Structure-function Studies and

in silico Drug Discovery Targeting the Liver

Oncoprotein Gankyrin

by

Iain Murchland, B. Biotech. (Hons)

A thesis submitted in fulfillment of the requirements

for the degree of Doctor of Philosophy

School of Molecular and Biomedical Science

Discipline of Biochemistry

August 2012

Table of Contents

Abstract

Gankyrin is an ankyrin repeat protein known for its oncogenic effects and its up-regulation in the early stages of almost all cases of hepatocellular carcinoma, as well as other cancers. It mediates its oncogenic effects via a series of protein-protein interactions with several key regulators of the cell cycle including, but not limited to, pRb, CDK4 and Mdm2. A better understanding of these interactions, including their structural details, would facilitate the development of novel therapeutics targeting gankyrin, as well as improving our understanding of gankyrin and the cell cycle control mechanisms it alters.

The nature of gankyrin's interaction with pRb is the subject of some dispute within the literature, including whether gankyrin is in a folded or unfolded state when interacting with pRb. Previous studies have identified that hydroxylation of ankyrin repeat proteins by FIH-1 requires substrate unfolding. Therefore, an investigation of the hydroxylation of gankyrin and other substrates demonstrates that gankyrin does not unfold when in complex with the C-terminal domain of the proteasomal S6 ATPase subunit. Molecular dynamics simulations of FIH-1 in complex with substrates derived from ankyrin repeats suggest that local sequence effects, as well as the tendency of the repeat to unfold, can determine whether or not an ankyrin repeat protein can be hydroxylated by FIH-1. Using hydroxylation to detect protein unfolding, this study provides further evidence that gankyrin interacts with pRb in a folded state, and via the use of in silico methods, proposes a structural model of the interaction. Mutations in gankyrin and pRb that are predicted to disrupt the interaction on the basis of the model do not result in altered binding when subjected to *in vitro* testing. An *in silico* screen to identify small molecules that can disrupt the gankyrin-CDK4 interaction was undertaken to identify lead compounds for the development of novel therapeutics for hepatocellular carcinoma and as a case study in the development of protein-protein interaction inhibitors. No lead molecules were identified, but the study provides evidence that the structural properties of gankyrin are amenable to the discovery of lead

molecules with affinities in at least the low micromolar range. By attempting to recapitulate the findings of an *in vitro* lead discovery screen targeting the ankyrin repeat domain of RNaseL using in silico techniques, the conclusion was drawn that such techniques, without further improvements, are unlikely to be successful in identifying small molecule ligands of ankyrin repeat proteins like gankyrin.

On the basis of this study, we can conclude that the discovery of small molecule inhibitors of gankyrin remains a valid and plausible objective, but that typical in silico techniques are unlikely to fulfil it. While the details of the interaction between gankyrin and pRb remain uncertain and thus inaccessible to drug design efforts, it can now be concluded with greater certainty that gankyrin participates in this interaction in its folded form. It therefore must do so via a mechanism that differs from other proteins containing an LXCXE motif.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Iain Murchland

Friday, September 7, 2012

Acknowledgements

First I would like to thank Associate Professor Grant Booker for the opportunity to undertake this PhD in his laboratory, for his tutelage during the course of the program, and his patience while the thesis was being written. In addition I thank the School of Molecular & Biomedical Sciences, and in particular the Discipline of Biochemistry, as well as my co-supervisor Associate Professor Simon Pyke for their roles in helping to make me the scientist I am today, and for providing a thoroughly enjoyable environment in which to work.

On the business end of things, I thank the Australian Government for providing my scholarship, eResearch SA for the use of their high-performance computational and visualization resources, and the School support staff for lightening the load a little with their administrative assistance and buffer and media preparations. I also thank Phil Clements and the School of Chemistry and Physics for maintenance of and access to the NMR spectrometer.

I am grateful for all the friends I have made in the Booker and Wallace labs over the years (in particular Huang-Hui Chen, Ethan Chen, Steven Polyak, Belinda Ng, Karina Martin, Lungisa Mayende, Tatiana Soares Da Costa, John Sharley, John Wallace, Briony Forbes, Carlie Delaine, Clair Alvino, Kerrie McNeil, Shee Chee Ong and Harinda Rajapaksha) for their willingness to help – whether that be with the serious matters of science, or the even more serious of matters of cake and coffee over morning tea, a laugh or two, or the odd philosophical and ideological discussion. Special mentions go to Phillippa Smith for a thoroughly enjoyable year working together and her contribution to the work on the gankyrin-CDK4 interaction and to Cvetan Stojkoski for his tutelage in all things computational and a great friendship. I also thank the Shearwin lab for the friendship and continued professional development they have given so generously over my peri-doctoral period. To my friends and collaborators in the Peet lab (Dan Peet, Sarah Linke, Sarah Wilkins and Rachel Hampton-Smith) – thanks for your infectious enthusiasm and for being so easy to work with, and of course for your contribution to this work via protein preparations and hydroxylation

Chapter 1

Introduction

1.1 Ankyrin repeat proteins

Ankyrin repeats are a protein structural motif, with a consensus sequence length of 33 amino acids. They are found in a wide variety of predominantly eukaryotic organisms, and are abundant in the human proteome, being present in an estimated 6% of predicted protein sequences, and representing the third most abundant class of protein domain/motif (Bjorklund, Ekman et al. 2006).

1.1.1 Structural properties of ankyrin repeat proteins

The ankyrin repeat consensus sequence (Figure 1.1A) shows strong conservation of a number of buried hydrophobic residues that are structurally important alongside large sequence variations in positions corresponding to solvent-exposed residues. Structurally, the motif forms an L-shape, wherein the horizontal "stroke" is composed of two anti-parallel alpha helices, and a beta-hairpin-like structure forms the vertical stroke (Figure 1.1B). The repeats then stack together creating a long, slightly curved structure with a characteristic concave surface comprised of the "inner" faces of the L-shaped repeats (Figure 1.1C).

The N- and C-terminal motifs from an ankyrin repeat domain often do not conform well to the consensus sequence, since hydrophobic residues that would otherwise form key interactions with neighbouring repeats in the hydrophobic core become solvent-exposed in terminal repeats and thus have often evolved to hydrophilic residues in a strong departure from the consensus. Further, terminal repeats are often only partial repeats, leaving out the regions responsible for linking consecutive repeats. Together, these two factors mean that terminal repeats are often not successfully identified by the Hidden Markov Models of domain/motif discovery projects like Pfam (Sonnhammer, Eddy et al. 1997; Punta, Coggill et al. 2012) or SMART (Schultz, Milpetz et al. 1998; Letunic, Doerks et al. 2012). Thus, despite some proteins being identified by such methods as containing only a single ankyrin repeat, they are thought to most likely represent instances of two or three repeats in which the terminal repeat(s) have evaded detection.

A

Binz et al (2003)	`D--G-TPLHLAA-- GHLEIVEVLLKZGA DVNA-`
Tripp & Barrick (2007)	`SKDGNTPLHLAAKN GHAEIVKLLLAKGA SVNAR`
Mosavi et al (2002)	`DKNGRTPLHLAARN GHLEVVKLLLEAGA DVNAK`
SMART (50%)	`sGpTsLHhAsps sshcllchLlspus slst`

loop   loop

Degenerate amino acid codes

z	H, N, Y	p	polar (C,D,E,H,K,N,Q,R,S,T)
c	charged (D,E,H,K,R)	s	small (A,C,D,G,N,P,S,T,V)
h	hydrophobic (A,C,F,G,H,I,K,L,M,R,T,V,W,Y)	t	turn-like (A,C,D,E,G,H,K,N,Q,R,S,T)
l	aliphatic (I,L,V)	u	tiny (A,G,S)

B   **C**

90°

concave surface

Figure 1.1 (A) The ankyrin repeat consensus sequence. The top three lines show the consensus sequences used for the design of soluble, stable ankyrin repeat proteins, and are consensus sequences in the traditional sense. These represent something approximating the most-favoured amino acid at each position of the ankyrin repeat sequence. The underlined sequence appears in the original reference as the beginning of the consensus. For more details about the derivation of these sequences, refer to (Mosavi, Minor et al. 2002; Binz, Stumpp et al. 2003; Tripp and Barrick 2007). The fourth line represents the statistical consensus that describes 50% of the ankyrin repeat entries in the SMART database, thus giving a sense of the level and locations of diversity found in ankyrin repeat sequences. (B) The fold of a single ankyrin repeat, with the characteristic L-shaped structure (bottom panel). (C) A seven ankyrin repeat protein in ribbon (top) and surface (bottom) representations, demonstrating the stacking of individual ankyrin repeats with slight curvature, and the formation of a concave surface formed by the insides of each L-shaped repeat.

The existence or otherwise of such single ankyrin repeat-containing proteins is significant, since thus far single ankyrin repeats have been found not to fold independently, thus defining them as motifs rather than domains. Instead, there seems to be a requirement for the cooperative folding of at least two, or more commonly three or more motifs to form a stably folded structure (Zhang and Peng 2000; Mosavi, Minor et al. 2002; Mosavi, Cammett et al. 2004). The mechanism of folding of ankyrin-repeat proteins – and the related question of the "foldedness" of ankyrin repeat proteins at equilibrium – has become an active area of research. This on the one hand is driven by biological questions related to the functions of ankyrin repeat proteins and how our understanding of them may vary if there are significant populations of unfolded or partially folded proteins of this kind in the cell. On the other hand, it is also pursued as a path toward a more fundamental understanding of protein folding in general. Being composed of repeats in which residue-residue contacts are confined to within a single repeat or its two direct neighbours reduces the complexity of the folding problem from a theoretical perspective and correspondingly allows access to a simpler set of experimental investigations than would otherwise be possible (Barrick, Ferreiro et al. 2008).

