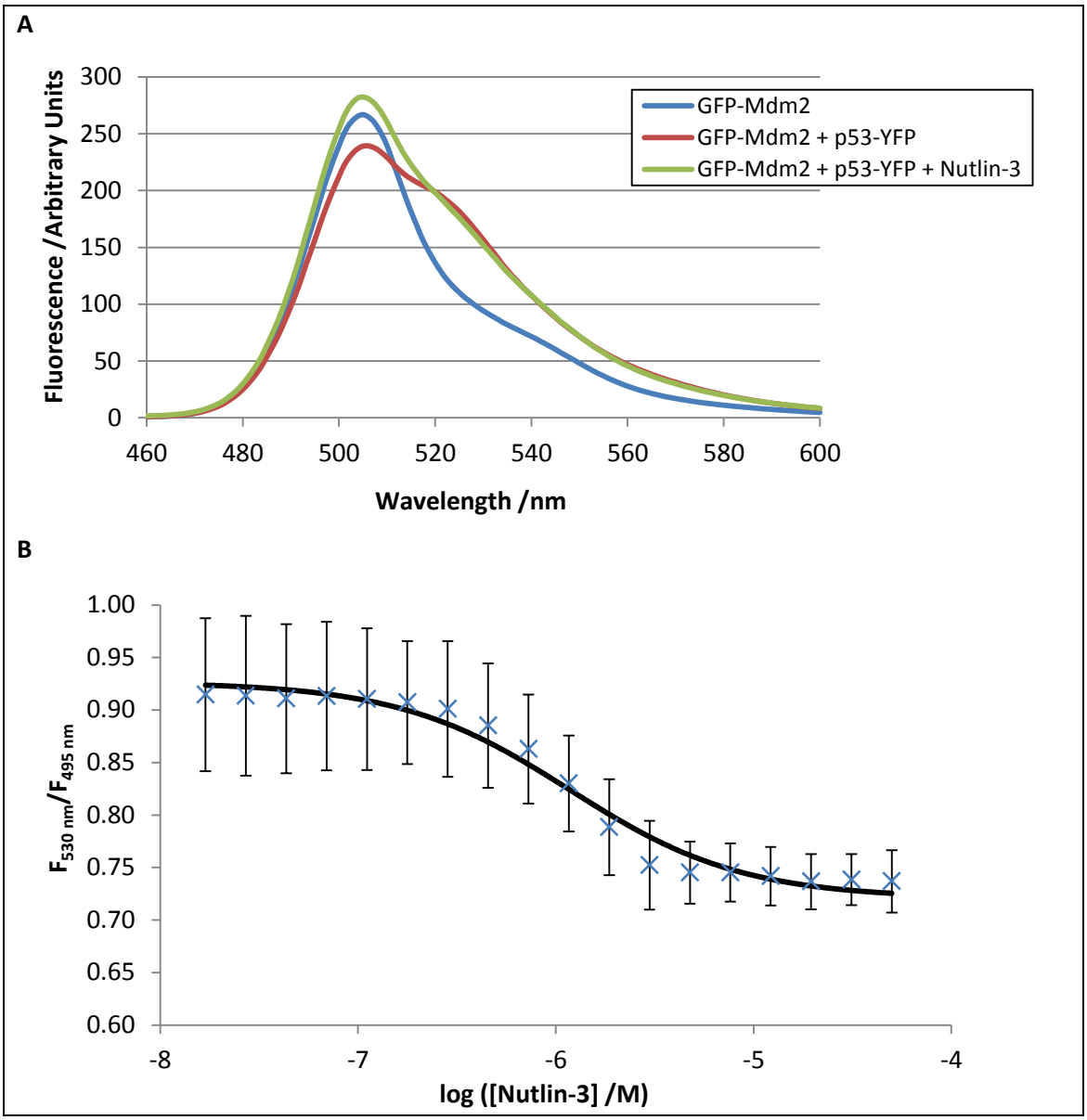


Figure 3.6: Titration of Nutlin-3 into a GFP-Mdm2 and p53-YFP solution. A) The spectra of three solutions: 2 µM GFP-Mdm2; 2 µM GFP-Mdm2 and 2 µM p53-YFP; 2 µM GFP-Mdm2, 2 µM p53-YFP and 50 µM Nutlin-3. B) The ratio of the fluorescence at 530 nm to the fluorescence at 495 nm of a solution containing 2 µM p53-YFP and 2 µM GFP-Mdm2 as Nutlin-3 was titrated in.

3.4.3. EXPERIMENTS WITH P53-EOSIN

It might be possible to develop an *in vivo* assay involving co-expression of the GFP-Mdm2 and p53-YFP hybrid proteins in a cell. However, because these proteins have a low affinity for each other, large amounts of each protein would need to be expressed to achieve a significant concentration of the complex. Furthermore, the considerable overlap of the GFP and YFP emission spectra would mean that FRET had little effect on the emission spectrum.

A possible strategy to deal with the overlap of FRET donor and acceptor emission spectra would be to record the fluorescence at multiple wavelengths and use mathematics to separate the fluorescence of donor and acceptor[281]. However, in the case of the oligobenzamides, the colour of the compounds themselves would make this approach challenging to implement.

3.4.3.1. Titration of p53-eosin into GFP-Mdm2 solution

Eosin was the third FRET acceptor to be tested. It was chosen for two reasons. Firstly, there is a significant distance between the peak in its emission spectrum (542 nm)[282] and that of the GFP spectrum (509nm), so isolation of the GFP emission from a mixture is easier. Secondly, being a small molecule as opposed to a large protein, eosin is less likely to hinder binding of the labelled peptide to GFP-Mdm2. George Preston (University of Leeds) supplied p53 with an eosin label on the C-terminal cysteine (Figure 3.7).

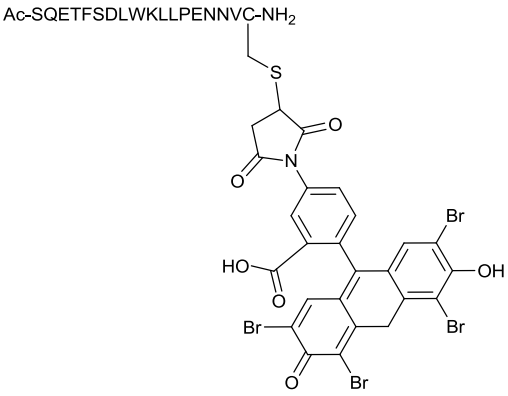


Figure 3.7: p53-eosin

The addition of p53-eosin to GFP-Mdm2 causes a reduction in the GFP fluorescence (Figure 3.9A). Then adding Nutlin-3 to a concentration of 50 μM reverses the drop in GFP fluorescence, to the extent that the intensity of the GFP emission peak is almost restored to its original level (Figure 3.9A). The remaining discrepancy could be explained by an inner filter effect: absorption of the GFP-emitted light by the eosin.

Titration of p53-eosin into GFP-Mdm2 (Figure 3.9B) yielded a dissociation constant of 0.68 ± 0.16 μM. This is close to the K_d of 0.89 μM measured by Czarna *et al.*[273] for the binding of $p53_{15-33}$ (residues 15 to 33 of human p53) to Mdm2 but is not consistent with the significantly lower K_d of 74 nM determined by Campbell *et al.*[98] for the binding of $p53_{15-31}$-fluorescein to Mdm2. It therefore suggests that steric hindrance by the GFP label might interfere with the binding of p53-eosin to the Mdm2.

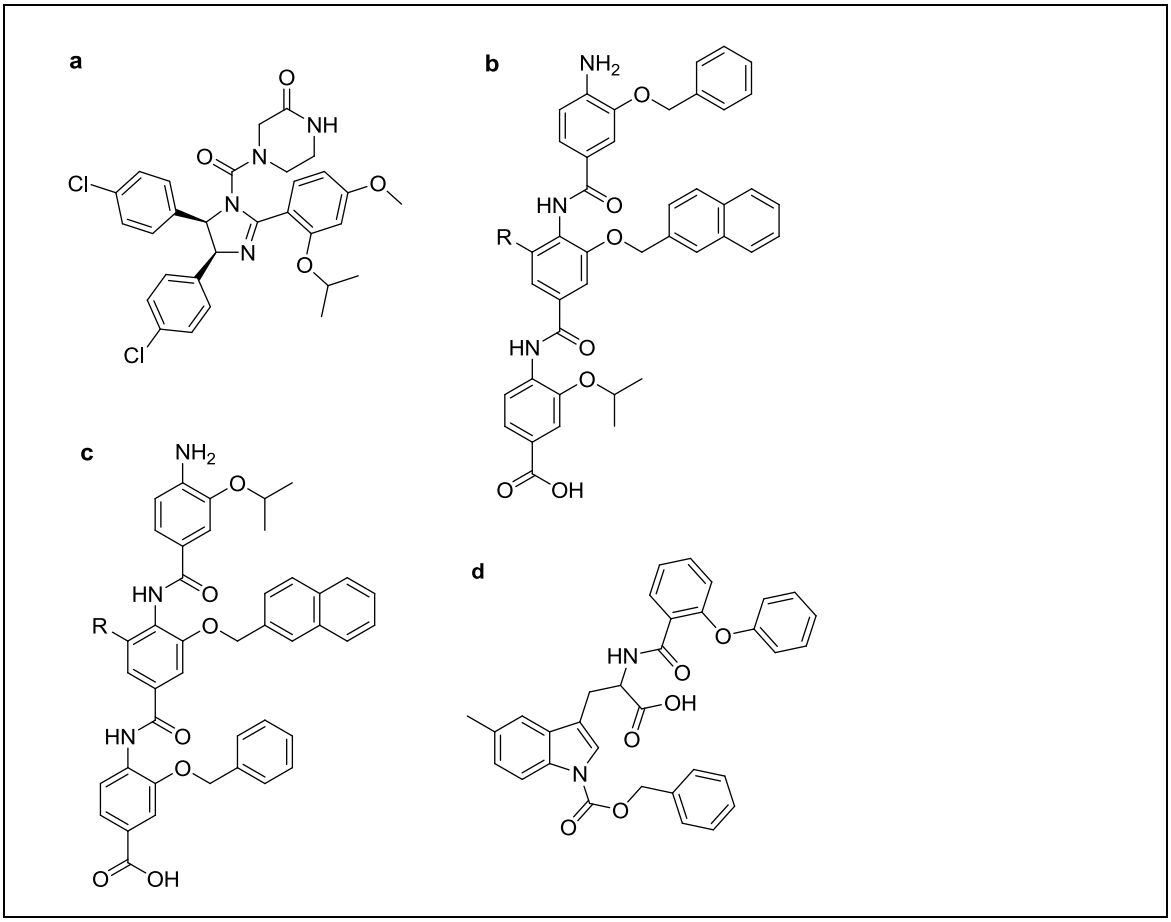


Figure 3.8: Inhibitors used to test the FRET assays. A) Nutlin-3[40]. The active isomer (Nutlin-3a) is shown. The racemic mixture was used in experiments. B) PPY-2-75. C) PPY-2-95. D) R5C3[275].

Adding oligobenzamide PPY-2-75 (Figure 3.8) to a GFP-Mdm2 and eosin-p53 solution reduces the eosin-p53 fluorescence and causes some recovery of the GFP-Mdm2 fluorescence (Figure 3.9C), consistent with the compound being an inhibitor of the p53-eosin-GFP-Mdm2 interaction and the FRET reducing upon its addition.

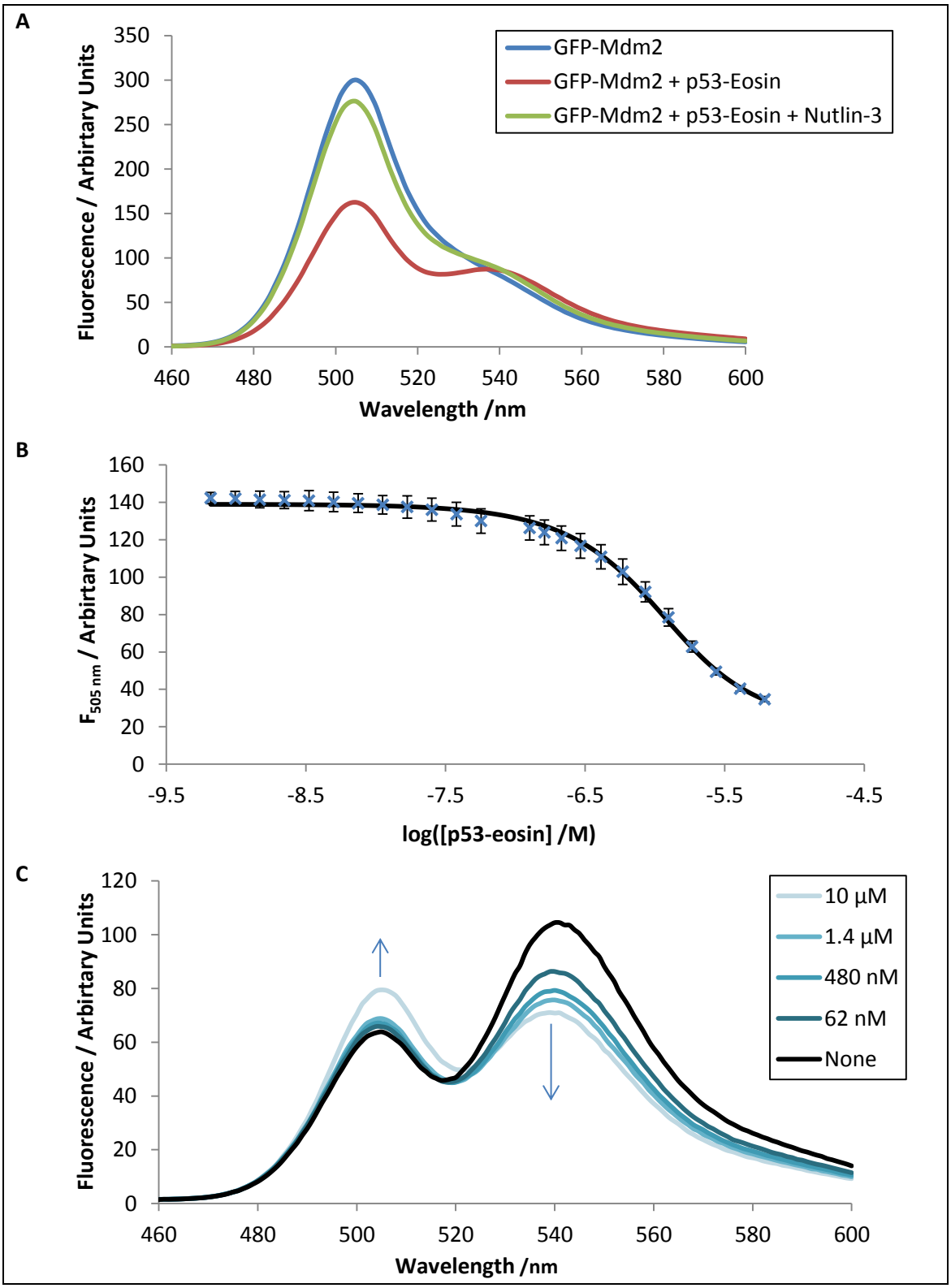


Figure 3.9: Graphs to illustrate the effect of adding GFP-Mdm2 to p53-eosin. 450 nm light was used to excite the GFP. A) Emission spectra for 2 µM GFP-Mdm2 alone, in the presence of 2 µM p53-eosin and with both 2 µM p53-eosin and 2 µM Nutlin-3. B) The effect of p53-eosin addition on the GFP fluorescence (at 505 nm). C) Emission spectra to show the effect of oligobenzamide addition on a solution of 1 µM GFP-Mdm2 and 6 µM p53-eosin. The final oligobenzamide concentrations are shown in the key.

3.4.3.2. Fluorescence anisotropy with p53-eosin

To provide further validation of the GFP-Mdm2 and p53-eosin system, single point assays were performed. Several known inhibitors were added to a GFP-Mdm2 and p53-eosin mix and both the FRET (as indicated by the GFP fluorescence) and the p53-eosin anisotropy were measured. The fluorescence anisotropy of the eosin can be determined in the presence of GFP-Mdm2 provided it is excited at a wavelength which is too high to excite the GFP-Mdm2. 535 nm was used in this case. The FRET and fluorescence anisotropy results showed the expected correlation (results not shown) supporting the hypothesis that the change in FRET is due to separation of the GFP-Mdm2 and p53-eosin.

Although the single point assay results suggest that the amount of FRET is a measure of the amount of GFP-Mdm2-p53-eosin complex, if the bulky GFP label hinders compound binding, this could affect the relative affinities of compounds and thus make the assay unsuitable for ranking compounds in terms of their affinity for unlabelled Mdm2.

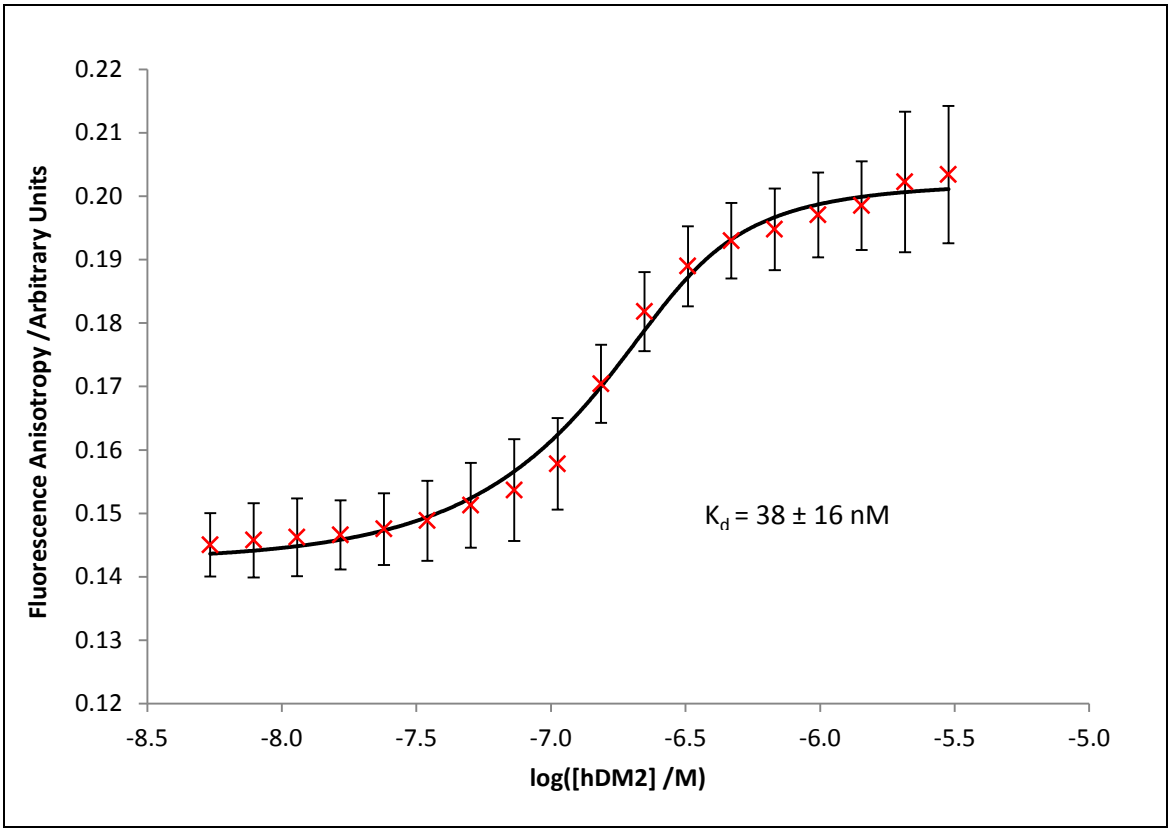


Figure 3.10: Titration of unlabelled Mdm2 into 250 µM p53-eosin

To test whether the high K_d obtained with GFP-Mdm2 and p53-eosin was due to the GFP label, unlabelled Mdm2 was titrated into 250 µM p53-eosin and the fluorescence anisotropy of the eosin was monitored (Figure 3.10). This yielded a much lower K_d of 38 ± 16 nM for the $p53_{15-31}$-eosin peptide binding to $Mdm2_{17-126}$ L33E, a value which is consistent with the K_d of 50 ± 10

nM observed by Lai *et al.*[283] using Oregon green-labelled $p53_{17-26}$ and close to the K_d of 74 ± 4 nM observed by Campbell *et al.*[98] using $p53_{15-31}$-fluorescein. This result suggested that although the GFP does not prevent binding of p53-eosin to GFP-Mdm2, it does weaken the interaction.

In the complex, the part of p53-eosin closest to the GFP should be the eosin label. Consequently, rather than restricting access to the p53 binding site, the GFP might prevent interactions which unlabelled Mdm2 can make with the eosin. If true, this would probably not affect the ranking of different oligobenzamides.

3.4.3.3. Titration of compounds into a p53-eosin and GFP-Mdm2 solution

To investigate whether the GFP-Mdm2 and p53-eosin FRET assay could be used to determine successfully the affinity of some known inhibitors of the interaction, they were each titrated into a mixture of p53-eosin and GFP-Mdm2.

The concentration at which an inhibitor causes 50% of the maximal possible inhibition of a biological system is known as the IC_{50} value. A lower IC_{50} value indicates stronger binding of the inhibitor to its target protein.

IC_{50} values were determined for Nutlin-3 and unlabelled p53 peptide by carrying out, for each, two titrations. Firstly, the inhibitor was added to a solution of GFP-Mdm2 and eosin-p53, in which FRET can occur (the main titration). Secondly, the inhibitor was titrated into a solution with an identical concentration of GFP-Mdm2 but no eosin-p53 and, consequently, no FRET. Excitation was at 450 nm and emission was measured at 495 nm so that no direct eosin fluorescence (eosin fluorescence not due to FRET) was detected; only GFP fluorescence was observed. The ratio of fluorescence with eosin-p53 to that without was calculated and then plotted against inhibitor concentration. The resulting curve should, in theory, plateau at a ratio of 1 because at high inhibitor concentrations there is no FRET in either solution and, therefore, the GFP fluorescence should be the same in both. In practice, in the cuvette titration, the results were not consistent with complete convergence of the main and control fluorescences at high inhibitor concentration (Figure 3.11A). This could be attributable to irreversible aggregation of some GFP-Mdm2 and p53-eosin over the course of the experiment which no amount of inhibitor could disrupt.

The IC_{50} for Nutlin-3 was 2.4 ± 1.2 μM when the compound was titrated into 1 μM GFP-Mdm2 and 2 μM p53-eosin in the spectrofluorimeter. This corresponds to a Ki of 470 ± 330 nM. The published K_i for Nutlin-3 is 36 nM for the active Nultin-3 enantiomer[266] so a K_i of twice this, 72 nM, would be expected for the racemic Nutlin-3 mixture tested. The observed affinity of Nutlin-3 for the p53 binding site of Mdm2 was evidently much lower than expected. Possible explanations include the presence of the GFP label or partial denaturation of the Mdm2 by the DMSO present.

The IC_{50} for unlabelled p53 peptide was found to be 13 ± 11 μM, which corresponds to a K_i of 3.1 ± 2.9 μM. In this case, the peptide's low binding affinity and high cost made it impossible to achieve the high inhibitor concentrations essential for the accurate determination of the IC_{50} value. Only the first part of the titration curve could be obtained, accounting for the large error in the IC50 value.

To decrease the amounts of unlabelled p53 peptide required to achieve high concentrations, the FRET assay was adapted for a microwell plate, with each titration corresponding to the use of a set of wells of increasing inhibitor concentration across the plate. Because the assay volume is smaller, higher concentrations can be achieved using less peptide.

Nutlin-3, the oligobenzamide PPY-2-75 and another inhibitor, R5C3 (molecule d in Figure 3.8)[275], were tested successfully using this plate assay (Figure 3.11B). However, preliminary results with unlabelled p53 peptide (not shown) suggested that the peptide was affecting the meniscus shape and that the size of the effect was different in the main and control titrations.

In contrast to the outcome of the cuvette assays, in the plate reader set-up, the results were consistent with convergence of the main and control GFP fluorescences at high inhibitor concentrations. The initial fluorescence ratio was also higher than observed in the spectrofluorimeter, as shown in Figure 3.11. Possible explanations for both of these observations are, firstly, that the shorter duration of the plate reader experiment allowed less time for denaturation and non-specific aggregation and, secondly, that the shorter light path through the solution reduced the eosin internal filter effect. Plates were read at 30°C, a significantly higher temperature than that used for cuvette measurements (10°C). This probably had an effect on the concentration of p53-Mdm2 before inhibitor was added so could also have contributed to the high fluorescence ratio at the start of the experiment.

The plate reader assay has notable advantages, namely the smaller amount of compound required to achieve high concentrations and the greater amenability of the plates to high-throughput testing.

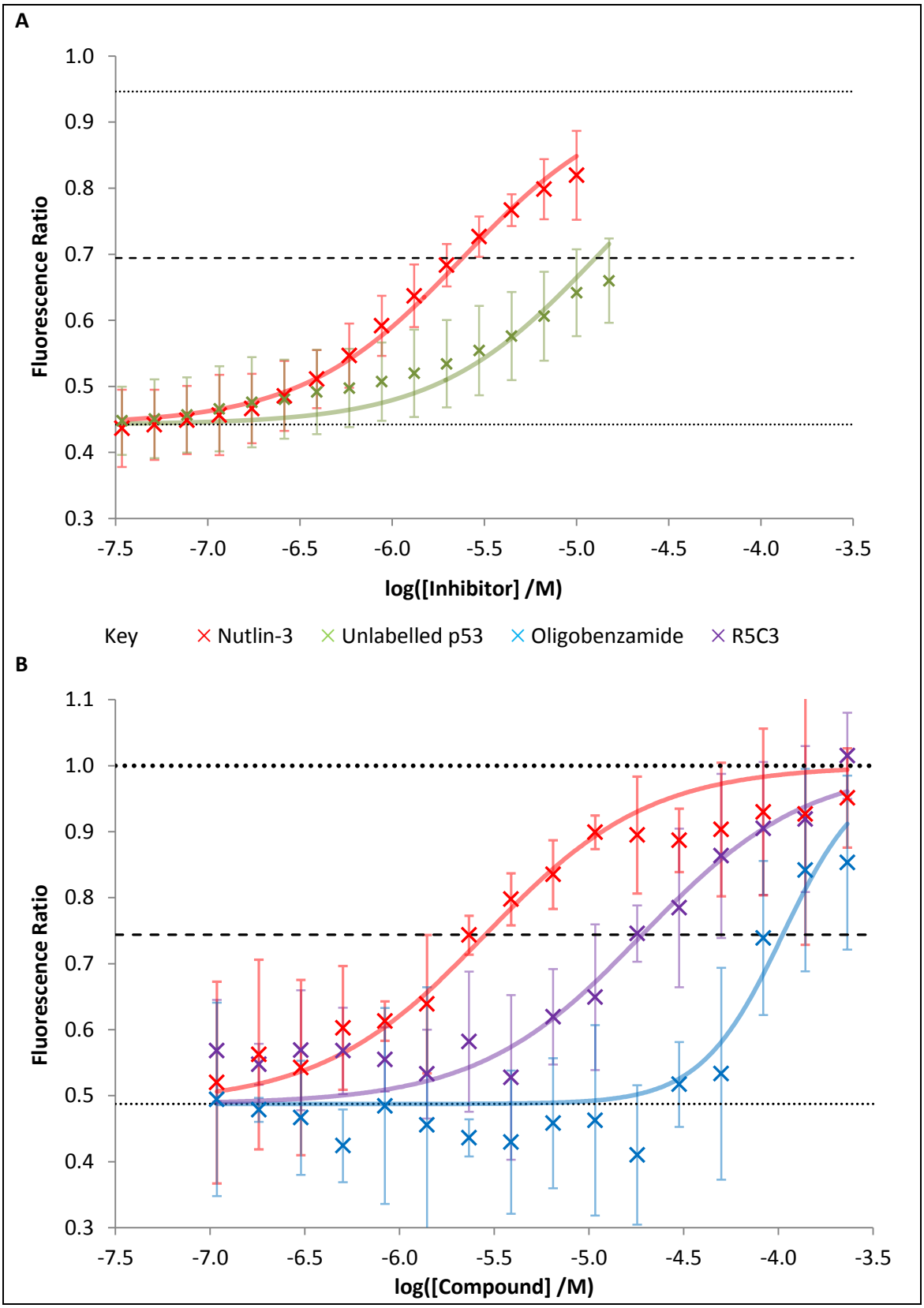


Figure 3.11: Graphs to show the effect on FRET ratios of titrating known and potential inhibitors into a solution of 1 µM GFP-Mdm2 and 2 µM p53-eosin. GFP fluorescence was measured at 492 nm. A) Spectrofluorimeter results. B) Plate reader results. Dotted lines show the maximum and minimum ratios obtainable in each experiment. Dashed lines show the fluorescence ratio at which there was 50% inhibition.

For Nutlin-3, the IC_{50} was found to be 3.9 μM ± 0.9 μM, which is similar to that obtained from the cuvette assay. For R5C3, an IC_{50} value of 25 ± 4 μM was observed, a higher value than obtained with Nutlin-3. This is consistent with the relative K_i values of these two compounds, 72 nM for Nutlin-3 and a higher value, 100 nM for R5C3[275]. The oligobenzamide PPY-2-75 appeared to bind to GFP-Mdm2 even more weakly than R5C3, which is consistent with previously published fluorescence anisotropy results which suggest that its K_i is micromolar.

3.4.4. PRODUCTION OF P53-MCHERRY

Although eosin is a better FRET acceptor than YFP because its emission peak is further away from the emission peak of GFP, it is harder to synthesise and purify p53-eosin than it is to express and purify p53-YFP. Furthermore, it would be difficult to introduce p53-eosin into a cell to carry out an *in vivo* assay. Consequently, an effort was made to produce another labelled peptide with a protein fluorophore, p53-mCherry. mCherry has an emission peak at 610 nm[284], much higher than that of GFP (509nm[282]). The excitation maximum of mCherry is at 587 nm[284].

Preliminary work was performed by Jill Augustine (University of Leeds). She produced a construct coding for p53-mCherry based on the pET28bTEV vector, where the encoded p53 peptide was of a mutated version which had previously been found to have a much higher affinity than the wild-type[279]. In the event, the mutant p53-mCherry DNA supplied was flanked by not one, but two histidine-tag-encoding regions, one at either end. Site-directed mutagenesis primers designed by me for this project (see methods section) allowed Jill Augustine to supply two further constructs: wild-type-p53-mCherry with two histidine tags and mutant-p53-mCherry with one histidine tag.

To produce a wild-type-p53-mCherry construct with one histidine tag, both of the new constructs were cut with PstI, which cuts in the mCherry open reading frame, and *ClaI*, which cuts in the middle of the *kan* antibiotic resistance gene (Figure 3.12). The p53-mCherry front end of the wild-type construct was ligated to the no-histidine-tag C-terminal end from the mutant vector and DH5α cells were transformed with the reaction mixture. Those transformed with correctly re-ligated vector were selected using media containing the antibiotic kanamycin.

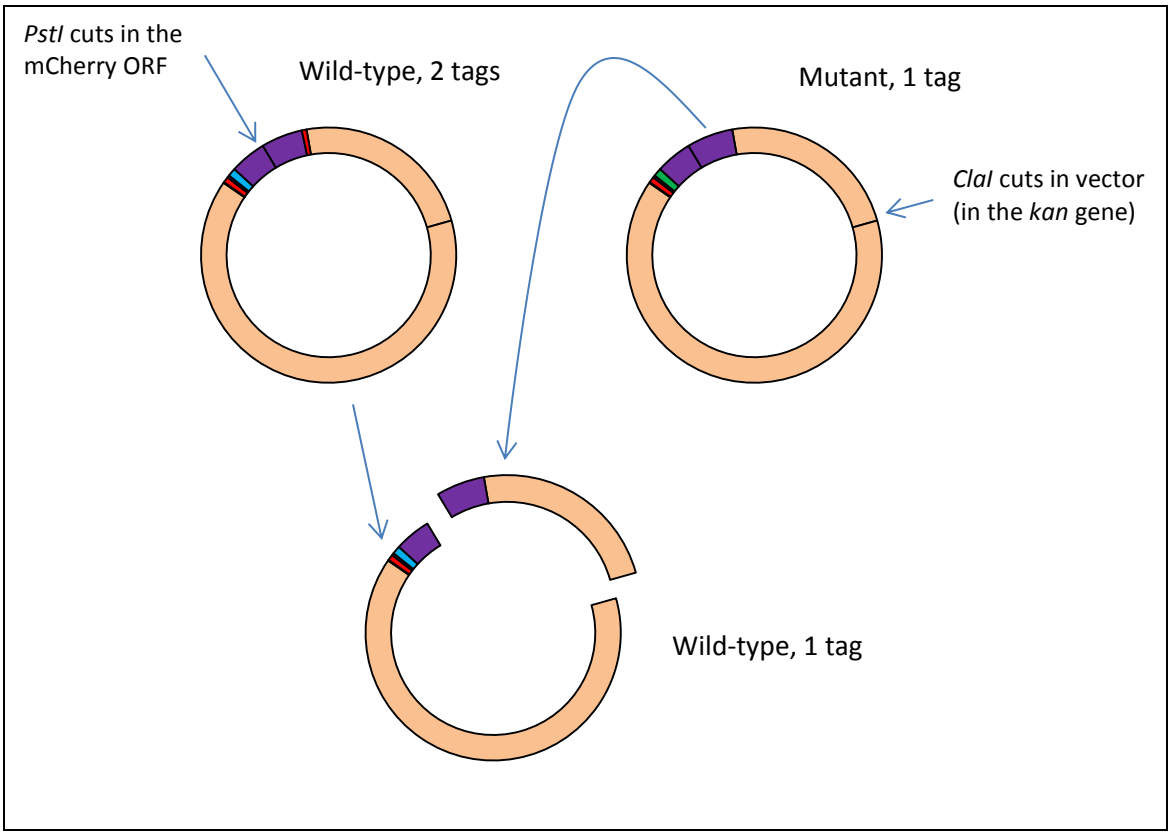


Figure 3.12: Diagram to illustrate the production of a wild-type-p53-mCherry construct with one oligo-histidine tag from the double His-tagged wild-type construct and one from the single His-tagged mutant construct. Note that in the final construct, the C-terminal-tag is not actually absent but is preceded by a STOP codon.

Figure 3.13 below shows agarose gels made during production of the wild-type construct with only one His-tag. Part A shows the gel used for purification following cutting of the two starting constructs. Part B shows the colony PCR results after re-ligation. All of the colonies tested had the correct construct, as predicted given that the PstI cut site is in the selectable marker.

Both wild-type p53-mCherry and mutant p53-mCherry (Figure 3.14) were expressed, purified and then tested in the spectrofluorimeter. The results of purification by metal affinity chromatography are shown in Figure 3.15. The red mCherry label made it easy to identify which elution fractions contained labelled protein.

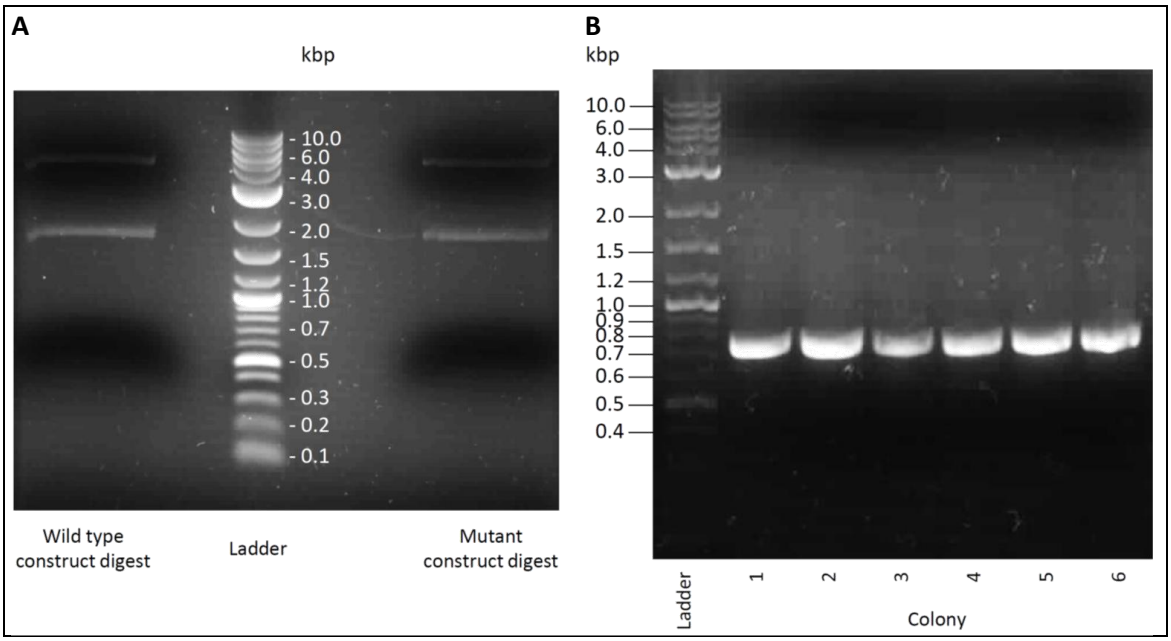


Figure 3.13: Agarose gels from the production of the single-His-tag wild-type p53-mCherry-encoding construct. A) The results of cleavage of the wild-type double His-tagged construct (left) and the mutant construct with a STOP codon at the end (right). The apparent fragment sizes are consistent with the predicted sizes of 4.28 kb and at 1.77 kb. The bands were cut out after taking this image B) Results for colony PCR of the p53-cherry open reading frame in the course of checking for correctly ligated constructs.

A

MGHHHHHHHHHHSSGHENLYFQGSQLTFEHYWAQLTSSKGEEDNMAIIKEFMRFKVHMEGSVN
GHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLS
FPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSER
MYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVE
QYERAEGRHSTGGMDELYK

B

MGHHHHHHHHHHSSGHENLYFQGSQETFSDLWKLLPENSKGEEDNMAIIKEFMRFKVHMEGSV
NGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKL
SFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSE
RMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIV
EQYERAEGRHSTGGMDELYK

Blue – Oligo-histidine tags
Green – p53 or p53-like sequence
Yellow - TEV protease cut site
Red - mCherry

Figure 3.14: The final sequences of the expressed mutant (A) and wild-type (B) p53-mCherry proteins

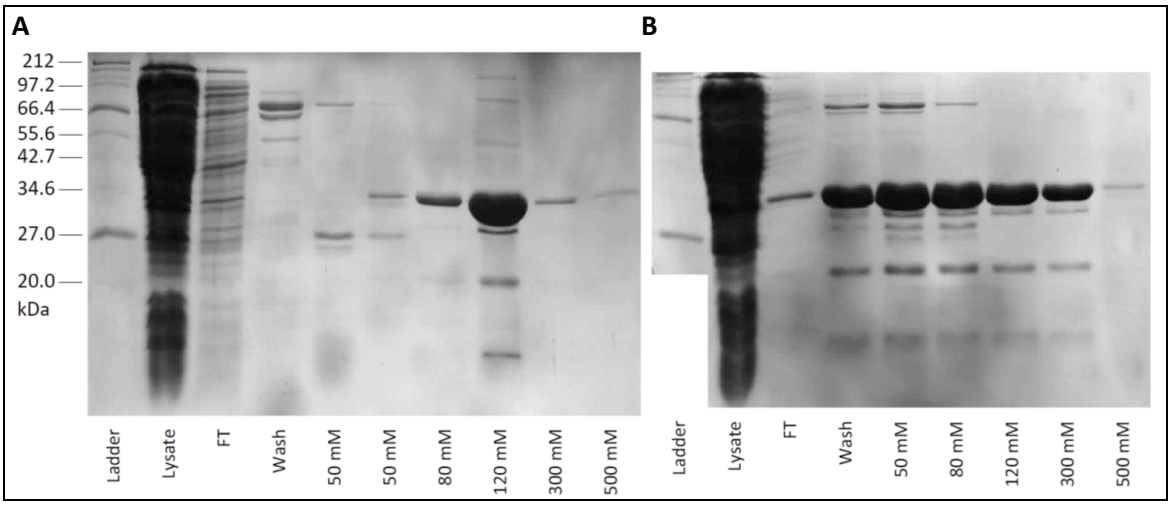


Figure 3.15: Metal ion affinity chromatography for separation of His-tagged p53-mCherry from the remainder of the cell lysate. A) Wild type. B) Mutant.

The 300 mM imidazole samples were further purified by gel-filtration (Figure 3.16). Two peaks were apparent in the results for each protein, the first implying the presence of aggregated protein and the second, the presence of a smaller species than expected. Polyacrylamide gel electrophoresis (Figure 3.17) revealed that the protein producing the second peak was, in the case of both the mutant and wild-type titrations, the intact, 31 kDa monomer, consistent with the red colour of the solutions.

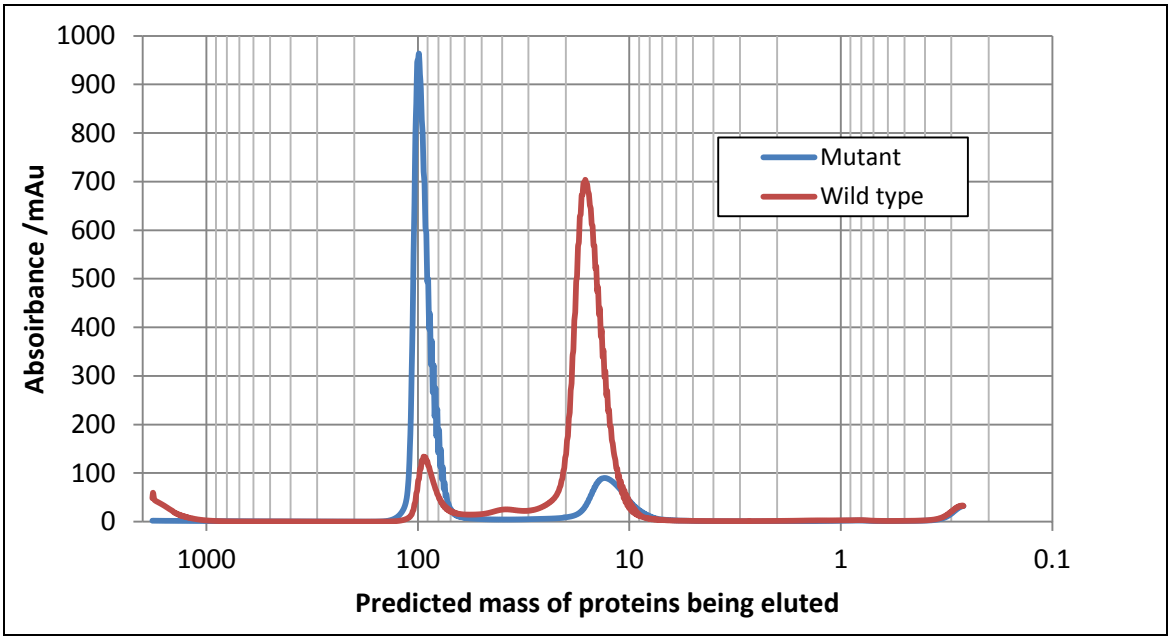


Figure 3.16: Trace showing the UV absorbance during gel filtration of His-tagged p53-mCherry protein following metal ion affinity chromatography. The mutant protein sample (blue) was run first. Purification of the wild-type (red) was carried out immediately after all of the wild-type protein had cleared the column. There is a linear relationship between the logarithm of the molecular weight and the elution volume so this graph resembles a trace of absorbance against elution volume. The absorbance is in milli-absorbance units.

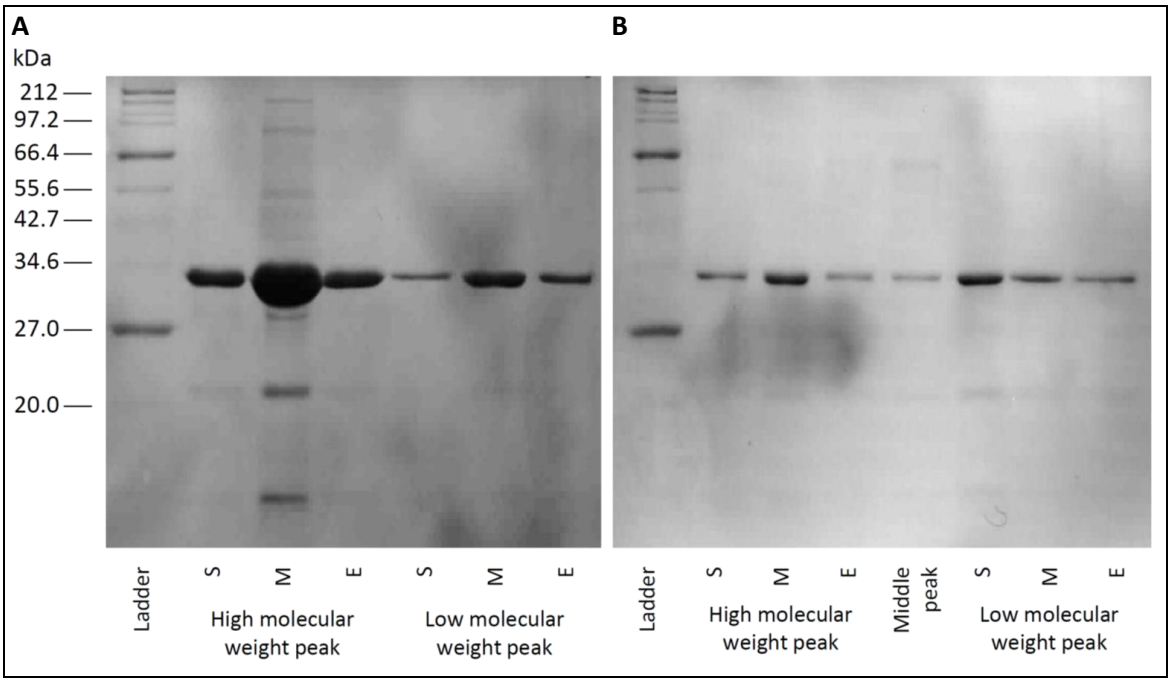


Figure 3.17: Polyacrylamide gels with fractions from purification of both the wild-type (A) and mutant (B) p53-mCherry proteins revealing that all samples contain an intact, 31 kDa p53-mCherry protein. Samples were taken at the start (S), middle (M) and end (E) of each of the peaks shown in Figure 3.16 above.

3.4.5. FRET EXPERIMENTS WITH P53-MCHERRY

3.4.5.1. Titration of p53-mCherry into a GFP-Mdm2 solution

The mutant peptide was found to bind more strongly to GFP-Mdm2 than the wild-type. Titration of p53-mCherry protein into GFP-Mdm2 and measurement of the GFP emission at 505 nm yielded titration curves for both the wild-type and mutant (Figure 3.18), from which the dissociation constants for binding were found to be 3.7 ± 0.2 µM and 0.21 ± 0.11 µM respectively.

Both the wild-type and mutant constructs have potential advantages. The wild-type, having the correct p53 sequence, should theoretically provide the best indication of which compounds compete most effectively with p53 for Mdm2. The higher affinity mutant, because it binds more strongly, can be used at a lower concentration, reducing the amount of protein needed for assays.

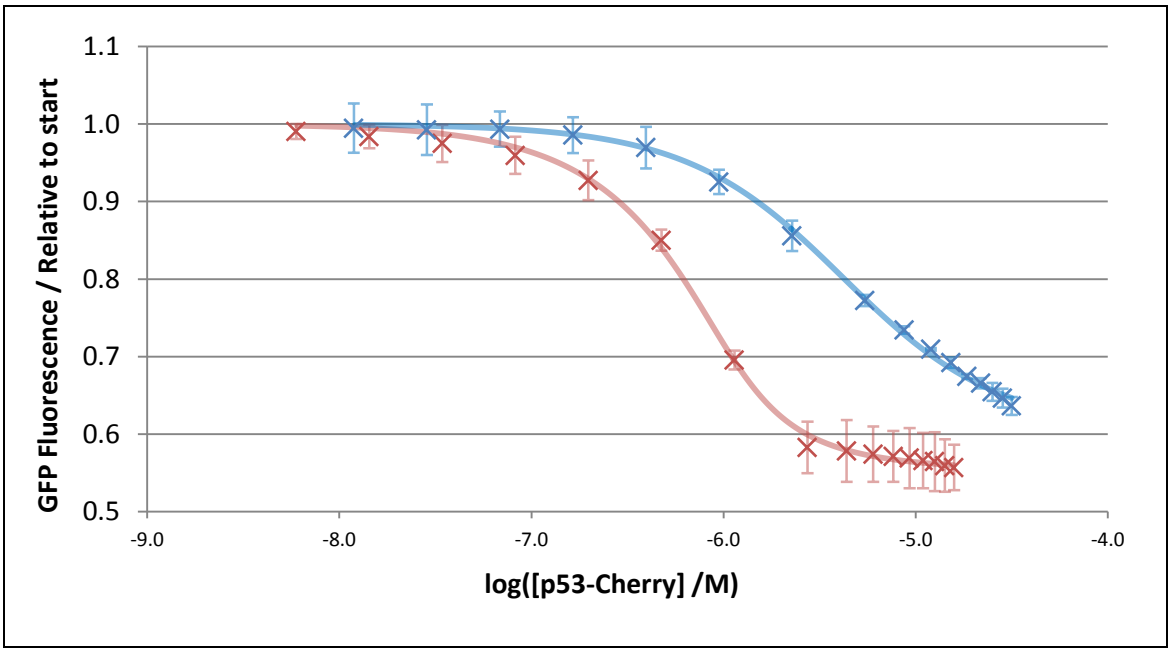


Figure 3.18: Titration of p53-mCherry into GFP-Mdm2. Because the curves were unlikely to plateau at the very highest concentrations of p53-mCherry, many high concentrations were tested to ensure that the dissociation constant could still be determined accurately. Blue) Wild-type p53-mCherry. Red) Mutant p53-mCherry.

3.4.5.2. Titration of potential inhibitors into a p53-mCherry and GFP-Mdm2 solution

After the K_d of the p53-mCherry GFP-Mdm2 interaction had been found, Nutlin-3, R5C3 and unlabelled p53 peptide were each added to solutions containing GFP-Mdm2 and p53-Cherry to see if the system could be used to distinguish their relative binding affinities for Mdm2. Figure 3.19 shows the effects of Nutlin-3 and R5C3 on the emission spectrum. As expected, adding the known inhibitor R5C3 abolished the FRET indicating that R5C3 displaced the p53-Cherry from the p53 binding site of GFP-Mdm2 (Figure 3.19B). However, adding a large amount of Nutlin-3 only partially reduced the FRET (Figure 3.19A). The original spectrum of GFP-Mdm2 in isolation might not have been restored due to Nutlin-3 precipitation causing scattering, reducing the intensity of the fluorescence measured.

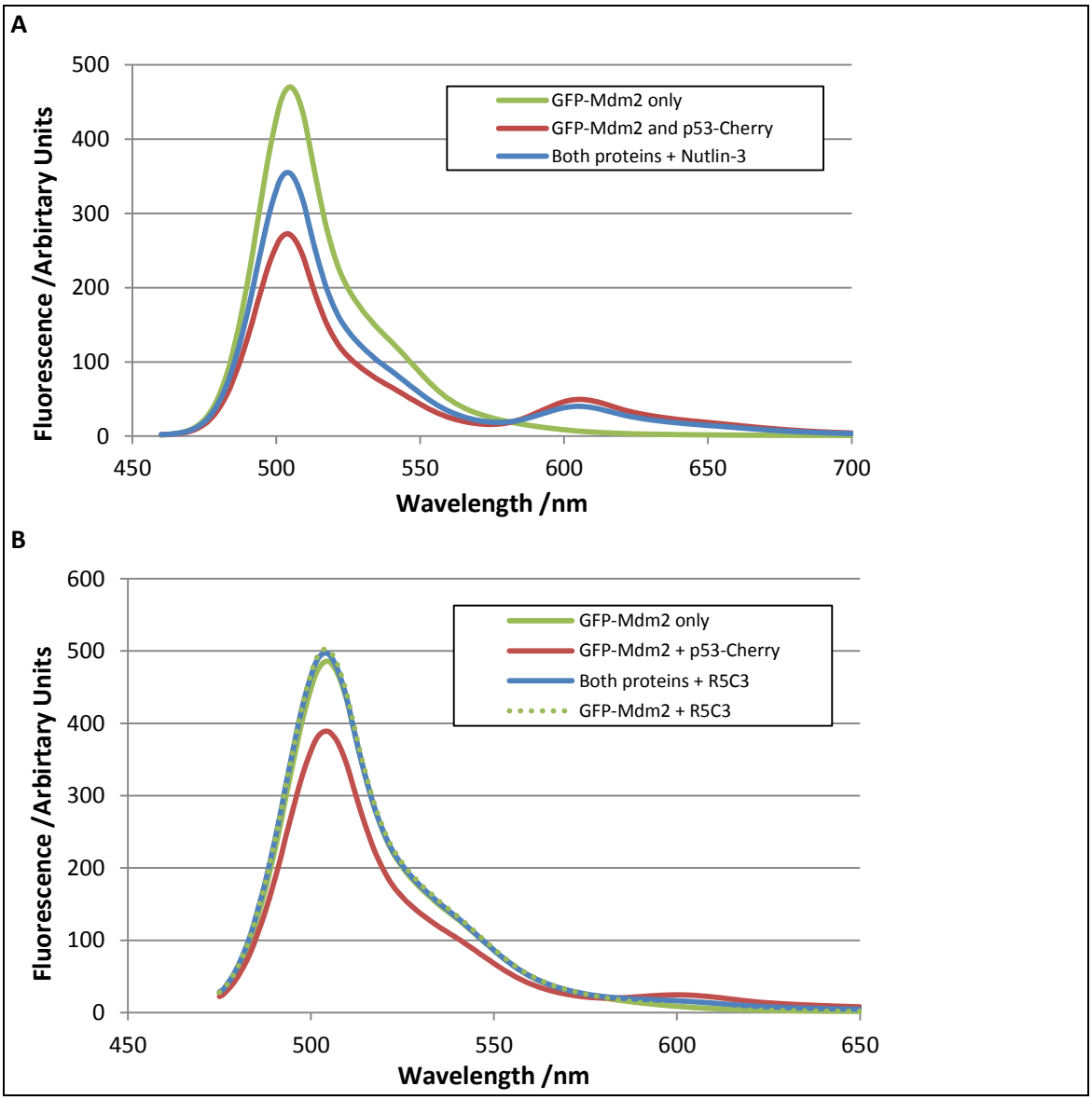


Figure 3.19: Emission spectra showing the effects of introducing Nutlin-3 and R5C3 into the GFP-Mdm2 mutant p53-mCherry system. A) 13.5 µM Nutlin-3 added to 2 µM GFP-Mdm2 and 5 µM p53-mCherry. B) 50 µM of R5C3 added to 1 GFP-Mdm2 µM and 1 µM p53-mCherry. In both A and B, the initial spectrum of GFP-Mdm2 is shown in green. The addition of p53-mCherry produced the red spectrum and addition of the inhibitor generated the blue spectrum. The dotted green line in B shows the spectrum of GFP-Mdm2 in the presence of R5C3 but no p53-mCherry, which indicates that the difference between the solid green and blue lines reflects absorption by the mCherry.

Figure 3.20 shows the results of titrations carried out into GFP-Mdm2 and wild-type p53-mCherry. Using 1 µM GFP-Mdm2 and 5 µM p53-mCherry, IC_{50} values of 1.4 ± 0.2 µM, 2.5 ± 0.3 µM and 17 ± 5 µM were obtained for Nutlin-3, R5C3 and unlabelled p53 peptide respectively. These results correctly reflect the relative binding affinities of these inhibitors, which have K_i values of 72 nM[266], 100 nM[275] and 890 nM[273] respectively.

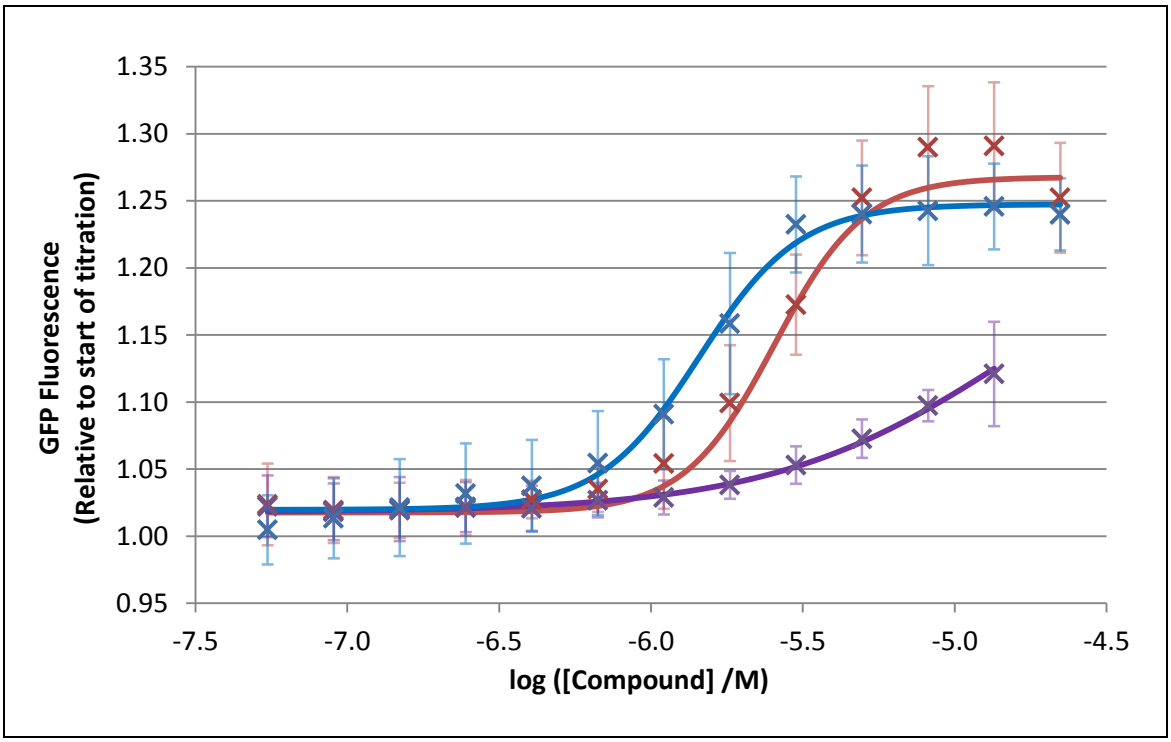


Figure 3.20: GFP fluorescence during the titration of inhibitors in to a solution of 1 µM GFP-Mdm2 and 5 µM p53-mCherry (wild-type). Results for Nutlin-3 are shown in blue, results for R5C3 are in red and results for the unlabelled p53 peptide are in purple.

As described above, titration of the oligobenzamide PPY-2-75 into a mixture of GFP-Mdm2 and p53-eosin caused an increase in GFP fluorescence. However, the presence of the oligobenzamide PPY-2-75 (at a concentration of 50 µM) in a mixture of 2 µM GFP-Mdm2 and 5 µM p53-mCherry (Figure 3.21, p234) causes a drop in the GFP fluorescence of the mixture, rather than the expected increase. A control run was performed with no p53-mCherry and, in this test, the oligobenzamide did not reduce the GFP fluorescence, suggesting that the drop observed in the presence of p53-mCherry reflects an increase in the amount of FRET, an observation for which there is no obvious explanation. The increase is seen using both mutant and wild-type p53-mCherry and comparable results were obtained using the similar compound PPY-2-95 in place of PPY-2-75. (See Figure 3.8 for the structure of PPY-2-95.) These findings warrant further research.

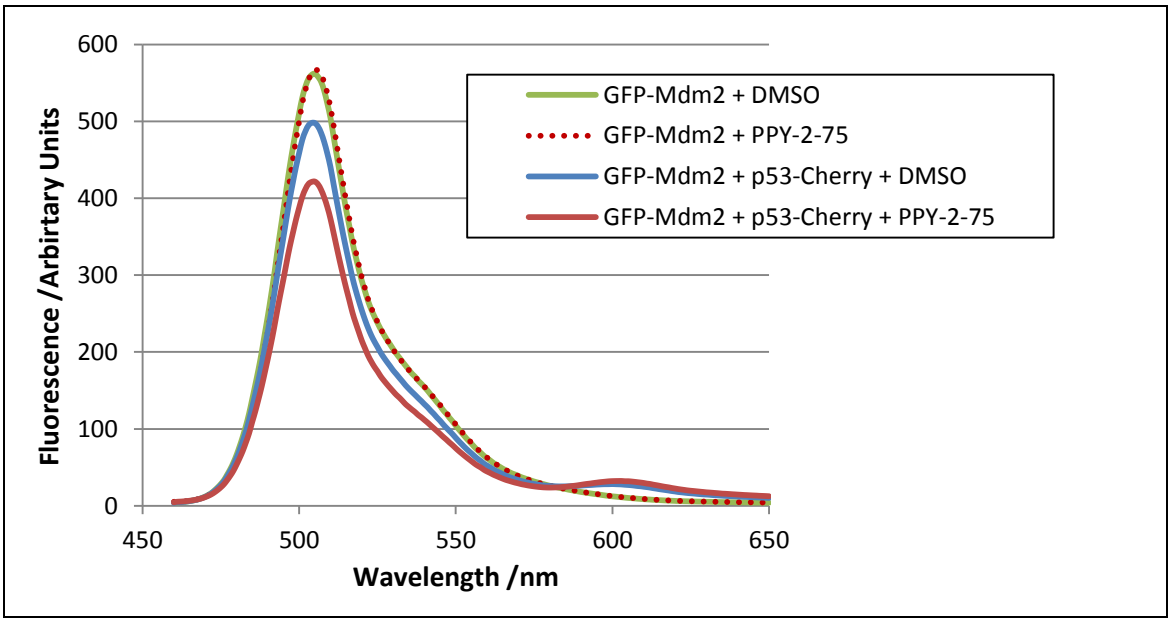


Figure 3.21: The effect of oligobenzamide PPY-2-75 on the degree of FRET between GFP-Mdm2 and p53-eosin. The emission spectrum of a 2 μM GFP-Mdm2, 5 μM wild-type p53-mCherry and 60 μM PPY-2-75 (an oligobenzamide) solution is shown in red. The other lines are controls which lack either the p53-mCherry component (replaced with buffer) or the PPY-2-75 (replaced with DMSO). PPY-2-75 causes a decrease in GFP-Mdm2 fluorescence, but only in the presence of p53-mCherry.

3.4.6. INVESTIGATION OF WHETHER OLIGOBENZAMIDES BIND TO P53-EOSIN AS WELL AS TO GFP-MDM2

The results involving PPY-2-75 so far described present an apparent inconsistency. This oligobenzamide causes a decrease in FRET when added to GFP-Mdm2 and p53-eosin but seems to cause an increase in FRET when added to GFP-Mdm2 and p53-mCherry. In the former case, the assumption has been that the PPY-2-75 causes the dissociation of GFP-Mdm2 and p53-eosin by binding to GFP-Mdm2 but similar results would be obtained if the oligobenzamide binds to the p53-eosin instead. This could explain why no reduction in FRET is apparent in the GFP-Mdm2-p53-mCherry system and also implies that PPY-2-75 is unlikely to be an effective inhibitor of the p53-Mdm2 interaction.

To test the hypothesis that the oligobenzamide interacts with p53-eosin, PPY-2-75 was titrated into a solution of p53-eosin alone (with no GFP-Mdm2) across a microwell plate (Figure 3.22). A 40% drop in eosin fluorescence was observed, which suggests that oligobenzamides in some way lead to quenching of the fluorescence. Given that saturation is observed at high oligobenzamide concentrations, interaction of the oligobenzamide with the p53-eosin molecule is almost certainly the cause of the drop in fluorescence. This would be consistent with previously published results which suggest that oligobenzamides bind to p53 labelled with fluorescein, a fluorophore similar to eosin[42].

Any interaction between oligobenzamides and p53-eosin should not directly influence FRET measurements in the GFP-Mdm2 and p53-eosin assay because the amount of FRET is

monitored using the GFP-Mdm2 fluorescence alone. However, if the interaction significantly reduces the amount of oligobenzamide available to bind to GFP-Mdm2 or contributes to the dissociation of GFP-Mdm2 and p53-eosin, then the calculated binding affinities would be affected.

Establishing the K_d of the oligobenzamide-p53-eosin complex would provide a useful measure of the potential of this interaction to affect the validity of assays. In theory, it should be possible to derive K_d from results of the type shown in Figure 3.22, using Equation 3.1 (p205). Unfortunately, in this case, the apparent high affinity of oligobenzamide for p53-eosin coupled with the high concentration of p53-eosin required to produce detectable fluorescence in the plate reader combined to make this impossible in practice.

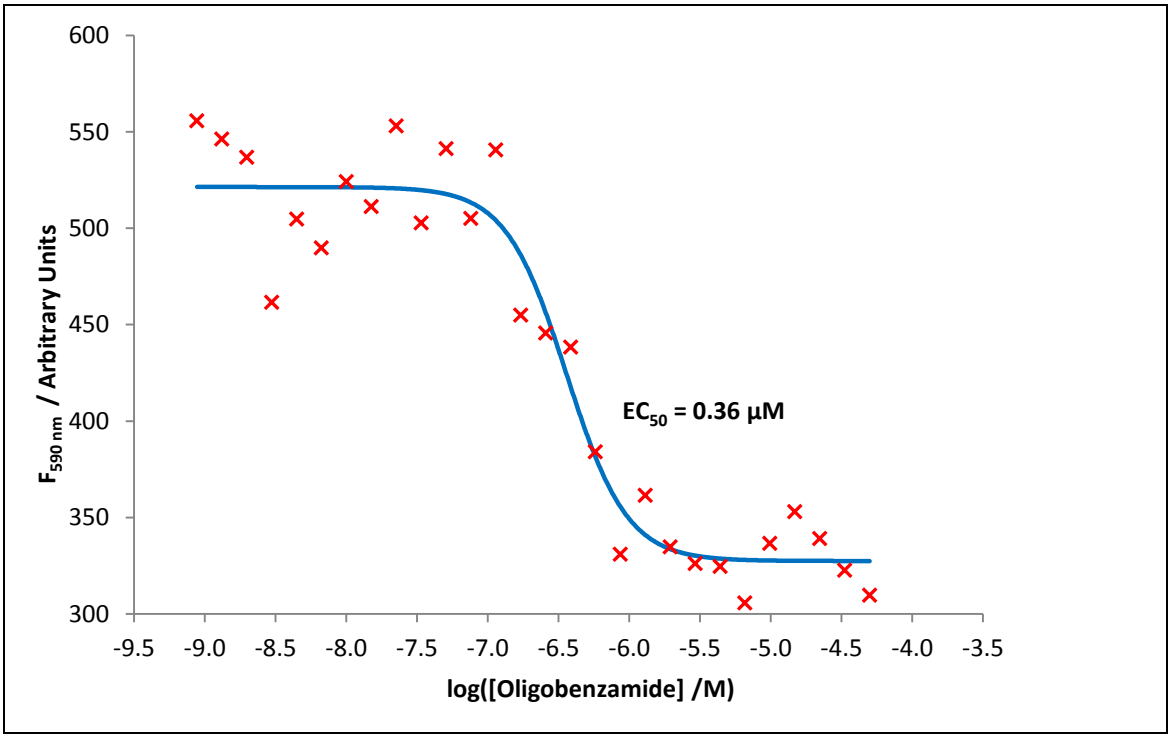


Figure 3.22: The effect of PPY-2-75 concentration on the apparent fluorescence of microwell plate well solutions containing 1.6 µM p53-eosin. Excitation was at 492 nm and emission was measured at 590 nm.

Figure 3.11B (p225) shows the results of titrating inhibitors in to a mixture of p53-eosin and GFP-Mdm2. Whereas curves based on the Hill equation with the Hill coefficient fixed at 1 could be fitted to the results obtained with both Nutlin-3 and R5C3, the oligobenzamide results were not consistent with a Hill coefficient of 1. A curve with the Hill coefficient fixed at 2 (shown in the figure) fitted the oligobenzamide results much better. This suggests that two molecules of oligobenzamide might be involved in the dissociation of GFP-Mdm2 and p53-eosin, one to bind to the GFP-Mdm2 and one to bind to the p53-eosin.

3.5. Conclusions and Further Work

The aim of this practical section of the project was to produce a FRET assay which would be independent of the fluorescence anisotropy assay for the purpose of measuring the binding affinity of p53-Mdm2 inhibitors. Three different FRET acceptors were tested: p53-YFP, p53-eosin and p53-mCherry. The first of these was found to be of limited use due to low FRET efficiency and the significant overlap of the GFP and YFP emission spectra, which meant that FRET caused only a small change in the fluorescence spectrum.

Changing the acceptor fluorophore from YFP to eosin reduced the emission spectrum overlap. The GFP-Mdm2 and p53-eosin assay was tested using both a temporal titration in a spectrofluorimeter cuvette and a spatial titration across the wells of a microwell plate. The demonstrated success of the second method is crucial because it makes the assay amenable to high-throughput screening by a plate-handling robot and plate reader.

Unfortunately, p53-eosin, like any fluorescently labelled peptide, is difficult to purify from unlabelled peptide following synthesis and cannot be expressed in a cell making it unsuitable for *in vivo* assays. Consequently, another attempt was made to find an efficient FRET pair involving a protein FRET acceptor. Two constructs of p53-mCherry were made, one based on wild-type p53 peptide and one based on a mutated p53 peptide. The wild-type protein produced good results with the known p53-Mdm2 inhibitors Nutlin-3 and R5C3. The mutant also produced good preliminary results but was not extensively tested, partly because very little of the protein was available.

While the oligobenzamide PPY-2-75 inhibited FRET in the p53-eosin system, in contrast, it appeared to promote FRET when p53-mCherry was used as the acceptor. This was surprising given that the protein to which the oligobenzamide is thought to bind, GFP-Mdm2, is the same in both the eosin and mCherry assays. The possibility that dissociation of the GFP-Mdm2-eosin-p53 complex could be caused primarily by the binding of the inhibitor to the labelled peptide and not the GFP-Mdm2 was considered. Oligobenzamides had previously been found to bind to p53-peptide[42] and a drop in p53-eosin fluorescence was observed when compound was added to p53-eosin in the absence of GFP-Mdm2. Such a situation would make it impossible to predict the relative abilities of different oligobenzamides to inhibit of the p53-Mdm2 interaction using molecular dynamics simulations of Mdm2 alone. A method was therefore required to confirm that oligobenzamides bind to Mdm2, a method not requiring the use of competing p53 peptide. Nuclear magnetic resonance (NMR) can be used to test for small molecule binding and can potentially reveal where binding is likely to occur. This is the subject of the next chapter.

Many inhibitors of protein-protein interactions have a low solubility. Consequently, assays which can be used to determine the affinity of potential inhibitors without the need to test them at high concentrations would be useful. The p53-eosin assay developed in this project for testing oligobenzamides is such an assay. The GFP fluorescence of each solution of GFP-Mdm2, eosin-p53 and potential inhibitor is determined relative to that of a control lacking the p53-eosin. To ensure that the unbound FRET acceptor, which is present in the mixture but not the control, does not affect the fluorescence, the donor is excited at a wavelength at which there is no acceptor absorbance and its fluorescence is read at a wavelength at which there is no acceptor emission. If the small molecule causes dissociation, the fluorescence of the mixture and the control converge at high inhibitor concentrations, so the ratio of the fluorescences approaches 1. This known titration endpoint allows a curve to be fitted to titration results without the requirement to test very high compound concentrations. In future, this new method could be used to test other low solubility or high absorbance inhibitors which could then be developed into more soluble compounds.

In the method above, the wavelengths used to excite the GFP and detect its fluorescence were chosen to avoid detection of eosin, not to produce the strongest signal. A quencher molecule could be used instead of an acceptor fluorophore, eliminating all background fluorescence from the acceptor[285]. This would make it easier to select the best measurement wavelengths, particularly when restricted to the small set of filters available in a plate reader.

If colourless oligobenzamides are produced in the future then FRET could be monitored by observing the ratio of acceptor to donor fluorescence. This increases the signal to noise ratio in a FRET assay because variations in the absolute fluorescence during the experiment are taken into account.

When measuring the ratio of donor to acceptor fluorescence, emission due to direct excitation of the acceptor can reduce the effect of FRET on the acceptor fluorescence. BRET (bioluminescence resonance energy transfer) is a solution to this problem. The luminescent donor molecule (typically *Renilla reniformis* luciferase) does not need to be excited with light so there can be no direct excitation of the acceptor[87,286].

4. The Assignment of Mdm2 L33E NMR Peaks and Their Use to Characterise the Binding of Peptidomimetics to Mdm2

4.1. Summary

This chapter describes the production of ^{15}N-labelled Mdm2 L33E for NMR, the assignment of its ^{15}N-Heteronuclear Spin Quantum Correlation (HSQC) spectrum peaks and the movement of these peaks upon the addition of oligoamides.

Addition of PPY-2-75 to Mdm2 L33E or the same molecule but with a polyethylene glycol (PEG) chain added to increase solubility (AB062) was found to have little if any effect on the HSQC peak positions, suggesting that there was no binding. The latter compound is soluble, so the lack of a 1D ^{1}H-NMR spectrum for this compound, if not also PPY-2-75, reflects the formation of a large slowly-tumbling, soluble aggregate with a very broad spectrum rather than precipitation of the compound to the extent that there is none left in the solution to detect.

A third compound, similar in structure but lacking the middle benzene ring (VA3-192) gave a visible proton-NMR spectrum and did, in contrast to PPY-2-75, cause peaks to shift on the HSQC spectrum. These corresponded to residues close to the p53 binding site and the binding position of the compound predicted by docking, a position in which the compound is acting as a peptidomimetic. Taken together, these findings suggest that the difficulty obtaining a proton spectrum with the other compounds might reflect a solubility or aggregation problem unique to some oligobenzamides rather than an inability of oligobenzamides to bind to Mdm2 in general. They suggest that if this problem can be resolved, oligobenzamides could bind in the p53 binding site.

The assignments for the L33E mutant will be useful for future work and the findings of this research, most notably the instability of the protein in DMSO concentrations above 5%, the formation of a large aggregate by some soluble oligobenzamides and the evidence of VA3-192

binding in the p53 binding site, provide insight into the type of oligoamides likely to be most useful for future inhibitor development.

Key research outcomes described in this chapter:

- Assignment of the HSQC peaks for Mdm2 L33E, a protein fixed in the open conformation potentially facilitating more accurate determination of where ligands bind in future.
- Confirmation that VA3-192 binds in the p53 binding site of Mdm2, contacting residues which are consistent with the compound acting as a peptidomimetic.
- Discovery that Mdm2 L33E starts to lose its native structure at DMSO concentrations above 6% and identification of an aggregation problem involving oligobenzamides which means that they are not free in solution.

4.2. Introduction

The docking results described in chapter 2 suggest that oligobenzamide compounds are able to bind to Mdm2 in many ways other than in the paradigmatic binding pose proposed at the start of this thesis, where the compound lies extended in the p53-binding site with each side chain mimicking that of a p53 residue.

The fluorescence polarisation results of the previous chapter suggest that oligobenzamides bind in the p53 binding groove but these results are compromised by evidence of some oligobenzamides binding to the p53-fluorescein probe. The chapter describes how FRET-based assays were developed in an attempt to produce an orthogonal method, which could yield supporting results. While in FRET assays, the O-linked compound PPY-2-75 appeared to displace p53-eosin from the p53 binding groove, there was also evidence of the compound binding to the p53-eosin probe. Also, preliminary results with p53-Cherry were less clearly in favour of the compound binding to Mdm2. Given the potential impact of binding position on affinity and on the ability of a compound to prevent p53 binding, more evidence supporting the predicted binding location of oligobenzamides in the p53 binding site would be desirable.

The movement of peaks in an HSQC NMR spectrum can indicate which residues a compound binds close to on the surface of a protein. In an HSQC spectrum, there is one peak per N-H bond so, with the exception of proline, which does not show up, and residues with amine or amide-containing side chains, which have multiple peaks, there is one peak per residue. The position of each peak depends on the environment of the hydrogen and nitrogen atoms so is different for each amino acid. Conformational changes in the protein or direct binding of a compound can

affect the position of the residue peaks on the spectrum. Mapping residues associated with the shifting peaks to the 3D structure of the protein can therefore indicate where a small molecule might binds.

An HSQC NMR titration has previously been performed using an oligobenzamide dimer with carbon-linked side chains and wild-type Mdm2[118]. This project builds on this research, testing oligoamide trimers which are much more easily synthesised because they have oxygen-linked side chains and which should bind more strongly because they are larger and the side chains are prearranged by intramolecular hydrogen bonds.

This work is also novel because it uses the L33E mutant of Mdm2, a protein stabilised in the open, unbound conformation of the protein. Interpretation of HSQC spectrum peak movements can be complicated by the propagation of conformational change through a protein to residues some distance from a ligand binding site. The binding of peptides to wild-type Mdm2 is known to cause the movement of HSQC peaks created by residues some distance from the p53 binding site[126]. The L33E mutant of Mdm2 could therefore be useful because, if it does not undergo long-range conformational changes, only the peaks of residues directly involved in binding should move when a compound binds. Consequently, use of this protein in future, facilitated by its assignment in this project, could enable a better understanding of where potential inhibitors of the p53-Mdm2 interaction bind to Mdm2.

4.3. Methods

4.3.1. CELLS

E.coli DH5α and expression strains transformed with Novagen pET14b plasmid (Merck) carrying DNA encoding Mdm2 L33E were supplied as glycerol stocks by Geoffrey Plante and Alice Bartlett (University of Leeds). The plasmid was extracted from DH5α overnight stock cultures by Miniprep (Qiagen) according to the manufacturer's protocol. The expression strains CodonPlus (Stratagene), GOLD (Stratagene) and Rosetta 2 (Merck) were transformed using heat shock according to the manufacturer's instructions.

4.3.2. MEDIA

Overnight cultures were grown in 2xYT rich media comprising 16 gl^{-1} tryptone soya broth (Oxoid), 10 gl^{-1} yeast extract and 5 gl^{-1} sodium chloride solution.

Large scale expression was performed in 99% standard minimal medium, 1% LB rich medium with 100 gl^{-1} (260 mM) Carbenicillin. Expression tests were performed using a variety of different LB concentrations as detailed in the results section, also using 100 gl^{-1} Carbenicillin. LB comprised 10 gl^{-1} tryptone soya broth, 5 gl^{-1} yeast extract and 10 gl^{-1} sodium chloride solution.

Standard minimal medium contained 50 mM sodium phosphate, 0.4% w/w (22 mM) D-glucose, 20 mM potassium phosphate, 1 gl^{-1} (18 mM) ammonium chloride (^{15}N-labelled in large scale expressions), 0.5 gl^{-1} (8.6 mM) sodium chloride and 1 mM magnesium sulphate with trace elements and BME vitamins (Sigma Aldrich). The trace elements were, in their final concentrations, 50 µM iron(III) chloride, 20 µM calcium chloride, 10 µM manganese(II) chloride, 10 µM zinc sulphate, 2 µM cobalt(II) chloride, 2 µM copper(II) chloride, 2 µM nickel(II) chloride, 2 µM sodium molybdate and 2 µM boric acid. BME vitamin solution from Sigma Aldrich comprises 85 mgl^{-1} NaCl, 2 mgl^{-1} myo-inositol, 1 mgl^{-1} D-Biotin, 1 mgl^{-1} choline chloride, 1 mgl^{-1} folic acid, 1 mgl^{-1} niacinamide, 1 mgl^{-1} D-pantothenic acid-(50% Ca), 1 mgl^{-1} pyridoxal-HCl, 1 mgl^{-1} thiamine-HCl and 0.1 mgl^{-1} riboflavin.

Autoinduction medium was also used where indicated. This comprised 5% w/v (54 mM) glycerol, 50 mM potassium phosphate, 50 mM sodium phosphate, 2.5 gl^{-1} (46 mM) ^{15}N-ammonium chloride, 5 mM sodium sulphate, 2 gl^{-1} (5.8 mM) α-lactose, 0.5 gl^{-1} (2.8 mM) D-glucose, 1 mM magnesium sulphate, trace elements as above and BME vitamins (Sigma Aldrich) as above.

4.3.3. EXPRESSION TESTING

CodonPlus, GOLD or Rosetta 2 (see above) with pET14b carrying cDNA for Mdm2 L33E were grown up in 10 ml 2xYT broth overnight at 37°C, shaken at 200 rpm. 150 µl of the saturated culture was transferred into 15 ml of the growth medium containing 100 gl^{-1} Carbenicillin for testing. Cultures were grown as before at 37°C until the optical density of the solution at 600 nm exceeded 0.5. Expression was then induced by the addition of isopropyl thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM and cells were grown at 18°C for 16 hours unless otherwise stated. Cells were pelleted using 15 minutes' centrifugation at 3000 × g and were re-suspended in 20 mM Tris buffer (pH 7.9) with 0.5 M sodium chloride and 20 mM imidazole. Lysozyme and DNAse were added in excess. Then, the cells were lysed using repeated freezing in liquid nitrogen and thawing at room temperature until the viscosity of the solution appeared to drop indicating that lysis had occurred. The cell debris was pelleted by centrifugation at 16200 × g. The lysate was removed from each sample and analysed using polyacrylamide gel electrophoresis. Where stated, the composition of the cell debris pellet was similarly analysed after resuspension in 150 µl of phosphate buffered saline. Soluble proteins will be present in the supernatant. If the protein is not present in the supernatant, running the pellet on a gel in addition to the supernatant allows a lack of protein expression to be distinguished from expression followed by precipitation of the protein.