Experiments looking at the folding transitions of ankyrin repeat proteins have found varied results. Looking at the totality of results, it appears that domains containing a small number of repeats fold via a simple two-state pathway (unfolded and folded) with no observable partially folded intermediate. The likelihood that folding proceeds via a more complex multi-state pathway is correlated with, but not exclusively determined by, a larger number of repeats. For instance, p16^{INK4A} and myotrophin, both contain 4 ankyrin repeats and have been shown to fold in a cooperative manner via a two-state pathway (Tang, Guralnick et al. 1999; Mosavi, Williams et al. 2002; Lowe and Itzhaki 2007). In contrast, studies of p19^{INK4D} and the ankyrin repeat domain of ankyrinR, containing 5 and 12 ankyrin repeats, respectively, have found evidence of intermediate states and thus a multistate folding process (Zeeb, Rosner et al. 2002; Low, Weininger et al. 2007; Werbeck and Itzhaki 2007). However the ankyrin repeat domain of Notch disrupts these

beginnings of a pattern, exhibiting simple two-state folding of 7 ankyrin repeats (Zweifel, Leahy et

al. 2003). Interestingly, simulations and atomic-force microscopy data suggest that the force

required to unfold (or refold) ankyrin repeat proteins is greater if initiated at the N-terminus

proceeding to the C-terminus, as compared to the opposite direction (Lee, Zeng et al. 2012). From

a theoretical perspective, this does not alter the folding equilibrium itself, but rather the rates of

conversion via the different pathways. Thus while one could expect folding/unfolding to proceed

preferentially via the C-terminally initiated pathway, one could also expect transitions via this

pathway to proceed more quickly. There is therefore no clear conclusion to be made about the

relative cellular abundance of partially folded intermediates derived from either pathway, or any

other biological consequences of this finding.

Biologically speaking, probably the more important question is the extent to which each of the

folded, unfolded or intermediate states of the ankyrin repeat proteins are populated at

equilibrium. Probing this question directly in a cellular context is difficult, but a combination of

biophysical analysis of isolated ankyrin repeat domains along with biological information provides

some insight into the matter. Firstly, it is notable that naturally-occurring ankyrin repeat proteins

tend to have lower thermodynamic stability than repeat domains that have been designed to the

consensus (Tripp and Barrick 2003). The implications of this alone are unclear, especially in light of

the fact that in the cellular context, factors such as molecular crowding, post-translational

modification, the presence of general binding partners (such as chaperones) or specific binding

partners will mean that the intrinsic thermodynamic properties of these isolated, unmodified

proteins will not be the sole determinant of where the equilibrium between the folded and

unfolded forms lies. It may be that the level of stability found in most native proteins is sufficient

for the equilibrium to lie strongly in favour of the folded form, and that there is therefore little or

no selective pressure encouraging the evolution of more stable proteins. Alternatively, it is equally

possible that there is a biological function for the unfolded form of the protein as well as (or

instead of) the folded form, or that there is a biological need for some structural plasticity that is

absent in more stable ankyrin repeat domains. Either of these scenarios would lead specifically to

the evolution of ankyrin repeat proteins with lower or even marginal stability.

This remains an area of uncertainty, though recent evidence that many ankyrin repeat

proteins are hydroxylated by FIH-1 (Cockman, Webb et al. 2009; Cockman, Webb et al. 2009) and

that the ankyrin repeat protein substrate is hydroxylated in an unfolded form (Coleman,

McDonough et al. 2007) suggest that unfolded ankyrin repeats are not entirely biologically

irrelevant. The fact that hydroxylation increases with the age of the protein shows that this

phenomenon is more than just the hydroxylation of nascent, unfolded peptides being

hydroxylated shortly after translation (Singleton, Trudgian et al. 2011). Additional evidence for the

biological importance of some partially unfolded ankyrin repeat proteins comes from the

NFκB/IκBα interaction. Multiple lines of evidence suggest that ankyrin repeats five and six of IκBα

are not well structured and that this forms an important part of its function in contributing to its

rapid proteolytic degradation, plasticity in the recognition of binding partners, and its ability to

promote dissociation of NFκB from its DNA binding site (Komives 2012).

It is now well established that mutation towards the consensus at conserved positions almost

always improve the stability of a repeat domain, while mutations away from the consensus are

destabilizing and mutations at positions that tolerate diversity have diverse effects (Tang,

Guralnick et al. 1999; Binz, Stumpp et al. 2003; Zweifel, Leahy et al. 2003). On the biological side, it

has been shown that many cancer-derived mutants of p16^{INK4A} are destabilizing (Tang, Guralnick et

al. 1999), and that artificially introducing stabilizing mutations into p16^{INK4A} make it resistant to the

effects of the cancer-derived destabilizing mutations as measured by a yeast two-hybrid

interaction assay with CDK4 (Cammett, Luo et al. 2003). Crucially, however, the stabilized version

of p16^{INK4A} did not exhibit any detectable increase in interaction with CDK4 in the absence of other

destabilizing mutations. From these data, it is clear that the folded form of p16^{INK4A} is the

functional form (at least in terms of CDK4/6 inhibition), and further that the unfolded state is not

significantly populated for the wild-type protein, since a stabilized form is no more active. On the other hand, a destabilizing mutation isolated via directed evolution of a designed ankyrin repeat protein has been shown, counter-intuitively, to *increase* its affinity for Her2, its binding partner (Zahnd, Wyler et al. 2007). In this case, strongly stabilized ankyrin repeat proteins exhibited slower on-rates than would be expected for diffusion-limited protein-protein interactions, whereas the destabilized mutant achieves a higher affinity by increasing the on-rate into the time-scales typical of such interactions. This, together with crystal structure data, led the authors to interpret this as evidence that the conformation of the protein that binds Her2 is sampled relatively rarely in the more stable ankyrin repeat proteins, and that the destabilized mutant samples the active conformation more frequently.

It is likely that there is no general rule that can be applied to ankyrin repeats and the biological implications of their foldedness. Rather, the ability to "tune" the stability of ankyrin repeat proteins likely represents another dimension of flexibility to allow the evolution of proteins suited to their particular individual purposes. In all likelihood, this will lead to a wide variety of structural properties among ankyrin repeat proteins. A myriad outcomes can be imagined, but some examples may be the evolution of highly stable ankyrin repeat folds in some cases, in others to folds with a degree of instability or plasticity to accommodate conformational changes associated with binding, and in yet others to highly unstable repeat domains which only fold upon binding to a partner protein, presenting opportunities to modulate the spatial arrangement of domains or the availability of signal or substrate peptides via the interaction of the partner protein.

1.1.2 Functional diversity & conservation among ankyrin repeat proteins

The ankyrin repeat domain itself has been firmly established to be a protein-protein interaction domain. This is thought to be driven by the fact that the ankyrin repeat consensus sequence tolerates a high degree of variation in the solvent-exposed residues while retaining structural integrity. This fact, combined with the naturally forming concave surface of the domain,

make it ideal for the evolution of a wide variety of binding epitopes that target a wide variety of interaction partners. Harnessing this idea, researchers have been able to use design and/or directed evolution approaches to develop wholly novel ankyrin repeat proteins (known as Designed Ankyrin Repeat Proteins or DARPins) with affinities in the nanomolar or even picomolar range to their protein of interest (Binz, Stumpp et al. 2003; Forrer, Stumpp et al. 2003; Zahnd, Wyler et al. 2007).

Of course, being a domain with such a versatile function, it is found in proteins with very diverse functions, as demonstrated by a cursory interrogation of the Gene Ontology (GO) annotations of ankyrin repeat proteins. Some 353 entries in the UniprotKB database (Uniprot-Consortium 2012) have GO associations out of the 934 human entries that contain ankyrin repeats as detected by SMART. According to this incomplete set of data, 54 out of a total 70 high-level entries in the "biological process" ontology are associated with at least one ankyrin repeat-containing protein, with the most common process annotation of ankyrin-repeat proteins being "signal transduction" (Table 1.1).

Mosavi (2004) and co-workers assembled a list of 27 verified protein-protein interactions involving ankyrin repeat domains. Aside from noting a similar diversity of functions associated with ankyrin repeat proteins, their collection also demonstrates that approximately half of the ankyrin repeat proteins listed are known to interact with more than one partner – and potentially a number of those listed with only one interaction partner also participate in other, as yet undiscovered, protein-protein interactions. The ability to evolve a single, rigid domain that interacts with multiple partners presents abundant opportunities for competition or cooperativity in the binding of partner proteins. This may in part explain the apparent prevalence of ankyrin repeat proteins involved in signal transduction, since competition and cooperativity are each capable of generating ultrasensitivity, a property that is necessary to produce the switch-like (off or on) behavior that is characteristic of many signaling responses (Buchler and Louis 2008). Equally, competition and cooperativity provide simple but important mechanisms for positive or negative cross-talk between signaling pathways (Sneppen, Krishna et al. 2010).

Table 1.1 Annotations of ankyrin repeat proteins in the Biological Process Gene Ontology. The annotations of 353 human ankyrin repeat proteins from the UniprotKB database were classified into 70 categories, and the number of ankyrin repeat proteins represented in each category (and percentage of all 353 proteins) is shown. Please note that a single protein may have multiple process annotations, meaning that the total number of entries across all categories exceeds 353.