4.3.4. POLYACRYLAMIDE GELS

Protein samples were diluted using an equal volume of SDS reducing buffer (0.3 M Tris-HCl buffer (pH 6.8), 10% SDS, 0.05% w/v bromophenol blue and 50% glycerol). They were placed in a 95°C heat block for at least 3 minutes prior to loading. 15% polyacrylamide gels (described p214) were used, run in SDS running buffer comprising 25 mM Tris base, 192 mM glycine and 0.1% w/v SDS. Gels were run using a constant current of 30 mA.

4.3.5. LARGE SCALE EXPRESSION

GOLD cells with pET14b carrying cDNA for Mdm2 L33E were grown up in 40 ml 2xYT broth overnight at 37°C, shaken at 200 rpm. Large scale expression was performed using 2 litres of 99% standard minimal medium, 1% LB rich media with 100 gl^{-1} (260 mM) Carbenicillin. This was split between 4 baffled flasks. 10 ml of the 2xYT overnight culture was added to each flask. The cultures were grown as before at 37°C until the optical density of the solution at 600 nm exceeded 0.5. Expression was then induced by the addition of IPTG to a final concentration of 0.5 mM and cells were grown at 18°C for 16 hours. Cells were pelleted using 15 minutes' centrifugation at 6000 × g and were re-suspended in 20 mM Tris buffer (pH 7.9) with 0.5 M sodium chloride and 20 mM imidazole. Lysozyme and DNAse were added in excess. The cells were then lysed by sonication. The cell debris was subsequently pelleted by centrifugation at 15000 × g leaving the expressed protein in the supernatant.

4.3.6. PROTEIN PURIFICATION

Protein purification was carried out as described for p53-Cherry in chapter 3 (p214) but with two key differences. Firstly, the final storage buffer, hereinafter referred to as "NMR buffer", was pH 7.3, 73 mM Na_2HPO_4, 27 mM NaH_2PO_4, 1 mM DTT and 2.5% v/v (340 mM) glycerol, so constituted to ensure that the pH and ionic strength of the solution were similar to those used in the production of previously published chemical shifts for Mdm2[287]. Secondly, ^{15}N-Mdm2 is colourless so the protein concentration was determined using the UV absorbance, measured using a Nanodrop 1000 (Thermo Scientific) spectrophotometer. The construct contains no tryptophan but it has 7 tyrosine residues and a cysteine. The ProtParam tool[288] (http://web.expasy.org/protparam/, Swiss Institute of Bioinformatics) predicted the extinction coefficient to be 10430 M^{-1} cm^{-1} at 280 nm.

4.3.7. COMPOUNDS

The compounds PPY-2-75, AB062 and VA3-192 (Figure 4.13, p263) were supplied dissolved in DMSO by Panchami Prabhakaran, Anna Barnard and Valeria Azzarito (University of Leeds) respectively. PPY-2-75 is an oligobenzamide which fluorescence polarisation results suggest binds to Mdm2[42]. AB062 is identical to PPY-2-75 but for the addition of a polyethylene glycol chain intended to increase solubility. VA3-192 closely resembles PPY-2-75 but it has three changes intended to significantly increase solubility. Firstly, the central phenyl ring of the scaffold has been replaced with a single carbon which the middle side chain is attached to. Secondly, this middle side chain has been reduced in size to a benzyl group. The middle unit of the oligoamide has therefore become phenylalanine residue. Finally, a glycine residue has been added to the C-terminal end of the molecule.

4.3.8. TESTING COMPOUND SOLUBILITY

UV spectrophotometry was used to test the solubility of PPY-2-75 and AB062. In the case of PPY-2-75, 2 µl of 10 mM PPY-2-75 in DMSO was added to 3.3 ml of phosphate buffered saline with 10% DMSO in an attempt to make a 6 µM aqueous solution. After thorough vortex mixing, 1 ml was loaded into a cuvette, and further 1.1 ml aliquots were loaded into each of two Eppendorf tubes. Using a Cary 300 UV-Vis spectrophotometer (Varian), a UV spectrum (from 240 nm to 440 nm) of the sample in the cuvette was taken immediately. The other two samples were left at room temperature for 3 hours. One of these was then centrifuged at 16200 × g for 1 minute. 1 ml of the supernatant was removed from the top of each Eppendorf tube and loaded into a separate cuvette. The cuvettes were scanned like the first. The buffer was used for blanking.

The solubility of AB062 was also tested as follows. 338 µl of the NMR buffer (see "Protein purification" above) and 18.6 µl of pure DMSO were vortex mixed. Then, 3 µl of 10 mM AB062 was added in an attempt to make 360 µl of a 6% DMSO, 10 µM AB062 solution. After a

second vortex mix, 150 µl of the solution was removed and mixed into 3 ml of DMSO in a cuvette allowing the UV spectrum to be read from 280 nm to 400 nm. The remaining 210 µl was immediately centrifuged at 16200 × g for 1 minute. 150 µl of the supernatant was then removed and this was mixed into 3 ml of DMSO in a cuvette. The spectrum was read again. 150 µl of NMR buffer was added to 3 ml of DMSO in a cuvette to make the blank.

The experiment with AB062 was repeated using a DMSO concentration of only 1%. Again 360 µl of solution was made; 3 µl of 10 mM AB062 in DMSO was added to 356 µl of NMR buffer and 0.6 µl DMSO.

The spectra were compared to see if centrifugation had reduced the concentration of the compound in the solution, causing a drop in its UV absorption. Unless they have aggregated to form larger particles, soluble compounds stay in the solution so the spectra should be identical.

4.3.9. NMR

As described in the results section, three dimensional NOESY-HSQC and TOCSY-HSQC spectra were used to assign the amide peaks in the HSQC spectrum of Mdm2 L33E to their respective amino acids. The NOESY-HSQC and TOCSY-HSQC were collected on a 750 MHz NMR spectrometer (Varian) using 16 transients per increment. After assignment had been completed, potential ligands were titrated into a sample of the protein. HSQC spectra were collected after selected additions using either the 750 MHz spectrometer or a 500 MHz machine (Varian) with typically 12 to 16 transients; however the number of transients was increased for the last point of a titration if the signal to noise ratio was poor.

Pulse sequences from the BioPack pulse sequence library (Varian) were used. Phasing was performed using NMRpipe (National Institutes of Health)[289] and spectra were viewed using CcpNmr Analysis (CCPN)[290].

All measurements were taken at 15°C and, unless otherwise stated, spectra were obtained using solutions of 90% NMR buffer (see "Protein purification" above) and 10% deuterium oxide (Cambridge Isotope Laboratories). (The protein was stored in NMR buffer so the protein solution added formed part of this component.) Fresh dithiothreitol was added to the buffer used.

For the 3D assignment spectra, 300 µl of solution was used in a Shigemi tube with a protein concentration of 300 µM. For 2D spectra, the protein concentration was reduced to 100 µM. In titrations, 300 µl was the initial volume but the volume increased as compound was added. The change in solution volume was factored in when calculating the volume of compound solution to add at each step of the titration.

Ligands were added from 10 mM stock solutions with DMSO as the solvent with the exception of a few additions which would have required adding less than 0.5 µl of the stock solution. In these cases, a 100 times (0.1 mM) dilution of the stock solution was made using NMR buffer and was used to add the required amount of compound to the protein solution. As described in the methods, in an experiment testing AB062, the detergent CHAPS (3-[(3-cholamidopropyl)dimethyl-ammonio]-1-propanesulfonate) was used at a concentration of 20 mM (before any dilution due to compound addition). This was incorporated by adding 10% w/v CHAPS when preparing the titrant and reducing the volume of NMR buffer accordingly.

To quantify the difference in peak positions between two spectra, a distance metric, d, was defined, as shown in equation 4.1[291]. This is the root mean square of the chemical shift deviations, δ_H and δ_N for hydrogen and nitrogen respectively, after the hydrogen shifts have been multiplied by 5 to account for the difference in magnitude of the shifts for the two atom types.

$$d = \sqrt{\frac{1}{N}\sum_{peaks}\left[(5\delta_H)^2 + {\delta_N}^2\right]} \qquad \textbf{4.1}$$

Where N is the number of peaks.

4.4. Results and Analysis

4.4.1. EXPRESSION AND PURIFICATION OF MDM2 L33E

4.4.1.1. Expression testing

Expression of Mdm2 was tested in three different host cell lines under a variety of conditions to identify which cell line and conditions would lead to most effective production of soluble Mdm2. The small size of the energy differences probed in NMR experiments leads to a high signal to noise ratio, necessitating the use of high protein concentrations. However, production of labelled protein necessitates the use of minimal media, which can lead to poor expression. Expression testing enables the optimal conditions for expression of soluble protein to be determined prior to the use of large volumes of media containing expensive labelled compounds.

Cells of each cell line were transformed with recombinant pET14b containing Mdm2 L33E cDNA. This encoded a 6 His-tagged version of the protein. The ability of the cells to express soluble Mdm2 was assessed as described in the methods section. Good expression had previously been observed using Condon Plus cells with rich media; however, in minimal media, the results obtained with CodonPlus cells were inconsistent, suggesting that there was some tendency for plasmid loss. In addition, subsequent testing under a variety of conditions (Figure 4.1) implied that Mdm2 was insoluble. (After centrifugation of the cell contents, polyacrylamide gel electrophoresis showed Mdm2 was present in the pellets but not the lysates.)

No expression was evident in Rosetta cells. However, there was good expression of soluble protein in GOLD cells using several media. Initially GOLD cell expression was not observed in pure minimal medium (Figure 4.2); however, in further trials to establish the minimum concentration of LB necessary for expression in GOLD (Figure 4.3), expression was observed in entirely minimal medium. The latter suggested that the former result was anomalous and therefore that the LB medium was probably superfluous; however, large scale expression was carried out using cells from this last experiment with 1% LB in case the concentration of a trace LB medium constituent was a limiting factor.

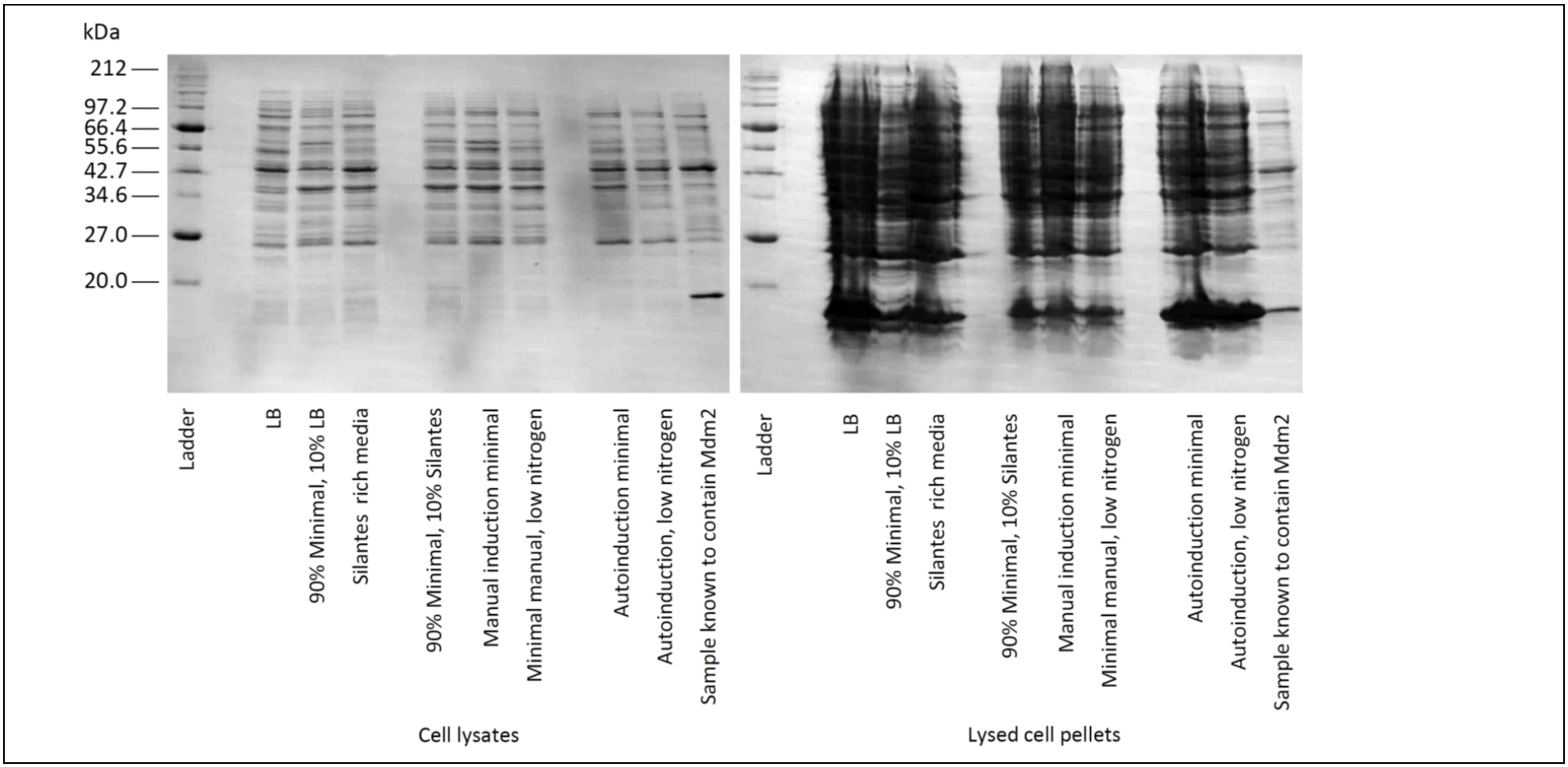


Figure 4.1**: Expression of Mdm2 in CodonPlus: trials using different media**. After cells had been lysed, both the lysates (left) and the cell debris pellets (right) were analysed by SDS-PAGE. The results show that Mdm2 was expressed but was insoluble.

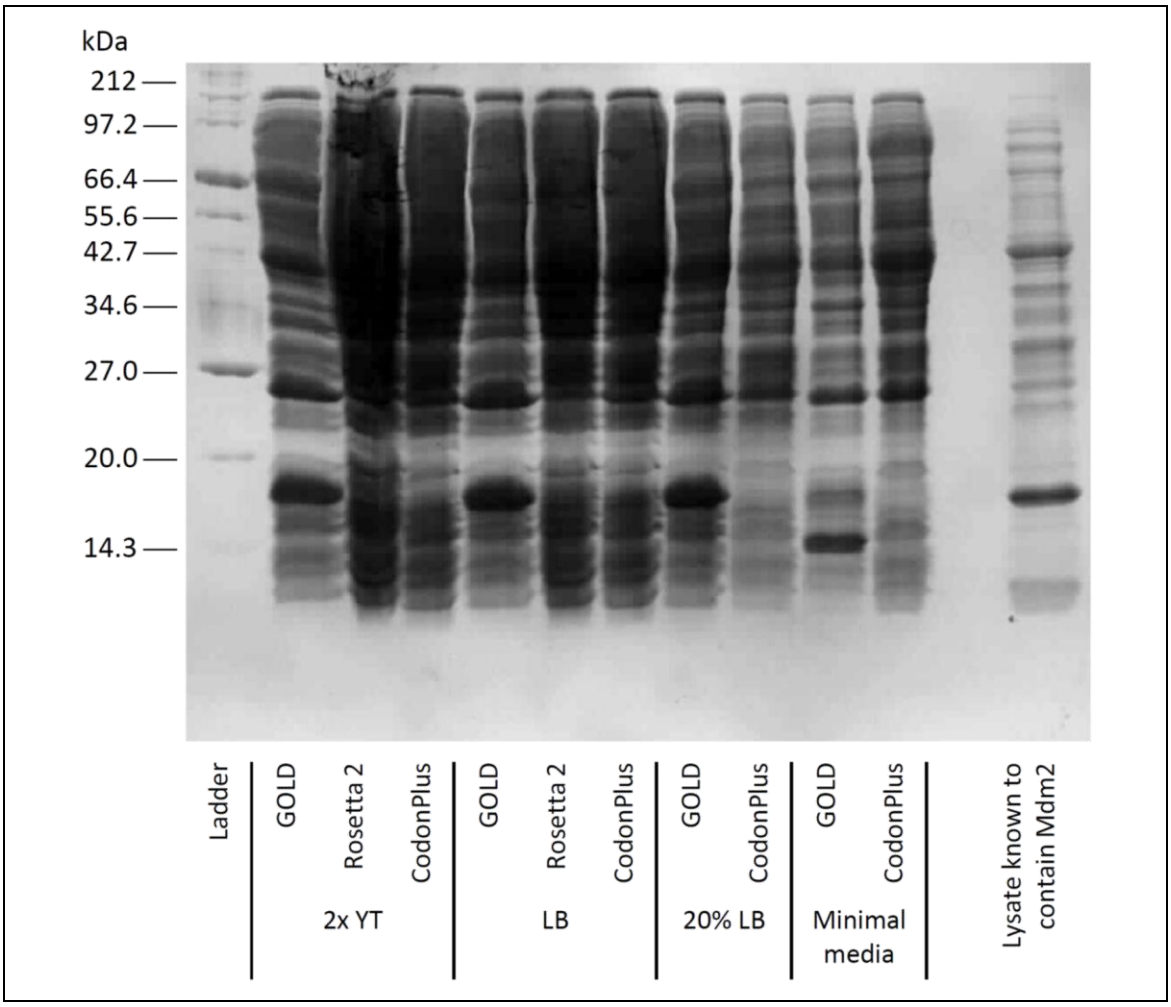


Figure 4.2: SDS-PAGE gel showing samples from lysed GOLD, CodonPlus and Rosetta 2 cells after expression testing in different media. Soluble Mdm2 is present in every GOLD sample except the one from cells grown in minimal medium. The intense band in the GOLD minimal medium lane around 15 kDa is lysozyme, which was added (in excess) to all samples prior to sonication to aid lysis.

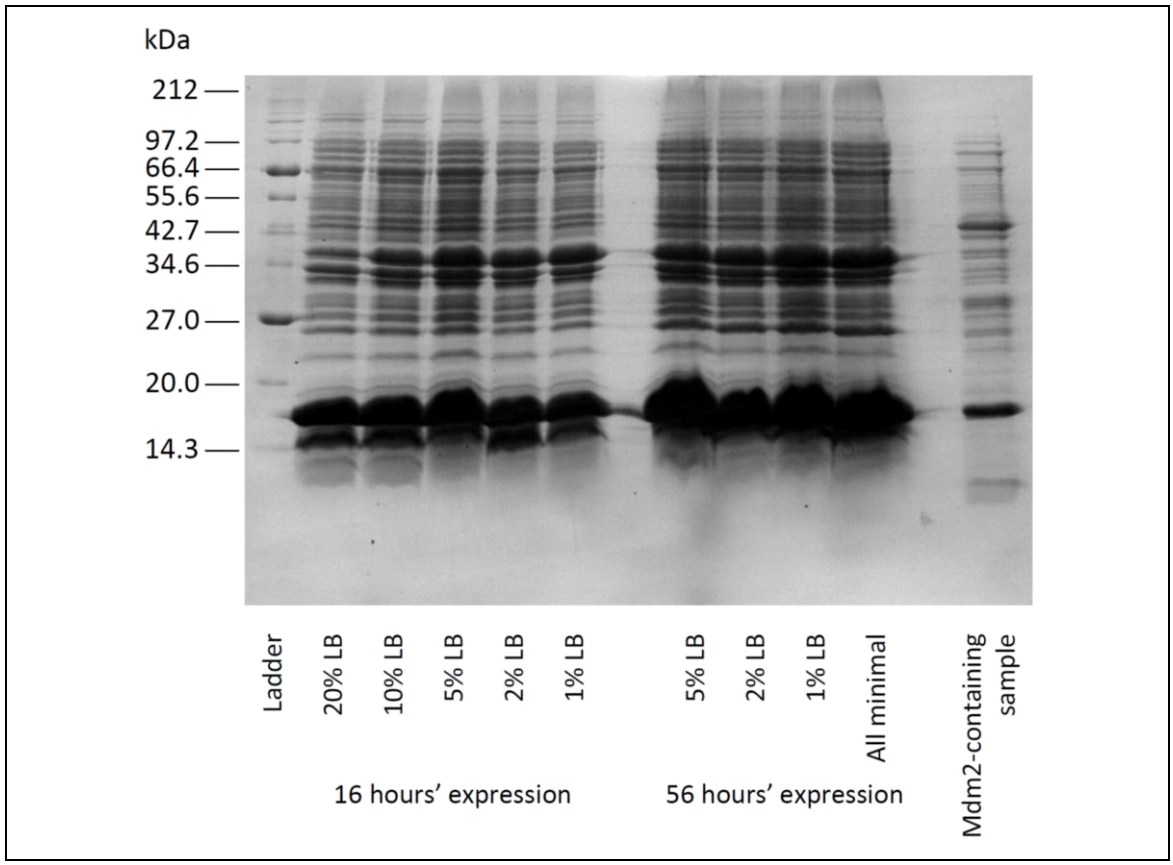


Figure 4.3: SDS-PAGE gel showing samples from lysed GOLD cells grown in part-minimal, part-LB medium. Two different expression times were tried and a variety of LB concentrations. Expression was observed under all conditions.

4.4.1.2. Large scale expression of Mdm2

Based on the expression tests, a large scale expression of Mdm2 was set up using minimal medium supplemented with 1% LB to ensure all necessary trace elements and vitamins were present. For further details see the relevant methods section (p242). For the purpose of protein purification, GOLD cells were lysed by sonication and the homogenate was filtered. Subsequently, the His-tagged Mdm2 was separated from the filtrate by nickel affinity chromatography. Elution occurred at an imidazole concentration of 120 mM (Figure 4.4). After centrifugal concentration of the eluate, gel filtration, which separates proteins based on their size, was used to further purify the protein solution. The resulting UV absorption trace of the eluate (Figure 4.5A) was consistent with the presence of some aggregated material but it showed that a significant amount of protein had been eluted at the expected elution volume for Mdm2 of around 200 ml (based on its molecular weight of 15 kDa). Samples of the eluate responsible for the first peak revealed that Mdm2 did not contribute to this aggregate peak (Figure 4.5B). Contaminating DNA could have been responsible for the peak because the UV absorbance was measured at 280 nm, a frequency which is absorbed by DNA as well as protein.

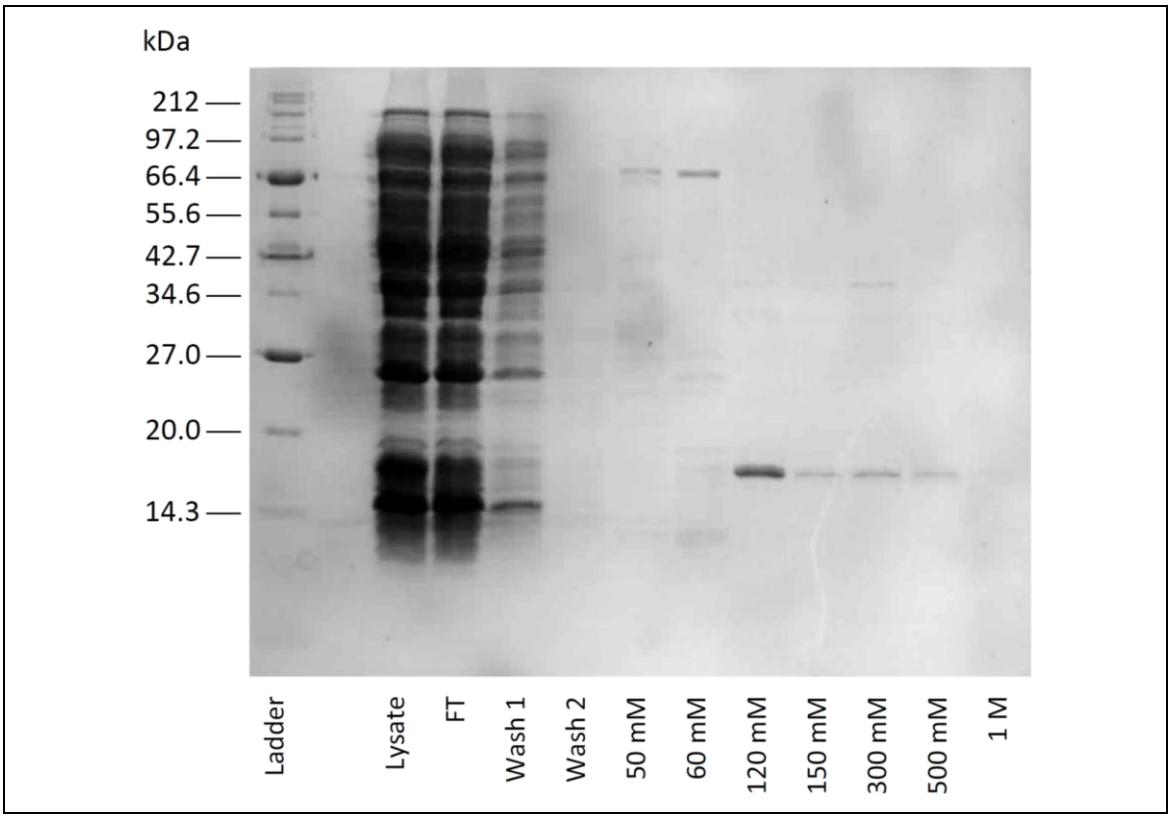


Figure 4.4: SDS-PAGE gel showing samples eluted at different imidazole concentrations during the first part of the 15N-Mdm2 purification process, nickel-affinity chromatography. The eluate at 120 nM was concentrated and then carried forward to the gel filtration stage. (FT=Flow through.)

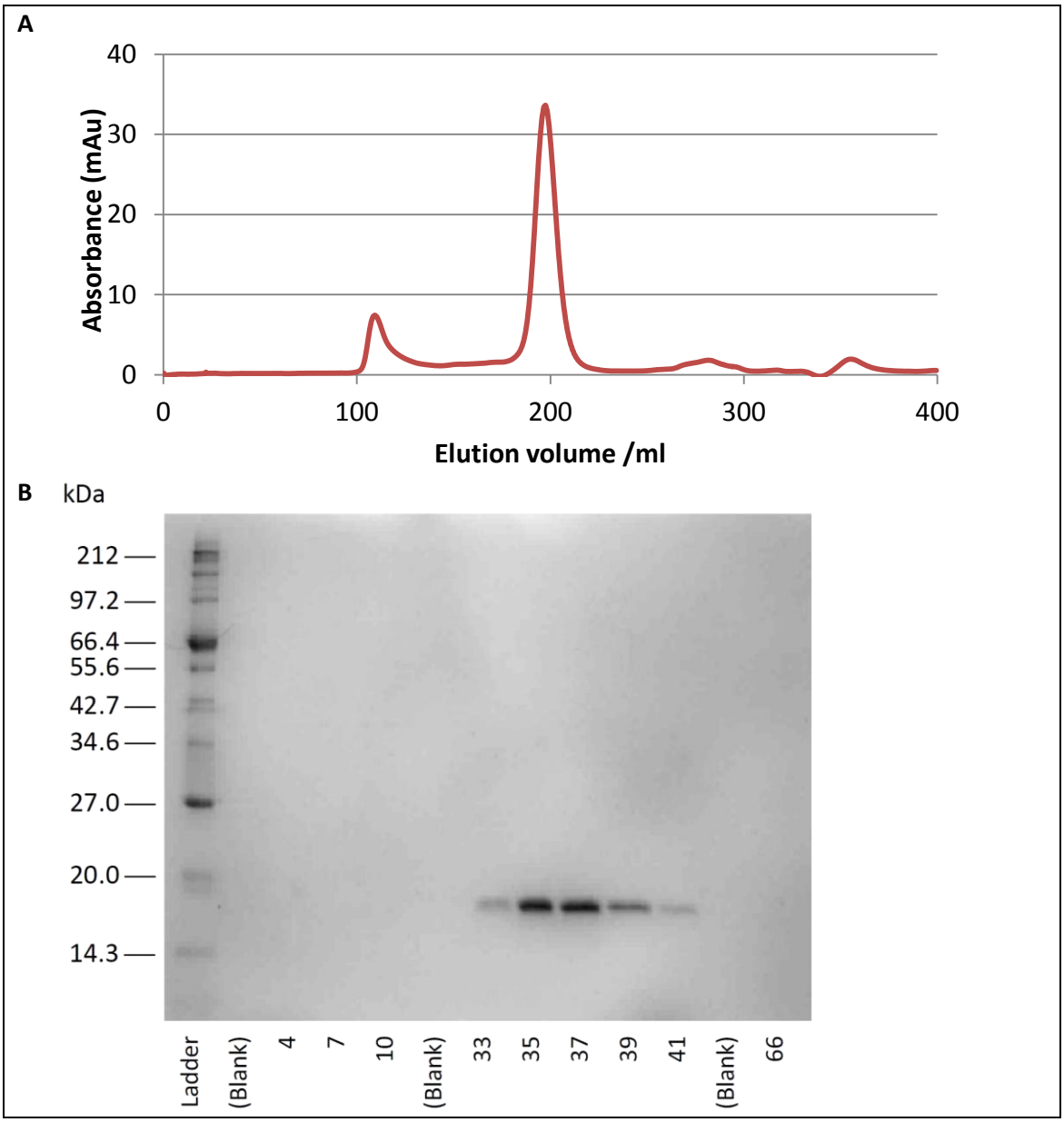


Figure 4.5: Gel filtration of ^{15}N-Mdm2. A) Graph of the eluate absorbance at 280 nm during the purification. B) SDS-PAGE gel showing samples from selected fractions of the eluate. 5.5 ml samples were taken so the first few lanes feature samples from the initial aggregate peak, the subsequent lanes with bands correspond to the main 200 ml peak and the sample in the final lane was eluted around the 360 ml point. The gel confirms that Mdm2 came off in fractions 33 to 41.

4.4.2. ASSIGNMENT OF MDM2 HSQC SPECTRUM PEAKS TO RESIDUES

In order to interpret the movement of HSQC spectrum peaks upon the addition of a potential peptidomimetic it is necessary to assign the amide peaks in the HSQC spectrum of Mdm2 L33E to their respective amino acids.

Previously published amino-nitrogen and hydrogen resonances for wild-type Mdm2 were obtained from Biological Magnetic Resonance Bank (BMRB) (http://www.bmrb.wisc.edu/)[292] entry 6612[287] (the Mdm2 N-terminal domain with no ligand).

A Heteronuclear Spin Quantum Correlation (HSQC) spectrum of the purified Mdm2 L33E protein was obtained. Comparison of peak positions with those of the wild-type revealed many shifted resonances preventing full assignment from the wild-type resonances alone. Because the protein is fairly small, assignment of Mdm2 L33E could be completed using only ^{15}N-labelled protein, without recourse to double labelling.

Assignment was carried out primarily using a 3D ^{1}H-^{15}N NOESY-HSQC spectrum, a 2D projection of which is shown in Figure 4.6. This indicates where the nuclear Overhauser effect (p55) occurs between atoms of the protein and consequently, which atoms are in close proximity. The nuclear Overhauser effect has a distance dependence inversely proportional to the sixth power of the distance between the interacting nuclear dipoles[107(p26)] so the atoms must be very close.

Two types NOE of peak are visible in the spectrum. Firstly, there are sequential amide-amide NOE peaks. These occur in the 6-9 ppm region of the spectrum and are caused by magnetisation transfer between the alpha amino groups of adjacent residues. Secondly, there are amide-alpha proton NOEs. These are produced by magnetisation transfer between alpha amino groups and the aliphatic protons in the side chain of both the same residue and the preceding residue in the polypeptide chain. Consequently, NOEs can be identified between amide-alpha-1 proton and amide-beta-1 protons in addition to the aforementioned sequential amide-amide NOEs.

Figure 4.7 illustrates the use of sequential amide-amide NOESY-HSQC peaks to identify adjacent residues.

The expressed region of Mdm2 (Figure 4.8) is a mixed alpha/beta protein. Due to the different interatomic distances within alpha helices and beta-sheets, amide-amide NOESY peaks are particularly clear in alpha-helical regions whereas amide-aliphatic proton NOESY peaks appear strongest in beta-sheets.

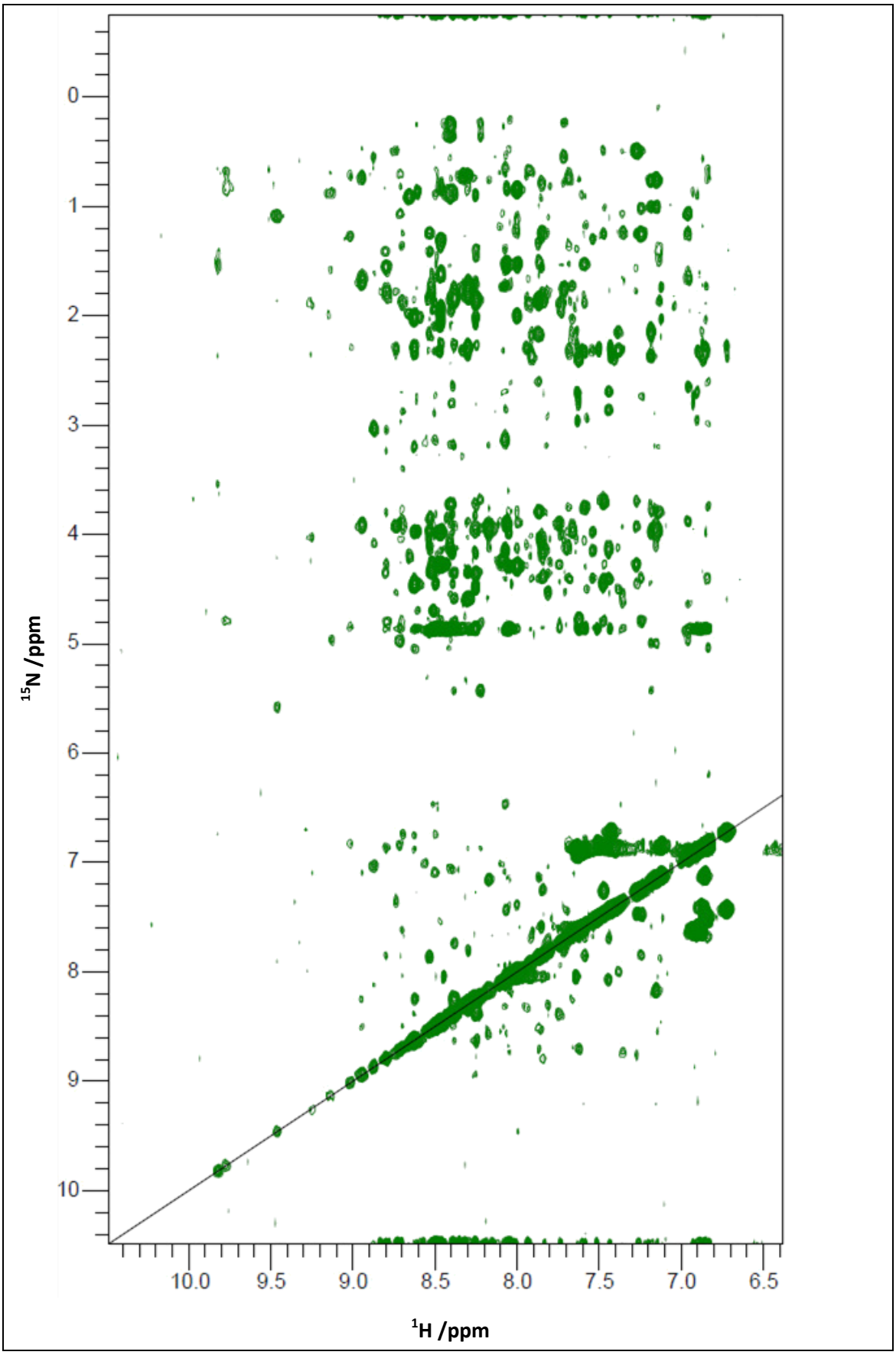


Figure 4.6: A 2D projection of the 3D NOESY-HSQC spectrum of Mdm2 L33E, looking down the ^{15}N axis

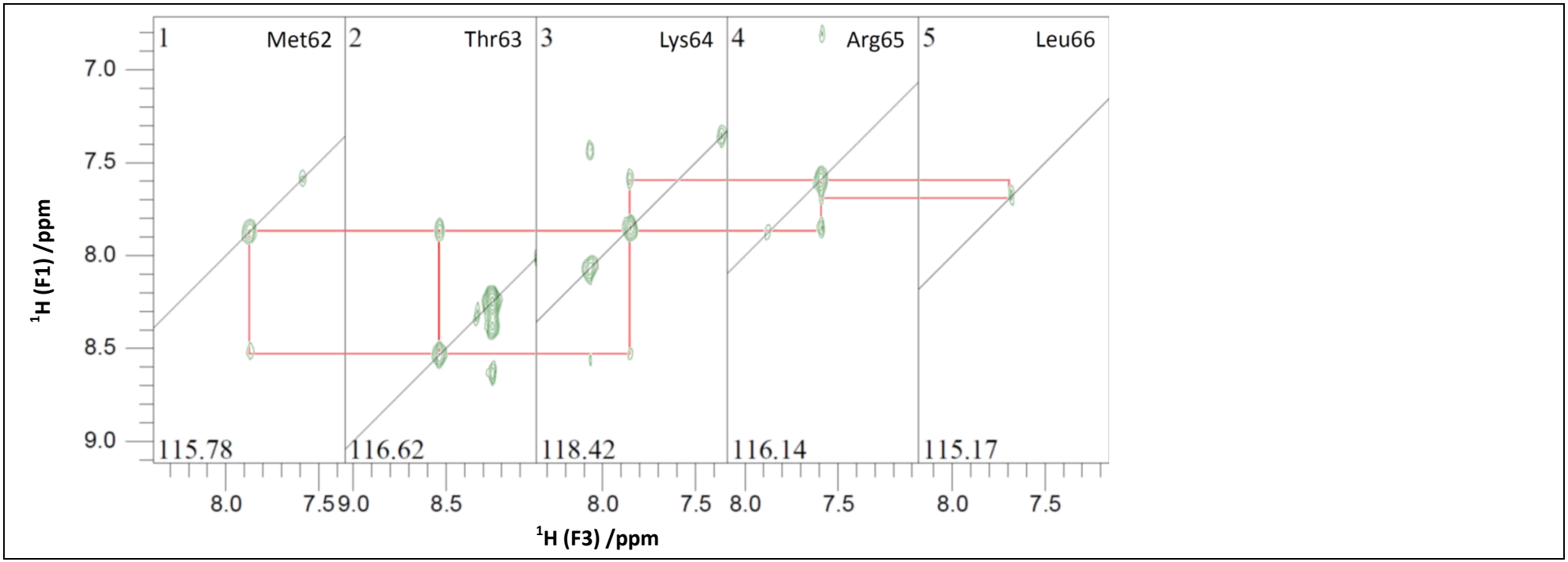


Figure 4.7: NOESY-HSQC spectra strip projections for residues 62 to 66 showing cross-peaks. The position of the peaks on the ^{15}N axis is shown at the bottom of each strip. The unfiltered HSQC peaks lie on the diagonal. The cross-peaks indicate HSQC peaks caused by the transfer of magnetisation between adjacent residues, thereby facilitating their assignment of the residues.

Figure 4.8: The anticipated secondary structure of the ^{15}N-labelled protein. The amino acids corresponding to residues 17-125 of Mdm2 are shown in colour. The regions which fold to make alpha-helices and beta strands in Mdm2 are shown in orange and red respectively. This secondary structure annotation is based on UniProt entry Q00987 (wild-type Mdm2) and is consistent with the author secondary structure annotation of PDB X-ray structures 4DIJ[293] and 3LBK[294], which feature the L33E mutant.

To aid identification of the amino acid residue responsible for each HSQC peak, a TOCSY (Total Correlation Spectroscopy)-HSQC spectrum was used. TOCSY-HSQC spectrum peaks result from the transfer of magnetisation between the amide group and protons within the same residue. Because the transfer is confined to a single residue, the resulting peaks can facilitate identification of the residue type. Furthermore, they enable the aliphatic proton peaks of residues in the NOESY-HSQC spectrum to be distinguished from the cross-peaks in this region created by the side chain of the preceding residue in the sequence. A 2D projection of the 3D TOCSY-HSQC spectrum is shown in Figure 4.9.

By using the NOE based assignment strategy described above, 84 of the 109 p53 residues present in the construct could be assigned. The residues that are part of the His-tag and linker at the N-terminus could not be assigned and neither could the last 7 residues at the C-terminus. The other unassigned residues are distributed around the ends of the protein sequence, the longest unassigned stretch being residues 111 to 114.

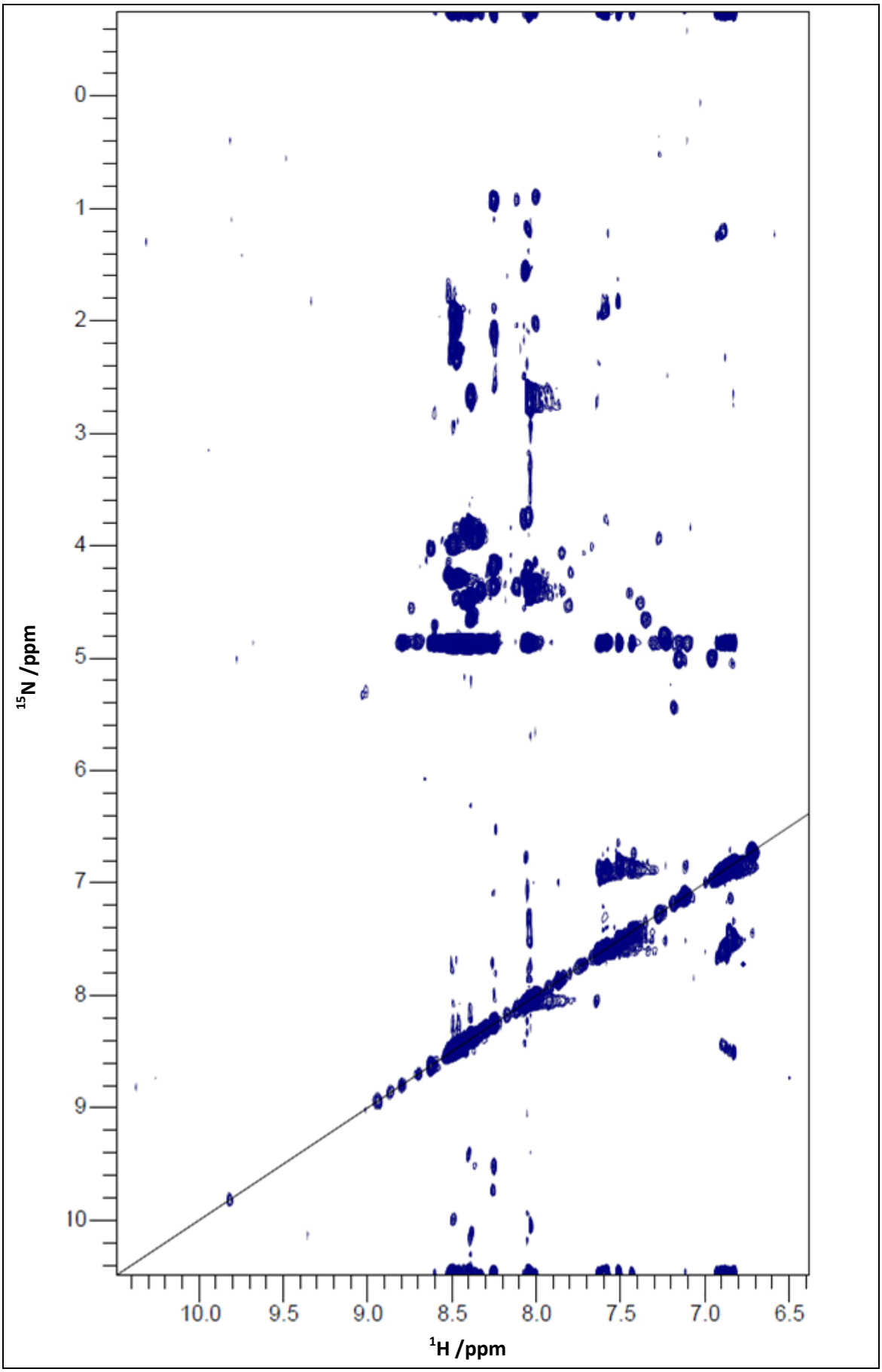


Figure 4.9: A 2D projection of the 3D TOCSY-HSQC spectrum of Mdm2 L33E, looking down the ^{15}N axis

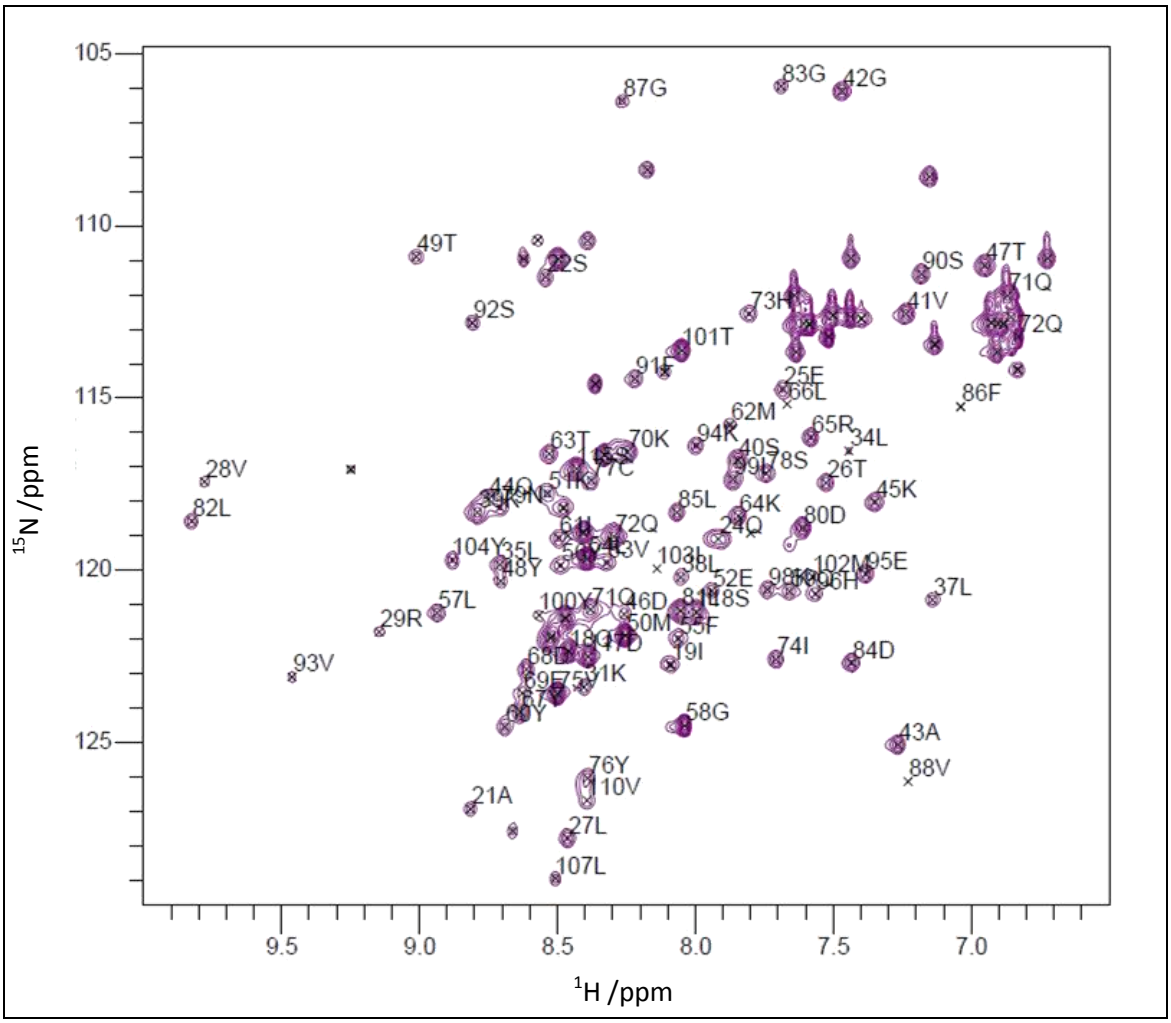


Figure 4.10: The HSQC spectrum of Mdm2 L33E spectrum with peaks assigned. The Phe86, Val88 and Leu103 peaks are not visible above because the cut-off is set high to remove all of the noise; these peaks are apparent if the cut-off is reduced and their positions are indicated.

There were significant differences between the previously published wild-type shifts and those observed with the mutant (Figure 4.11). The location of these shifted residues is consistent with the L33E mutation being responsible for the movement; the affected residues are either near to the site of the mutation or are closely associated with residues near to the site. For example, Leu81 is next to and faces towards Leu33 in structure 1T4F and it has moved significantly. The reason why the Asp46 peak has moved is unclear.

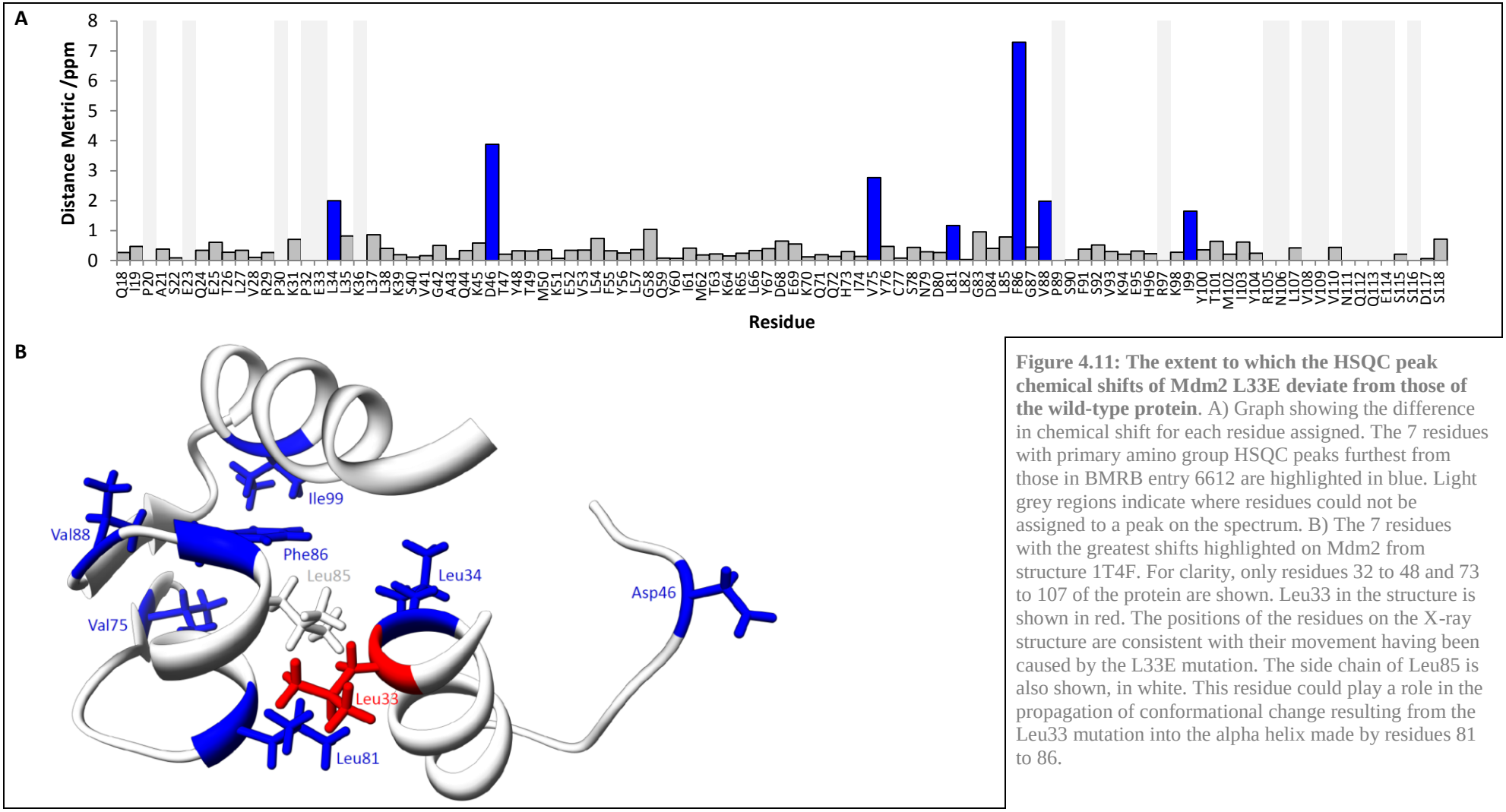


Figure 4.11: The extent to which the HSQC peak chemical shifts of Mdm2 L33E deviate from those of the wild-type protein. A) Graph showing the difference in chemical shift for each residue assigned. The 7 residues with primary amino group HSQC peaks furthest from those in BMRB entry 6612 are highlighted in blue. Light grey regions indicate where residues could not be assigned to a peak on the spectrum. B) The 7 residues with the greatest shifts highlighted on Mdm2 from structure 1T4F. For clarity, only residues 32 to 48 and 73 to 107 of the protein are shown. Leu33 in the structure is shown in red. The positions of the residues on the X-ray structure are consistent with their movement having been caused by the L33E mutation. The side chain of Leu85 is also shown, in white. This residue could play a role in the propagation of conformational change resulting from the Leu33 mutation into the alpha helix made by residues 81 to 86.

4.4.3. TESTING OF POTENTIAL P53-MDM2 INTERACTION INHIBITORS

Once peaks in the HSQC spectrum had been identified, small molecules were titrated in to the Mdm2 solution. Table 4.1 shows the concentrations used in each titration and summarises the degree of peak movement observed upon the addition of each compound.

Initially DMSO was tested because it must be added with many potential inhibitors. Subsequently, three potential inhibitors were tested. To summarise the results, little peak movement was observed upon the addition of the oligobenzamides PPY-2-75 and AB062 considering that these had to be added with DMSO and movement was observed when DMSO alone was added. In contrast, some peaks were seen to move a significant distance upon addition of the oligoamide VA3-192. The movement of peaks was assessed using the distance metric described by Equation 4.1 (p245). While there was little movement overall, some peaks moved a significant distance in a dose-dependent manner, with one peak seen to move a distance corresponding to a metric of 0.755 upon addition of VA3-192 to a concentration of 200 μM.

In the subsequent sections, the results are discussed in more detail. Specifically, the locations of residues with moving peaks are identified in PDB structure 1T4F for the purpose of assessing their position relative to the p53 binding site.

Table 4.1: The root mean square HSQC peak movement during titrations into Mdm2 L33E

Addition	Concentration (/μM unless otherwise stated)	Compared to start of titration		
		Shared α-amino peaks	RMS distance metric /ppm	Maximum distance metric /ppm
DMSO	2%	83	0.0866	0.1962
	4%	71	0.1219	0.3609
	6%	68	0.1734	0.4184
	8%	42	0.2266	0.3987
PPY-2-75	140	82	0.0379	0.0953
	270	89	0.0646	0.1568
	395	90	0.0875	0.3199
AB062	0.16	89	0.0667	0.1443
(no CHAPS)	4	76	0.0850	0.1721
	100	50	0.1340	0.2962
VA3-192	1.56	66	0.0670	0.2346
	6.25	61	0.0679	0.3112
	25	51	0.0880	0.4607
	100	71	0.0984	0.5714
	200	70	0.1254	0.7551

The root mean square (RMS) of the peak distance metrics (calculated using equation 4.1, p245) was used to assess the similarity of the spectrum obtained after each addition to the spectrum obtained at the start of the titration before any titrant had been added. The "Shared α-amino peaks" column shows the number of peaks used to calculate the root mean square (RMS) distance metric. The rightmost column shows the distance metric for the peak which moved the most. Little movement was observed when PPY-2-75 and AB062 were titrated in; however, some peaks moved a significant amount and in a dose-dependent manner upon addition of VA3-192.

4.4.3.1. Premiliary testing with DMSO

The potential inhibitors tested in this project were added from 10 mM stock solutions with DMSO as the solvent. Using DMSO ensures that the compounds are soluble both in these concentrated solutions and when diluted using an aqueous protein solution.

To investigate the effect of DMSO alone on the position of the HSQC spectrum peaks, a titration was performed in which the DMSO concentration was increased from 0% to 8% inclusive, in increments of 2%. Changes in chemical shift were observed for residues all over the surface of the protein (Figure 4.12), particularly when the DMSO concentration was increased from 6% to 8%. As indicated in Table 4.2 (p262), this final increase was accompanied by a decrease in the total number of peaks which were visible above the background noise and an increase in the proportion of these peaks which could not be identified (because they had moved significantly). These results suggest that the protein is becoming denatured at concentrations above 6% DMSO.

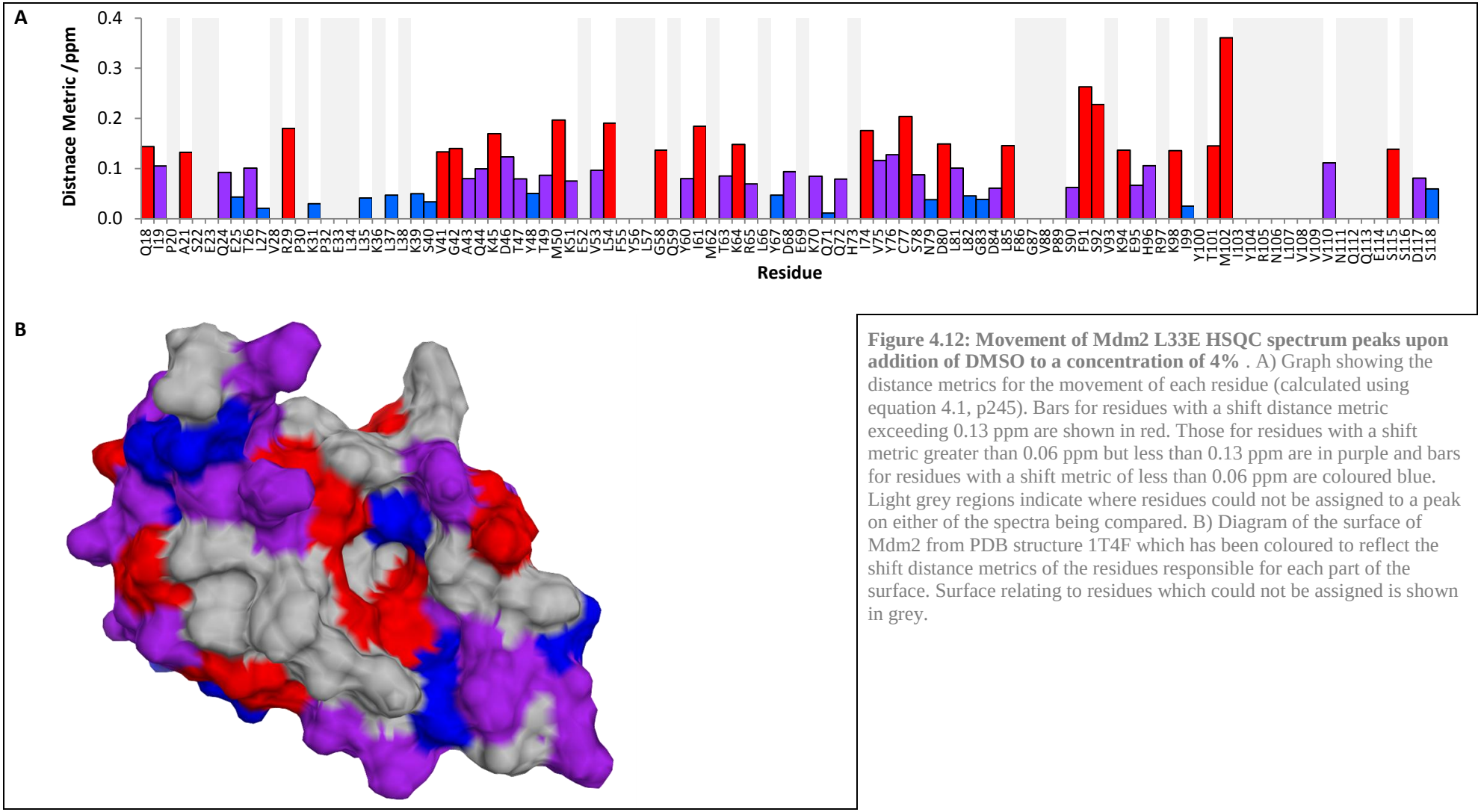


Figure 4.12: Movement of Mdm2 L33E HSQC spectrum peaks upon addition of DMSO to a concentration of 4% . A) Graph showing the distance metrics for the movement of each residue (calculated using equation 4.1, p245). Bars for residues with a shift distance metric exceeding 0.13 ppm are shown in red. Those for residues with a shift metric greater than 0.06 ppm but less than 0.13 ppm are in purple and bars for residues with a shift metric of less than 0.06 ppm are coloured blue. Light grey regions indicate where residues could not be assigned to a peak on either of the spectra being compared. B) Diagram of the surface of Mdm2 from PDB structure 1T4F which has been coloured to reflect the shift distance metrics of the residues responsible for each part of the surface. Surface relating to residues which could not be assigned is shown in grey.

Table 4.2: Peak statistics for the titration spectra analysed

Experiment	Titrant concentration (μM unless %)	Total peaks	Recognised primary amino peaks*	Assigned primary amino group peaks
Initial HSQC	N/A	140	100	84
DMSO titration	0%	174	92	79
	2%	135	86	74
	4%	113	73	64
	6%	156	70	60
	8%	82	43	34
PPY-2-75 titration	0	155	95	79
	140	113	83	72
	270	137	90	76
	395	176	93	80
AB062 titration	0	148	94	80
	0.16	141	91	77
	4	138	76	71
	100	87	50	43
VA3-192 titration	0	130	79	72
	1.56	112	71	62
	6.25	101	64	57
	25	96	52	47
	100	176	83	71
	200	177	81	70

*"Recognised peaks" were peaks thought to be caused by primary amino groups that had been observed in the initial HSQC spectrum and which were tracked, where possible, through each experiment. They included assigned peaks but also some which could not be attributed to a particular residue.

4.4.3.2. Inhibitor 1: PPY-2-75

PPY-2-75 (Figure 4.13, compound **a**) was titrated into an Mdm2 solution to a concentration of 395 μM. Because it was added from a 10 mM stock solution with DMSO as the solvent, the Mdm2 solution was 4% DMSO at the end of the titration.

If PPY-2-75 were binding to Mdm2, the peaks representing residues around the predicted binding site would be those most likely to move. In contrast, the observed peak movement (Figure 4.14) suggests residue perturbation all over the protein, as was seen after the addition of pure DMSO. Five of the fifteen residues whose peaks had shifted most by the end of the PPY-2-75 titration were among the 15 peaks which had moved the most by the end of the DMSO-only titration: Ser92, Cys77, Met50, Leu54 and Arg29. Peaks corresponding to the compound PPY-2-75 were not visible on a 1D proton spectrum of PPY-2-75 in 4% DMSO with no protein (not shown). This suggests that the HSQC peak movement was caused by the DMSO in which the compound was dissolved, rather than binding of the compound itself.

The distance moved by the peaks over the course of the titration was similar to that observed in the DMSO-only titration over the 0% to 4% range. However, the PPY-2-75 (and 4% DMSO)

spectrum bears more similarity to the spectrum taken at the start of the DMSO titration (the 0% DMSO spectrum) (distance metric 0.044 ppm) than the 4% DMSO spectrum in the DMSO-only titration (distance metric 0.105 ppm). This suggests that the observed peak movement when PPY-2-75 is added reflects random peak movement caused by DMSO-driven Mdm2 denaturation. To summarise, the movement of the peaks does not constitute evidence that PPY-2-75 would inhibit the p53-Mdm2 interaction.

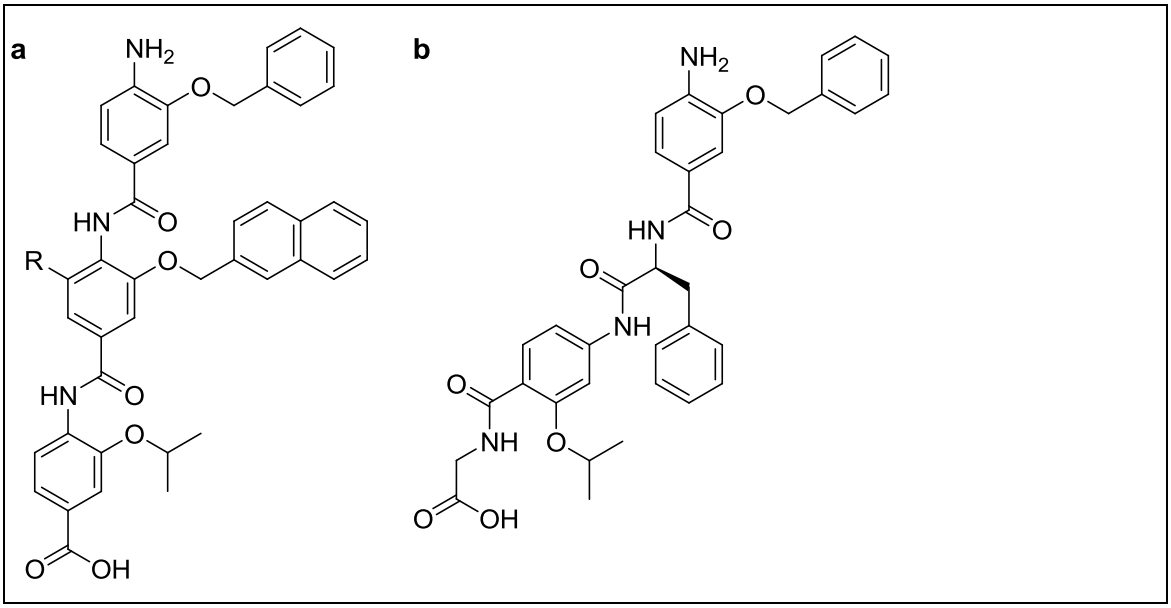


Figure 4.13: The structures of the potential inhibitors titrated into Mdm2. a) PPY-2-75 (R=H) and AB062, the same molecule but with a polyethylene glycol (PEG) chain attached ($R=O(CH_2CH_2O)_3CH_3$). b) VA3-192.

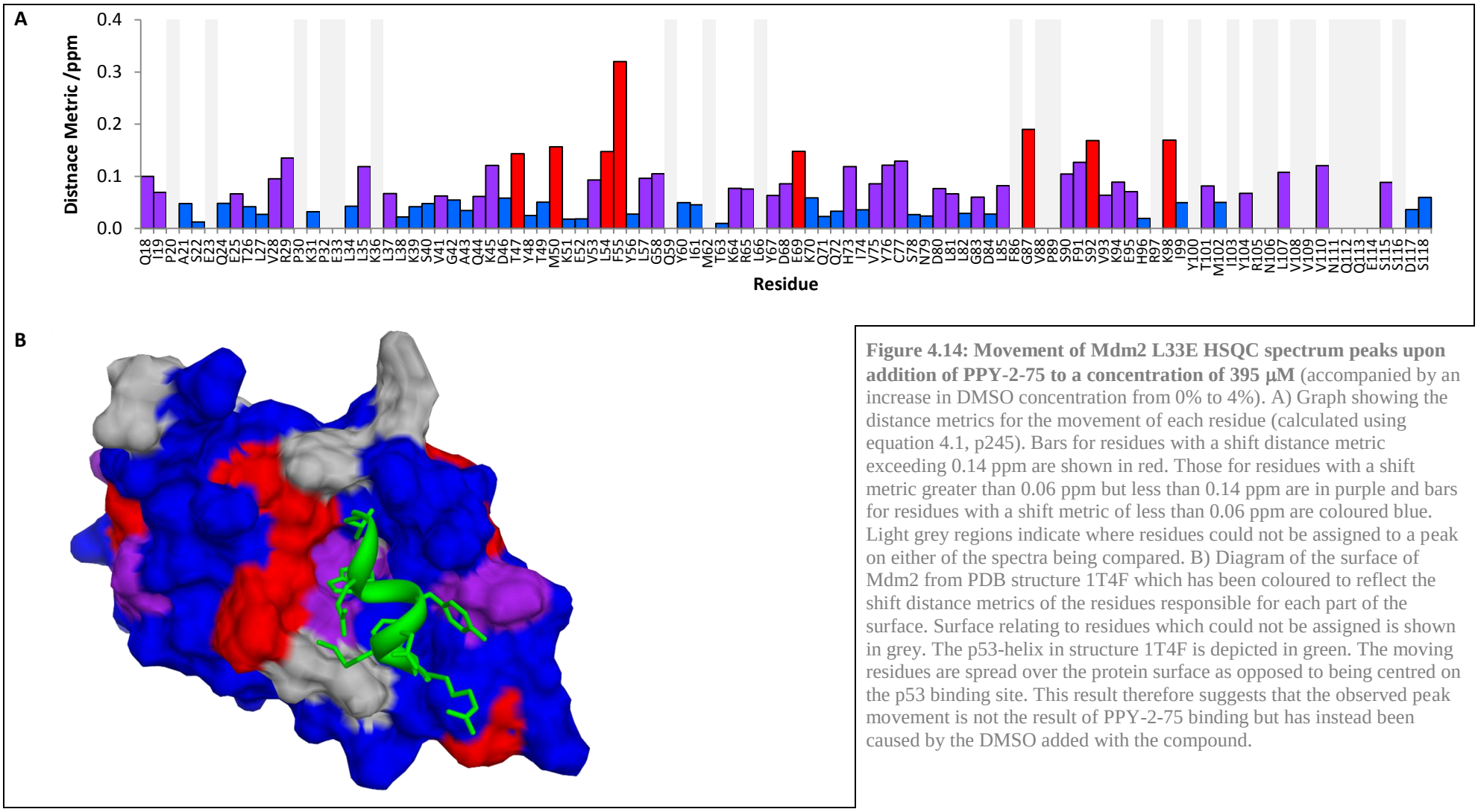


Figure 4.14: Movement of Mdm2 L33E HSQC spectrum peaks upon addition of PPY-2-75 to a concentration of 395 µM (accompanied by an increase in DMSO concentration from 0% to 4%). A) Graph showing the distance metrics for the movement of each residue (calculated using equation 4.1, p245). Bars for residues with a shift distance metric exceeding 0.14 ppm are shown in red. Those for residues with a shift metric greater than 0.06 ppm but less than 0.14 ppm are in purple and bars for residues with a shift metric of less than 0.06 ppm are coloured blue. Light grey regions indicate where residues could not be assigned to a peak on either of the spectra being compared. B) Diagram of the surface of Mdm2 from PDB structure 1T4F which has been coloured to reflect the shift distance metrics of the residues responsible for each part of the surface. Surface relating to residues which could not be assigned is shown in grey. The p53-helix in structure 1T4F is depicted in green. The moving residues are spread over the protein surface as opposed to being centred on the p53 binding site. This result therefore suggests that the observed peak movement is not the result of PPY-2-75 binding but has instead been caused by the DMSO added with the compound.

To test the solubility of PPY-2-75, an attempt was made to make a 6 µM aqueous solution of the compound. Samples were taken from the top of the resulting solution immediately, after 3 hours and after a subsequent centrifugation. Each sample was immediately scanned using a UV spectrophotometer.) The significant reduction in absorbance evident in the centrifuged sample suggests that roughly half of the compound molecules initially in solution had precipitated.

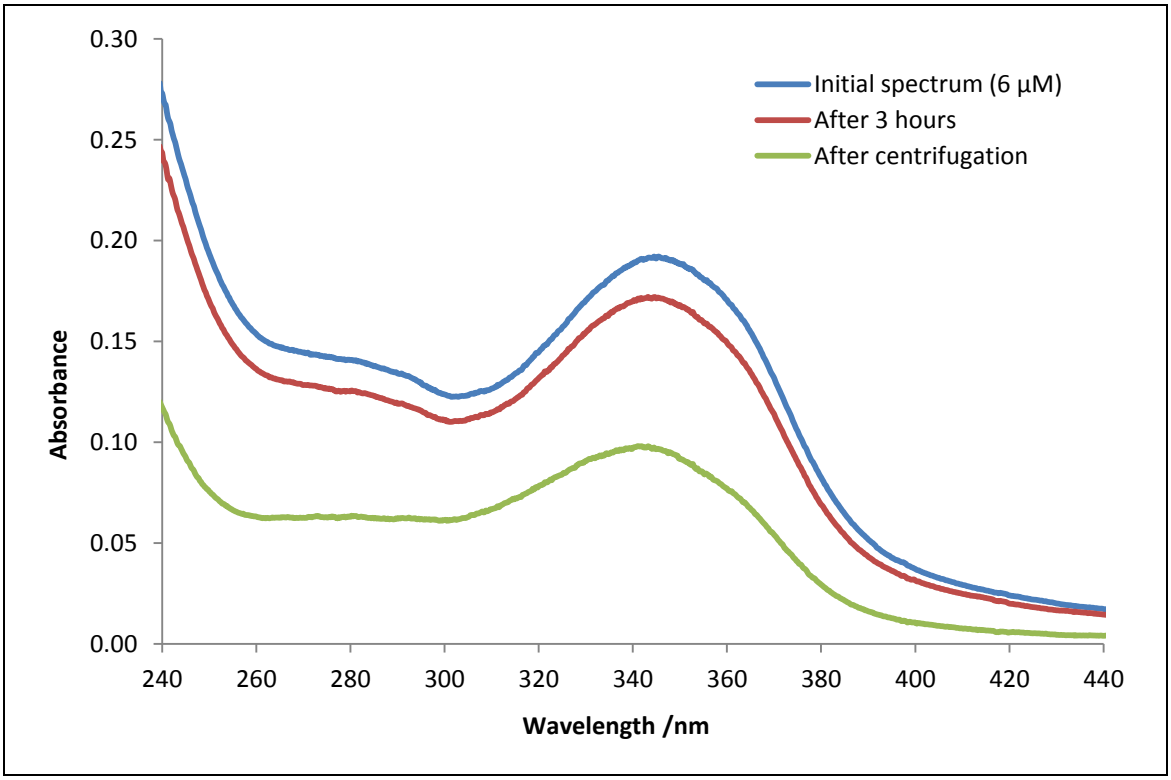


Figure 4.15: Testing the solubility of PPY-2-75: The absorption spectrum of an aqueous PPY-2-75 solution before and after centrifugation. The initial spectrum is shown in blue. The spectrum after the sample had been left for three hours is shown in red. The sample was then immediately centrifuged at 16200 × g for 1 minute before a final spectrum, shown in blue, was taken.

4.4.3.3. Inhibitor 2: AB062

A modified version of PPY-2-75 (AB062 supplied by Anna Barnard, University of Leeds), which differs from PPY-2-75 due to the presence of a short polyethylene glycol chain attached to the scaffold (Figure 4.13, p263), was subsequently tested and this appeared to be soluble in the NMR buffer (p243) (Figure 4.16 below). However, titration of AB062 into Mdm2 still did not cause significant HSQC peak movement (Figure 4.17).

Despite its solubility, a 1D proton NMR spectrum of AB062 could not be obtained suggesting that it existed in a slowly-tumbling, soluble but aggregated form when in the NMR buffer. In case adding detergent could disrupt this aggregated form, the detergent CHAPS (3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate) was titrated into Mdm2 in preparation for the addition of AB062. However, significant peak movement was observed when CHAPS was added (results not shown) so this idea was abandoned.

AB062 is an amphipathic molecule; it has a polar side and non-polar side. Aggregation of AB062 molecules could be energetically favourable if this hides their hydrophobic side from the solvent.

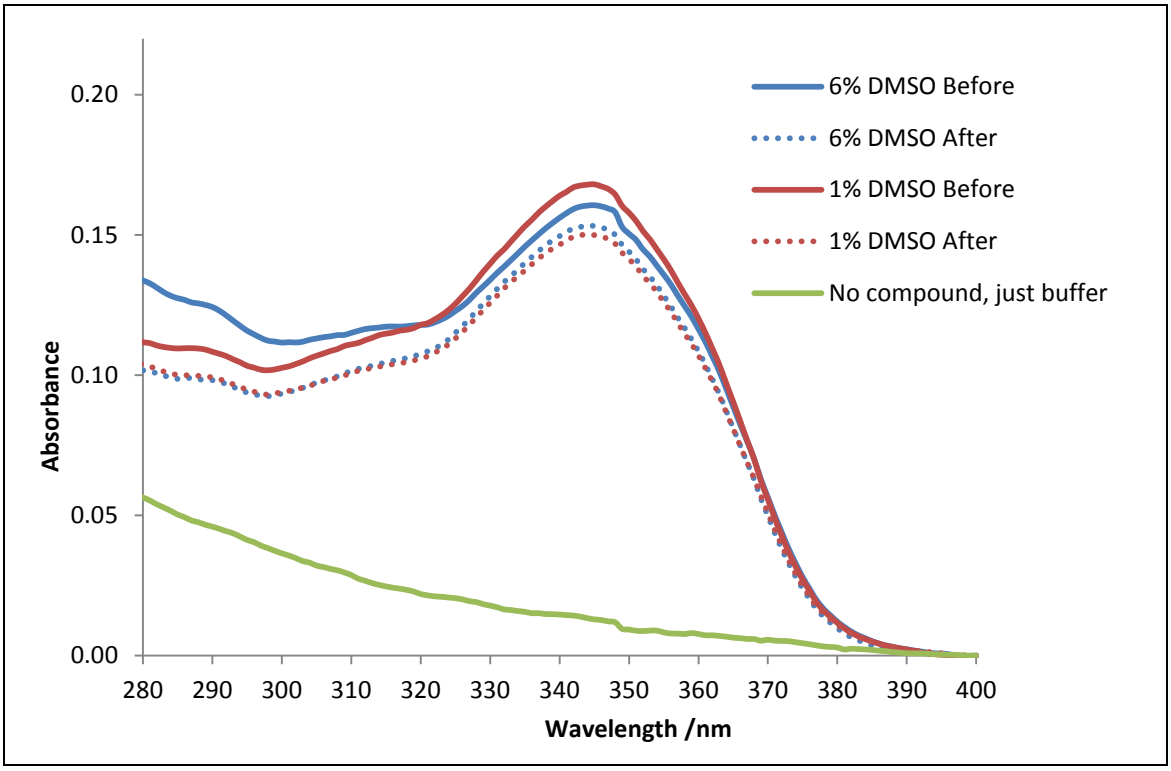


Figure 4.16: Testing the solubility of AB062: The absorption spectrum of an aqueous AB062 solution before and after centrifugation. Attempts were made to dissolve the compound in the NMR buffer (p243) with 6% or with 1% DMSO (just the DMSO of the solution containing the compound in the latter case). Samples from each solution and a buffer blank were diluted with DMSO and then UV spectra were taken (the solid lines). After centrifugation of the 6% and 1% DMSO solutions at 16200 × g for 1 minute, samples were taken again, diluted with DMSO and scanned to produce two more spectra (the dotted lines). AB062 appears to be soluble in the assay buffer because centrifugation did not appear to cause a large drop in the absorbance of the AB062 solutions.