Biological Process	Entries (%)	Biological Process	Entries (%)
signal transduction	143 (40.5%)	vesicle-mediated transport	7 (2%)
cellular nitrogen compound metabolic process	88 (24.9%)	developmental maturation	5 (1.4%)
anatomical structure development	61 (17.3%)	cell adhesion	5 (1.4%)
small molecule metabolic process	51 (14.4%)	biological process	5 (1.4%)
transport	48 (13.6%)	cell motility	5 (1.4%)
cell differentiation	48 (13.6%)	extracellular matrix organization	3 (0.8%)
biosynthetic process	47 (13.3%)	mRNA processing	3 (0.8%)
catabolic process	47 (13.3%)	aging	3 (0.8%)
protein modification process	41 (11.6%)	cell junction organization	3 (0.8%)
nucleobase-containing compound catabolic process	37 (10.5%)	membrane organization	3 (0.8%)
response to stress	37 (10.5%)	mitosis	2 (0.6%)
cell death	29 (8.2%)	translation	2 (0.6%)
cell cycle	23 (6.5%)	plasma membrane organization	2 (0.6%)
cellular component assembly	22 (6.2%)	circulatory system process	2 (0.6%)
transmembrane transport	21 (5.9%)	mitochondrion organization	1 (0.3%)
cell morphogenesis	18 (5.1%)	chromosome segregation	1 (0.3%)
homeostatic process	17 (4.8%)	carbohydrate metabolic process	1 (0.3%)
immune system process	17 (4.8%)	cofactor metabolic process	1 (0.3%)
growth	17 (4.8%)	cell division	1 (0.3%)
neurological system process	15 (4.2%)	ribonucleoprotein complex assembly	0 (0%)
cytoskeleton organization	15 (4.2%)	cell wall organization or biogenesis	0 (0%)
cell proliferation	15 (4.2%)	protein folding	0 (0%)
nucleocytoplasmic transport	14 (4%)	nitrogen cycle metabolic process	0 (0%)
locomotion	14 (4%)	generation of precursor metabolites and energy	0 (0%)
embryo development	13 (3.7%)	transposition	0 (0%)
DNA metabolic process	13 (3.7%)	secondary metabolic process	0 (0%)
lipid metabolic process	12 (3.4%)	photosynthesis	0 (0%)
cell-cell signaling	12 (3.4%)	sulfur compound metabolic process	0 (0%)
anatomical structure formation involved in morphogenesis	12 (3.4%)	ribosome biogenesis	0 (0%)
protein targeting	10 (2.8%)	symbiosis, encompassing mutualism through parasitism	0 (0%)
macromolecular complex assembly	10 (2.8%)	cytoskeleton-dependent intracellular transport	0 (0%)
reproduction	10 (2.8%)	tRNA metabolic process	0 (0%)
protein complex assembly	10 (2.8%)	vacuolar transport	0 (0%)
chromosome organization	10 (2.8%)	pigmentation	0 (0%)
cellular amino acid metabolic process	8 (2.3%)	protein maturation	0 (0%)

Some level of conservation in the binding of ankyrin repeat proteins to partner proteins has previously been noted. Mosavi *et al* note all 8 ankyrin repeat-containing complex structures in the PDB at the time of writing (2004) depict an interface composed almost exclusively of residues on the concave surface of the ankyrin repeat protein. The rapid expansion of PDB entries in recent years permits further analysis of this type. There are now 41 such structures in the database, including two in which the ankyrin repeats represented do not actually participate in a protein-protein interaction. Of the other 39, it remains true that residues on the concave surface of the ankyrin repeat domain contribute substantially to the interaction interface in most cases. But while this contact commonly forms all or most of the interaction surface, there are also numerous cases in which up to half of the interface lies outside the concave face. There are only two examples of structures in which an ankyrin repeat protein interacts with a second protein via an interface that lies entirely outside its concave surface. Both of these (PDB IDs 1YCS, 4A63) involve P53BP2, and the majority of the intermolecular contacts are provided by the adjacent SH3 domain, rather than the ankyrin repeat domain itself. In a couple of cases, such as the Notch-CSL-Mastermind complex or the IκBα/NF-κB heterodimer complex (PDB IDs 2F0I, 3NBN, 1NFI), one of the components of a ternary complex interacts with the ankyrin repeat protein without contacting the concave surface. However, these interactions are cooperative or even dependent on the interaction of the third component of the complex, which in turn occurs via the concave face. Thus, while the near-exclusive involvement of the concave surface of ankyrin repeat proteins in intermolecular interactions may not be as ubiquitous as previously thought, the available structural evidence supports the notion that this is the preferred site for protein binding. What is less clear is whether this preference is as absolute as implied by the available entries in the PDB, or whether it is more the case that interactions outside of this region tend to be less stable and thus less amenable to co-crystallisation.

Yet despite this apparent conservation in how ankyrin repeat proteins interact with other proteins, the same survey reveals significant diversity in another aspect of their function as

protein-protein interaction modules. That is, there appears to be little or no conservation of the characteristics of the proteins to which ankyrin repeats bind. Represented among the set of co-crystal structures are ankyrin repeat proteins bound to random coil peptides, as well as both beta sheet and alpha helical structural elements. Further, even among substrates with similar secondary structures at the interface, the orientation of these elements relative to the ankyrin repeats varies considerably. This diversity underscores the versatility of ankyrin repeat domains as protein-protein interaction modules, and adds to the case for the value and broad applicability of DARPins. In this context, it is unsurprising that in the search for DARPins that bind Her2, Zahnd and colleagues were able to generate DARPins that bound to distinct epitopes of the target (Zahnd, Pecorari et al. 2006).

1.2 Cell cycle regulation and cancer

1.2.1 Cell cycle overview

The process of cell division is conceptually divided into four phases: the S (synthesis) phase, in which the cell's DNA is synthesized; the M (mitosis) phase, in which the cell physically divides in two; and two gap phases, G1 and G2, preceding S and M phases, respectively. Progression through these various stages is regulated directly by the activity of a series of cyclin-dependent kinases (CDKs). The activity of these kinases is regulated via multiple mechanisms, but most fundamentally by the presence or absence of their requisite partner cyclin protein, and their phosphorylation state. Different CDKs are active at different stages of the cell cycle in pulses of activity that initiate the various steps required for replication (such as DNA synthesis, cell growth, mitosis and cytokinesis) before fading, allowing tight, sequential control of these processes.

Of course in most biological contexts, continual cell division via unrestricted progression through the cell cycle is not desirable. In order to control this, a major check point or restriction point exists at the transition from the G1 phase to the S phase. This is the logical place to implement overall control of the cycle, since passage through S phase results in duplication of the

chromosomes, thus committing the cell to another division. A second check point at the G2/M

transition prevents entry into mitosis if DNA replication has not successfully completed or DNA

damage is detected.

There are a variety of factors and signals that serve to inhibit progression past the G1/S check

point. Some are part of the ordinary functioning of cellular growth, such as the extracellular signals

that are responsible for so-called "contact inhibition" of growth that allows cells in a tissue to grow

to an appropriate density but no further. Others represent safeguards against abnormal

functioning, such as the inhibition of growth in response to DNA damage.

1.2.2 Key molecular regulators of the cell cycle

In terms of the overall control of the cell cycle, the transition from G1 into S phase is most

critical, as described above. This transition is most directly controlled by the cyclin D-dependent

kinases CDK4 and CDK6 (Baldin, Lukas et al. 1993; Resnitzky and Reed 1995). This class of CDK is

directly inhibited by two families of proteins. The INK4 family, comprised of $p16^{INK4A}$, $p15^{INK4B}$,

$p18^{INK4C}$ and $p19^{INK4D}$, specifically bind to and inhibit the cyclin D-dependent kinases both by

distorting their active site and forcing them into a conformation that is unable to bind cyclin D

(Hirai, Roussel et al. 1995; Parry, Bates et al. 1995; Russo, Tong et al. 1998), thus inhibiting passage

through the G1/S checkpoint. Their key role in controlling this checkpoint means that loss-of-

function mutations in these genes are commonly found in various types of cancer (Kamb, Gruis et

al. 1994; Nobori, Miura et al. 1994). In keeping with this, one might expect that higher levels of

p16 expression in tumour cells might be associated with slower tumour progression or a better

patient prognosis, however studies probing this question have come to varying conclusions

(Rocco, Schandl et al. 2002; Tong, Sun et al. 2011; Bammidi, Neerukonda et al. 2012; Zhao, Huang

et al. 2012).

In contrast, the p21 family of CDK inhibitors, comprised of p21[CIP1], p27[KIP1] and p57[KIP2] bind to a variety of CDKs(Sherr and Roberts 1995), including CDK4 and CDK6, but are thought to primarily exert their effect through inhibition of the cyclin E/CDK2 complex (Gartel and Radhakrishnan 2005). The ability of the p21 family to inhibit CDK4/6 remains somewhat uncertain and may be context dependent, with some studies suggesting that they promote cyclin D-CDK4 assembly and activity (LaBaer, Garrett et al. 1997; Cheng, Olivier et al. 1999), with others confirming the inhibitory effect on CDK4 observed in initial studies (He, Siddik et al. 2005). Interestingly, p21 family mutations are not a common feature of cancers, while a p21-deficient mouse does exhibit an increased rate of tumorigenesis, but this phenotype only becomes apparent at the relatively late stage of 16 months (Martin-Caballero, Flores et al. 2001). This is despite the fact that inhibition of CDKs is not the only mechanism by which p21 can inhibit cell cycle progression, with evidence that it can inhibit DNA replication directly (Chen, Peters et al. 1996; Oku, Ikeda et al. 1998), and despite p21 playing a key role in the DNA damage response. Paradoxically, some studies have suggested that p21 activity can accelerate tumour development and promote metastasis (Cheng, Xia et al. 2010) and is associated with poor patient prognosis (Yang, Klos et al. 2003). Recent work has suggested that p21 is required for cell migration, and that therefore proliferative and migratory cell states are mutually exclusive (Qian, Hulit et al. 2012). In this context, the role of p21 is considered as a means of continually switching between the two states, promoting alternate waves of proliferation and migration, thus reconciling the anti-proliferative functions of p21 and its association with particularly aggressive tumour types.