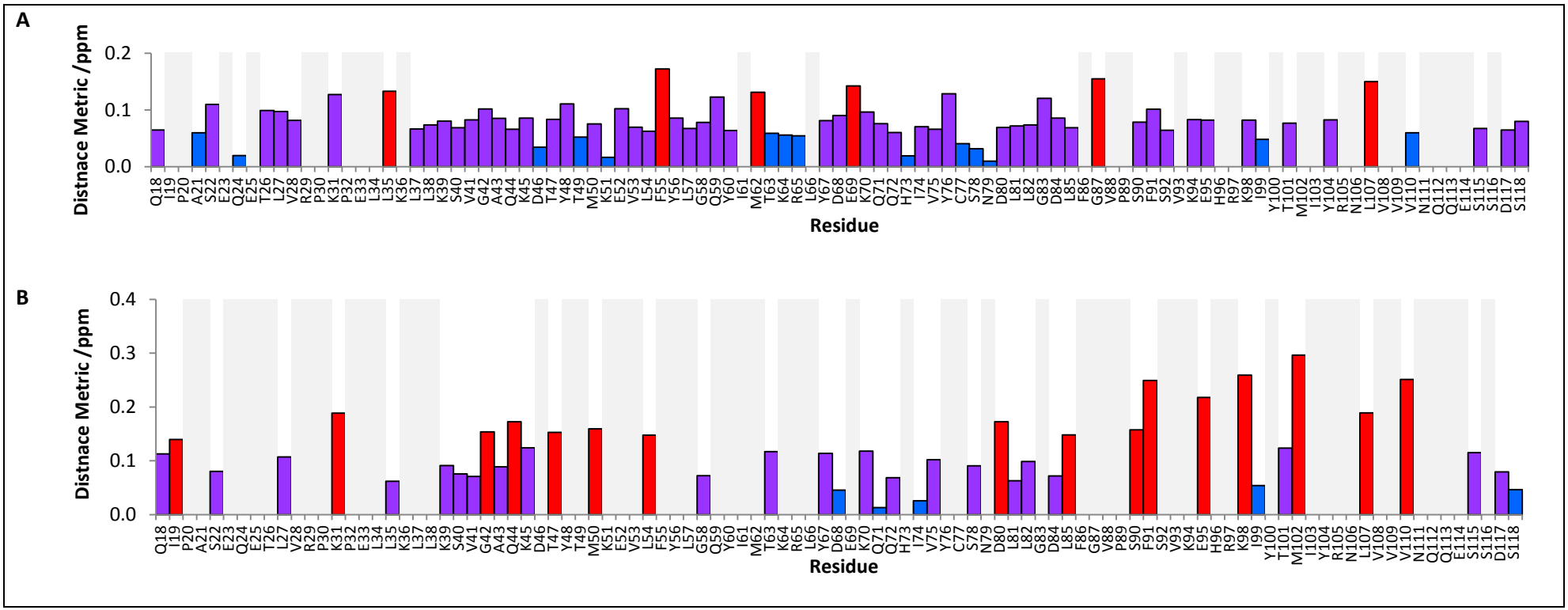


Figure 4.17: Distance metrics for the movement of each Mdm2 L33E alpha-amino group HSQC spectrum peak upon addition of AB062 to a concentration of 4 μM (A) and 100 μM (B). No detergent was used. The distances are relative to the spectrum before any compound had been added. Bars for residues with a shift distance metric exceeding 0.13 ppm are shown in red. Those for residues with a shift metric greater than 0.06 ppm but less than 0.13 ppm are in purple and bars for residues with a shift metric of less than 0.06 ppm are coloured blue. Light grey regions indicate where residues could not be assigned to a peak on either of the two spectra being compared.

4.4.3.4. Inhibitor 3: VA3-192

The next attempt to make a more soluble molecule entailed modification of the backbone itself. VA3-192 (Figure 4.13b, supplied by Valeria Azzarito, University of Leeds) lacks the central benzene ring of an oligobenzamide, a single carbon taking its place, as in a usual amino acid residue. A 1D spectrum was observed in the NMR buffer and titration of the compound into the protein caused significant peak movement (Figure 4.18). Figure 4.19 (p270) shows HSQC spectra before and after the addition overlaid and Figure 4.20 (p271) shows more clearly the movement of some selected peaks. Spectra taken after each addition facilitated identification of peaks in the final spectrum where they had moved a significant distance.

In a study published in 2006[118], an ortho-substituted oligobenzamide dimer (with carbon-linked side chains) was titrated into wild-type Mdm2 in a similar ^{15}N-^{1}H HSQC experiment. The most mobile peaks in this study were those produced by residues His73 and Phe55. In the VA3-192 titration described here, the three most mobile peaks were, in order of decreasing distance metric were those of, Phe55, Leu81 and His73. Furthermore, the fourth most mobile peak in the VA3-192 titration was Ser22 and, in the previous research, when p53 peptide was titrated into the protein solution, the fourth most mobile peak was also Ser22. Consequently, the results of this work are consistent with previously published results.

As shown in Figure 4.21 (p272), the residues with moving peaks are close to the p53 binding site, with the exception of Leu81 which is on the opposite side of the protein. As indicated at the bottom of the figure, interaction of positively charged Lys64 with negatively charged Asp80 might facilitate the transfer of any conformational change in the helix containing Gln59 and Phe55 into the helix holding Leu81 when VA3-192 binds. The NMR results are therefore consistent with VA3-192 binding in the p53 binding pocket.

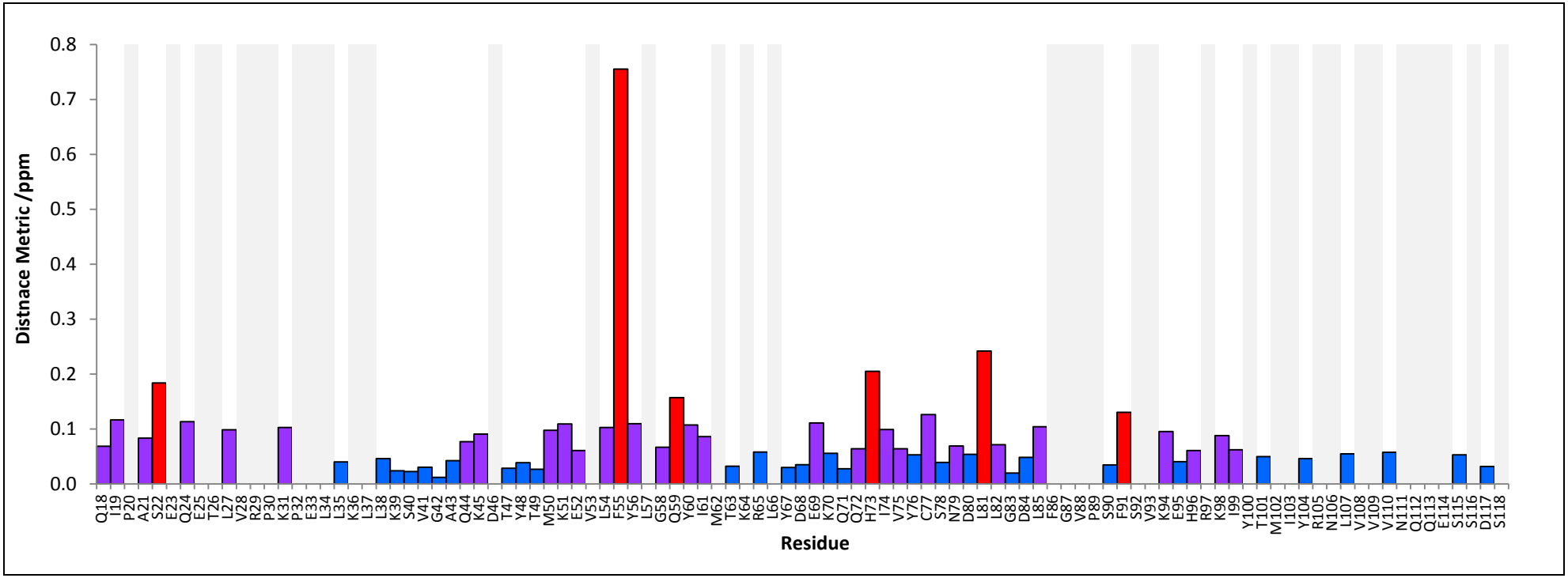


Figure 4.18: Distance metrics for the movement of each Mdm2 L33E alpha amino group HSQC spectrum peak upon addition of VA3-192 to a concentration of 200 μM. The distances are relative to the spectrum before any compound had been added. Bars for residues with a shift distance metric exceeding 0.13 ppm are shown in red. Those for residues with a shift metric greater than 0.06 ppm but less than 0.13 ppm are in purple and bars for residues with a shift metric of less than 0.06 ppm are coloured blue. Light grey regions indicate where residues could not be assigned to a peak on either of the spectra being compared.

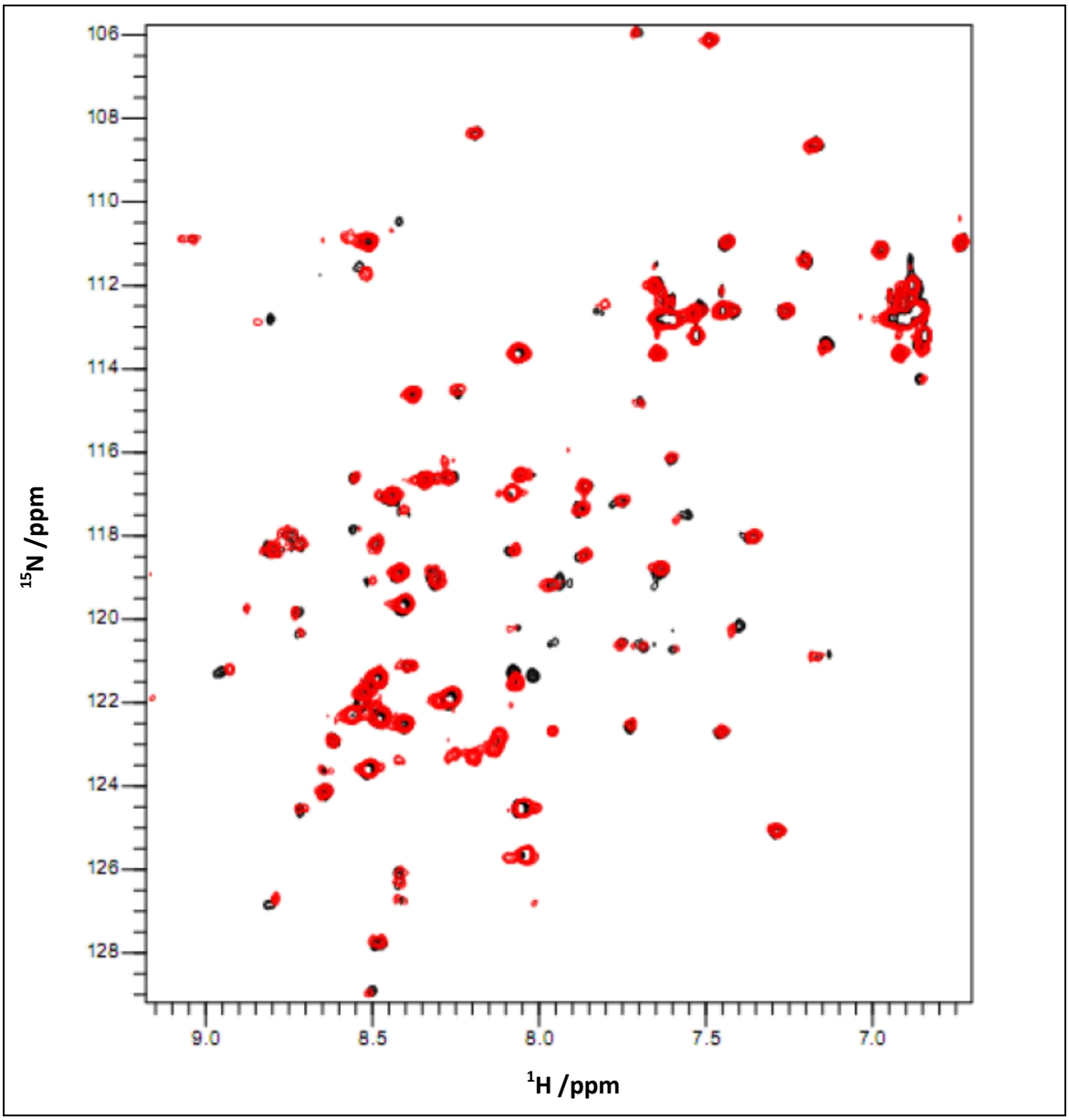


Figure 4.19: The HSQC spectrum of ^{15}N-Mdm2 before and after addition of VA3-192 to a concentration of 200 µM. The initial spectrum is shown in black and the final spectrum is shown in red.

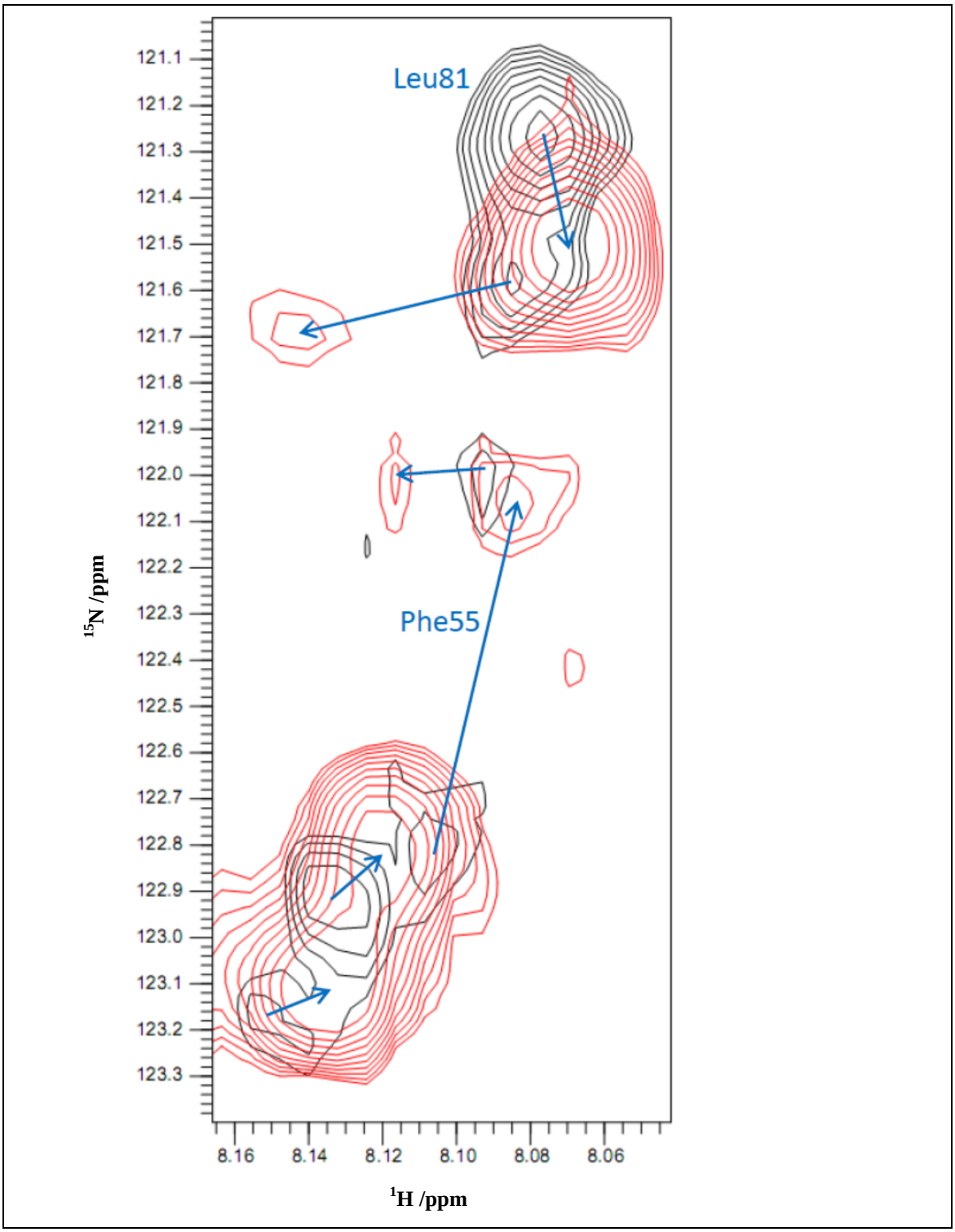


Figure 4.20: The HSQC spectrum around the Phe55 peak before and after addition of VA3-192 to a concentration of 200 µM. The initial spectrum is shown in black and the final spectrum is shown in red.

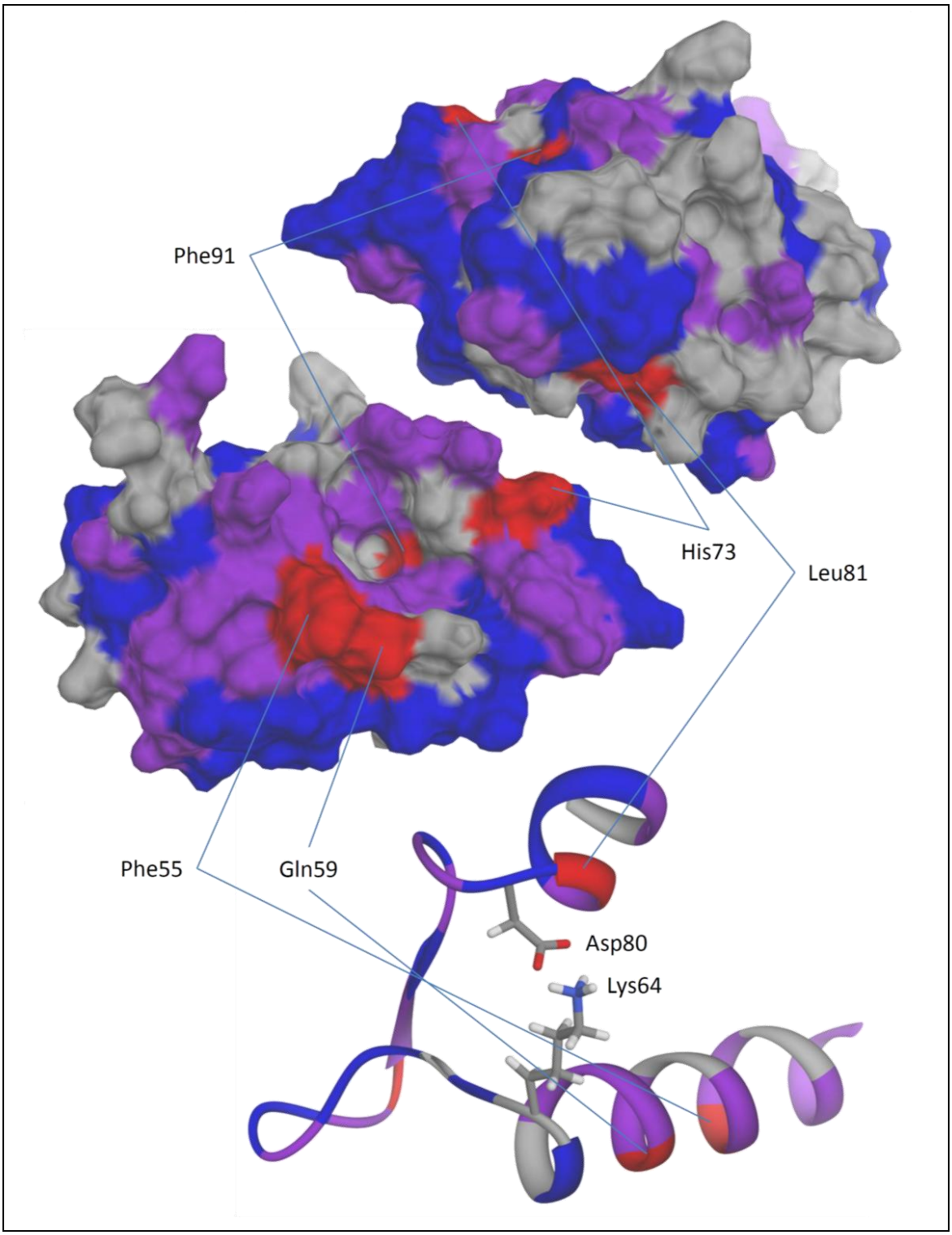


Figure 4.21: Diagram showing the position of Mdm2residues with peaks moving in response to addition of VA3-192 to a concentration of 200 µM. Mdm2 from structure 1T4F is shown in space-filling form from both the back (top right) and front (middle). Phe91 is at the back of the p53 binding site. Residues for shifts with a distance metric exceeding 0.13 ppm are shown in red. Those with a shift metric greater than 0.06 ppm but less than 0.13 ppm are in purple and residues with a shift metric of less than 0.06 are coloured blue. Residues with no assigned peak in either the spectrum before or the spectrum after compound addition are shown in grey. At the bottom is a ribbon representation of part of Mdm2 which shows how movements in the helix holding Phe55 and Gln59 might be propagated through Lys64 and Asp80 to Leu81, a residue some distance from the p53-binding site.

Docking of VA3-192 in AutoDock Vina places the compound in close proximity to those residues seen to move in the NMR spectra; in the top two poses (Figure 4.22, A and B) the N-terminal end of the helix is close to Phe55 and the negatively charged C-terminal end is close to His73, which could have a positive charge. These top poses show VA3-192 acting as a peptidomimetic (Figure 4.23, p275). Although the alignment is antiparallel to the position of p53-helix in structure 1T4F, the first side chain of the compound is in the Leu26 binding pocket and the middle side chain is in the Leu23 binding site. The third side chain is close to the Phe19 pocket but the phenylalanine ring appears to be being mimicked by the C-terminal benzene ring of the scaffold rather than the isopropyl side chain. Consequently, the NMR results are not just consistent with VA3-192 binding in the p53 binding groove but also with it acting as a peptidomimetic. While VA3-192 is not itself an oligobenzamide, these results suggest that oligobenzamides are also likely to bind in such a way to Mdm2 once they have been sufficiently modified to increase their solubility and decrease their propensity for aggregation.

Although spectra of VA3-192 and AB062 in aqueous solution could not be obtained and the location of the mobile peaks in the titrations of these compounds were not centred on the p53 binding site, the most mobile peak when PPY-2-75 was added was that of Phe55 (Figure 4.14A) and the Phe55 peak was also the most mobile upon addition of AB062 to a concentration of 4 μM (Figure 4.17A). Consequently, some binding could have been occuring in the p53-binding site.

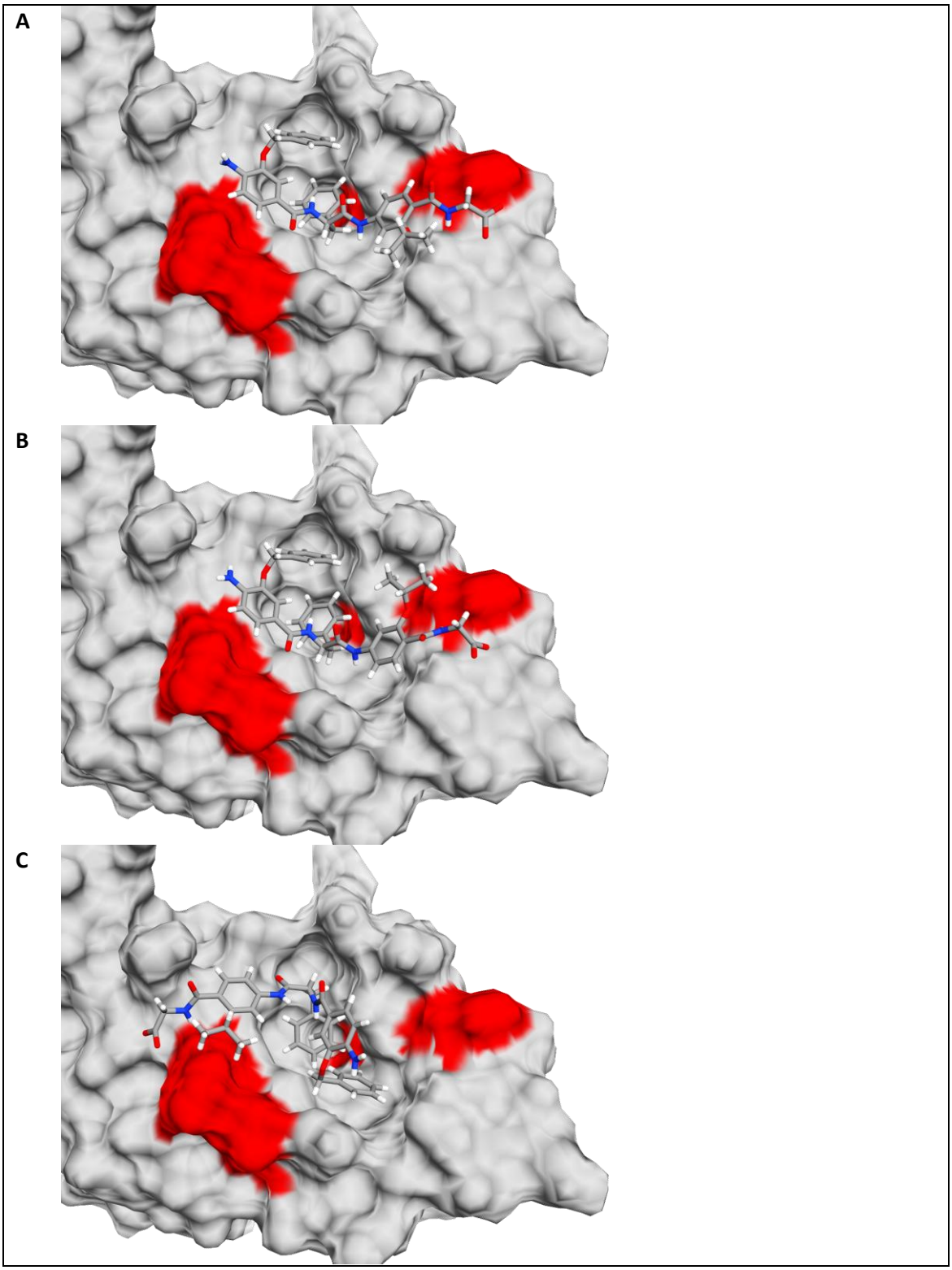


Figure 4.22: Computer generated images showing the three highest-scoring AutoDock Vina docking poses for compound VA3-192 and how they are close to the residues seen to move most in the NMR HSQC spectrum upon addition of the compound. The two highest scoring poses, (A and B) are antiparallel to the p53-helix orientation whereas the third highest-scoring pose (C) is arranged parallel to it. The residues seen to move most in the NMR HSQC spectrum are highlighted in red. In the top two poses the compound is seen to be close to all of the highlighted residues around the binding site. In the highest scoring pose (A) the first two side chains are in the binding pockets of p53 residues Phe19 and Trp23 and the third side chain is close to the Leu26 pocket, consistent with the compound acting as a peptidomimetic.

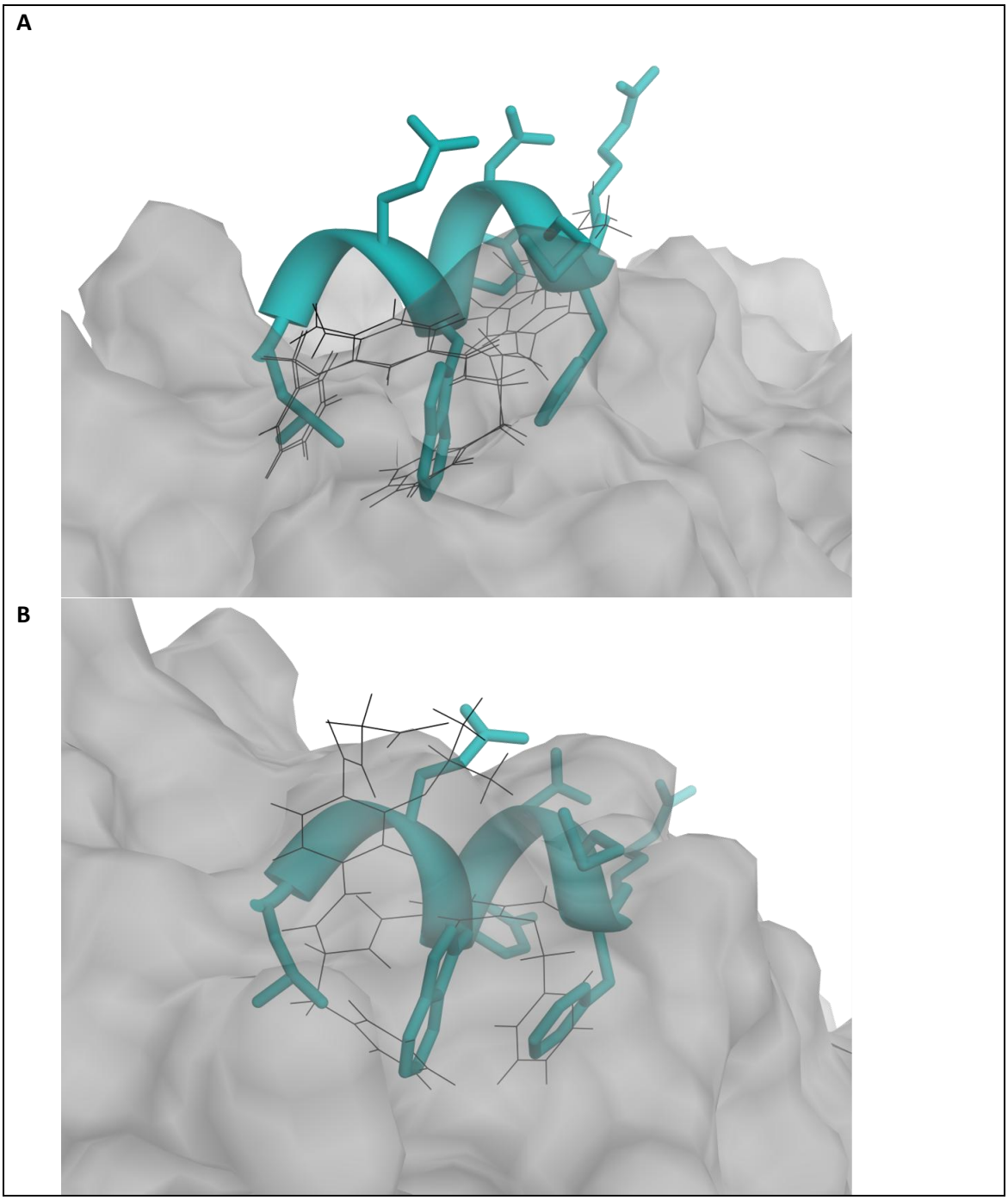


Figure 4.23: Computer generated images that show the extent to which the side chains of VA3-192 mimic, in the highest scoring poses predicted by Autodock Vina, those of the bound p53 helix. The same poses are shown in Figure 4.22. A) The two highest scoring poses. B) The third highest scoring pose. The protein is shown in grey and semi-transparent so that the compound (black) and the p53-helix (cyan) side chains can be seen in the binding pockets.

4.5. Conclusions and Further Work

4.5.1. EXPRESSION, PURIFICATION AND ASSIGNMENT OF ^{15}N-MDM2 L33E

In this chapter, the successful expression and purification of ^{15}N-labelled Mdm2 with the L33E mutation has been described. Purification was straightforward and yielded adequate amounts of protein. The solubility of the expressed protein was sufficient for it to be concentrated sufficiently for use in 3D NMR experiments. Further work is required to determine the optimal conditions for effective and reliable production.

There are no published amino acid peak assignments for Mdm2 L33E. There appear to be moderately large differences between the chemical shifts of the peaks for the mutant and the published chemical shifts for the wild-type protein, particularly for peaks corresponding to residues around the position of the mutation. Assignment was facilitated by NOESY-HSQC and TOCSY-HSQC spectra although the high signal to noise ratio made assignment challenging. The assignments determined should be useful in future work, especially if a higher protein concentration could be achieved to make peak identification easier. Double labelling with both ^{15}N and ^{13}C would allow CBCA(CO)NH and CBCANNH spectra[295] to be obtained. These 3D spectra could potentially facilitate full assignment without recourse to previously published resonances, allowing every residue to be identified.

4.5.2. SOLUBLE OLIGOBENZAMIDES MIGHT AGGREGATE WHEN THEY DISSOLVE

During the NMR work, evidence emerged of significant changes to the structure of Mdm2 L33E at higher concentrations of DMSO. These changes occurred at concentrations below 10% DMSO, the concentration used in the FRET assays described earlier in this thesis. However, testing of the oligobenzamide PPY-2-75 suggested that a significant DMSO concentration is needed for this compound to dissolve, especially to produce free monomers. Thus, there may be no concentration of DMSO which will dissolve unmodified oligobenzamides while maintaining Mdm2 in its native conformation. If this is the case, the molecular dynamics simulations carried out prior to these experiments may have been modelling situations which cannot actually occur *in vitro*. Further work to produce more soluble compounds should resolve this issue.

The NMR work carried out indicates that addition of a soluble chain to the scaffold, creating an amphipathic molecule, might not be the best way to increase oligobenzamide solubility. AB062, the oligobenzamide with a polyethylene glycol (PEG) chain attached to the scaffold core appears to be soluble at lower DMSO concentrations. However, the absence of a spectrum when it dissolves suggests that this compound forms a large structure which tumbles slowly. Large complexes do not show up on an NMR spectrum because they do not rotate fast enough for

time-averaging of the external magnetic field over all orientations of the structure. Instances of the complex in different orientations have different chemical shifts creating a very broad peak which is lost in the noise.

There was no evidence of AB062 affecting the positions of the peaks in the HSQC spectrum of the protein. The intensity of peaks on the spectrum decreased with successive additions. This would be consistent with the protein being bound and incorporated into a large, slow-moving structure with a very broad peak. This could leave only the unbound protein visible in the HSQC spectrum. However, an alternative explanation is that, because it is amphipathic, this oligobenzamide causes denaturation and subsequent precipitation of the protein. In either case, it seems that though the attachment of a soluble group to an oligobenzamide might render the compound more soluble, it could also create an amphipathic molecule with adverse properties. Further work involving size exclusion chromatography of Mdm2 in the presence of oligobenzamide might reveal the size of any soluble aggregate formed. This might suggest how AB062 could be modified to prevent aggregation, for example, by the addition of bulky groups to create steric hindrance or charged groups, such as amino or carboxylic acid groups, to cause the repulsion of adjacent molecules.

Both the FRET experiments and computational methods discussed earlier in this project assume a simple equilibrium between the free protein and compound in solution and the complex. The measured binding constant in FRET experiments will reflect any other, complicating phenomena such as binding of the ligand to the displaced p53 peptide or precipitation of the compound. The computational results will not however, because such factors are not included in the computational model. Such effects could therefore be responsible for the poor correlation between predicted and experimental results encountered with oligobenzamides in chapter 2 and the NMR experiments in this chapter with the oligobenzamides PPY-2-75 and PPY-2-95 support this hypothesis in that they suggest that little of the unbound ligand is actually free in the solution.

4.5.3. VA3-192 RESEMBLES AN OLIGOBENZAMIDE AND APPEARS TO ACT AS A PEPTIDOMIMETIC

The compound VA3-192 resembles an oligobenzamide but for the fact that the central benzene ring has been replaced by a methyl group. This molecule appears to bind to Mdm2 L33E because the HSQC spectrum peaks corresponding to residues around the p53 binding site move when this compound is added to a solution of the protein in the spectrometer and the most mobile peaks are consistent with previously published work using an oligobenzamide dimer with carbon-linked side chains[118].

The inference from this evidence is that if precipitation and aggregation can be prevented, oligobenzamides should bind in this position. The HSQC peak movements observed specify the

residues around which the environment changes when VA3-192 binds but not whether these residues are in close contact with the compound. Titration of p53 into Mdm2 L33E would indicate which residues normally move when the p53-binding pocket is used and confirm whether or not VA3-192 attaches to the p53 binding groove of Mdm2.

4.5.4. ADDITIONAL FURTHER WORK

HSQC peak movement during repeated addition of a compound could be used to produce a titration curve from which the binding constant could be determined[296]. The RMSD of peaks from their position prior to compound addition could be plotted on the y-axis.

Saturation transfer difference NMR can reveal which parts of a small molecule bind to a protein[297]. The protein is saturated, meaning that its nuclei are excited to excess so that the populations of nuclei in each spin quantum state become equal and no further net absorption by the system is possible. Magnetisation spreads to the ligand when it transiently contacts the protein, causing the nuclei of ligand atoms that interact with the protein to also become saturated. The spectrum of the bound ligand is subtracted from a reference spectrum of the unbound ligand to get a spectrum where the peaks made by these interacting ligand atoms have an increased intensity. Estevao *et al.*[298] used saturation transfer to study indole based cyclooxygenase inhibitors of COX-2. They saw that the dealkylation at two adjacent positions on a compound was sufficient to change the binding pose significantly, moving a sulphonamide group out of a key pocket which confers selectivity.

5. Final Conclusions

In many human tumours, aberrant down-regulation of p53 activity occurs due to overexpression of Mdm2. An effective inhibitor of the p53-Mdm2 interaction could potentially be developed into a pharmaceutical, in addition to being a useful laboratory tool. The original aim of this project was to demonstrate that computational methods could be used to predict the relative binding affinities of oligobenzamides for the p53 binding site of Mdm2. The use of computer predictions to streamline identification of possible inhibitors could make the process of drug development faster, more efficient and cheaper.

To some extent, this primary aim was achieved. Binding energies predicted for oligobenzamides binding to Mdm2 showed a moderate correlation with experimental results when implicit solvent methods were used. Spearman's rank correlation coefficients typically between 0.5 and 0.6 were obtained. This finding suggests that such methods may be valuable for making future predictions.

Despite its statistical significance, the correlation between computational predictions and experimental data was not particularly strong, reinforcing conclusions from earlier work which show that it is challenging to make reliable *in silico* predictions [299]. Consequently, the scope of the project expanded to address the question of why it is difficult to establish correlations, and, by investigating the causes, provide insight into how the underlying problems might be solved.

It became clear that the problem is two-fold; predicted affinities must be accurate but laboratory measurements must be equally so. Thus, the final project involved a mixture of computational and laboratory work.

5.1. Computational Work

In this project, two computational approaches to predicting relative binding affinity were assessed, thermodynamic integration and implicit solvent methods.

5.1.1. THERMODYNAMIC INTEGRATION

The relative affinities of 31 oligobenzamides were predicted using TI. This method estimates the effect of small side chain modifications on the binding energy of a ligand. A prerequisite is that the two ligand structures involved bind in roughly the same place. Preliminary docking results were consistent with the proposal that all oligobenzamides do indeed bind to Mdm2 in the same position, namely the p53 binding site, with their side chains mimicking the key p53 residues Phe19, Trp23 and Leu26. Nonetheless, there was a poor correlation between the TI predictions and the experimentally determined binding affinities.

Subsequent analysis of more extensive docking results with approximately 25,000 oligobenzamides revealed a possible explanation for the limited success of TI: oligobenzamides can actually bind in many different positions in and around the p53 binding site. Modification of the side chain combination appears to cause changes in molecular orientation and frame, straddling of binding pockets in some cases and, sometimes if the side chains are small, binding of the scaffold to Mdm2 as opposed to the side chains. If compounds bind in completely different poses, the contribution of functional groups to the binding energy of each side chain may not be additive and the influence of each side chain on affinity is unlikely to be independent of the other side chains present.

5.1.2. IMPLICIT SOLVENT METHODS

Using implicit solvent methods, a much stronger correlation between the predicted and laboratory-determined affinities of different oligobenzamides was observed. The MM-PBSA and MM-GBSA methods provide a useful compromise between speed and accuracy[129] and their ability to yield an estimated binding energy from a single trajectory proved to be particularly useful in this project. Simulations were kept short to allow a large combinatorial library of compounds to be tested with the limited computing power available.

To investigate whether the correlation between predicted and experimental results could be further improved, the MM-PBSA and MM-GBSA methods were tested with different settings to assess their robustness. The effect of changes was typically small, for example, variation in ionic strength had little effect. However, the success of the MM-PBSA method was dependent on the use of a low dielectric constant for the inside of the protein.

Further improvement can sometimes be achieved by using Normal mode analysis to estimate the changes in entropy when the protein and small molecule bind. However, this analysis is resource intensive and the results obtained in this project support earlier findings[249] that factoring in the entropy component in this way typically does not significantly improve the accuracy of predictions. Similarly, simulation of the ligand in isolation, as opposed to using an approach based entirely on the trajectory of the complex, did not appear to improve the predictions, again consistent with earlier studies[265].

5.1.3. REFINING COMPUTATIONAL PREDICTION METHODS

Analysis of variance (ANOVA) using the MM-PBSA results, treating the side chains as factors, confirms that there is a significant statistical interaction between the side chains of an oligobenzamide. This means that one side chain affects the affinity of those adjacent to it. The interaction is manifest in many of the components of the binding energy.

There are two possible causes of this interaction. Firstly, a change in side chain could influence where binding occurs, changing the part of Mdm2 to which the side chains bind and therefore their affinity. Secondly, rather than changing the binding site, a modification could control how the other side chains bind within a particular binding position.

The extensive docking results obtained in this project support the first option because they show that modification of a side chain affects where binding occurs, and by implication, the side chain affinity. Treatment of the pose as a factor in the MM-PBSA prediction ANOVA analysis confirms that the binding pose affects the affinity of the side chains.

In support of the second effect, an interaction between the side chains remains when poses all associated with the same binding location are compared. However, the magnitude of the interactions in this case is smaller and, depending on the circumstances, could be ignored.

To perform the analyses described herein, simulations of oligobenzamides with many different side chain combinations had to be performed. This was computationally time consuming and expensive. The small degree of interaction between the side chains of molecules within each binding position suggests that, to compare compounds binding within a particular location, it would be valid to predict the effect of each side chain individually, saving simulation time. In contrast to the difficulty of binding affinity prediction, docking can often find the correct pose of a ligand reasonably quickly and accurately[161,162]. Docking could thus be used to select, from a range of positions for which structure-activity relationships had been identified, the right one for predicting the affinity of a new ligand. Identification of quantitative relationships[134] at each site could allow ligands binding at different sites to be compared to produce an overall ranking.