The DNA damage response and its master regulator p53 represent another significant mechanism of cell cycle control. p53 protein, a transcription factor, is stabilized and activated by a variety of signaling pathways upon detection of DNA damage in the cell, leading to the expression of a variety of genes in response, leading to cell cycle arrest (one of which is p21, as discussed above) and the activation of DNA repair mechanisms. In the case of excessive DNA damage, activation of p53 can induce apoptosis. Under normal circumstances, p53 is present at very low

concentrations in the cell, because it is actively degraded via the ubiquitin-proteasome pathway. The E3 ubiquitin ligase Mdm2 (often referred to as Hdm2 in humans) is directly responsible for ubiquitylation of p53, and thus in many cases, p53 stabilization is achieved by inactivation of Mdm2, although some Mdm2- and ubiquitin-independent mechanisms for activating p53 have been described (Di Conza, Mancini et al. 2012). Because of its central role in both inhibiting cell cycle progression and in preventing genomic instability, as well as its apoptosis-inducing activity, p53 is very commonly found to be mutated in cancers, since both of these characteristics are considered necessary to overcome in the transformation to cancerous growth.

The other family of transcription factors that is known to play a major role in the control of the cell cycle is the E2F family, whose members are functionally almost a mirror image of p53. E2Fs, in contrast to p53, are positive regulators of the cell cycle and are responsible for activating the genes necessary for passage through the G1/S checkpoint, including those required for DNA replication as well as the cyclins and CDKs required in the next stages of the cell cycle. E2Fs are also negatively and constitutively regulated by a partner protein, but in this case are sequestered in an inactive, extracellular form by their obligate partner pRb. Inhibition of E2Fs is relieved by the phosphorylation of pRb by the cyclin E- and cyclin D-dependent kinases at a multitude of sites, leading to both a disruption of the E2F-pRb protein-protein interaction, and to degradation of pRb via the proteasome. The E2Fs are then able to translocate to the nucleus, and activate its target genes, thus potentiating the G1/S transition. The availability of structural information for phosphorylated forms of pRb is beginning to reveal that inhibition of the E2F-pRb interaction occurs by both direct and allosteric mechanisms via only a subset of the known pRb phosphorylation sites (Burke, Hura et al. 2012). Additionally, pRb is now thought to similarly sequester and inhibit a range of other factors such as chromatin modifying enzymes as well as the previously characterized E2Fs (Gordon and Du 2011), suggesting that phosphorylation events at different sites are likely to have differing, but potentially quite precise effects on the behavior of pRb and its binding partners.

14

1.3 Gankyrin

1.3.1 Oncogenic properties of gankyrin

Gankyrin is a seven ankyrin repeat protein, identified as an oncoprotein that was up-regulated in clinical samples of hepato-cellular carcinoma (HCC). Functional studies showed that expression of gankyrin in NIH/3T3 cells increased cell growth rate and enabled formation of colonies in soft agar, indicating a tumour-like phenotype. The tumorigenic activity of gankyrin was confirmed by inoculation of nude mice with gankyrin-expressing NIH/3T3 clones, leading to the development of tumours within 30 days (Higashitsuji, Itoh et al. 2000). Since that time, gankyrin has also been implicated as a factor in the onset or progression of several other types of cancer, including colorectal cancer (Tang, Yang et al. 2010), pancreatic cancer (Meng, He et al. 2010), gastric cancer (Li, Zhang et al. 2009) and epithelial cancers (Ortiz, Ito et al. 2008; Rickman, Millon et al. 2008; Li, Knobloch et al. 2011). A further analysis of clinical samples of early stage HCC identified that gankyrin is up-regulated in cancerous tissue not only relative to healthy liver tissue, but also relative to non-cancerous yet abnormal growths that appear in the cirrhotic liver tissue of Hepatitis C sufferers (Llovet, Chen et al. 2006), making gankyrin a useful marker for differentiating between these two types of growth that are difficult to distinguish by other methods. Independent studies utilising immunohistochemistry first confirmed the finding that gankyrin expression is strongly correlated specifically with the early stages of carcinogenesis (Fu, Tan et al. 2004; Umemura, Itoh et al. 2008), while also demonstrating that gankyrin expression is not strongly predictive of cancerous growth overall (only 62.8% of HCC samples exhibited specific cytoplasmic staining)(Umemura, Itoh et al. 2008).

1.3.2 Normal cellular functions of gankyrin

Before the discovery of its oncogenic properties, gankyrin was first identified as a non-ATPase subunit of the regulatory 19S complex of the 26S proteasome and at the time was named p28 – which also led to the alternative name p28(GANK) as a combination of the two prior names. Gene

knockout studies of the homologous protein Nas6p in *S. cerevisiae* demonstrated that unlike other proteasome components, Nas6p (and presumably gankyrin) is not essential for cell survival (Hori, Kato et al. 1998). Nonetheless, analysis of the expression pattern of gankyrin in the same study found evidence of gankyrin mRNA in all human tissues examined, including lung, liver, kidney, pancreas, spleen, thymus, prostate, ovary, small intestine, colon and peripheral blood, with higher levels in the heart, brain, placenta, skeletal muscle and testis. In all cases, the level of gankyrin was lower than another novel 19S subunit, p40.5, though their level of expression appeared to be correlated.

All of this data is consistent with the idea that gankyrin is a non-essential, peripheral component of the ubiquitous 26S proteasome. It is rather difficult to reconcile, however, with gankyrin's apparent oncogenic role in early hepatocellular carcinoma. As an oncoprotein, expression of gankyrin would be expected to be correlated with cell proliferation, yet some of the tissues in which gankyrin was found to be most highly expressed (heart, brain, skeletal muscle) are thought to be largely non-proliferative.

Iwai *et al* (2003) and Qin *et al* (2003) both sought to explain the normal cellular role of gankyrin in light of its pro-proliferative properties, hypothesizing that gankyrin plays a role in promoting hepatocyte proliferation during liver regeneration. These studies provide substantial evidence that up-regulation of gankyrin is a feature of regenerative hepatocyte growth. Gankyrin was found to be up-regulated at the mRNA level in cultures of primary rat hepatocytes in response to a variety of mitogenic signals, as well as in clinical samples obtained from patients presenting with fulminant hepatic failure (used as a model of human liver regeneration) relative to samples from a healthy control group (Iwai, Marusawa et al. 2003). Up-regulation of gankyrin in the clinical samples was independent of the patients' serological reactivity to Hepatitis A, B or C antigens. The second study used partial hepatectomy in rats as a simpler, more direct model of liver regeneration, albeit in a non-human system. The findings complement those of the first study well,

showing that gankyrin is up-regulated in response to partial hepatectomy both by Northern blot and immunohistochemistry, with gankyrin expression peaking around 72 hours after the operation and coinciding with the peak of cell proliferation (Qin, Fu et al. 2003). Samples from rats subjected to a sham operation exhibited no statistically significant up-regulation of gankyrin, but, consistent with the findings of Hori *et al* (1998) gankyrin mRNA was still detected at a constant, low level in the control group.

1.3.3 Protein interaction partners of gankyrin

As a protein composed entirely of ankyrin repeats, it is no surprise that gankyrin has been found to interact with several protein partners – a fact which makes the molecular mechanisms of its oncogenicity all the more intriguing, given that it stands in contrast to most existing examples of oncoproteins, which typically either have a catalytic activity (including kinases of various types such as HER2/neu, Src-family kinases and c-Raf as well as GTPases like Ras) or directly drive cell cycle progression via transcriptional regulation (such as myc). The major exceptions to this classification are secreted mitogenic factors like TGF-α, which arguably also exert their effect via a protein-protein interaction, but as direct, secreted stimulants of their cognate receptors and their downstream signaling cascades, these are conceptually quite different to gankyrin, as will shortly be made clear. Rather than directly activating a particular signaling pathway, gankyrin acts via a series of protein-protein interactions that modulate the activities of several cellular factors (summarised in Figure 1.2). Mechanistically, the most similar oncoproteins are the adaptor proteins Grb2 and Crk, which similarly act without catalytic activity via protein-protein interactions, but interact with significantly fewer partners than gankyrin. The various interactions of gankyrin, to the extent currently known, are summarized below.