The accuracy of predicted binding affinities would be improved by the use of longer simulation times. This would be facilitated by the independent treatment of each specific side chain in

every possible binding site because this would reduce the total number of simulations required and thus allow more time to be spent on each one. Oligobenzamides are flexible enough to bind in many different conformations but their rigidity is sufficient to extend the time needed for complete sampling of the conformational space to tens of nanoseconds[172]. The adjusted R-squared values of many of the ANOVA models in this project suggest that much of the variation in the data is not accounted for by the model in question. Consequently, there is room for improvement, which might be achieved if the accuracy of the model parameters is increased by use of longer simulations or more repeats.

The number of simulations required could also be reduced through use of a new method known as lambda dynamics[272]. This method could allow a single simulation to encompass both testing at multiple side chain attachment positions and assessment of numerous side chains at each position. Although very long simulations might be needed, the nature of this method means that simulation time would be spent testing the most potentially useful side chains. Furthermore, less simulation time would be spent on the equilibration of a large number of systems.

5.1.4. WHICH OLIGOBENZAMIDES PROVIDE THE GREATEST POTENTIAL FOR FUTURE RESEARCH?

The FlexX[149] docking scores of an oligobenzamide library were used to investigate the specific properties likely to increase the affinity of ligands which bind in the expected, idealised pose occupying the Phe19, Trp23 and Leu26 binding sites of Mdm2. The results indicated that the first side chain should be small, the second large and the third flexible, in accordance with the side chains of the p53 residues they mimic.

Of the three oligobenzamide scaffolds investigated, the sulphur-linked scaffolds could provide the greatest scope for future work. TI results suggest that the sulphur link atoms make more of a contribution to the binding energy of oligobenzamides based on the ortho-substituted sulphur scaffold than the corresponding oxygen atoms make to the binding of oligobenzamides with oxygen-linked side chains. This could be because the strength of the intramolecular hydrogen bonding in the latter prevents the molecule from achieving the optimum conformation for binding. While compounds based on the nitrogen-linked scaffold might bind strongly, the side chains in these compounds are attached to the amide nitrogen atoms so the parameters of the scaffold could be affected by the nature of the side chains present. Rigorous investigation with different scaffold torsion parameters for different side chains would not be straightforward and might yield results that were difficult to compare.

The sulphur-containing scaffolds discussed in this work have received little attention in the laboratory. They should be investigated further.

Whichever scaffolds are the focus of future research, evidence should be obtained to confirm that the parameters used generate predictions that reliably agree with experimentally determined affinities. Such validation relies on accurate experimental data from laboratory testing.

5.2. Laboratory Work

There are two stages to the computational prediction of a ligand's binding affinity. The first is the determination of where and how the ligand binds. The second is the estimation of the strength with which the ligand will bind in the site identified. In the experimental part of this project, each of these stages was considered, with a view to the development of methods for the validation of computational predictions.

5.2.1. NMR: A METHOD OF DETERMINING WHERE BINDING OCCURS

Docking results suggest that oligobenzamides could bind to the p53 binding site of Mdm2 in various different ways. Correct establishment of the precise manner in which a particular oligobenzamide binds is essential for the prediction of accurate affinities for the molecule itself or similar oligobenzamides. A laboratory method to identify where oligobenzamides bind would consequently be useful.

Unfortunately, it was not possible to perform NMR experiments with the oligobenzamides available. They either precipitated or were soluble but could not be detected by NMR, probably due to their aggregation. Even in the presence of 10% DMSO, a spectrum of them in aqueous solution could not be obtained. Given that the FRET assays used to test oligobenzamides involve aqueous buffer, this evidence of more complex interactions is of relevance to the interpretation of FRET results. A simple equilibrium is assumed in the FRET assays. If not all of the compound is in solution then the K_d for binding will be overestimated.

Reliable availability of oligobenzamides soluble at a DMSO concentration in which the protein is stable would greatly facilitate future work. To this end, work to improve compound solubility has continued within the collaborating Chemistry group. A possible method being investigated is the addition of soluble groups to the backbone of the scaffold. The NMR results suggest that the propensity for compound aggregation must also be minimised. Amphipathic molecules with polar side chains to increase solubility on one side and hydrophobic side chains to bind to Mdm2 on the other might have a tendency to interact with each other.

In light of the solubility problems associated with the oligobenzamides, NMR was used to study the interaction of a compound resembling an oligobenzamide, VA3-192 (Figure 4.13B, p263) with Mdm2. The location of the residues which appeared to be perturbed suggested that VA3-

192 bound in a manner consistent with the two highest scoring docking poses generated by Autodock Vina[152]. In these poses, the ligand is acting as a peptidomimetic. Due to their similarity, this is evidence that oligobenzamides also have the potential to act as peptidomimetics if the challenges of solubility and aggregation can be overcome.

Crystals containing Mdm2 and VA3-192 or, ideally, an oligobenzamide could usefully be produced for X-ray diffraction. X-ray structures might provide more conclusive and detailed information about how oligobenzamides really bind and such an X-ray structure would be an excellent starting point for future computational work.

The L33E form of Mdm2 was used for NMR because this mutant is stabilised in the open conformation. Upon compound binding, only the resonances of residues near to the site of binding are likely to change so those that do reveal the binding position. Because the amide nitrogen and amide proton resonances of L33E Mdm2 differ significantly from those of the wild type protein, their assignment was necessary. These assignments will be valuable for future work.

5.2.2. FRET: A METHOD OF DETERMINING BINDING AFFINITY

This project involved the development of a Förster resonance energy transfer (FRET) assay to test the affinity of p53-Mdm2 interaction inhibitors. This assay is orthogonal to the fluorescence polarisation assay currently in use[42]. It involves titration of the potential inhibitor into a solution of fluorescently labelled Mdm2 and fluorescently labelled p53 peptide. Before the titration, a sizeable proportion of the p53 molecules are bound to Mdm2 molecules causing FRET to occur between their fluorescent labels. During the titration, binding of an inhibitor to the Mdm2 displaces any p53 peptide. This is observed as a decrease in FRET.

Three possible labelled p53 peptides were tested, wild type p53-eosin, wild type p53-mCherry and a mutant p53-mCherry. Titration results show that the known inhibitor Nutlin-3 displaces all three of these proteins from GFP-Mdm2. This indicates that the fact these proteins are labelled does not prevent them from binding in the expected position. Further evidence of correct binding comes from comparison of the binding affinities of the Cherry-labelled mutant peptide and labelled wild type peptide with those of the two unlabelled peptides. In both cases, the mutant binds more strongly. The binding strength of the mutant peptide construct could be particularly useful for testing the efficacy of high affinity inhibitors.

A useful characteristic of the new FRET assay developed using p53-eosin is that it works in the presence of 10% DMSO. This allows it to be used with less soluble inhibitors and where a potential inhibitor has a significant absorbance. Furthermore, its adaptation for use on a microwell plate brings high throughput characterisation of accurate oligobenzamide affinities a step closer.

Although the FRET method is a useful orthogonal assay, the fluorescence polarisation assay has distinct advantages. p53-eosin binds significantly more strongly to unlabelled Mdm2 (measured by fluorescence polarisation) than to GFP-Mdm2 (measured by FRET) suggesting that the GFP label in the FRET assay hinders access to the Mdm2 binding pocket. Furthermore, the high K_d means that high concentrations of GFP-Mdm2 and p53 peptide are required to form enough of the complex for compound testing. Comparably high concentrations of each potential inhibitor are needed to compete with the peptide. In contrast, in the fluorescence polarisation assay, far less labelled peptide is required. Because concentrations of peptide and (unlabelled) Mdm2 do not need to be equivalent, the equilibrium can be shifted using a high concentration of unlabelled Mdm2 alone. Consequently, less inhibitor is needed. High IC_{50} values are a problem in the FRET assay when using oligobenzamides because of their low solubility and colour. In contrast, in the fluorescence polarisation assay, the results are unaffected by the absorbance of fluorescence by the inhibitor. Consequently, the FRET assay is not a good substitute for the fluorescence polarisation assay when testing compounds *in vitro*. Rather, it is a useful, complimentary method for use in conjunction with the polarisation method, to support its findings.

In vivo, fluorescence polarisation cannot be used because there is no easy method of introducing the chemically labelled peptide into cells. In contrast, the p53-Cherry proteins could be co-expressed with GFP-Mdm2 in a cell so the FRET assay could potentially be developed into an assay for *in vivo* testing. In this respect, the FRET assay is a significant advance.

5.3. A combined approach

The complexity of oligobenzamide binding advocates closer integration of *in silico* and *in vitro* work. These complimentary approaches would ideally form parts of a cyclical strategy such as that shown in Figure 5.1 where computational predictions are assisted by reliable laboratory data and synthetic work is driven by the resulting predictions.

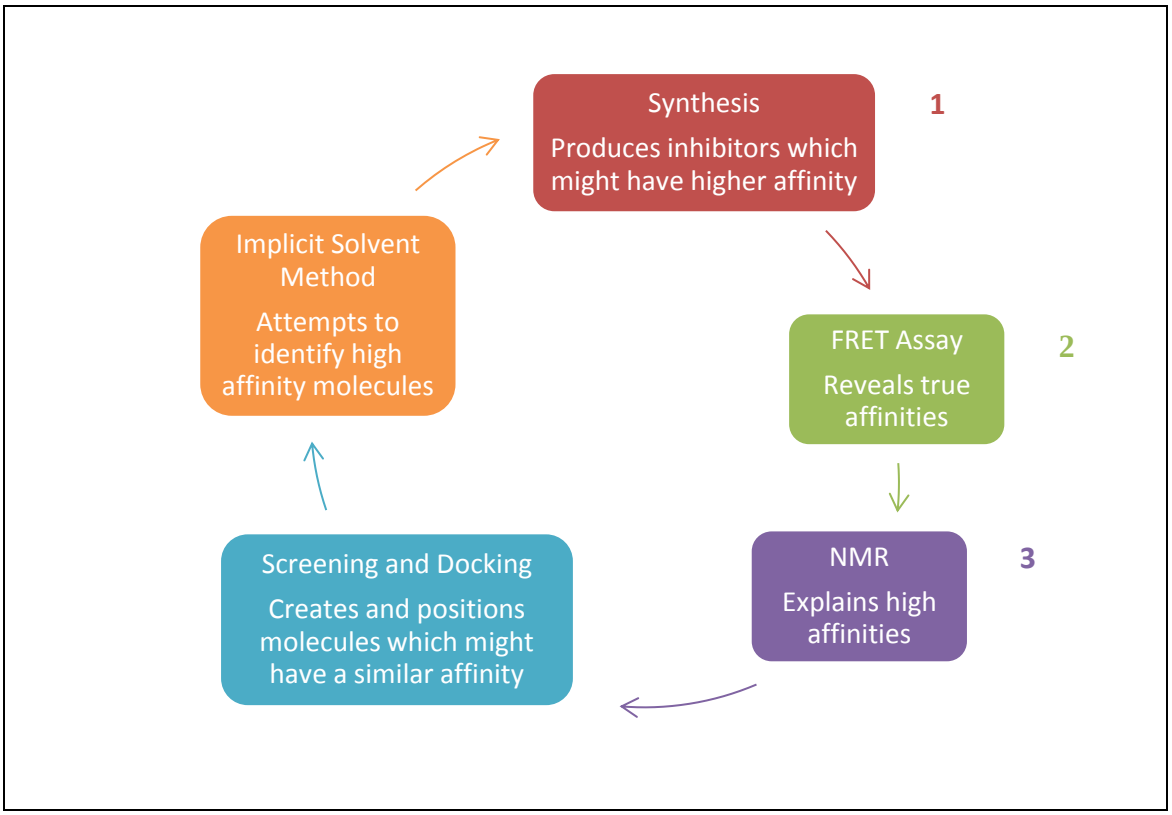


Figure 5.1: A cyclical strategy for the development of oligobenzamide inhibitors. Each box represents a method which could be used and describes what its contribution to the process would be. The numbers relate to the points in the text below, which describe how progress in each area has been advanced by the findings or developments of this project.

The unusual breadth of this project means that its findings and developments could be applicable to all stages of this broad, interdisciplinary problem, as shown in Figure 5.1:

1. This project highlighted which scaffolds it might be most useful to use in future synthetic work on the grounds of their ease of parameterisation and expected stronger binding. Work with FlexX demonstrated how the properties of side chains best suited to a particular binding pose might be identified. Specific properties pertaining to the Phe19, Trp23 and Leu26 binding sites were determined.
2. FRET-based assays were developed for *in vitro* analysis of compounds once they had been synthesised. A microwell plate assay was developed for high-throughput testing and p53-Cherry protein constructs were produced which could be used in the course of future *in vivo* work.
3. The HSQC peaks of L33E Mdm2 were assigned to its residues. These assignments could enable future testing to determine the binding position of inhibitors. Structural binding information can indicate how an oligobenzamide might best be modified to generate a new set of possible inhibitors from which some can be selected for synthesis using *in silico* methods.

4. The project demonstrated a method of computationally evaluating the broad range of binding poses obtained when docking a library of oligobenzamide peptidomimetics.
5. MM-PBSA and MM-GBSA methods were shown to be of use for determining affinities *in silico* from docking poses. Various settings were tested and useful approaches were identified. A method was also suggested which would reduce the number of simulations required while still taking into account the effect of side chain choice on binding pose. *In silico* prediction allows compounds which are unlikely to bind to be eliminated. It thus reduces the number of compounds which must be synthesised for further testing.

6. Appendices

6.1. Appendix A: Statistical Methods

6.1.1. REGRESSION

6.1.1.1. Measuring the Correlation between Continuous Variables

6.1.1.1.1. Pearson's Product Moment Correlation Coefficient

Pearson's Product Moment Correlation Coefficient[250] for a population, ρ, is a measure of the extent to which there is a linear correlation between two variables, each normally distributed (distributed as a Gaussian distribution). For a sample from a population, the coefficient is denoted by r. In this work coefficients were calculated using the PEARSON function in Microsoft Excel (2010):

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \qquad 6.1$$

Where the macron (line) above x and y indicates use of the arithmetic mean of that variable.

Where there is a positive linear correlation, r is 1. Where there is a perfect negative linear correlation, r is -1. (If one of the variables is constant, making a perfect linear correlation which is neither positive nor negative, r is 0.) Where there is no correlation, r has a symmetrical probability distribution centred on 0. Testing to see if a correlation is significant involves testing the null hypothesis that ρ=0.

6.1.1.1.2. Spearman's Coefficient of Rank Correlation

When one of two variables is not distributed according to a normal distribution or the relationship between the variables is not linear, Pearson's Product Moment Correlation Coefficient is not appropriate. Here, Spearman's Rank correlation coefficient[251], a non-parametric statistic, can be calculated. This is the Pearson's Product Moment Correlation

Coefficient of the data where the values have been replaced by their rank, their order if the dataset were to be sorted by the variable in question.

In this project, ranking was carried out in Microsoft Excel, using a VBA macro to ensure the correct treatment of ties if they were present.

6.1.1.1.3. Confidence limits

All confidence limits throughout this thesis are 95% confidence limits and are, unless otherwise stated, calculated using the standard error on x and the t distribution.

$$95\%\ \text{confidence limits} = \pm t_{n-1} \times SE_x = \pm \frac{t_{n-1} s_x}{\sqrt{n}} \quad \textbf{6.2}$$

Where SE_x is the standard error on x, s_x is the standard deviation of the sample, n is the number of values and t_{n-1} is Student's t for the probability 0.975 with n-1 degrees of freedom (which tends towards 1.96 as n increases). Standard deviations quoted in this project are the square root of the unbiased sample variance (the standard deviation obtained by dividing by N-1).

Correlation coefficients are not normally distributed. This complicates the calculation of their confidence limits. Fisher[300] derived a formula to transform correlation coefficients into a normally distributed variable, Fisher's z:

$$z = \frac{1}{2} \ln\left(\frac{1 + \mathrm{r}}{1 - \mathrm{r}}\right) \quad \textbf{6.3}$$

Where z is Fisher's z for the sample correlation coefficient r.

The standard error on any correlation coefficient can be calculated on Fisher-transformed coefficients as it can with any variable which is normally distributed and confidence limits can be converted back to correlation coefficients using the inverse of the formula above. This is described by Caruso and Cliff[301].

In this work, the variance of Fisher's z was estimated using the formula of Bonett and Wright[302]:

$$\sigma^2(z) = \frac{1 + \rho}{2n - 6} \quad \textbf{6.4}$$

Where z is Fisher's z and n is the sample size. The population correlation coefficient ρ was assumed to equal the sample correlation coefficient for the purpose of estimating the variance of Fisher's z.

For calculating confidence limits on correlation coefficients, the cumulative normal distribution was used, as is standard practice. (1.96 was used in place of t in Equation 6.2.)

6.1.1.2. Logistic regression

In this project, some docking results were analysed using logistic regression[252] as detailed in the results section of the computational results chapter. This analysis was performed using IBM SPSS Statistics 20.

Binary logistic models have the form:

$$P = \frac{1}{1 + e^{-y}} \quad \textbf{6.5}$$

P is the probability of "success", which could be, for example, the probability of side chain 2 being in the middle, Trp23 binding pocket.

In this project,

$$y = \beta_{constant} + \beta_{MW}x_{MW} + \beta_{logP}x_{logP} + \beta_{H}x_{H} + \beta_{R}x_{R} \quad \textbf{6.6}$$

The β variables are constants describing how each factor, x, affects the probability.

Use of an equation in the form of Equation 6.5 ensures that P is always predicted to be between 0 and 1, regardless of the values for the parameters.

Rearrangement of Equation 6.5 and substitution of Equation 6.6 into it yields Equation 6.9. On the left side of Equation 6.9 is the odds ratio, the probability of success divided by the probability of failure, and on the right is the product of a term for each factor. e^{β} is a useful value to calculate for each factor because it indicates what the odds ratio is multiplied by if x_i increases by 1 (and what it is divided by if x_i decreases by 1).

$$\frac{1}{P} = 1 + \exp\left(-\sum_i \beta_i x_i\right) \quad \textbf{6.7}$$

$$\frac{1-P}{P} = \exp\left(-\sum_i \beta_i x_i\right) \quad \textbf{6.8}$$

$$\frac{P}{1-P} = \exp\left(\sum_i \beta_i x_i\right) = \prod_i \left(e^{\beta_i}\right)^x \quad \textbf{6.9}$$

Fitting a binary logistic model involves finding the constants denoted by β. Making a prediction with the model here involves calculating P for a particular molecular weight, logP, number of heteroatoms and number of rings (substituting for x in the formula) and then looking to see whether P > 0.5. Testing the statistical significance of each factor involves testing the null hypothesis that β = 0 for the factor in question (after the effects of the other variables have been accounted for).

When fitting, SPSS identifies the maximum likelihood (most likely) value of β, $\hat{\beta}$. SPSS also gives the Wald statistic, W, for β. This is the square of the maximum likelihood estimate of β divided by the estimated variance of this estimate (the standard error squared)[252].

$$W = \frac{\hat{\beta}^2}{\hat{\sigma}^2_{\hat{\beta}}} \qquad \textbf{6.10}$$

The variance of the likelihood at the observed maximum can be estimated from the second derivative of the function at this maximum[303]. The significance of the maximum likelihood estimate, $\hat{\beta}$, can be assessed by comparing the Wald statistic with a critical value from the chi-squared distribution with 1 degree of freedom (for large samples) or the F-distribution with 1 as the first degree of freedom (for small samples).

6.1.2. ANALYSIS OF VARIANCE (ANOVA)

6.1.2.1. The t-Test

The t-test can be used to test whether two samples are drawn from populations with the same mean. In this project, the t-test was used to compare the predicted binding affinities of oligobenzamides differing in whether oxygen or sulphur was used to connect the side chains to the scaffold.

To perform the test on two groups (sets of repeats in this context) a test statistic, t, is calculated and compared to the critical value for t from tables of Student's t-distribution.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}} \qquad \textbf{6.11}$$

Where $\bar{X}_1$ and $\bar{X}_2$ are the means of the two groups and $s_{\overline{X}_1 - \overline{X}_2}$ is the standard deviation of $\overline{X}_1 - \overline{X}_2$ which would be obtained if the two groups were to be repeatedly sampled from their parent populations.

$$s_{\overline{X}_1 - \overline{X}_2} = \frac{\sqrt{s_1^2 + s_2^2}}{\sqrt{n}} \qquad \textbf{6.12}$$

Where n is the size of each of the two groups and s_1 and s_2 are the unbiased sample standard deviations of the two groups.

Usually, the two groups are assumed to have the same variance and the t statistic is evaluated using the t-distribution with 2n-2 degrees of freedom. However the assumption of equal variances does not need to be made if the t-distribution is used with the number of degrees of freedom, d, given by the Welch–Satterthwaite equation[261], which for two equally sized samples simplifies to

$$d = \frac{(s_1^2 + s_2^2)^2(n-1)}{s_1^4 + s_2^4}. \quad \textbf{6.13}$$

A t-test is a test of whether the variation between two groups can be explained solely by the variation within the groups, as would be the case if the two groups were samples drawn from the same population. It is an analysis of the variance and the t-test is an example of a more general set of tests which can be applied to any number of factors and groups, tests known collectively as analysis of variance (ANOVA).

6.1.2.2. Introduction to ANOVA

There were many hypotheses which could be tested using the multivariate data produced in this project. ANOVA[215] allows a large number of hypotheses to be assessed in a single test, ensuring rigor whilst allows statistical power to be maintained.

ANOVA is the standard method for testing whether different nominal factors (factors with discrete unordered values) affect a normally distributed interval (continuous) variable. Performing an ANOVA significance test is similar to performing any statistical test in that it involves the calculation of a test statistic (F in the case of ANOVA) and then the comparison of this calculated statistic with a critical value which can be looked up from tables (of the inverse cumulative F-distribution function in this case).

The first step is the calculation of what is referred to as the sums of squares for each source of variation (factor and factor combination). This is the sum of the squares of the deviations of a set of values from a particular mean. For a two factor ANOVA there are four sums of squares. Firstly, there is a sum of squares (SS) for each of the two factors:

$$SS_A = br\sum_{i=1}^{a}(\bar{X}_i - \bar{X})^2 \quad \textbf{6.14}$$

$$SS_B = ar\sum_{j=1}^{b}\left(\bar{X}_j - \bar{X}\right)^2 \quad \textbf{6.15}$$

Where a and b are the numbers of possible values for two factors, A and B and r is the number of repeats per cell (combination of A and B values).

Secondly, there is a sum of squares for the interaction between the variables, SS_{AB}.

$$SS_{AB} = r\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{X_{ij}} - \bar{X}_i - \bar{X}_j + \bar{X}\right)^2 \quad \textbf{6.16}$$

Thirdly, there is a sum of squares for the within group variation (SS_{within})

$$SS_{within} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left(X_{ijk} - \bar{X}\right)^2 \quad \textbf{6.17}$$

For each sum of squares the "mean sum of the squares" must be calculated by dividing each sum of squares by the number of degrees of freedom for that sum of squares. Finally, F values can be calculated by dividing each between-group mean sums of squares by the within-group mean sums of squares. This is summarised in Table 6.1 for two factors, A and B.

A significant interaction term F statistic indicates the presence of a statistical interaction between the factors. In the case of a two factor interaction term, this means that the manner in which each of the factors affects the dependent variable varies according to the value of the other factor. When the highest order interaction term in an ANOVA involving three or more factors is significant, this indicates that there is a deterministic relationship between the factors and the dependent variable which cannot be accounted for by the summation of independent effects caused by each of the factors individually, their pairwise interactions or any other possible lower order interaction.

Table 6.1: A two-way ANOVA table

Source of variation	Degrees of freedom	Sums of squares	"Mean sums of squares"	F statistic
A	$a-1$	SS_A	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_{within}}$
B	$b-1$	SS_B	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_{within}}$
A × B	$(a-1)(b-1)$	SS_{AB}	$MS_{A\times B} = \frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_{within}}$
Within group	$ab(r-1)$	SS_{within}	$MS_{within} = \frac{SS_{within}}{ab(r-1)}$	

A and B are factors. An interaction term and a residual variance are included in this model. a is the number of classes of factor A, b is the number of classes of factor B and r is the number of replicates within each cell (for each combination of values of A and B).

There are different methods of calculating the sums of squares in an ANOVA. The different methods produce identical results for the last term in the model (SS_{AB} above) and the same results for all terms (SS_A, SS_B, SS_{AB}) where there is the same number of replicates for each combination of factor values. However, for the main effects (A and B), if the number of replicates differs per cell (an unbalanced design), the choice of formulae is important because it subtly changes the hypotheses tested.

The expressions shown in the table are for type II sum of squares. Calculation of the type III sums of squares for each source of variation requires the formulation of a separate table similar to that above with the term in question placed last in the model. Type III sums of squares therefore indicate the degree of variation which can be explained by a term in the model after all of the other terms have been taken into account.

6.1.2.2.1. Missing values

There were two types of missing data encountered when processing binding energies for different compounds. Firstly, there were differing numbers of repeats for each compound for each pose; docking was carried out three times but sometimes there was no suitable pose from one or two of the three repeats. Secondly, there were some occasions where there was no result because none of the three repeats yielded a suitable pose.

The first type of imbalance was not a problem. Type III sums of squares (discussed above) were used, which give the size of a particular effect after all of the other factors of the model have been taken into account. These are not biased by such missing values and consequently are the most widely used type of sums of squares in ANOVA and the default in SPSS.

The second type of imbalance is a significant problem and two strategies were used to overcome it. Firstly, the number of factors used in any single ANOVA was kept small. Where necessary, more than one ANOVA was performed. Problems with this method are that results for separate ANOVAs cannot be compared, reducing the number of questions which can be investigated, and also that, when multiple ANOVAs are performed, multiple hypotheses are being tested, so care must be taken to avoid the incorrect rejection of null hypotheses due to chance.

The problem of missing values was resolved by impution (backward interpolative prediction) of a value for each empty cell. The model used for impution in this project (Equation 6.18) was one with only main effects to prevent the introduction of pairwise interactions.

$$Y_{ijk} = m + a_i + b_j + c_k + \epsilon_{ijk} \quad \textbf{6.18}$$

Where Y_{ijk} is the predicted binding energy for the combination of factor values i, j and k, m is the mean binding energy over all of the cells (combinations), a_i, b_j and c_k are coefficients describing the (assumed independent) contribution of each factor and ϵ_{ijk} is an error term which accounts for effects not modelled by the parameters.

To impute values, the coefficients are first calculated using the available data. There is one coefficient a for every possible side chain at position 1, one coefficient b for every possible side chain at position 2 and one coefficient c for every possible side chain at position 3. The model formula (Equation 6.18) is then used for each value to be imputed assuming ϵ_{ijk} is 0.

$$m = \frac{1}{N}\sum_{i,j,k} y_{ijk} \quad \textbf{6.19}$$

Where y_{ijk} is mean of the observations for the factor value combination i, j, k and N is the total number of cells (combinations). m is, as in Equation 6.18, the mean binding energy over of the cells.

$$a_i = \frac{1}{N}\sum_{j,k} y_{ijk} - m \quad \textbf{6.20}$$

$$b_j = \frac{1}{N}\sum_{i,k} y_{ijk} - m \quad \textbf{6.21}$$

$$c_k = \frac{1}{N}\sum_{i,j} y_{ijk} - m \quad \textbf{6.22}$$

Where a_i, b_j and c_k are coefficients of the model, as shown in Equation 6.18.

When testing for two-way interactions, impution increases the chance of a type II statistical error (failure to reject the null hypothesis).

To reduce the chance of a three-way interaction between two side chains and a third factor (the pose or scaffold choice) being created, values were deleted prior to impution such that an equal number of cell values were imputed for each value of the third factor (each pose or each scaffold). However, it is an assumption of the tests for three-way interactions following impution described in this thesis that (for oligobenzamides based on a particular scaffold in a particular pose) the probability of a prediction missing (due to docking not generating a pose) does not depend on the choice of side chains. This is known as the MCAR (missing completely at random) assumption[304].

Impution and ANOVA were performed using IBM SPSS Statistics 20.

6.1.2.2.2. R-squared values

When an ANOVA or logistic regression is carried out in SPSS, the software calculates an R-squared value (R^2, Equation 6.23) which indicates the goodness of fit of the model, more specifically, the proportion of the variation in the data described by the model. Standard (unadjusted) R-squared values range from 0 (no explanation of the results) to 1 (full explanation of the variation in the results).

For ANOVA analyses:

$$R^2 = 1 - \frac{SS_{Error}}{\sum(x - \bar{x})^2} \quad \textbf{6.23}$$

SS_{Error} is the sum of the squares of the residuals, the residual being the error term in the model, the difference between each value and its predicted value based on the fitted terms of the model. This sums of squares is, as shown, divided by the sum of the squares of the differences of each value, x, from the overall mean of all the data. In SPSS this denominator is referred to as the total adjusted sums of squares (SS_{Total}).

Models with more terms are intrinsically more likely to fit data better than models with fewer terms. Adjusted R-squared values (R^2_{adj}, Equation 6.24), are also calculated by SPSS. These

take into account the number of coefficients in the model and thus allow a fair comparison of models with different numbers of terms.

$$R^2_{adj} = 1 - \frac{(1 - R^2)(N - 1)}{N - k - 1} \quad \textbf{6.24}$$

N is the number of data points and k is the number of coefficients in the model.

6.1.2.2.3. Necessary Assumptions for ANOVA

It is an assumption of ANOVA that the residuals (ϵ_{ijk} in Equation 6.18) are distributed normally (with no skew or excess kurtosis) and that the variance of the residuals is the same in each cell (the data are heteroscedastic).

The Fisher g statistics g_1 and g_2 can be used to assess the skew and kurtosis (pointedness) of a distribution respectively[305].

$$g_1 = \frac{n}{n-2} \frac{\sum(X_i - \bar{X})^3}{(\sum(X_i - \bar{X})^2)^{3/2}} \quad \textbf{6.25}$$

$$g_2 = \frac{n-1}{(n-2)(n-3)} \left[n(n+1) \frac{\sum(X_i - \bar{X})^4}{(\sum(X_i - \bar{X})^2)^2} - 3(n-1) \right] \quad \textbf{6.26}$$

The g_2 of samples randomly drawn from a normal distribution is distributed as a normal distribution with a variance approximately equal to 24 divided by the sample size, n[306(p89)].

$$g_2 \sim N\left(0, \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}\right) \quad \textbf{6.27}$$

In the computational results chapter, a kurtosis of 0.517 is described as small. This is because using Equation 6.27 it can be shown that, at the 5% significance level ($|z| > 1.96$), the sample size must exceed 237 for such a kurtosis to be statistically significant. In the results described, the kurtosis was statistically significant because the sample size exceeded 237. However, the larger the sample size, the more robust ANOVA is to deviations from normality[307] so the results of the ANOVA subsequently described in the results are as reliable as they would have been had the sample been smaller and thus had a statistically insignificant kurtosis.

Similarly, the homoscedasticity of data was assessed prior to ANOVA using Levene's test and in some cases a statistically significant result (indicating heteroscedasticity) was obtained but ignored because it was small in magnitude and only significant due to the large number of cells. Levene's test generates F statistics, which are assessed using the F distribution. If the number of ANOVA cells (the first degree of freedom) is 25 (a typical value for an ANOVA study) or less, statistical tables for the F distribution show that, at the 5% significance level, the critical value is never less than 1.5, however large the sample size (which affects the second degree of freedom). Therefore, Levene's test F values of 1.5 were considered sufficiently small for the

results of ANOVA tests to be reliable. Levene's test is robust to deviations from normality (skew and kurtosis)[269].

Because tests for homogeneity of variance such as Levene's test are very sensitive and the ANOVA test is reasonably robust to homogeneity of variances when there is a large amount of data, the importance of testing for homoscedasticity before ANOVA is controversial[270(p308)].

6.2. Appendix B: NMR In Terms of Product Operators

The magnetic moments of the nuclei in a sample combine to create an overall magnet field parallel to the large external magnetic field in an NMR experiment. The radio frequency pulses applied cause rotation of the average magnetic moment around an axis in the plane perpendicular to the magnetic field.

The position of the average magnetic moment can be described in terms of x, y and z components, Cartesian angular momentum operators[107(p150)] which describe the components of the angular momentum in the x, y and z dimensions. They are operators because they describe how the spin system is changing with time.

When a sample is placed in a magnetic field, a hydrogen nucleus within the system can be described as having an operator (a single-element operator) of $\widehat{H_z}$. The z indicates that the magnetisation is aligned with the z-axis, which, by definition is always parallel to the external magnetic field.

If the net magnetisation is rotated into the xy-plane by a radio-frequency magnetic pulse then it will rotate around the z-axis as the magnetic moments of the nuclei do likewise, a process known as precession. This would complicate mathematical working so, by definition, the xy-plane is treated as a rotating frame of reference, rotating such that nuclei precessing at the zero chemical shift frequency are stationary relative to it[107(p32)].

Pulses can be described using superoperators, which are operators acting on the operators. Superoperators can be parallel to the x-axis or y-axis, for example, for the H nucleus, $\widehat{\widehat{H_x}}$ or $\widehat{\widehat{H_y}}$. Applying a 90° pulse, a pulse of the length required to rotate the hydrogen nucleus 90° (π/2 radians) away from the z-axis, shifts the magnetisation into the xy-plane as shown here (using H to represent a hydrogen nucleus):

$$\widehat{H_z} \xrightarrow{\widehat{\widehat{H_x}}\left(\frac{\pi}{2}\right)} -\widehat{H_y} \qquad \textbf{6.28}$$

$$\widehat{H_z} \xrightarrow{\widehat{\widehat{H_y}}\left(\frac{\pi}{2}\right)} \widehat{H_x} \qquad \textbf{6.29}$$

The length of time over which the superoperator acts is shown in brackets.

180° pulses reverse the direction of the magnetisation.

$$\widehat{H_z} \xrightarrow{\widehat{\widehat{H_x}}(\pi)} -\widehat{H_z} \qquad \textbf{6.30}$$

$\widehat{\widehat{H_z}}(\omega_H t_H)$ is the superoperator for the evolution of chemical shift, the rotation of the nucleus around the z-axis relative to the xy-plane caused by the chemical shift of the nucleus not being

equal to 0. Chemical shift evolution occurs with respect to any magnetisation in the xy-plane whenever there is a delay between pulses.

The effect of a 90°pulse followed by evolution of chemical shift in the xy-plane is shown below.

$$\widehat{H_z} \xrightarrow{\widehat{H_x}\left(\frac{\pi}{2}\right)} \widehat{H_z}\cos\left(\frac{\pi}{2}\right) - \widehat{H_y}\sin\left(\frac{\pi}{2}\right) = -\widehat{H_y} \xrightarrow{\widehat{H_z}(\omega_H t_H)} \widehat{H_x}\sin(\omega_H t_H) - \widehat{H_y}\cos(\omega_H t_H) \quad \textbf{6.31}$$

Detection occurs in the xy-plane. The $-\widehat{H_y}\cos(\omega_H t_H)$ term is the part of the resulting magnetisation which is detected here. A Fourier transform is performed on the oscillating magnetic field detected, turning the $\cos(\omega_H t_H)$ wave into a peak at ω_H, the chemical shift of the nucleus.

2D and 3D NMR spectra are produced by combining many 1D spectra produced using different length delays (t_1 (2D) or t_1 and t_2 (3D)) in the pulse sequence. The length of other delays in the pulse sequence is typically also important. For example, often two identical times are found either side of a 180° pulse. This refocusing sequence prevents chemical shift evolution because evolution in the first delay is reversed in the second part.

The Zeeman splitting of the spin states of a nucleus is affected by the spins of adjacent nuclei that share bonding electrons with it. This is J-coupling, otherwise known as scalar coupling[107(p35)] and it occurs because the spin of electrons is influenced by and influences the spin of their nuclei. It can be represented by the superoperator $\widehat{N_z}\widehat{H_z}$. (N and H are shown here for coupling between a nitrogen atom and a hydrogen atom.)

Like chemical shift, scalar coupling can be refocused. This requires a delay of the right length (which depends on the coupling constant (strength) of the coupling) with a 180° pulse somewhere in the middle of this period on each channel to be refocused.

As well as scalar coupling, nuclei can experience dipolar coupling[107(p37)]. This is a through-space rather than through-bond effect resulting from the magnetic field of adjacent nuclei changing the overall field felt by each nucleus. Dipolar coupling is usually entirely cancelled out due to the isotropic tumbling of molecules in solution; however, partial arrangement of the molecules in a sample causes cancellation to be incomplete leaving residual dipolar couplings (RDCs).

Table 6.2, Table 6.3 and Table 6.4 show using product operator notation, how the pulses in an HSQC, a NOESY-HSQC and a TOCSY-HSQC experiment affect the magnetisation of the nuclei. The reader might find it helpful to refer to Figure 1.10, Figure 1.11 and Figure 1.12 (pp54-56) which show graphically the pulse sequences.

Table 6.2: The pulses and delays typically used to generate an HSQC spectrum

	Step	Super-operator	Result
	The initial state. (Nitrogen polarisation ignored.)		$\widehat{H_z}$
P1	Transfer of polarisation from the sensitive proton nucleus to the insensitive nitrogen nucleus.	$\widehat{\widehat{H_x}}\left(\frac{\pi}{2}\right)$	$-\widehat{H_y}$
D1		$\widehat{\widehat{N_z}}\widehat{\widehat{H_z}}\left(\frac{\pi}{4}\right)$	$-\widehat{H_y}\cos\left(\frac{\pi}{4}\right) + 2\widehat{N_z}\widehat{H_x}\sin\left(\frac{\pi}{4}\right)$ $= -\frac{1}{2}\sqrt{2}\widehat{H_y} + \sqrt{2}\widehat{N_z}\widehat{H_x}$
P2		$\widehat{\widehat{H_x}}(\pi) + \widehat{\widehat{N_x}}(\pi)$	$\frac{1}{2}\sqrt{2}\widehat{H_y} - \sqrt{2}\widehat{N_z}\widehat{H_x}$
D2		$\widehat{\widehat{N_z}}\widehat{\widehat{H_z}}\left(\frac{\pi}{4}\right)$	$\frac{1}{2}\sqrt{2}\widehat{H_y}\cos\left(\frac{\pi}{4}\right) - \sqrt{2}\widehat{N_z}\widehat{H_x}\cos\left(\frac{\pi}{4}\right)$ $- \sqrt{2}\widehat{N_z}\widehat{H_x}\sin\left(\frac{\pi}{4}\right)$ $- \frac{1}{2}\sqrt{2}\widehat{H_y}\sin\left(\frac{\pi}{4}\right)$ $= \frac{1}{2}\widehat{H_y} - \widehat{N_z}\widehat{H_x} - \widehat{N_z}\widehat{H_x} - \frac{1}{2}\widehat{H_y}$ $= -2\widehat{N_z}\widehat{H_x}$
P3		$\widehat{\widehat{H_y}}\left(\frac{\pi}{2}\right) + \widehat{\widehat{N_x}}\left(\frac{\pi}{2}\right)$	$= -2\widehat{N_y}\widehat{H_z}$
D3	Evolution of chemical shift on the nitrogen.	$\widehat{\widehat{N_z}}\left(\frac{1}{2}\Omega_N t_1\right)$	$-2\widehat{N_y}\widehat{H_z}\cos\left(\frac{1}{2}\Omega_N t_1\right) + 2\widehat{N_x}\widehat{H_z}\sin\left(\frac{1}{2}\Omega_N t_1\right)$
P4		$\widehat{\widehat{H_y}}(\pi)$	$2\widehat{N_y}\widehat{H_z}\cos\left(\frac{1}{2}\Omega_N t_1\right) - 2\widehat{N_x}\widehat{H_z}\sin\left(\frac{1}{2}\Omega_N t_1\right)$
D4		$-\widehat{\widehat{N_z}}\left(\frac{1}{2}\Omega_N t_1\right)$	$2\widehat{N_y}\widehat{H_z}\cos^2\left(\frac{1}{2}\Omega_N t_1\right)$ $- 2\widehat{N_x}\widehat{H_z}\sin\left(\frac{1}{2}\Omega_N t_1\right)\cos\left(\frac{1}{2}\Omega_N t_1\right)$ $+ 2\widehat{N_x}\widehat{H_z}\sin\left(\frac{1}{2}\Omega_N t_1\right)\cos\left(\frac{1}{2}\Omega_N t_1\right)$ $+ 2\widehat{N_y}\widehat{H_z}\sin^2\left(\frac{1}{2}\Omega_N t_1\right)$ $= 2\widehat{N_y}\widehat{H_z}\left(\cos^2\left(\frac{1}{2}\Omega_N t_1\right) - \sin^2\left(\frac{1}{2}\Omega_N t_1\right)\right)$ $= 2\widehat{N_y}\widehat{H_z}\cos(\Omega_N t_1)$
P5	Transfer of polarisation back to the hydrogen for detection.	$\widehat{\widehat{H_x}}\left(\frac{\pi}{2}\right) + \widehat{\widehat{N_x}}\left(\frac{\pi}{2}\right)$	$-2\widehat{N_z}\widehat{H_y}\cos(\Omega_N t_1)$
D5		$\widehat{\widehat{N_z}}\widehat{\widehat{H_z}}\left(\frac{\pi}{4}\right)$	$-\sqrt{2}\widehat{N_z}\widehat{H_y}\cos(\Omega_N t_1) + \frac{1}{2}\sqrt{2}\widehat{H_x}\cos(\Omega_N t_1)$
P6		$\widehat{\widehat{H_x}}(\pi) + \widehat{\widehat{N_x}}(\pi)$	$-\sqrt{2}\widehat{N_z}\widehat{H_y}\cos(\Omega_N t_1) + \frac{1}{2}\sqrt{2}\widehat{H_x}\cos(\Omega_N t_1)$
D6		$\widehat{\widehat{N_z}}\widehat{\widehat{H_z}}\left(\frac{\pi}{4}\right)$	$-\widehat{N_z}\widehat{H_y}\cos(\Omega_N t_1) + \frac{1}{2}\widehat{H_x}\cos(\Omega_N t_1)$ $+ \frac{1}{2}\widehat{H_x}\cos(\Omega_N t_1)$ $+ \widehat{N_z}\widehat{H_y}\cos(\Omega_N t_1)$ $= \widehat{H_x}\cos(\Omega_N t_1)$

The leftmost column indicates whether the step relates to the application of a pulse (P) or to what happens in a delay (D). The application of the superoperator on row n transforms the operator on row n-1 in the rightmost column into the operator on row n. Evolution of chemical shift is not shown where it is refocused. The product operator expressions were derived with the help of Cavanagh *et al.*[109(p417)].

Table 6.3: The pulses and delays typically used to generate a NOESY-HSQC spectrum

	Step	Super-operator	Resulting product operators
	Initial state (Nitrogen polarisation ignored.)		$\widehat{H_z}$
P1	Magnetisation brought into xy-plane.	$\widehat{\widehat{H_x}}\left(\frac{\pi}{2}\right)$	$-\widehat{H_y}$
D1	Evolution of J coupling.	$\widehat{\widehat{N_z}}\widehat{\widehat{H_z}}\left(\frac{1}{2}\Omega_H t_1\right)$	$-\widehat{H_y}\cos\left(\frac{1}{2}\Omega_H t_1\right) + 2\widehat{N_z}\widehat{H_x}\sin\left(\frac{1}{2}\Omega_H t_1\right)$
P2	Refocusing pulse	$\widehat{\widehat{H_x}}(\pi) + \widehat{\widehat{N_x}}(\pi)$	$\widehat{H_y}\cos\left(\frac{1}{2}\Omega_H t_1\right) - 2\widehat{N_z}\widehat{H_x}\sin\left(\frac{1}{2}\Omega_H t_1\right)$
D2	Evolution of J-coupling cancelling the previous evolution.	$\widehat{\widehat{N_z}}\widehat{\widehat{H_z}}\left(\frac{1}{2}\Omega_H t_1\right)$	$\widehat{H_y}\cos^2\left(\frac{1}{2}\Omega_H t_1\right)$ $- 2\widehat{N_z}\widehat{H_x}\sin\left(\frac{1}{2}\Omega_H t_1\right)\cos\left(\frac{1}{2}\Omega_H t_1\right)$ $+ 2\widehat{N_z}\widehat{H_x}\sin\left(\frac{1}{2}\Omega_H t_1\right)\cos\left(\frac{1}{2}\Omega_H t_1\right)$ $- \widehat{H_y}\sin^2\left(\frac{1}{2}\Omega_H t_1\right)$ $= \widehat{H_y}\left(\cos^2\left(\frac{1}{2}\Omega_H t_1\right) - \sin^2\left(\frac{1}{2}\Omega_H t_1\right)\right)$ $= \widehat{H_y}\cos(\Omega_H t_1)$
P3	Returning magnetisation to the z-axis	$\widehat{\widehat{H_x}}\left(\frac{\pi}{2}\right)$	$\widehat{H_z}\cos(\Omega_H t_1)$
D3	Magnetisation transfer	$\hat{\hat{E}}\,\sigma_{\mathrm{H,H2}}(\tau_{\mathrm{m}})\cos(\Omega_{H2} t_1)$	$\widehat{H_z}\,\sigma_{\mathrm{H,NH}}(\tau_{\mathrm{m}})\cos(\Omega_H t_1)\cos(\Omega_{H2} t_1)$
Result after P1 to D6 from HSQC			$-\widehat{H_x}\,\sigma_{\mathrm{H,NH}}(\tau_{\mathrm{m}})\cos(\Omega_H t_1)\cos(\Omega_{H2} t_1)\cos(\Omega_N t_2)$

Evolution of chemical shift is not shown where it is refocused. The product operator expressions were derived with the help of Cavanagh *et al.*[109(p448)]. $\sigma_{\mathrm{H\cdot NH}}$ is the cross relaxation rate constant[109(p289)].

Table 6.4: The pulses and delays typically used to generate a TOCSY-HSQC spectrum

	Step	Super-operator	Resulting product operators
	The initial state.		$\widehat{H_{1z}}$
P1	Bringing the net magnetisation into the xy-plane.	$\widehat{\widehat{H_{1x}}}\left(\frac{\pi}{2}\right)$	$-\widehat{H_{1y}}$
D2	Evolution of chemical shift	$\widehat{\widehat{H_{1z}}}\left(\frac{1}{2}\Omega_H t_1\right)$	$-\widehat{H_{1y}}\cos\left(\frac{1}{2}\Omega_{H1} t_1\right)$ $+ 2\widehat{H_{1x}}\sin\left(\frac{1}{2}\Omega_H t_H\right)$ (Latter term removed by phase cycling)
	Simultaneous evolution of J coupling between protons	$\widehat{\widehat{H_{1z}}}\widehat{\widehat{H_{2z}}}(\Omega_{H2} t_1)$	$-\widehat{H_{1y}}\cos\left(\frac{1}{2}\Omega_{H1} t_1\right)\cos\left(\frac{1}{2}\Omega_{H2} t_1\right)$ $+ 2\widehat{H_{1x}}\widehat{H_{2z}}\cos\left(\frac{1}{2}\Omega_H t_1\right)\sin\left(\frac{1}{2}\Omega_{H2} t_1\right)$ (Latter term removed by interference during the composite pulse decoupling)
CPD	Transfer of magnetisation in the xy-plane	$\hat{\hat{E}}\left(1+\cos\left(2\pi \mathrm{J}_{\mathrm{H1,H2}}\tau_{\mathrm{m}}\right)\right)$	$-\widehat{H_{1y}}\cos\left(\frac{1}{2}\Omega_{H1} t_1\right)\cos\left(\frac{1}{2}\Omega_{H2} t_1\right)$
Result after P2 to D6 from HSQC		$-\widehat{H_{1x}}\left(1+\cos\left(2\pi \mathrm{J}_{\mathrm{H1,H2}}\tau_{\mathrm{m}}\right)\right)\cos(\Omega_{H1} t_1)\cos(\Omega_{H2} t_1)\cos(\Omega_N t_2)$	

The product operators for one of the two hydrogen atoms are shown. The end result would also contain a similar H_{2x} product operator component. $J_{H1,H2}$ is the J coupling parameter for the two hydrogen atoms in this example.

6.3. Appendix C: Additional methods diagrams

The figures in this section show in detail how the methods described in the computational results chapter were used to generate the results shown. Figure 6.2 focusses on the methods. It shows what software was used and how the output of some programs was the input for others. Some steps in this diagram were performed multiple times with different data. Figure 6.3 focusses on the results. It shows specifically how each set of results was generated.

In these figures, each box represents a piece of data and the arrows between these nodes represent the application of methods to the data. Figure 6.1 below is a key to aid interpretation of the box styles in Figure 6.2 and Figure 6.3.

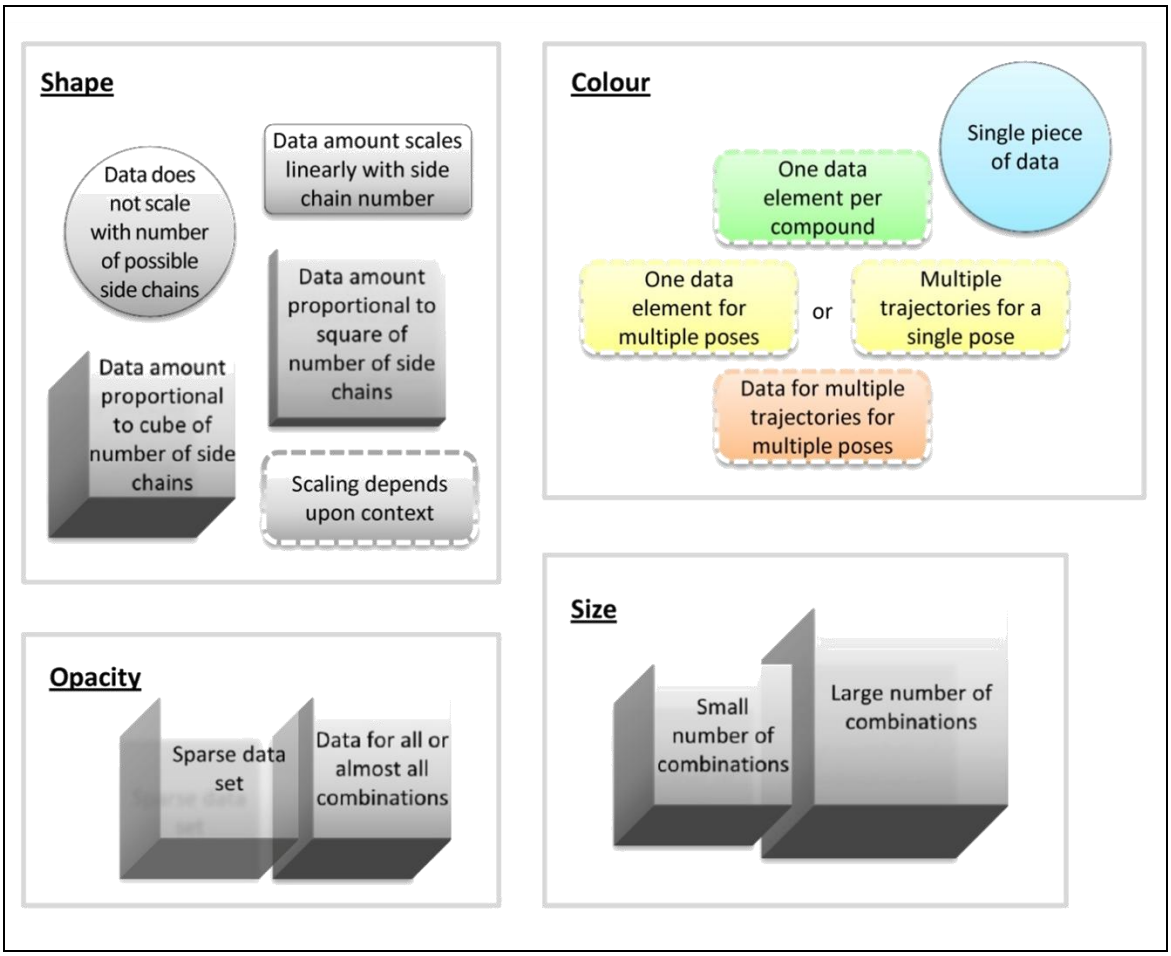
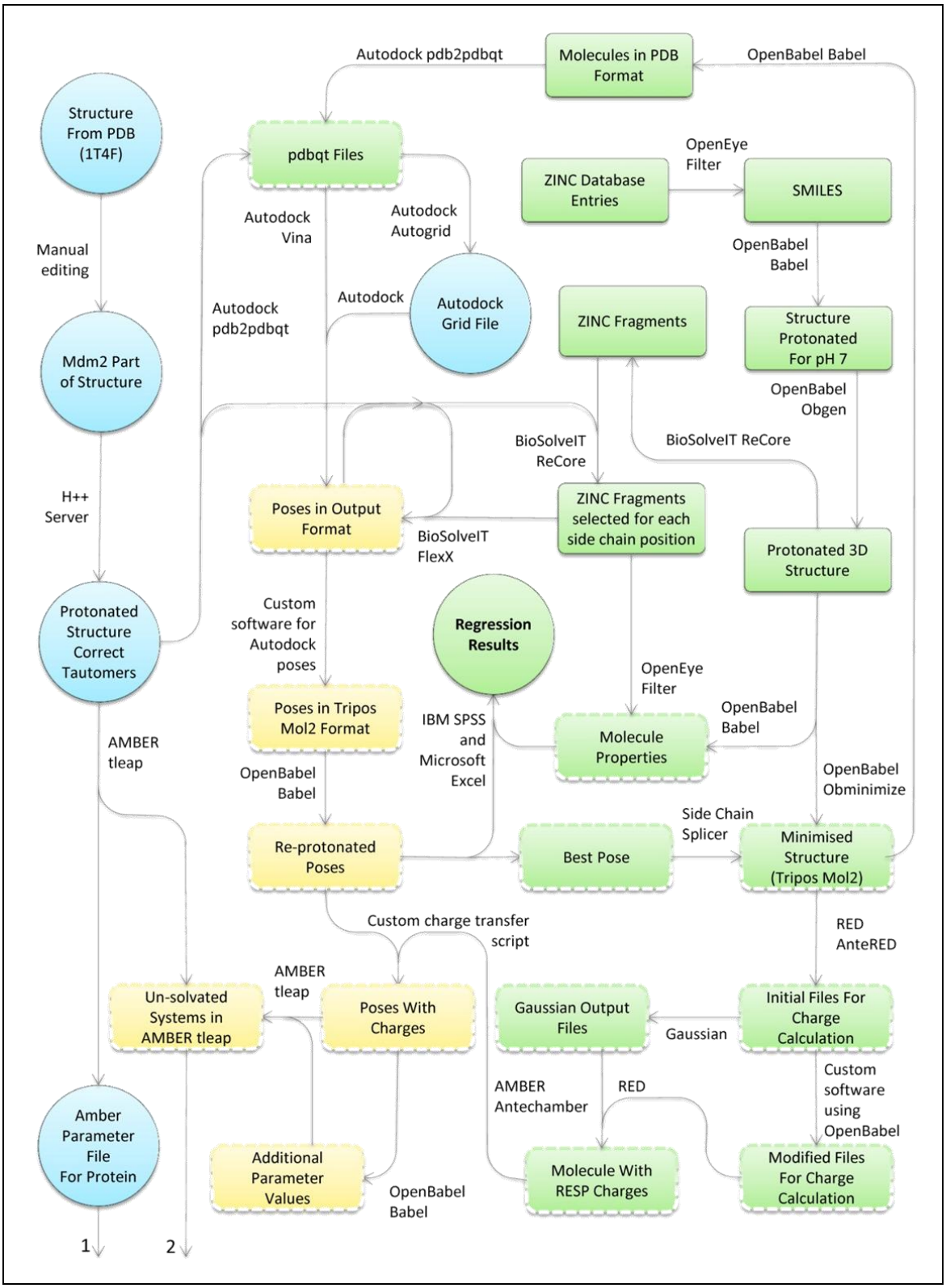


Figure 6.1: Key for Figure 6.2 and Figure 6.3 showing how the shape and colour of the nodes reflects the amount and structure of the data at each point



Continued on the next page…

Figure 6.2: The computational methods used in this project.

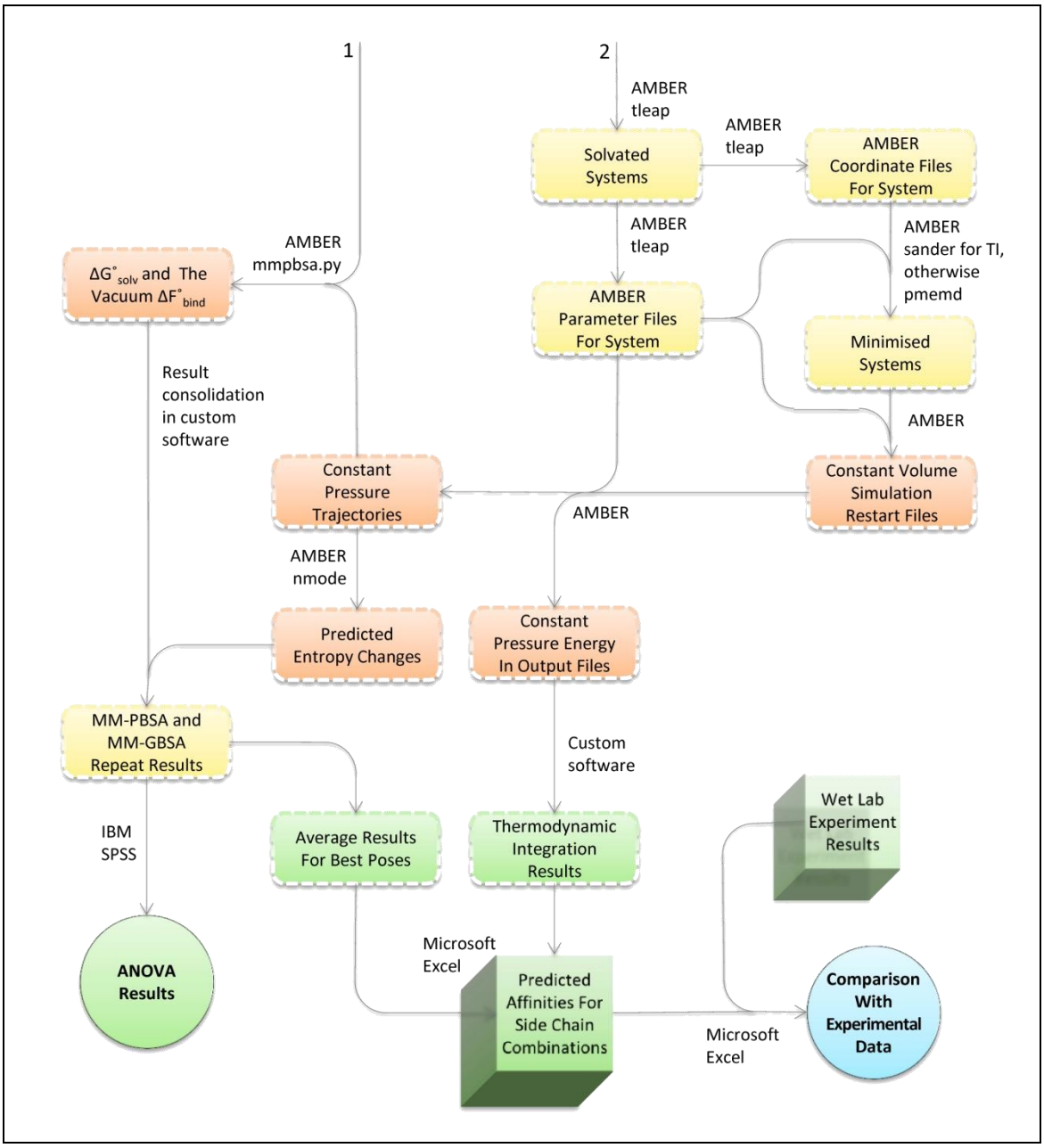
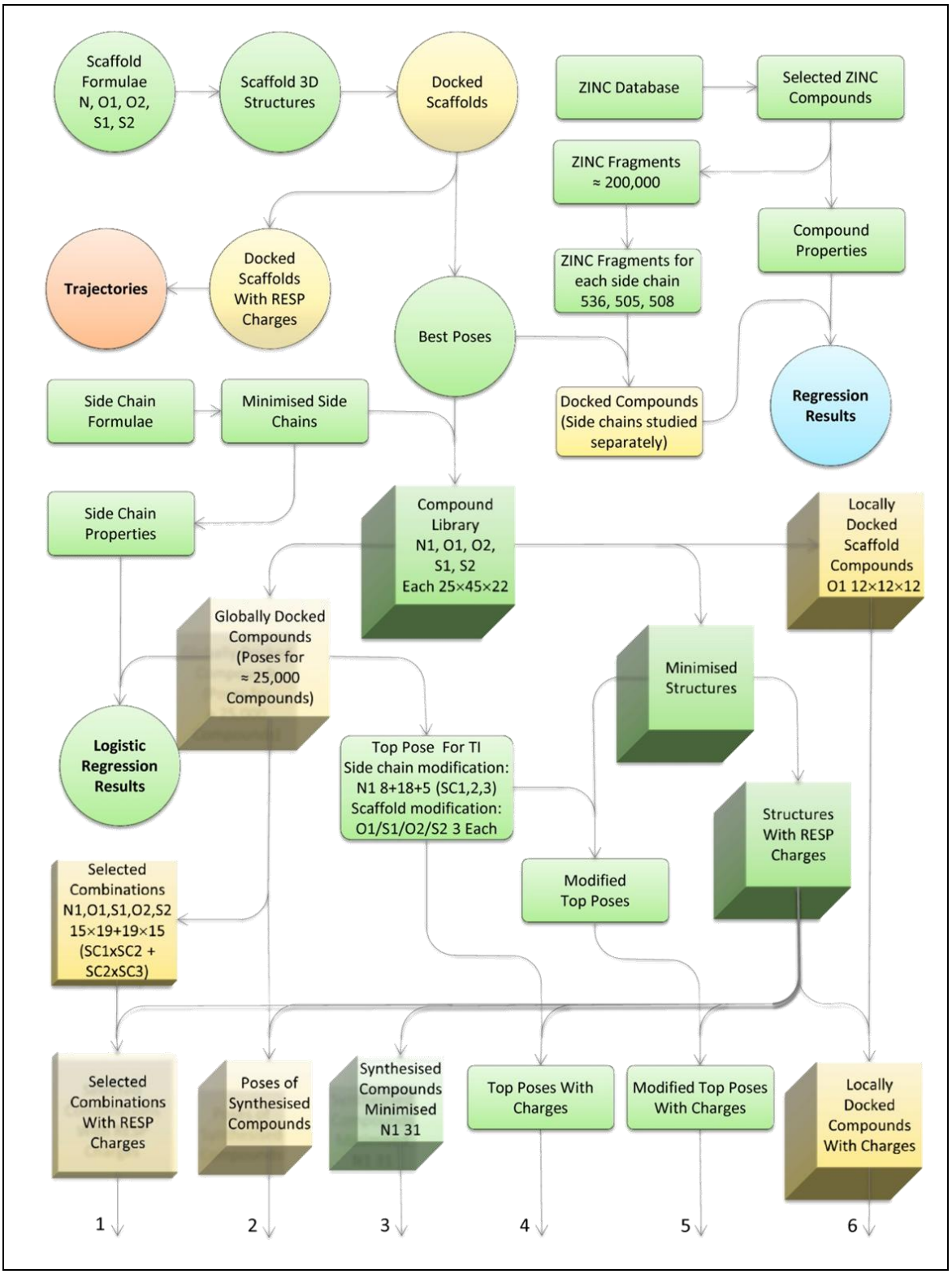


Figure 6.2 continued: The computational methods used in this project. The nodes (boxes) represent pieces of data. Arrows show how data was combined and processed to get from one node to the next. The diagram is generic, showing what methods were used but not exactly to what data sets they were applied and when; this additional detail is shown in Figure 6.3. Blue nodes depict a single piece of data. Green nodes indicate where there was a single data element or record per compound. Yellow nodes indicate multiple elements per compound, either multiple poses or multiple simulation trajectories from a single pose. Orange nodes indicate data for multiple trajectories started from multiple poses. Merging arrows indicate that two pieces of data were used together to generate the next, in contrast to where two separate arrows end at the same box, which indicates where there is more than one way by which the data in the box was produced. A similar distinction does not exist between multiple arrows coming from the same box and arrows resulting from the divergence of a single arrow; both show different uses of the data concerned. Arrows crossing the page boundary are numbered for clarity. A key for this figure and Figure 6.3 is shown in Figure 6.1.



Continued on the next page…

Figure 6.3: Data processing in this project.

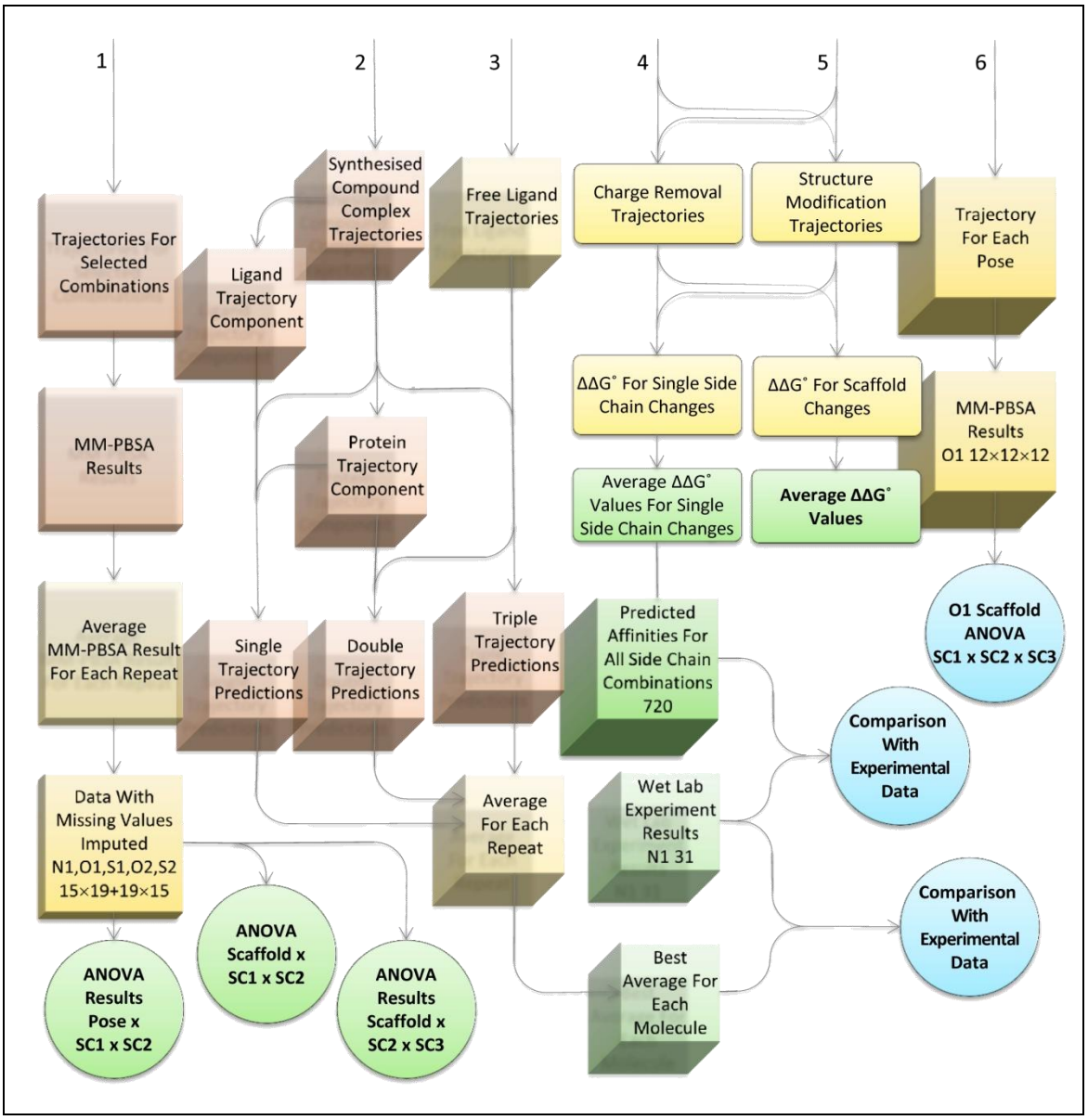


Figure 6.3 continued: Data processing in this project. The nodes represent sets of data and the arrows represent processes converting each piece of data to the next. Further details of the methods used to carry out each conversion are shown in Figure 6.2. This figure shows more specifically how the methods were applied in this project. The end results are highlighted in bold. A key for this figure and Figure 6.2 is shown in Figure 6.1. Blue nodes show a single piece of data. Green nodes indicate where there is a single data element or record per compound. Yellow nodes indicate multiple elements per compound, either multiple poses or multiple simulation trajectories from a single pose and orange nodes indicate data for multiple trajectories from multiple poses. There can be one compound per scaffold (circles), or multiple compounds, with the number scaling linearly (rounded rectangle) or with the square (thin square blocks) or cube (cube) of the number of side chains at each side chain position. The number of side chains is indicated qualitatively by the dimensions for the squares and cubes. The opaque 3D nodes indicate where there was data for all combinations. 50% transparency signifies a degree of data sparsity.

6.4. Appendix D: Additional results figures for computational work

6.4.1. TORSION PARAMETER CALCULATION

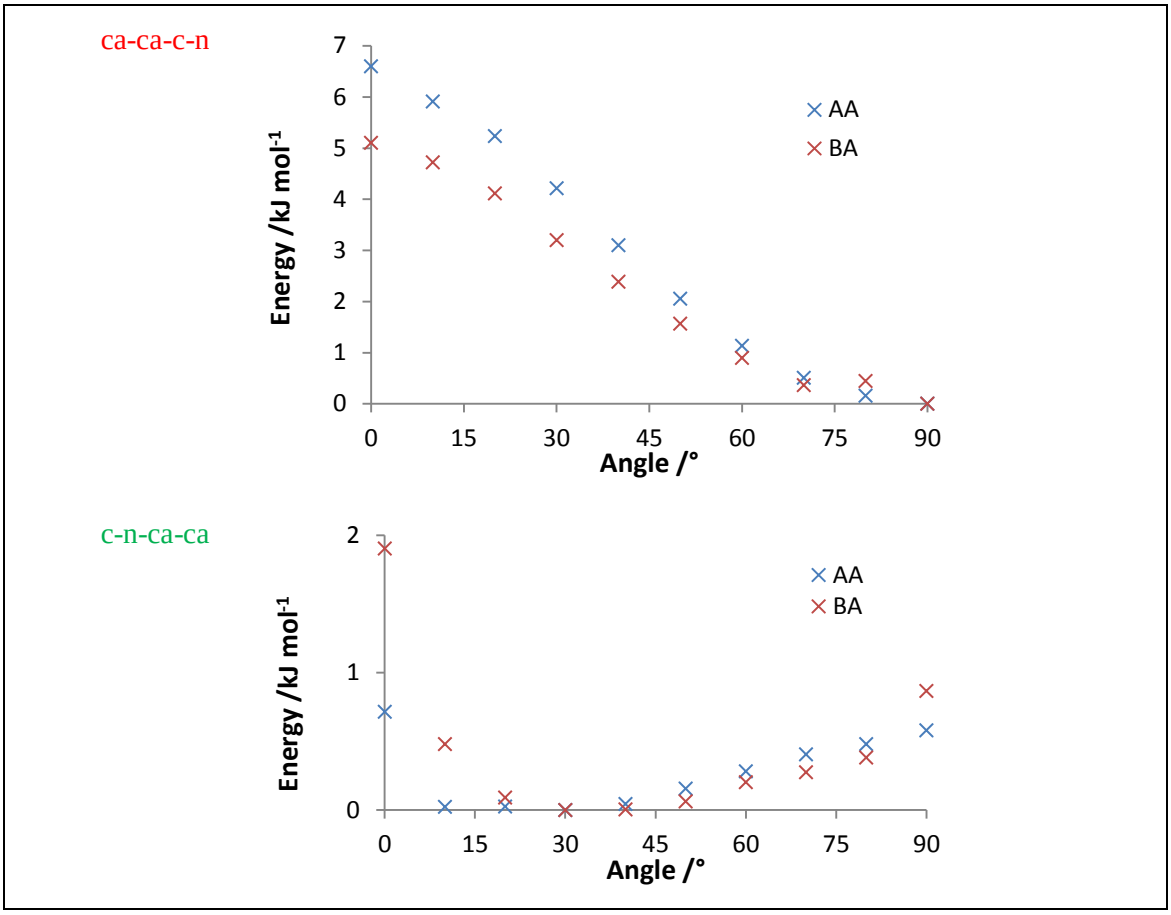


Figure 6.4:1-4 interaction energies for rotation around the amide carbon and adjacent aromatic ring carbon (c-ca) bond and the amide nitrogen and adjacent ring carbon (n-ca) bond of benzanilide (BA) and 4-(4-aminobenzamide)benzoic acid (AA). The structure of AA is shown in Figure 2.6 B (p114). Torsion parameters were calculated using the results for AA. In the simulations of this project, the ca-ca-c-n and c-n-ca-ca torsions were modelled using the AMBER force field by way of a single trigonometric term peaking at 90°. (Oligobenzamides are most stable when flat or close to flat.) The barrier height parameter was calculated using published barrier heights. In simple terms, the effect of the through-space interactions between the 1st and 4th atoms of each torsion (the electrostatic interaction and Leonard-Jones potentials which together constitute the 1-4 interaction energy) must be subtracted from the published barrier height to get the height for use in parameters (because the 1-4 interaction energy is added on again later by Amber during simulations). However, care must be taken to consider the phase and therefore sign of the 1-4 interaction energy with respect to the overall torsional barrier. With regard to the c-n-ca-ca torsion, the peak in 1-4 interaction energy at exactly 0° was ignored; the c-n-ca-ca 1-4 interaction energy was assumed to vary with angle like a sine wave, peaking at 90°. Details of the parameter calculation are shown in Table 6.5 below. The torsion angles are indicated on an oligobenzamide in Figure 2.6A (p114) by red and green arrows (corresponding to the colours of the labels above). The torsion parameters calculated here were used where the c-ca and n-ca bonds were not restricted by intramolecular hydrogen bonding. For example, the ca-ca-c-n barrier height parameter was used in the O1 and S1 scaffolds and the c-n-ca-ca parameter was used in the O2 and S2 scaffolds. Previously published parameters were used for the torsions complicated by intramolecular hydrogen bonding, as shown in Table 2.2 (p115).

Table 6.5: Subtraction of the 1-4 interaction energy from the torsional barrier heights of the bonds either side of the ring in 3-aminobenzoic acid in order to generate estimated AMBER torsion parameters for the corresponding bonds in oligobenzamides

	ca-ca-c-n torsion	c-n-ca-ca torsion
Total barrier height from literature[175] /kJ mol^{-1}	22.3	26.2
Electrostatic 1-4 interaction energy /kJ mol^{-1}	-1.01	1.63
Van der Waals 1-4 interaction energy /kJ mol^{-1}	-5.58	-1.05
Total 1-4 interaction energy /kJ mol^{-1}	-6.60	0.58
Barrier height /kJ mol^{-1}	28.9	25.6
PK parameter /kJ mol^{-1}	14.4	12.8

In Amber, the energy of through-space interactions (the electrostatic interaction and Leonard-Jones potentials) between atoms 3 bonds apart (1-4 interactions) are calculated and included in the non-bonded energy. Consequently, to calculate AMBER force field torsion angle parameters from the known barrier height for a particular torsion, the change in 1-4 interaction energy in going from the bottom to the top of the barrier must be subtracted so that, when it is added again during the simulation, the barrier height is correct. GAFF atom types are used here. "ca" represents aromatic carbon and "c" and "n" represent sp^2-hybridised carbon and nitrogen respectively. The torsion angles are indicated in Figure 2.6 (p114) by red and green arrows. The two parameters calculated here were used for the N-linked oligobenzamide and the bonds in the O and S-linked scaffold where there is not restriction due to intramolecular hydrogen bonding. The PK parameters are entered into AMBER in kcal mol^{-1}. (1 J = 0.239 cal.)

6.4.2. DOCKING EXHAUSTIVENESS EVALUATION

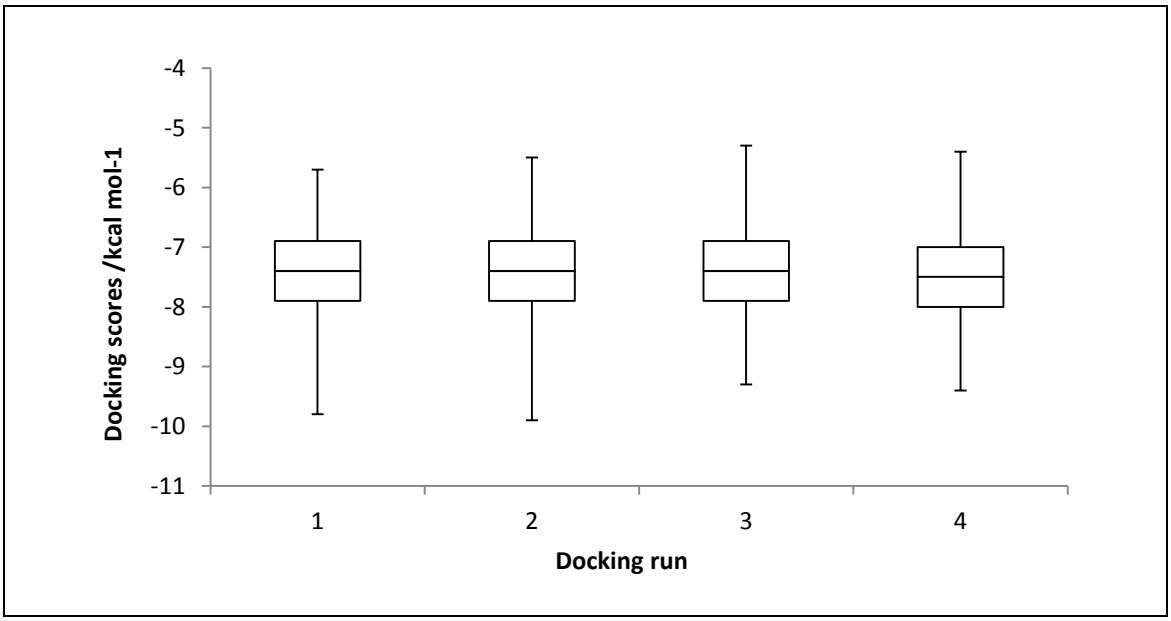


Figure 6.5: The effect of the Autodock Vina "exhaustiveness" setting on docking score. Box and whisker plots for all docking scores obtained in four different Autodock Vina docking runs are shown. Runs 1, 2 and 3 were with the exhaustiveness set to 8: run 4 was with the exhaustiveness set to 16. There is a statistically significant improvement in the docking score between the first three runs and the fourth ($P<0.01$) but the increase is small (0.09).