In exploring the oncogenicity of gankyrin, Higashitsuji *et al* (2000) identified that gankyrin promotes the degradation of the retinoblastoma protein (pRb) and contains the LXCXE motif in common with other pRb-binding proteins. Co-immunoprecipitation and GST pull-down

experiments showed that gankyrin and pRb interact in cell lysates in a LXCXE-dependent manner, and this was followed later with clear evidence of a direct interaction between the two by GST pull-down between the purified proteins alone, along with evidence that the pocket domain of pRb was the region responsible for the interaction with gankyrin (Li and Tsai 2002). While the existence of this interaction and its importance for the oncogenic activity of gankyrin is now well established, the molecular details of the interaction remain uncertain. In particular, as additional data about the interaction has become available, it appears less likely that the gankyrin-pRb interaction occurs in the same way as that between pRb and other LXCXE-containing proteins. This will be discussed in greater detail in Chapter 3.
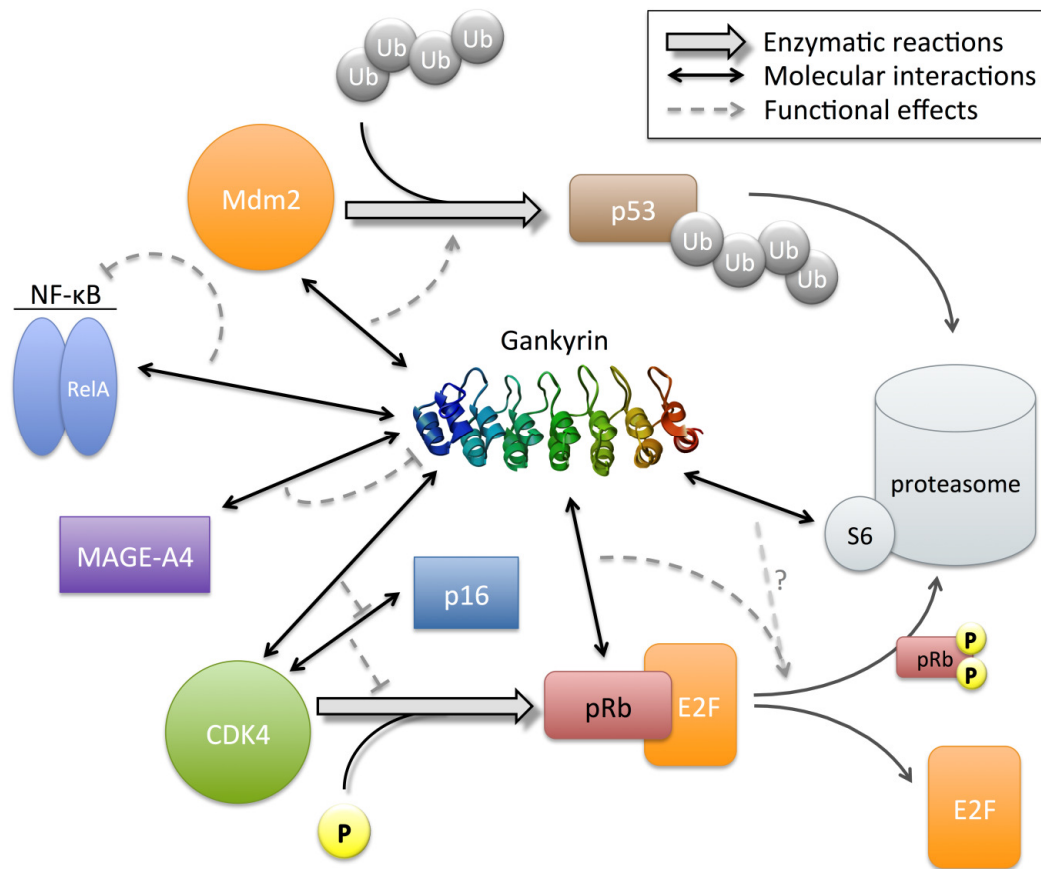
Figure 1.2 A schematic summary of the protein-protein interactions of gankyrin, and their effect on various cell cycle regulatory processes.

Having been isolated independently as a component of the proteasome, gankyrin obviously interacts with one or more other components of this large protein complex. Indeed, a yeast two-hybrid screen using the S6 ATPase (referred to as MS73) subunit of the proteasome as bait identified gankyrin (referred to as 73BP) as an interaction partner (Dawson, Hastings et al. 1997), and further evidence for the interaction was obtained by demonstrating that both endogenous and HA-tagged S6 co-purify with His-tagged gankyrin when isolated from co-expressing HEK293 cells in a gankyrin-dependent manner, and that the interaction occurs via the C-terminal domain of the S6 ATPase (Dawson, Apcher et al. 2002). It was further shown that both S6 and gankyrin are present in the 19S regulatory complex of the proteasome, although gankyrin does not appear to be stably associated with this complex, and can be detected in other complexes as well (Hori, Kato et al. 1998; Dawson, Apcher et al. 2002). In light of the already established association of gankyrin with the proteasome, Higashitsuji *et al* (2000) also demonstrated that gankyrin accelerates the degradation of pRb *in vitro* which implies a role for gankyrin's interactions with both pRb and the proteasome. However, the dependence of this phenomenon on gankyrin's association with the proteasome has not been shown directly.

Dawson *et al* (2002) also found that CDK4 co-precipitates with His-FLAG-tagged gankyrin (and vice-versa) from U2-OS cell lysates, building a picture of gankyrin acting at several stages in the degradation of pRb, since CDK4 is one of the kinases responsible for hyper-phosphorylation of pRb. "Complexes" of gankyrin, S6 and CDK4 were also reportedly observed, though because these were obtained by pull-down of gankyrin, there is in fact no evidence that what was observed was not two discrete populations of gankyrin, bound to CDK4 and S6, respectively. Independently, the gankyrin-CDK4 interaction was shown to be direct by both yeast two-hybrid assay and *in vitro* pull-down assays. More importantly, however, it was shown that gankyrin competes with the CDK4 inhibitors p16 and p18 (also ankyrin repeat proteins), thus de-repressing CDK4/Cyclin-D2 activity. Significantly, though, addition of gankyrin alone to a CDK4 activity assay does not increase the phosphorylation of pRb relative to a no-gankyrin control, indicating that binding of gankyrin does

not intrinsically increase the activity of CDK4, but rather acts by releasing the inhibition of the INK4 family of CDK inhibitors. Another interesting insight gained by this series of experiments is that C-terminally truncated gankyrin constructs are equally able to de-repress CDK4 activity. These observations offer important insights into what the gankyrin/CDK4 complex may look like (discussed further in Chapter 5) while also revealing that gankyrin's interaction with CDK4 is structurally independent of that with pRb, since constructs that have been shown to lack binding to pRb retain the ability to bind CDK4. An additional insight here is that since such gankyrin constructs are as effective as wild-type gankyrin at promoting pRb phosphorylation (in competition with p16), we can see that there is no evidence for any cooperativity (nor competition) in the binding of CDK4 and pRb to gankyrin. Due to this series of interactions with CDK4, pRb, the proteasomal S6 ATPase, and competition with the INK4 inhibitors, gankyrin justifiably has come to be seen by some as an additional component of the INK4-CDK4/6-pRb cell-cycle regulatory pathway.

Other studies have exposed a broader role for gankyrin than simply being a modulator of pRb degradation, though. Gankyrin was found to immunoprecipitate in a complex containing Mdm2 and p53 from U-2 OS cells. *In vitro* pull-down experiments using purified GST-gankyrin and *in vitro*-translated Mdm2 or recombinantly expressed p53 provided evidence of a direct interaction between gankyrin and Mdm2, but no interaction between gankyrin and p53. The interaction between Mdm2 and gankyrin was shown to increase mono- and poly-ubiquitylation of p53 in an *in vitro* ubiquitylation assay in an Mdm2-dependent fashion, and to de-stabilise p53 in cell culture. In common with its interactions with S6 and pRb but not CDK4, the gankyrin-Mdm2 interaction was not robust to deletions of any part of gankyrin, whereas deletions of various portions of Mdm2 showed that the interaction is dependent on the region adjacent to the RING domain. Based on the experiments conducted, the region encompassing the Zinc-finger domain to the beginning of the RING domain (residues 322-437) is the minimum region sufficient to interact with gankyrin,

though truncations at the N-terminal end of this region have not been thoroughly explored

(Higashitsuji, Itoh et al. 2005; Higashitsuji, Liu et al. 2005).

In a further diversification of the pathways in which gankyrin is seen to act, a direct protein-

protein interaction with RelA has also been observed, both by pull-down from cell lysates and

using purified proteins (Higashitsuji, Liu et al. 2007). RelA (also known as p65) is, along with p50,

one component of the most common form of the NF-κB family of heterodimeric transcription

factors. Hyper-activation of NF-κB has been shown to inhibit apoptosis (Burstein and Duckett

2003; Kucharczak, Simmons et al. 2003) and has been identified as a key element of the

development of some cancers (Basseres and Baldwin 2006; Kang, Ji et al. 2012; Yang, Kantrow et

al. 2012). Intriguingly, however, the interaction between gankyrin and RelA promotes cytoplasmic

localization of NF-κB (Chen, Li et al. 2007; Higashitsuji, Liu et al. 2007), thus casting gankyrin as a

negative regulator of NF-κB, in contrast to gankyrin's other pro-mitotic functions. While this seems

somewhat paradoxical, there is some evidence that in hepatocytes specifically, inactive NF-κB is

associated with increased hepatocarcinogenesis (Maeda, Kamata et al. 2005; Luedde, Beraza et al.

2007).

To date, the only evidence of a negative regulation of gankyrin, is its interaction with MAGE-

A4. The function of the MAGE super-family of proteins remains largely unknown. They remain the

subject of particular interest in the field of immune-targeted anti-tumour therapy though, since

they are not expressed in most adult tissues, but are displayed as tumour antigens by some

cancers. Initially identified in a yeast two-hybrid screen, the interaction was shown to reduce

anchorage-independent growth of gankyrin-transformed NIH/3T3 cells *in vitro* and to reduce

tumour volumes when these cells were used to inoculate athymic nude mice (Nagao, Higashitsuji

et al. 2003). A C-terminal fragment of MAGE-A4 that is the proteolytic product of the expression of

full-length MAGE-A4 in human cells has been shown to induce apoptosis by both p53-dependent

and p53-independent pathways. In this context, the notion that MAGE-A4 may stabilize p53 and

induce apoptosis by disrupting the gankyrin-Mdm2 interaction is attractive, but is refuted by

evidence that p53 stabilization is induced only by C-terminally truncated derivatives of MAGE-A4

that lack the ability to bind gankyrin (Sakurai, Itoh et al. 2004). Given that MAGE proteins are not

normally expressed in adult tissues, it might be expected that the MAGE-A4/gankyrin interaction is

biologically irrelevant, however several studies have identified expression of MAGE-A4 in

hepatocellular carcinomas (Roch, Kutup et al. 2010; Hussein, Ghareib et al. 2011; Hussein, Morad

et al. 2012).