7. References

1 McMillan, K. *et al.* (2000) Allosteric inhibitors of inducible nitric oxide synthase dimerization discovered via combinatorial chemistry. *Proc. Natl. Acad. Sci.* 97, 1506–1511

2 Leung, D.K. *et al.* (2000) Selective disruption of protein aggregation by cyclodextrin dimers. *Proc. Natl. Acad. Sci.* 97, 5050–5053

3 Smith, A. and Hirschmann, R. (1997) An orally bioavailable pyrrolinone inhibitor of HIV-1 protease: computational analysis and X-ray crystal structure of the enzyme complex. *J. Med. Chem.* 40, 2440–2444

4 Burgess, K. (2001) Solid-phase syntheses of β-turn analogues to mimic or disrupt protein-protein interactions. *Acc. Chem. Res.* 34, 826–835

5 Jochim, A.L. and Arora, P.S. (2009) Assessment of helical interfaces in protein-protein interactions. *Mol. Biosyst.* 5, 924–926

6 Keskin, O. *et al.* (2008) Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.* 108, 1225–1244

7 Blazer, L.L. and Neubig, R.R. (2009) Small molecule protein-protein interaction inhibitors as CNS therapeutic agents: current progress and future hurdles. *Neuropsychopharmacology* 34, 126–141

8 Arkin, M.R. and Wells, J.A. (2004) Small-Molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* 3, 301–317

9 Murray, J.K. and Gellman, S.H. (2007) Targeting Protein–Protein Interactions: Lessons from p53/MDM2. *Biopolymers* 88, 657–686

10 Bird, G.H. *et al.* (2010) Hydrocarbon double-stapling remedies the proteolytic instability of a lengthy peptide therapeutic. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14093–14098

11 Scholtz, J.M. and Baldwin, R.L. (1992) The mechanism of alpha-helix formation by peptides. *Annu. Rev. Biophys. Biomol. Struct.* 21, 95–118

12 Edwards, T.A. and Wilson, A.J. (2011) Helix-mediated protein-protein interactions as targets for intervention using foldamers. *Amino Acids* 41, 743–754

13 Fasan, R. *et al.* (2004) Using a beta-hairpin to mimic an alpha-helix: cyclic peptidomimetic inhibitors of the p53-HDM2 protein-protein interaction. *Angew. Chem. Int. Ed. Engl.* 43, 2109–2112

14 Fasan, R. *et al.* (2006) Structure-activity studies in a family of beta-hairpin protein epitope mimetic inhibitors of the p53-HDM2 protein-protein interaction. *Chembiochem* 7, 515–526

15 Wells, J.A. and McClendon, C.L. (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 450, 1001–1009

16 Chène, P. (2006) Drugs targeting protein-protein interactions. *ChemMedChem* 1, 400–411

17 Suryadinata, R. *et al.* (2010) Control of cell cycle progression by phosphorylation of cyclin-dependent kinase (CDK) substrates. *Biosci. Rep.* 30, 243–255

18 Malumbres, M. and Barbacid, M. (2005) Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* 30, 630–641

19 Ekholm, S. V and Reed, S.I. (2000) Regulation of G(1) cyclin-dependent kinases in the mammalian cell cycle. *Curr. Opin. Cell Biol.* 12, 676–684

20 Luo, Y. *et al.* (1995) Cell-cycle inhibition by independent CDK and PCNA binding domains in p21(Cip1). *Nature* 375, 159–161

21 Tibbetts, R.S. *et al.* (1999) A role for ATR in the DNA phosphorylation of p53. *Genes Dev.* 13, 152–157

22 D'Orazi, G. *et al.* (2002) Homeodomain-interacting protein kinase-2 phosphorylates p53 at Ser 46 and mediates apoptosis. *Nat. Cell Biol.* 4, 11–19

23 Oda, K. *et al.* (2000) p53AIP1, a potential mediator of p53-dependent apoptosis, and its regulation by ser-46-phosphorylated p53. *Cell* 102, 849–862

24 Canman, C.E. *et al.* (1998) Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* 281, 1677–1679

25 Levine, A.J. (1997) p53, the Cellular Gatekeeper for Growth and Division. *Cell* 88, 323–331

26 Böttger, A. *et al.* (1997) Molecular characterization of the hdm2-p53 interaction. *J. Mol. Biol.* 269, 744–756

27 Joerger, A.C. and Fersht, A.R. (2010) The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harb. Perspect. Biol.* 2, a000919

28 Selivanova, G. and Wiman, K.G. (2007) Reactivation of mutant p53: molecular mechanisms and therapeutic potential. *Oncogene* 26, 2243–2254

29 Ardley, H.C. and Robinson, P.A. (2005) E3 ubiquitin ligases. *Essays Biochem.* 41, 15–30

30 Linke, K. *et al.* (2008) Structure of the MDM2/MDMX RING domain heterodimer reveals dimerization is required for their ubiquitylation in trans. *Cell Death Differ.* 15, 841–848

31 Dezi, C. *et al.* (2010) Molecular interaction fields and 3D-QSAR studies of p53-MDM2 inhibitors suggest additional features of ligand-target interaction. *J. Chem. Inf. Model.* 50, 1451–1465

32 Michelsen, K. *et al.* (2012) Ordering of the N-terminus of human MDM2 by small molecule inhibitors. *J. Am. Chem. Soc.* 134, 17059–17067

33 Dolezelova, P. *et al.* (2012) Mutational analysis reveals a dual role of Mdm2 acidic domain in the regulation of p53 stability. *FEBS Lett.* 586, 2225–2231

34 Yu, G.W. *et al.* (2006) Solution structure of the C4 zinc finger domain of HDM2. *Protein Sci.* 15, 384–389

35 Partridge, J.R. and Schwartz, T.U. (2009) Crystallographic and biochemical analysis of the Ran-binding zinc finger domain. *J. Mol. Biol.* 391, 375–389

36 Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242

37 Grasberger, B.L. *et al.* (2005) Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *J. Med. Chem.* 48, 909–912

38 Fry, D.C. *et al.* (2004) NMR structure of a complex between MDM2 and a small molecule inhibitor. *J. Biomol. NMR* 30, 163–173

39 Fischer, P.M. (2006) Peptide, Peptidomimetic, and Small-molecule Antagonists of the p53-HDM2 Protein-Protein Interaction. *Int. J. Pept. Res. Ther.* 12, 3–19

40 Vassilev, L.T. *et al.* (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303, 844–848

41 Bowman, A.L. *et al.* (2007) Small molecule inhibitors of the MDM2-p53 interaction discovered by ensemble-based receptor models. *J. Am. Chem. Soc.* 129, 12809–12814

42 Plante, J.P. *et al.* (2009) Oligobenzamide proteomimetic inhibitors of the p53-hDM2 protein-protein interaction. *Chem. Commun. (Camb).* 34, 5091–5093

43 Joseph, T.L. *et al.* (2010) Differential binding of p53 and nutlin to MDM2 and MDMX: computational studies. *Cell Cycle* 9, 1167–1181

44 Shangary, S. and Wang, S. (2009) Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. *Annu. Rev. Pharmacol. Toxicol.* 49, 223–241

45 Wade, M. *et al.* (2006) Hdmx modulates the outcome of p53 activation in human tumor cells. *J. Biol. Chem.* 281, 33036–33044

46 Hayashi, R. *et al.* (2009) N-acylpolyamine inhibitors of HDM2 and HDMX binding to p53. *Bioorg. Med. Chem.* 17, 7884–7893

47 Finch, R.A. *et al.* (2002) Mdmx Is a negative regulator of p53 activity in vivo. *Cancer Res.* 62, 3221–3225

48 Migliorini, D. *et al.* (2002) Mdm4 (Mdmx) regulates p53-induced growth arrest and neuronal cell death during early embryonic mouse development. *Mol. Cell. Biol.* 22, 5527–5538

49 Plante, J. *et al.* (2008) Synthesis of functionalised aromatic oligamide rods. *Org. Biomol. Chem.* 6, 138–146

50 Aimoto, S. (2001) Contemporary methods for peptide and protein synthesis. *Curr. Org. Chem.* 5, 45–87

51 Azzarito, V. *et al.* (2012) 2-O-Alkylated para-benzamide α-helix mimetics: The role of scaffold curvature. *Org. Biomol. Chem.* 10, 6469–6472

52 Kulikov, O. V. *et al.* (2013) Design and Synthesis of Oligoamide-Based Double α-Helix Mimetics. *European J. Org. Chem.* 2013, 3433–3445

53 Huc, I. (2004) Aromatic Oligoamide Foldamers. *European J. Org. Chem.* 2004, 17–29

54 Yamato, K. *et al.* (2012) Cavity-containing, backbone-rigidified foldamers and macrocycles. *Chem. Commun. (Camb).* 48, 12142–12158

55 Sánchez-García, D. *et al.* (2009) Nanosized hybrid oligoamide foldamers: aromatic templates for the folding of multiple aliphatic units. *J. Am. Chem. Soc.* 131, 8642–8648

56 Daranas, A.H. *et al.* (2004) Thermodynamics of binding of D-galactose and deoxy derivatives thereof to the L-arabinose-binding protein. *J. Am. Chem. Soc.* 126, 11870–11876

57 Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* 437, 640–647

58 Shimokhina, N. *et al.* (2006) Contribution of ligand desolvation to binding thermodynamics in a ligand-protein interaction. *Angew. Chemie - Int. Ed.* 45, 6374–6376

59 Chodera, J.D.J. and Mobley, D.L.D. (2013) Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Biophysics (Oxf).* 42, 121–142

60 Homans, S.W. (2007) Water, water everywhere - except where it matters? *Drug Discov. Today* 12, 534–539

61 Olsson, T.S.G. *et al.* (2011) Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein Sci.* 20, 1607–1618

62 Tsai, C.J. *et al.* (1997) Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci.* 6, 53–64

63 Tsai, C.J. *et al.* (1997) Structural motifs at protein-protein interfaces: protein cores versus two-state and three-state model complexes. *Protein Sci.* 6, 1793–1805

64 Jones, S. and Thornton, J.M. (1997) Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.* 272, 121–132

65 Sheinerman, F.B. and Honig, B. (2002) On the role of electrostatic interactions in the design of protein-protein interfaces. *J. Mol. Biol.* 318, 161–177

66 Norel, R. *et al.* (2001) Electrostatic contributions to protein-protein interactions : Fast energetic filters for docking and their physical basis. 10, 2147–2161

67 Kresheck, G.C. *et al.* (1995) Calorimetric studies on the interaction of horse ferricytochrome c and yeast cytochrome c peroxidase. *Biochemistry* 34, 8398–8405

68 Bogan, a a and Thorn, K.S. (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 280, 1–9

69 Winter, C. *et al.* (2006) SCOPPI: a structural classification of protein-protein interfaces. *Nucleic acids Res.* 34, D310–D314

70 Li, Y. *et al.* (2005) Magnitude of the hydrophobic effect at central versus peripheral sites in protein-protein interfaces. *Structure* 13, 297–307

71 Bordner, A.J. and Abagyan, R. (2005) Statistical analysis and prediction of protein-protein interfaces. *Proteins* 60, 353–366

72 Clackson, T. and Wells, J. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* 267, 383–386

73 Keskin, O. *et al.* (2005) Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* 345, 1281–1294

74 Hu, Z. *et al.* (2000) Conservation of polar residues as hot spots at protein interfaces. *Proteins* 39, 331–342

75 Rajamani, D. *et al.* (2004) Anchor residues in protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11287–11292

76 Smith, G.R. *et al.* (2005) The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. *J. Mol. Biol.* 347, 1077–1101

77 Li, X. *et al.* (2004) Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. *J. Mol. Biol.* 344, 781–795

78 Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208

79 Dastidar, S.G. *et al.* (2008) Multiple peptide conformations give rise to similar binding affinities: molecular simulations of p53-MDM2. *J. Am. Chem. Soc.* 130, 13514–13515

80 Grässlin, A. *et al.* (2009) Thermodynamic and computational studies on the binding of p53-derived peptides and peptidomimetic inhibitors to HDM2. *Chembiochem* 10, 1360–1368

81 Lea, W. and Simeonov, A. (2011) Fluorescence polarization assays in small molecule screening. *Expert Opin. Drug Discov.* 6, 17–32

82 Parola, A. and Piazza, R. (2004) Particle thermophoresis in liquids. *Eur. Phys. J. E* 15, 255–263

83 Lodish, H. *et al.* (2004) *Molecular Cell Biology*, 5th Ed.W. H. Freeman and Company, New York, NY, USA.

84 Fan, X. *et al.* (2008) Sensitive optical biosensors for unlabeled targets: a review. *Anal. Chim. Acta* 620, 8–26

85 Vidal, M.M.B. *et al.* (1993) Determination of probable alcohol yield in musts by means of an SPR optical sensor. 11, 455–459

86 Scarano, S. *et al.* (2010) Surface plasmon resonance imaging for affinity-based biosensors. *Biosens. Bioelectron.* 25, 957–966

87 Swann, M.J. *et al.* (2004) Dual-polarization interferometry: an analytical technique to measure changes in protein structure in real time, to determine the stoichiometry of binding events, and to differentiate between specific and nonspecific interactions. *Anal. Biochem.* 329, 190–198

88 Leavitt, S. and Freire, E. (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* 11, 560–566

89 Ladbury, J. *et al.* (2009) Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat. Rev. Drug Discov.* 9, 23–27

90 Lakowicz, J.R. (2007) *Principles of Fluorescence Spectroscopy*, 3rd Ed.Springer Science & Business Media, New York, NY, USA.

91 Clapp, A.R. *et al.* (2006) Förster resonance energy transfer investigations using quantum-dot fluorophores. *Chemphyschem* 7, 47–57

92 Förster, T. (1959) 10th Spiers Memorial Lecture. Transfer mechanisms of electronic excitation. *Discuss. Faraday Soc.* 27, 7

93 Patterson, G.H. *et al.* (2000) Förster distances between green fluorescent protein pairs. *Anal. Biochem.* 284, 438–440

94 Gadella, Theodorus, W., J. (2009) *FRET and FLIM Techniques (Laboratory Techniques in Biochemistry and Molecular Biology)*, Elsevier B.V., Oxford, UK.

95 Mere, L. *et al.* (1999) Miniaturized FRET assays and microfluidics: key components screening. 4, 363–369

96 Kon, T. *et al.* (2005) ATP hydrolysis cycle-dependent tail motions in cytoplasmic dynein. *Nat. Struct. Mol. Biol.* 12, 513–519

97 Kritzer, J.A. *et al.* (2004) Helical beta-peptide inhibitors of the p53-hDM2 interaction. *J. Am. Chem. Soc.* 126, 9468–9469

98 Campbell, F. *et al.* (2010) N-alkylated oligoamide alpha-helical proteomimetics. *Org. Biomol. Chem.* 8, 2344–2351

99 Yang, T.-T.T. *et al.* (1996) Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* 24, 4592–4593

100 Shaner, N.C. *et al.* (2007) Advances in fluorescent protein technology. *J. Cell Sci.* 120, 4247–4260

101 Pollok, B.A. and Heim, R. (1999) Using GFP in FRET-based applications. *Trends Cell Biol.* 9, 57–60

102 Barondeau, D.P. *et al.* (2006) Structural evidence for an enolate intermediate in GFP fluorophore biosynthesis. *J. Am. Chem. Soc.* 128, 3166–3168

103 Barondeau, D.P. *et al.* (2003) Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12111–12116

104 Zhang, L. *et al.* (2006) Reaction progress of chromophore biogenesis in green fluorescent protein. *J. Am. Chem. Soc.* 128, 4766–4772

105 Shaner, N.C. *et al.* (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* 22, 1567–1572

106 Gross, L. a *et al.* (2000) The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11990–11995

107 Pochapsky, T.C. and Pochapsky, S.S. (2007) *NMR for Physical and Biological Sciences*, Garland Science, Oxford, UK.

108 James, T.L. (1998) Chapter 1: Fundamentals of NMR. In *Biophysics Textbook Online (BTOL)* Biophysical Society, Rockville, MD, USA.

109 Cavanagh, J. *et al.* (1996) *Protein NMR Spectroscopy: Principles and Practice*, 2nd Ed.Academic Press, London, UK.

110 Mlynárik, V. (1986) Measurement of spin coupling constants to quadrupolar nuclei via relaxation studies. *Prog. Nucl. Magn. Reson. Spectrosc.* 18, 277–305

111 Yu, H. (1999) Extending the size limit of protein nuclear magnetic resonance. *Proc. Natl. Acad. Sci.* 96, 332–334

112 Williamson, M.P. (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* 73, 1–16

113 Gräslund, S. *et al.* (2008) Protein production and purification. *Nat. Methods* 5, 135–146

114 Takahashi, H. and Shimada, I. (2010) Production of isotopically labeled heterologous proteins in non-E. coli prokaryotic and eukaryotic cells. *J. Biomol. NMR* 46, 3–10

115 Andreev, B.M. (2001) Separation of Hydrogen Isotopes in H2O-H2S System. *Sep. Sci. Technol.* 36, 1949–1989

116 Safonov, M.S. *et al.* (2001) Fractionation Of 15N In Countercurrent HNO3-NO System At Low Temperature. *Sep. Sci. Technol.* 36, 1991–2025

117 Li, Z.Z.-T. *et al.* (2008) Peptide mimics by linear arylamides: A structural and functional diversity test. *Acc. Chem. Res.* 41, 1343–1353

118 Lu, F. *et al.* (2006) Proteomimetic libraries: design, synthesis, and evaluation of p53-MDM2 interaction inhibitors. *J. Comb. Chem.* 8, 315–325

119 Wang, W. *et al.* (2013) Benzimidazole-2-one: a novel anchoring principle for antagonizing p53-Mdm2. *Bioorg. Med. Chem.* 21, 3982–3995

120 Riedinger, C. *et al.* (2011) Understanding small-molecule binding to MDM2: insights into structural effects of isoindolinone inhibitors from NMR spectroscopy. *Chem. Biol. Drug Des.* 77, 301–308

121 Stoll, R. *et al.* (2001) Chalcone derivatives antagonize interactions between the human oncoprotein MDM2 and p53. *Biochemistry* 40, 336–344

122 Rothweiler, U. *et al.* (2008) Isoquinolin-1-one inhibitors of the MDM2-p53 interaction. *ChemMedChem* 3, 1118–1128

123 Schon, O. *et al.* (2002) Molecular Mechanism of the Interaction between MDM2 and p53. *J. Mol. Biol.* 323, 491–501

124 McCoy, M. a *et al.* (2003) Flexible lid to the p53-binding domain of human Mdm2: implications for p53 regulation. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1645–1648

125 Showalter, S.A. *et al.* (2008) Quantitative lid dynamics of MDM2 reveals differential ligand binding modes of the p53-binding cleft. *J. Am. Chem. Soc.* 130, 6472–6478

126 Schon, O. *et al.* (2004) Binding of p53-derived Ligands to MDM2 Induces a Variety of Long Range Conformational Changes. *J. Mol. Biol.* 336, 197–202

127 Yu, G.W. *et al.* (2006) The central region of HDM2 provides a second binding site for p53. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1227–1232

128 Czarna, A. *et al.* (2010) Robust generation of lead compounds for protein-protein interactions by computational and MCR chemistry: p53/Hdm2 antagonists. *Angew. Chem. Int. Ed. Engl.* 49, 5352–5356

129 Guvench, O. and MacKerell, A.D. (2009) Computational evaluation of protein-small molecule binding. *Curr. Opin. Struct. Biol.* 19, 56–61

130 Scior, T. *et al.* (2012) Recognizing pitfalls in virtual screening: a critical review. *J. Chem. Inf. Model.* 52, 867–881

131 Rush, T.S. *et al.* (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* 48, 1489–1495

132 Griewel, A. *et al.* (2009) Conformational sampling for large-scale virtual screening: accuracy versus ensemble size. *J. Chem. Inf. Model.* 49, 2303–2311

133 Maass, P. *et al.* (2007) Recore: a fast and versatile method for scaffold hopping based on small molecule crystal structure conformations. *J. Chem. Inf. Model.* 47, 390–399

134 Galatin, P.S. and Abraham, D.J. (2001) QSAR: hydropathic analysis of inhibitors of the p53-mdm2 interaction. *Proteins* 45, 169–175

135 Baudin, P. and Pacalet, A. (2002) Caveat: A tool for software validation. *Proc. 2002 Int. Conf. Dependable Syst. Networks*

136 Klebe, G. (2006) Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today* 11, 580–594

137 Hajduk, P.J. *et al.* (2005) Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* 48, 2518–2525

138 Schneider, G. and Fechner, U. (2005) Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* 4, 649–663

139 Sousa, S.F. *et al.* (2006) Protein-ligand docking: current status and future challenges. *Proteins* 65, 15–26

140 Hajduk, P.J. and Greer, J. (2007) A decade of fragment-based drug design: Strategic advances and lessons learned. *Nat. Rev. Drug Discov.* 6, 211–219

141 Rueda, M. *et al.* (2010) Recipes for the selection of experimental protein conformations for virtual screening. *J. Chem. Inf. Model.* 50, 186–193

142 Cheng, L.S. *et al.* (2008) Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. *J. Med. Chem.* 51, 3878–3894

143 Rueda, M. *et al.* (2009) Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. *J. Chem. Inf. Model.* 49, 716–725

144 Craig, I.R. *et al.* (2010) Ensemble docking into multiple crystallographically derived protein structures: An evaluation based on the statistical analysis of enrichments. *J. Chem. Inf. Model.* 50, 511–524

145 Armen, R.S. *et al.* (2009) An evaluation of explicit receptor flexibility in molecular docking using molecular dynamics and torsion angle molecular dynamics. *J. Chem. Theory Comput.* 5, 2909–2923

146 Huang, S.-Y.S.-Y. *et al.* (2010) Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* 12, 12899–12908

147 Eldridge, M.D. *et al.* (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided. Mol. Des.* 11, 425–445

148 Wang, R. *et al.* (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided. Mol. Des.* 16, 11–26

149 Rarey, M. *et al.* (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470–489

150 Muegge, I. and Martin, Y. (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* 42, 791–804

151 Morris, G.M. *et al.* (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791

152 Trott, O. and Olson, A. (2010) Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461

153 Kramer, B. *et al.* (1999) Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins* 241, 228–241

154 Kitchen, D.B. *et al.* (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3, 935–949

155 Bursulaya, B. and Totrov, M. (2003) Comparative study of several algorithms for flexible ligand docking. *J. Comput. Aided. Mol. Des.* 17, 755–763

156 Morris, G.G.M. *et al.* (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662

157 Baxter, J. (1981) Local optima avoidance in depot location. *J. Oper. Res. Soc.* 32, 815–819

158 Nocedal, J. *et al.* (1999) *Numerical Optimization*, 1st Ed.Springer, New York, NY, USA.

159 Schneider, G. (2010) Virtual screening: an endless staircase? *Nat. Rev. Drug Discov.* 9, 273–276

160 Warren, G.L. *et al.* (2006) A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 49, 5912–5931

161 Kontoyianni, M. *et al.* (2005) Evaluation of library ranking efficacy in virtual screening. *J. Comput. Chem.* 26, 11–22

162 Perola, E. *et al.* (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins Struct. Funct. Genet.* 56, 235–249

163 Merz, K.M. (2010) Limits of Free Energy Computation for Protein− Ligand Interactions. *J. Chem. Theory Comput.* 6, 1769–1776

164 Cheatham, T. and Miller, J. (1995) Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* 117, 4193–4194

165 Frenkel, D. and Smit, B. (2002) *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, San Diego, CA, USA.

166 Leach, Andrew, R. *Molecular modelling: principles and applications*, 2nd Ed.Prentice Education, Harlow, UK.

167 Kaplan, Ilya, G. (2006) *Intermolecular Interactions*, John Wiley & Sons Ltd, Chichester, UK.

168 Slivka, P.F. *et al.* (2009) A peptide antagonist of the TLR4-MD2 interaction. *Chembiochem* 10, 645–649

169 Senn, H.M. and Thiel, W. (2009) QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed. Engl.* 48, 1198–1229

170 Ponder, J.W. and Case, D. a (2003) Force fields for protein simulations. *Adv. Protein Chem.* 66, 27–85

171 Vemparala, S. *et al.* (2006) Ab initio calculations of intramolecular parameters for a class of arylamide polymers. *J. Comput. Chem.* 27, 693–700

172 Fuller, J.C. *et al.* (2012) Modeling of arylamide helix mimetics in the p53 peptide binding site of hDM2 suggests parallel and anti-parallel conformations are both stable. *PLoS One* 7, e43253

173 Liu, Z. *et al.* (2011) An ab initio molecular orbital study of intramolecular hydrogen bonding in ortho-substituted arylamides: implications for the parameterization of molecular mechanics force fields. *J. Comput. Chem.* 32, 1846–1858

174 Galan, J.F. *et al.* (2009) Intramolecular hydrogen bonding in ortho-substituted arylamide oligomers: a computational and experimental study of ortho-fluoro- and ortho-chloro-N-methylbenzamides. *J. Phys. Chem. B* 113, 12809–12815

175 Nishikawa, J. *et al.* (2005) Internal rotations of aromatic polyamides: a density functional theory study. *J. Mol. Struct.* 741, 221–228

176 Gasteiger, J. and Marsili, M. (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* 36, 3219–3228

177 Dupradeau, F. and Pigache, A. (2010) The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* 12, 7821–7839

178 Szabo, A. and Ostlund, N.S. (1996) *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 2nd Ed.Dover Publications, Mineola, NY, USA.

179 Bayly, C.I. *et al.* (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97, 10269–10280

180 Shirts, M.R. *et al.* (2007) Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J. Phys. Chem. B* 111, 13052–13063

181 Anslyn, Eric, V. and Dougherty, Dennis, A. (2005) *Modern Physical Organic Chemistry*, Universiy Science Books, South Orange, NJ, USA.

182 Ryckaert, J.-P.J. *et al.* (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341

183 Grossfield, A. and Zuckerman, D.M. (2009) Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. *Annu. Rep. Comput. Chem.* 5, 23–48

184 Lyman, E. and Zuckerman, D.M. (2006) Ensemble-based convergence analysis of biomolecular trajectories. *Biophys. J.* 91, 164–172

185 Hünenberger, P.H. (2005) Thermostat Algorithms for Molecular Dynamics Simulations. *Advcances Polym. Sci.* 173, 105–149

186 (2012) *Amber 12 reference manual*, The Amber Molecular Dynamics Consortium, University of California, Los Angeles, CA, USA.

187 Andersen, H.C. (1980) Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72, 2384–2393

188 Hoover, W.G. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* 31, 1695–1697

189 Hu, Y. and Sinnott, S.B. (2004) Constant temperature molecular dynamics simulations of energetic particle–solid collisions: comparison of temperature control methods. *J. Comput. Phys.* 200, 251–266

190 Tsai, D.H. (1979) The virial theorem and stress calculation in molecular dynamics. *J. Chem. Phys.* 70, 1375–1382

191 Berendsen, H.J.C. *et al.* (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690

192 Gilson, M.K. and Zhou, H.-X.H.-X. (2007) Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* 36, 21–42

193 Foloppe, N. and Hubbard, R. (2006) Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* 13, 3583–3608

194 Rizzo, R.C. *et al.* (2006) Estimation of absolute free energies of hydration using continuum methods: Accuracy of partial charge models and optimization of nonpolar contributions. *J. Chem. Theory Comput.* 2, 128–139

195 Moreira, I.S. *et al.* (2008) Protein–protein recognition: a computational mutagenesis study of the mdm2–p53 complex. *Theor. Chem. Acc.* 120, 533–542

196 Levy, R. and Zhang, L. (2003) On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.* 125, 9523–9530

197 Lum, K. *et al.* (1999) Hydrophobicity at small and large length scales. *J. Phys. Chem. …* 103, 4570–4577

198 Fogolari, F. *et al.* (1999) Biomolecular electrostatics with the linearized Poisson-Boltzmann equation. *Biophys. J.* 76, 1–16

199 Born, M. (1920) Volumen und hydratationswärme der ionen. *Zeitschrift für Phys.* 1, 45–48

200 Onufriev, A. *et al.* (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55, 383–394

201 Laurendeau, N.M. (2005) *Statistical thermodynamics*, Cambridge University Press, New York, NY, USA.

202 Cerjan, C. and Miller, W. (1981) On finding transition states. *J. Chem. Phys.* 75, 2800–2806

203 Fuller, J.C. *et al.* (2012) Configurational preferences of arylamide α-helix mimetics via alchemical free energy calculations of relative binding affinities. *J. Phys. Chem. B* 116, 10856–10869

204 Espinoza-Fonseca, L.M. and García-Machorro, J. (2008) Aromatic-aromatic interactions in the formation of the MDM2-p53 complex. *Biochem. Biophys. Res. Commun.* 370, 547–551

205 Verma, S. *et al.* (2013) Molecular Dynamics Investigation on the Inhibition of MDM2-p53 Interaction by Polyphenols. *Mol. Inform.* 32, 203–212

206 Zhong, H. and Carlson, H.A. (2005) Computational studies and peptidomimetic design for the human p53-MDM2 complex. *Proteins* 58, 222–234

207 Ding, Y. *et al.* (2008) Quantum mechanical studies of residue-specific hydrophobic interactions in p53-MDM2 binding. *J. Phys. Chem. B* 112, 11396–11401

208 Lm, E. *et al.* (2006) Conformational changes of the p53-binding cleft of MDM2 revealed by molecular dynamics simulations. *Biopolymers* 83, 365–373

209 Shan, B. *et al.* (2012) Competitive binding between dynamic p53 transactivation subdomains to human MDM2 protein: implications for regulating the p53·MDM2/MDMX interaction. *J. Biol. Chem.* 287, 30376–30384

210 Dastidar, S.G. *et al.* (2009) Modulation of p53 binding to MDM2: computational studies reveal important roles of Tyr100. *BMC Bioinformatics* 10, (15) S6

211 Dastidar, S.G. *et al.* (2011) Chemical states of the N-terminal "lid" of MDM2 regulate p53 binding: Simulations reveal complexities of modulation. *Cell Cycle* 10, 82–89

212 Northrup, S.H. *et al.* (1984) Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* 80, 1517

213 ElSawy, K.M. *et al.* (2013) On the interaction mechanisms of a p53 peptide and nutlin with the MDM2 and MDMX proteins: a Brownian dynamics study. *Cell Cycle* 12, 394–404

214 Chen, J. *et al.* (2011) Insight into mechanism of small molecule inhibitors of the MDM2-p53 interaction: molecular dynamics simulation and free energy analysis. *J. Mol. Graph. Model.* 30, 46–53

215 Ståhle, L. and Wold, S. (1989) Analysis of variance (ANOVA). *Chemom. Intell. Lab. Syst.* 6, 259–272

216 Daylight CIS Ltd. (2008) , Daylight Chemical Information Systems Tutorials. , *Daylight Website*. [Online]. Available: http://www.daylight.com/dayhtml_tutorials/. [Accessed: 30-Nov-2014]

217 Guha, R. *et al.* (2006) The blue obelisk - interoperability in chemical informatics. *J. Chem. Inf. Model.* 46, 991–998

218 Halgren, T. a. (1999) MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* 20, 720–729

219 Irwin, J.J. and Shoichet, B.K. (2005) ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182

220 Irwin, J.J. *et al.* (2012) ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* 52, 1757–1768

221 Ertl, P. *et al.* (2000) Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* 43, 3714–3717

222 Wang, R. *et al.* (1997) A new atom-additive method for calculating partition coefficients. *J. Chem. Inf. Comput. Sci.* 2338, 615–621

223 Pettersen, E.F. *et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612

224 D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.L. *et al.* (2008) *AMBER 10*, University of California, San Francisco, CA, USA.

225 Pearlman, D. a. *et al.* (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91, 1–41

226 Case, D. a *et al.* (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688

227 Kazius, J. *et al.* (2005) Derivation and validation of toxicophores for mutagenicity prediction. *J. Med. Chem.* 48, 312–320

228 Stahl, M. and Rarey, M. (2001) Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem.* 44, 1035–1042

229 Park, H. and Jeon, Y.H. (2008) Toward the virtual screening of Cdc25A phosphatase inhibitors with the homology modeled protein structure. *J. Mol. Model.* 14, 833–841

230 Solis, F. and Wets, R. (1981) Minimization by random search techniques. *Math. Oper. Res.* 6, 19–30

231 (2004) *AutoDock Vina Manual*, Molecular Graphics Laboratory, Scripps Research Institute, La Jolla, CA, USA. [Online]. Available: //vina.scripps.edu/manual.html. [Accessed: 30-Nov-2014].

232 Duan, Y. *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24, 1999–2012

233 Frisch, M.J. *et al.* (2013) *Gaussian 03 Revision D.02*, Gaussian, Inc., Wallingford, CT, USA.

234 Frisch, M.J. *et al.* (2009) *Gaussian 09 Revision D.01*, Gaussian, Inc., Wallingford, CT, USA.

235 Connolly, M.L. (1983) Analytical molecular surface calculation. *J. Appl. Crystallogr.* 16, 548–558

236 Jorgensen, W.L. *et al.* (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926

237 Mark, P. and Nilsson, L. (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* 105, 9954–9960

238 Steinbrecher, T. *et al.* (2011) Soft-core potentials in thermodynamic integration: comparing one- and two-step transformations. *J. Comput. Chem.* 32, 3253–3263

239 Bruckner, S. and Boresch, S. (2011) Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* 32, 1320–1333

240 Miller, B.R. *et al.* (2012) MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* 8, 3314–3321

241 Metz, A. *et al.* (2012) Hot spots and transient pockets: predicting the determinants of small-molecule binding to a protein-protein interface. *J. Chem. Inf. Model.* 52, 120–133

242 Xu, L. *et al.* (2013) Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *J. Phys. Chem. B* 117, 8408–8421

243 Srivastava, H.K. and Sastry, G.N. (2012) Molecular dynamics investigation on a series of HIV protease inhibitors: assessing the performance of MM-PBSA and MM-GBSA approaches. *J. Chem. Inf. Model.* 52, 3088–3098

244 Prabhakaran, P. *et al.* (2012) Conformational properties of O-alkylated benzamides. *Tetrahedron* 68, 4485–4491

245 Long, K. *et al.* (2013) Microwave assisted solid phase synthesis of highly functionalized N-alkylated oligobenzamide α-helix mimetics. *Bioorg. Med. Chem.* 21, 4034–4040

246 Saito, S. *et al.* (1996) Theoretical studies on cis-amide preference in N-methylanilides.

247 Campbell, F. *et al.* (2007) Macrocyclic scaffolds derived from p-aminobenzoic acid. *Chem. Commun. (Camb).* 2007, 2240–2242

248 Tsui, V. and Case, D. (2000) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* 56, 275–291

249 Genheden, S. *et al.* (2012) The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant. *J. Chem. Inf. Model.* 52, 2079–2088

250 Pearson, K. (1896) Research article: Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. *Philos. Trans. A. Math. Phys. Eng. Sci.* 187, 253–318

251 Spearman, C. (1904) The proof and measurement of association between two things. *Am. J. Psychol.* 15, 72–101

252 Bewick, V. *et al.* (2005) Statistics review 14: Logistic regression. *Crit. Care* 9, 112–118

253 Kussie, P.H. *et al.* (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 7, 948–953

254 Shoichet, B.K. *et al.* (1999) Ligand solvation in molecular docking. *Proteins* 34, 4–16

255 Humphrey, W. *et al.* (1996) VMD : Visual Molecular Dynamics. 7855, 33–38

256 Prlić, A. *et al.* (2010) Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* 26, 2983–2985

257 Prlić, A. *et al.* (2012) BioJava: an open-source framework for bioinformatics in 2012. *Bioinformatics* 28, 2693–2695

258 Shindyalov, I.N. and Bourne, P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng. Des. Sel.* 11, 739–747

259 Finkelstein, A. and Ptitsyn, O. (1987) Why do globular proteins fit the limited set of foldin patterns? *Prog. Biophys. Mol. Biol.* 50, 171–190

260 Platts, J. a. *et al.* (1996) Directionality of hydrogen bonds to sulfur and oxygen. *J. Am. Chem. Soc.* 118, 2726–2733

261 Welch, B. (1947) The generalization of "Student"s' problem when several different population variances are involved. *Biometrika* 34, 28–35

262 Clayden, J. *et al.* (2001) *Organic Chemistry*, Oxford University Press, New York, NY, USA.

263 Gonzalez-Lopez de Turiso, F. *et al.* (2013) Rational design and binding mode duality of MDM2-p53 inhibitors. *J. Med. Chem.* 56, 4053–4070

264 Nagelkerke, N. (1991) A note on a general definition of the coefficient of determination. *Biometrika* 78, 691–692

265 Hou, T. *et al.* (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* 51, 69–82

266 Lu, Y. *et al.* (2006) Discovery of a nanomolar inhibitor of the human murine double minute 2 (MDM2)-p53 interaction through an integrated, virtual database screening strategy. *J. Med. Chem.* 49, 3759–3762

267 Shaw, R. and Mitchell-Olds, T. (1993) ANOVA for unbalanced data: an overview. *Ecology* 74, 1638–1645

268 Levene, H. (1960) Robust tests for equality of variances. In *Contributions to Probability and Statistics* (Olkin, I., ed), pp. 278–292, Stanford University Press, Redwood City, CA, USA.

269 Anderson, M.J. (2006) Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62, 245–253

270 Glantz, S.A. and Slinker, B.K. (1990) *Primer of Applied Regression and Analysis of Variance*, McGraw-Hill, New York, NY, USA.

271 Zhou, R. (2007) Replica exchange molecular dynamics method for protein folding simulation. *Methods Mol. Biol.* 350, 205–223

272 Knight, J.L. *et al.* (2009) λ-Dynamics free energy simulation methods. *J. Comput. Chem.* 30, 1692–1700

273 Czarna, A. *et al.* (2009) High affinity interaction of the p53 peptide-analogue with human Mdm2 and Mdmx. *Cell Cycle* 8, 1176–1184

274 R Development Core Team (2005) *R: A language and environment for statistical computing.*, R Foundation for Statistical Computing, Vienna, Austria.

275 Zhang, R. *et al.* (2004) Fluorescence polarization assay and inhibitor design for MDM2/p53 interaction. *Anal. Biochem.* 331, 138–146

276 Nikolovska-Coleska, Z. *et al.* (2004) Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal. Biochem.* 332, 261–273

277 Bevington, P.R. and Robinson, D.K. (2002) *Data Reduction and Error Analysis for the Physical Sciences*, 3rd Ed.McGraw-Hill, New York, NY, USA.

278 Goutelle, S. *et al.* (2008) The Hill equation: a review of its capabilities in pharmacological modelling. *Fundam. Clin. Pharmacol.* 22, 633–648

279 Hu, B. *et al.* (2007) Efficient p53 activation and apoptosis by simultaneous disruption of binding to MDM2 and MDMX. *Cancer Res.* 67, 8810–8817

280 Dömling, A. (2014) Hot, hotter, hottest. *Cell Cycle* 8, 1112–1113

281 Zimmermann, T. *et al.* (2002) Spectral imaging and linear un-mixing enables improved FRET efficiency with a novel GFP2-YFP FRET pair. *FEBS Lett.* 531, 245–249

282 Dixon, J.M. *et al.* (2005) PhotochemCAD 2: A Refined Program with Accompanying Spectral Databases for Photochemical Calculations. *Photochem. Photobiol.* 81, 212–213

283 Lai, Z. *et al.* (2000) Thermodynamics of p53 binding to hdm2(1-126): effects of phosphorylation and p53 peptide length. *Arch. Biochem. Biophys.* 381, 278–284

284 Fan, J.-Y.J.-Y. *et al.* (2008) Split mCherry as a new red bimolecular fluorescence complementation system for visualizing protein-protein interactions in living cells. *Biochem. Biophys. Res. Commun.* 367, 47–53

285 Sapsford, K.E. *et al.* (2006) Materials for fluorescence resonance energy transfer analysis: beyond traditional donor-acceptor combinations. *Angew. Chem. Int. Ed. Engl.* 45, 4562–4589

286 Boute, N. *et al.* (2002) The use of resonance energy transfer in high-throughput screening: BRET versus FRET. *Trends Pharmacol. Sci.* 23, 351–354

287 Uhrinova, S. *et al.* (2005) Structure of free MDM2 N-terminal domain reveals conformational adjustments that accompany p53-binding. *J. Mol. Biol.* 350, 587–598

288 Gasteiger, E. *et al.* (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (Walker, J. M., ed), pp. 571–607, Humana Press, New York, NY, USA.

289 Delaglio, F. *et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293

290 Vranken, W.F. *et al.* (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687–696

291 Kalverda, A.P. *et al.* (2009) Poxvirus K7 Protein Adopts a Bcl-2 Fold: Biochemical Mapping of Its Interactions with Human DEAD Box RNA Helicase DDX3. *J. Mol. Biol.* 385, 843–853

292 Ulrich, E.L. *et al.* (2008) BioMagResBank. *Nucleic Acids Res.* 36, D402–D408

293 Furet, P. *et al.* (2012) The central valine concept provides an entry in a new class of non peptide inhibitors of the p53-MDM2 interaction. *Bioorg. Med. Chem. Lett.* 22, 3498–3502

294 Popowicz, G.M. *et al.* (2010) Structures of low molecular weight inhibitors bound to MDMX and MDM2 reveal new approaches for p53-MDMX/MDM2 antagonist drug discovery. *Cell Cycle* 9, 1104–1111

295 Grzesiek, S. and Bax, A. (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* 99, 201–207

296 Shan, J. and Zheng, J.J. (2012) Virtual ligand screening combined with NMR to identify Dvl PDZ domain inhibitors targeting the wnt signaling. *Methods Mol. Biol.* 928, 17–28

297 Venkitakrishnan, R.P. *et al.* (2012) Use of NMR saturation transfer difference spectroscopy to study ligand binding to membrane proteins. *Methods Mol. Biol.* 914, 47–63

298 Estevão, M.S. *et al.* (2012) Indole based cyclooxygenase inhibitors: synthesis, biological evaluation, docking and NMR screening. *Eur. J. Med. Chem.* 54, 823–833

299 Rastelli, G. *et al.* (2010) Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* 31, 797–810

300 Fisher, R.A. (1921) On the "probables error" of a coefficient of correlation reduced from a small sample. *Metron* 1, 3–32

301 Caruso, J.C. and Cliff, N. (1997) Empirical Size, Coverage, and Power of Confidence Intervals for Spearman's Rho. *Educ. Psychol. Meas.* 57, 637–654

302 Bonett, D.G. and Wright, T.A. (2000) Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika* 65, 23–28

303 Kang, S.-S. and Larsen, M.D. (2012) Tests of independence in incomplete multi-way tables using likelihood functions. *J. Korean Stat. Soc.* 41, 189–198

304 Baraldi, A.N. and Enders, C.K. (2010) An introduction to modern missing data analyses. *J. Sch. Psychol.* 48, 5–37

305 DeCarlo, L.T. (1997) On the meaning and use of kurtosis. *Psychol. Methods* 2, 292–307

306 Cramer, D. (1998) *Fundamental Statistics for Social Research: Step-by-Step Calculations and Computer Techniques Using SPSS for Windows*, Routledge, Oxford, UK.

307 Khan, A. and Rayner, G. (2003) Robustness to Non-Normality of Common Tests for the Many-Sample Location Problem. *J. Appl. Math. Decis. Sci.* 7, 187–206