Despite evidence of gankyrin participating in several different protein-protein interactions

with a variety of effects, there is relatively little data about the extent to which these different

interactions are competitive, cooperative or independent. We do know that the gankyrin/pRb

interaction is independent of interactions with both CDK4 and S6. However, the crystal structure

of the C-terminal domain of S6 in complex with gankyrin (PDB ID 2DVW) shows that S6 occupies

the entire concave surface of gankyrin (Nakamura, Nakano et al. 2007), which is usually favoured

as the site of intermolecular interactions among ankyrin repeat proteins. It would therefore be

expected that the CDK4 and S6 interactions would be competitive, and that it is possible or even

likely that the interactions with MDM2, RelA and/or MAGE-A4 may also compete with S6 and

possibly CDK4, and each other, or a subset of them could plausibly interact cooperatively.

1.3.4 Mechanisms of gankyrin's oncogenic activity

Assembling all of the available information about gankyrin's various interaction partners and

its effects on them, it becomes clear that gankyrin's oncogenic activity is primarily a result of its

destabilizing effect on the key tumour suppressor proteins pRb and p53. This may be either

augmented or counter-acted to some degree by its inhibition of NF-κB.

The data also suggest that up-regulation of gankyrin is likely a key and early step in the onset

of hepatocellular carcinoma, leading to increased cell proliferation, through a reduced threshold

for passage through the cell cycle checkpoints due to destabilization of p53 and pRb. However it is

likely that accumulation of further cellular defects are required in order to confer phenotypes associated with the later stages of cancer such as metastatic growth. Easier passage through cell cycle checkpoints presents a plausible mechanism to accelerate and facilitate the accumulation of such defects through, for instance, a diminished response to DNA damage. As more cancer-promoting genetic defects accumulate, the need to maintain high levels of gankyrin in order to promote cell division would be expected to dissipate, consistent with the observation that up-regulation of gankyrin is less commonly observed in such late-stage tumours.

The fact that inhibition of gankyrin by MAGE-A4 or knock-down by siRNA(Li, Fu et al. 2005) induces apoptosis via multiple pathways makes gankyrin an attractive target for therapeutic intervention. There could plausibly be opportunities for therapeutic intervention up-stream of gankyrin, aimed at correcting a presumed mis-regulation of gankyrin, rather than inhibiting it directly. However, based on the information available, it is not clear that such a target exists – or at least not in all cases. Given the close association between hepatocellular carcinoma and chronic hepatitis infection, a plausible hypothesis is that the up-regulation of gankyrin at the early stages of cancer is not the result of mis-regulation. Rather, it may be a direct, programmed result of the relatively constant proliferation of hepatocytes in response to the disease. In this proliferative state, with a reduced sensitivity to DNA damage, it may be other transformative mutations that ultimately represent a distinct departure from a healthy cellular state and perpetuate the proliferative state beyond its intended end. Nonetheless, the evidence that down-regulation of gankyrin is sufficient to arrest the cell cycle and induce apoptosis in HCC-derived cell lines suggest that gankyrin's role in helping to bypass DNA damage and cell cycle checkpoints remains a critical element in the development of HCC for some time, and that it remains an attractive target for therapeutic intervention.

1.4 Aims

Despite the existence of multiple experimental structures of gankyrin itself (Krzywda, Brzozowski et al. 2004; Yuan, Li et al. 2004), and studies providing insights into the structural nature of its interactions with CDK4 (Li and Tsai 2002; Mahajan, Guo et al. 2007; Guo, Mahajan et al. 2009) and the S6 ATPase proteasomal subunit (Nakamura, Nakano et al. 2007; Nakamura, Umehara et al. 2007), detailed structural information about most of gankyrin's other intermolecular interactions remains unknown. Given that these interactions are the very means by which gankyrin exerts its oncogenic effects, a lack of such information, and of whether any of these interactions compete with or reinforce each other represents a significant gap in our understanding of gankyrin's roles both in ordinary tissues and early-stage hepatocellular carcinoma. An understanding of the structural regions and epitopes of gankyrin that mediate its interactions and a quantitative view of its interaction network have the potential not only to contribute to a much richer understanding of gankyrin, but also to support and expand efforts for the development of therapeutics targeted towards hepatocellular carcinoma.

Thus, the aims of this work were to contribute to our knowledge of this structure-function relationship of gankyrin and to commence the development of a small molecule inhibitor of gankyrin. Specifically, this included the development and testing of a model of the interaction between gankyrin and pRb, and *in silico* and *in vitro* screening to identify lead compounds for development as gankyrin inhibitors. Lead compounds were sought not only for use in the development as a therapeutic, but also in order to add to the knowledge base of protein-protein interaction inhibitors and to contribute to our understanding of the challenges faced in this particular area of the broader drug discovery field. In order to support both the development of lead compounds and testing of the gankyrin-pRb binding model, the development of a quantitative gankyrin binding assay was a further objective of this work. The final aim was to investigate the mechanism and consequences of the hydroxylation of gankyrin as a means of

contributing to our understanding of the structure and function of gankyrin specifically, as well as ankyrin repeat proteins more generally.

1.5 A note on nomenclature

When gankyrin was first identified, six ankyrin repeats were found by sequence analysis, and designated repeats 1 to 6 (N-terminus to C-terminus). Subsequent structural studies revealed that the N-terminal region comprised an additional ankyrin repeat. Therefore, throughout this thesis, in order to preserve the original numbering of the ankyrin repeats, the most N-terminal ankyrin repeat (and the most recently identified) shall be referred to as ankyrin repeat 0, with subsequent repeats numbered 1 through 6. Thus full-length gankyrin, abbreviated as G0-6, contains seven ankyrin repeats.

Chapter 2

Materials and Methods

2.1 Abbreviations

2-5A	2',5'-linked oligoadenylate
A_{620}	Absorbance at 620nm
Amp	Ampicillin
Amp^R	Ampicillin resistance
bp	Base pairs
BSA	Bovine serum albumin
Chlor	Chloramphenicol
$Chlor^R$	Chloramphenicol resistance
DMSO	Di-methyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DOL	Degree of labeling
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylene diamine tetraacetic acid
FITC	Fluorescein isothiocyanate
GLB	Gel load buffer
GST	Glutathione S-transferase
HSQC	Heteronuclear Single Quantum Coherence
IMAC	Immobilised metal-ion affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobase pairs
K_d	Equilibrium dissociation constant
kDa	Kilo-Daltons
LB	Luria broth
MCS	Multiple cloning site
MQ	MilliQ™
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Enhancement Spectroscopy
OD_{600}	Optical density at 600nm wavelength
ONPG	o-nitrophenyl-β-D galactoside
PAGE	Poly-acrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PMSF	Phenylmethylsulfonylfluoride
ppm	Parts per million
pRbSP	pRb small pocket domain (residue 379-787)
psi	Pounds per square inch
RBS	Ribosome binding site
RE	Restriction endonuclease
SDS	Sodium dodecyl sulfate
TAE	Tris-acetate EDTA
TBS	Tris buffered saline
TE	Tris-EDTA
Tet	Tetracycline
Tet^R	Tetracycline resistance
TOCSY	Total Correlation Spectroscopy
Tris	Tris(hydroxymethyl)aminoethane
UV	Ultraviolet

2.2 Materials

2.2.1 Chemicals and Reagents

Chemical/Reagent	Supplier
^{13}C-glucose (uniformly labeled)	Spectra Stable Isotopes
^{15}NH$_4$Cl	Spectra Stable Isotopes
β-mercaptoethanol	Sigma
Acetic acid	BDH AnalaR
Agarose (DNA grade)	Progen
Ampicillin	Sigma
Bacto-agar	Difco
Bacto-tryptone	Difco
BigDye™ reaction mix	Applied Biosystems
BSA	Sigma
Bromophenol blue	Sigma
CaCl$_2$	Sigma
Casein	Sigma
Chloramphenicol	Amresco
Coomassie Brilliant Blue	Sigma
dNTPs	New England Biolabs
D$_2$O	Aldrich
DMSO	Sigma
DTT	Sigma
Ethanol	BDH AnalaR
Ethidium bromide	Sigma
EDTA	Sigma
FITC	Sigma
Glutathione agarose resin	ScientifiX
Glutathione (reduced form)	Sigma
Glycerol	Sigma
H$_2$NaPO$_4$	Sigma
HNa$_2$PO$_4$	Sigma
Hydrochloric acid	BDH AnalaR
Imidazole	Sigma
IPTG	BioVectra
Isopropanol	BDH AnalaR
Kanamycin	Sigma
KCl	BDH AnalaR
Methanol	BDH AnalaR
MgCl$_2$	BDH AnalaR
Mineral oil	Sigma
NaCl	BDH AnalaR

Chemical/Reagent	Supplier
NaN$_3$	Sigma
NaOH	BDH AnalaR
Ni-NTA agarose	GroPep
NiSO$_4$	Sigma
Pfu Turbo DNA polymerase	Stratagene
PMSF	Sigma
Protein Assay Dye reagent (Bradford reagent)	Bio-Rad
Restriction endonucleases	New England Biolabs
SDS	Sigma
T4 DNA ligase	New England Biolabs
Taq DNA polymerase	New England Biolabs
Tetracycline	Sigma
Thrombin	Sigma
Triton X-100	Sigma
Tween-20	Sigma
Xylene cyanol FF	Sigma

2.2.2 Bacterial Strains

All bacterial transformants used were derivatives of the following strains.

Name	Genotype	Use
DH5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	Maintenance and manipulation of plasmid DNA for cloning.
BL21	dcm ompT hsdS(r$_B^-$m$_B^-$) gal	Protein expression (general purpose)
BL21 (DE3)	dcm ompT hsdS(r$_B^-$m$_B^-$) gal (λDE3)	Protein expression of proteins controlled by a T7-promoter
KS1	MC1000 F'*lacIq* (*plac*O$_R$2-62.lacZ)	Reporter strain for bacterial two-hybrid (Dove, Joung et al. 1997)

2.2.3 Plasmid Vectors

Plasmid Name	Description/Features	Source
pGEX-4T1	pBR322 ori; Amp^R; For high level IPTG-inducible expression of thrombin-cleavable GST-fusion proteins	GE Healthcare
pGEX-4T1-G0-6	pGEX-4T1 derivative encoding GST-G0-6 fusion protein. Created by inserting DNA encoding residues 1-226 of gankyrin between the EcoRI and XhoI sites using primers 471 & 472	This work
pGEX-4T1-G0-3	pGEX-4T1 derivative encoding GST-G0-3 fusion protein. Created by inserting DNA encoding residues 1-138 of gankyrin between the EcoRI and XhoI sites using primers 471 & 478	This work
pGEX-4T1-pRbSP	pGEX-4T1 derivative encoding GST-pRbSP fusion protein. Created by inserting DNA encoding residues 379-787 of pRb between the EcoRI and XhoI sites using primers 473 & 476	This work
pET-32a(+)	pBR322 ori; Amp^R; For high level IPTG-inducible expression of thioredoxin fusion proteins with cleavable His-tag and S•Tag™ sequences	Merck Millipore
pET-32a-G0-6	pET-32a(+) derivative encoding Trx-G0-6 fusion protein. Created by sub-cloning small EcoRI/XhoI fragment of pGEX-4T1-G0-6 into EcoRI/XhoI sites of pET-32a(+)	This work
pET-32a-pRbSP	pET-32a(+) derivative encoding Trx-pRbSP fusion protein. Created by sub-cloning small EcoRI/XhoI fragment of pGEX-4T1-pRbSP into EcoRI/XhoI sites of pET-32a(+)	This work
pET-32a-S6C/G0-6	pET-32a-G0-6 derivative encoding Trx-S6C fusion protein and un-tagged G0-6 on a dicistronic transcript. Created by inserting DNA encoding residues 337-418 (plus C-terminal Cys residue intended to be used for chemical labelling, stop codon and RBS) into BamHI/EcoRI sites of pET-32a-G0-6.	This work
pTRG	ColE1 ori; Tet^R; rpoA fusion under lpp/placUV5 promoter	Agilent/Stratagene
pTRG-G0-6	pTRG derivative encoding rpoA-G0-6 fusion protein. Created by sub-cloning small BamHI/XhoI fragment of pGEX-4T1-G0-6 into BamHI/XhoI sites of pTRG.	This work
pTRG-pRbSP	pTRG derivative encoding rpoA-pRbSP fusion protein. Created by sub-cloning small BamHI/XhoI fragment of pGEX-4T1-pRbSP into BamHI/XhoI sites of pTRG.	This work
pBT	p15A ori; Chlor^R; λCI fusion under placUV5 promoter	Agilent/Stratagene
pBT-G0-6	pBT derivative encoding λCI-G0-6 fusion protein. Created by sub-cloning small BamHI/XhoI fragment of pGEX-4T1-G0-6 into BamHI/XhoI sites of pBT.	This work
pBT-pRbSP	pBT derivative encoding λCI-pRbSP fusion protein. Created by sub-cloning small BamHI/XhoI fragment of pGEX-4T1-pRbSP into BamHI/XhoI sites of pBT.	This work

2.2.4 Oligonucleotides

Synthetic oligonucleotides were supplied by Geneworks

Number	Name	RE Site	Sequence (5′ - 3′)
471	Gank_M1_5′	EcoRI	GACGAATTCATGGAGGGGTGTGTGTC
472	Gank_G226_3′	XhoI	CCACTCGAGTTAACCTTCCACCATTCTC
478	Gank_Y138_3′	XhoI	CTCGAGTTAATAATGGTCCTTAGCATC
502	Gank_L126R_fwd		CGCTGTCATGTTACGGGAAGGCGGGGCTAATCC
503	Gank_L126R_rev		GGATTAGCCCCGCCTTCCCGTAACATGACAGCG
578	Gank_K149D_fwd		CCGGGCAGCAGCCGACGGTAACTTGAAGATGATTC
579	Gank_K149D_rev		GAATCATCTTCAAGTTACCGTCGGCTGCTGCCCGG
560	Gank_E182K_fwd		CTACACTTAGCCTGTGATAAGGAGAGAGTGGAAG
561	Gank_E182K_rev		CTTCCACTCTCTCCTTATCACAGGCTAAGTGTAG
574	Gank_E186K_fwd		GATGAGGAGAGAGTGAAAGAAGCAAAACTGCTGGTG
575	Gank_E186K_rev		CACCAGCAGTTTTGCTTCTTTCACTCTCTCCTCATC
313	pBT_5′_SP		TCCGTTGTGGGGAAAGTTATC
314	pBT_3′_SP		GGGTAGCCAGCAGCATCC
284	pGEX_5′_SP		GGGCTGGCAAGCCACGTTTGGTG
285	pGEX_3′_SP		CCGGGAGCTGCATGTGTCAGAG
315	pTRG_5′_SP		TGGCTGAACAACTGGAAGCT
316	pTRG_3′_SP		ATTCGTCGCCCGCCATAA
473	Rb_M379_5′	EcoRI	AGGGAATTCATGAACACTATCCAACAATTA
476	Rb_R787_3′	XhoI	GTACTCGAGTCGAGGAATGTGAGGTAT
580	Rb_D718K_fwd		GGCATATGCAAAGTGAAGAATATAAAGCTTAAATTCAAAATCATTG
581	Rb_D718K_rev		CAATGATTTTGAATTTAAGCTTTATATTCTTCACTTTGCATATGCC
562	Rb_K720E_fwd		GTGAAGAATATAGACCTTGAATTCAAAATCATTGTAACAGC
563	Rb_K720E_rev		GCTGTTACAATGATTTTGAATTCAAGGTCTATATTCTTCAC
576	Rb_K765E_fwd		CGGTCTTCATGCAGAGACTGgAAACAAATATTTTGCAG
577	Rb_K765E_rev		CTGCAAAATATTTGTTTCCAGTCTCTGCATGAAGACCG
	S6_D337_5′	BamHI	CCAGGATCCGACCGCCGCCAGAAGAGATTG
	S6_K418_3′_RBS	EcoRI	CATGAATTCTGTTTCCTCAACACTTGTAAAACTCATGCTCCTG
	T7_Prom		TAATACGACTCACTATAGG
	T7_Term		GCTAGTTATTGCTCAGCGGTGG

2.2.5 Buffers, Solutions and media

Name	Composition
Blocking solution	5% (w/v) BSA in TBS
Coomassie blue staining solution	0.1% (w/v) Coomassie Brilliant Blue, 30% (v/v) methanol, 10% (v/v) acetic acid
Destain solution	50% (v/v) methanol, 5% (v/v) acetic acid
GLB (10×)	50% (v/v) glycerol, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol
IMAC elution buffer	50mM Tris, 500mM NaCl, 100mM NaCl, pH 8.0
LB	1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl (pH to 7.0 with NaOH)
L Agar	LB + 1.5% (w/v) bacto-agar
MinA	60 mM K_2HPO_4, 33mM KH_2PO_4, 1.7mM Na_3Citrate, 15mM NH_4Cl, 0.005% (w/v) thiamine, 0.2% (w/v) glucose, 0.8mM $MgSO_4$
PBS	8% (w/v) NaCl, 0.02% KCl, 0.02% (w/v) KH_2PO_4 (anhydrous), 0.115% (w/v) Na_2HPO_4 (anhydrous)
PBST	PBS + 0.1% Tween-20
Phosphate buffer	2.26 mM NaH_2PO_4, 7.74 mM Na_2HPO_4 (pH verified to be 7.4)
SDS-PAGE sample buffer	50mM Tris (pH 6.8), 4% (w/v) SDS, 12% (v/v) glycerol, 2% (v/v) β-mercaptoethanol, 0.01% (w/v) Coomassie Brilliant Blue
TAE	40mM Tris, 40mM sodium acetate, 10mM EDTA (pH 8.2)
TBS	20mM Tris (pH 7.4), 150mM NaCl
TBST	TBS + 0.1% Tween-20
TTBS	TBS + 0.1% Triton X-100
Western blocking solution	PBST + 5% (w/v) BSA
Western transfer buffer	190mM glycine, 25mM Tris, 15% (v/v) methanol

2.2.6 Molecular weight standards

Name	Size standards	Source
100bp DNA ladder	1517, 1200, 1000, 900, 800, 700, 600, 500/517, 400, 300, 200, 100 (bp)	New England Biolabs
1kb DNA ladder	10, 8, 6, 5, 4, 3, 2, 1.5, 1, 0.5 (kb)	New England Biolabs
SeeBlue® Plus 2	188, 98, 62, 49, 38, 28, 17, 14, 6, 3 (kDa, approx)	Invitrogen

2.2.7 Antibodies

anti-S•tag (polyclonal), raised in rabbit (Abcam)

anti-GST (monoclonal), raised in mouse was prepared from hybridoma cells

anti-Mouse Cy3 conjugate, raised in donkey (Rockland)

anti-Rabbit Cy5 conjugate, raised in donkey (Rockland)

2.2.8 Kits

Kit	Supplier
QIAprep Spin Miniprep	Qiagen
QIAquick Gel Extraction	Qiagen
QIAquick PCR Purification	Qiagen
Quikchange™ site-directed mutagenesis	Stratagene/Agilent
BacterioMatch II bacterial two-hybrid	Stratagene/Agilent
MatchMaker II yeast two-hybrid	Stratagene/Agilent

2.2.9 Miscellaneous materials

Material	Supplier
96-well plates (clear)	BD Biosciences
96-well plates (black)	BMG Technologies
Amicon Ultra-15 Centrifugal Filters	Millipore
Dialysis tubing (CelluSep T2)	Adelab Scientific
Hybond-LFP PVDF membrane	GE Healthcare
Minisart syringe-top filters (0.45 and 0.8 µm)	Sartorius
NMR tubes	Wilmad Glass co
NuPage 4-12% Bis-Tris polyacrylamide gels	Invitrogen
PD-10 columns	GE Healthcare
Syringes	BD Biosciences

2.2.10 Software

Software name	Use	Reference(s)
R	Statistical analysis (computational data)	
Graphpad Prism	Statistical analysis (experimental data)	
NMRPipe	NMR data processing	(Delaglio, Grzesiek et al. 1995)
CCPNMR Analysis	NMR data analysis	(Vranken, Boucher et al. 2005)
UCSF Chimera	Molecular visualization	(Pettersen, Goddard et al. 2004)
VMD	Molecular visualization	(Humphrey, Dalke et al. 1996)
Sybyl (version 7.3)	Molecular visualization, molecular editing	
Dock (version 5)	Molecular docking (small molecules)	(Moustakas, Lang et al. 2006)
3D-Dock	Molecular docking (protein-protein)	(Katchalski-Katzir, Shariv et al. 1992; Moont, Gabb et al. 1999)
OS v7.2	Calculation of occluded surface areas	(Fleming and Richards 2000)
NAMD	Molecular dynamics	(Phillips, Braun et al. 2005)
Modeller	Homology modeling	(Marti-Renom, Stuart et al. 2000)
MySQL	Data set manipulation	
ApE	DNA sequence manipulation	

2.3 Molecular Biology Techniques

2.3.1 Preparation of Plasmid DNA

Small-scale plasmid purifications were carried out using the Qiagen QIAprep™ Spin Miniprep

Kit. 5mL overnight cultures in LB, 100μg/mL ampicillin were processed according to the

manufacturer's instructions and plasmid DNA was eluted in 50μL Tris buffer as supplied.

2.3.2 Agarose gel electrophoresis

Agarose gels were prepared by pouring 12 mL of gel solution (1-2% (w/v) agarose in TAE) onto a 5.0 cm x 7.5 cm glass slide. Gels were submerged in TAE and samples containing GLB were electrophoresed at 75-100 Volts for 30-45mins dependent on application. DNA was stained in 5 μg/mL EtBr and visualised via exposure to medium wavelength UV light.

2.3.3 Polymerase Chain Reaction (PCR)

PCR reaction mix was prepared containing 1ng template DNA, 100ng each oligonucleotide primer, 2.5mM each dNTP, 2.5U *Pfu Turbo* DNA polymerase, and 1X *Pfu* reaction buffer. The reaction mix was overlaid with 30μL mineral oil and the reaction was carried out in a PTC-100 Programmable Thermal Cycler (MJ Research Inc) for 30 cycles preceded by 5min denaturation at 95°C and followed by 10min final extension at 72°C before being stored at 12°C for a maximum of 18 hours until further processing. Each cycle consisted of 95°C denaturation for 1min, annealing for 1min at a temperature dependent on the Tm of the primers in use as advised by the manufacturer and extension at 72°C for 1min per kb of expected product.

2.3.4 Restriction Enzyme Digest

0.5-5 μg of DNA was digested with appropriate enzymes in a total volume of 20μL or 50μL at 37°C for 1-4 hours. The buffer used was determined according to the supplier's instructions for each enzyme. Restriction digest reactions were analysed by agarose gel electrophoresis.

2.3.5 DNA purification by agarose gel

After carrying out agarose gel electrophoresis (Section 2.3.2), selected DNA bands were excised using a scalpel blade by visualisation under low intensity, long wavelength UV light. DNA was extracted from gel fragments using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. DNA was eluted in 30μL Tris buffer as supplied.

2.3.6 DNA purification by solid phase adsorption

Purification of DNA fragments (60-6000bp) from reaction mixtures was carried out using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions. DNA was eluted in 30μL Tris buffer as supplied.

2.3.7 DNA ligation

50 ng of purified, restricted plasmid and purified insert in at least 3-fold excess were incubated with 40 units T4 DNA ligase in ligation buffer, including 1mM ATP, for 2 hours at room temperature or overnight at 16°C.

2.3.8 Preparation of competent cells

A single bacterial colony was selected from a streak plate of stock cells and used to inoculate a 5mL LB culture that was incubated at 37°C overnight. The following day, the culture was diluted 1/100 with LB and incubated to an OD_{600} of 0.4-0.6. Cells were pelleted by centrifugation at 3000 × g, 4°C for 15 minutes. Cell pellets were resuspended in 50mL ice cold $CaCl_2$ solution (0.06M $CaCl_2$, 15% v/v glycerol) and centrifuged a further 15 minutes at 3000 × g, 4°C before being resuspended in 20mL ice cold $CaCl_2$ solution and snap frozen in 200μL aliquots using a dry ice/ethanol bath.

2.3.9 Plasmid transformation

DNA solution was added to 200 μL of competent *E. coli* DH5α or BL21 cells. The mixture was incubated on ice for up to 30 minutes and heat shocked at 42°C for 2 minutes. The cells were incubated on ice for a further 15 minutes. The cells were suspended in 150 μL of LB and were incubated at 37°C for 30 minutes. Cells were spread on an LB agar plate containing 100 μg/mL ampicillin. Plates were incubated overnight at 37°C.

2.3.10 Clone screening by colony PCR

After transformation of recombinant plasmids and overnight growth of transformation plates, single colonies were screened for insertion of DNA fragments at the multiple cloning site (MCS) of the vector by PCR using primers flanking the MCS and whole cell lysates as DNA template. Whole cell lysates were prepared by resuspending individual bacterial colonies in 50µL MQ water and incubating for 5 minutes at 95°C. PCR was carried out as described in section 2.3.3 using 4µL cell lysate as template and results were analysed by agarose gel electrophoresis (Section 2.3.2).

2.3.11 Cycle sequencing of DNA

50-250 ng circular plasmid DNA and 50ng sequencing primer were combined with 0.5µL BIGDYE™ Terminator Ready Reaction Mix (Applied Biosystems) in a total reaction volume of 10µL and overlaid with 20µL mineral oil. The reaction was carried out as follows: Cycles 1-31, 96°C 30 sec / 50°C 15 sec / 60°C 4 min on a PTC-100™ Programmable Thermal Controller. The DNA was precipitated by the addition of isopropanol to 60% v/v and pelleted by centrifugation at 20,000 × g for 10 minutes, washed with 75% isopropanol and air dried. DNA sequence information was obtained via the analysis service provided by the Institute of Medical and Veterinary Sciences.

2.3.12 Site-directed mutagenesis

Site-directed mutagenesis was carried out according to the method described in the Quikchange™ Site-Directed Mutagenesis kit (Stratagene). Mutagenesis PCR reaction mixes contained 100ng template DNA, 100ng each oligonucleotide primer, 2.5mM each dNTP, 2.5U Pfu Turbo DNA polymerase, and 1X Pfu reaction buffer to a final volume of 50µL and were topped with 30µL mineral oil. The reaction was carried out on a PTC-100TM Programmable Thermal Controller as follows: initial denaturation step 5 minutes at 95°C followed by 16-18 cycles of denaturation at 95°C for 1min, annealing at 55-60°C for 1 min and extension at 68°C for 8 min. After final extension 68°C for 10 min, PCR samples were stored at 12°C until further processing. PCR samples were

digested by DpnI to remove parental DNA and then transformed into DH5α competent cells by heat shock (Section 2.3.9).

2.3.13 LacZ reporter assays

LacZ assays were carried out based on the method described in (Dodd, Perkins et al. 2001). One-day-old colonies on selective LB plates were resuspended in LB and used to inoculate 100µL cultures in a 96-well microtitre plate. Plates were sealed and incubated overnight without shaking. Overnight cultures were diluted to a calculated OD_{600} of 0.003 into LB containing chloramphenicol (30µg/mL), tetracycline (10µg/mL) and IPTG (20µM) and grown at 30°C in an orbital shaker to an OD_{600} of 0.5-0.9 (log phase). All OD_{600} measurements were calculated based on A_{620} measurements using an empirically derived relationship and adjustment for light-scattering nonlinearity as described in (Bipatnath, Dennis et al. 1998). 50µL log-phase cells were added to 190µL assay solution (100 mM Tris-HCl, pH 8.0, 1 mM $MgSO_4$, 10 mM KCl, 0.27% (v/v) 2-mercaptoethanol, 50 µg/mL polymyxin B, 842µg/mL ONPG) and incubated for 1 hour at 28°C, taking A_{414} measurements every 2 minutes. Enzyme activity was determined as the slope of the line of best fit of A_{414} versus time (readings in the first 10 minutes or with $A_{414} > 2.5$ were ignored). LacZ units were calculated as $200,000 \times (A_{414}/min)/(OD_{600} \times 50)$ and were roughly equivalent to standard Miller units (Miller 1972).

2.4 Protein Chemistry Techniques

2.4.1 Expression of IPTG-inducible recombinant proteins

An overnight bacterial culture was diluted 1 in 100 into fresh LB containing 80µg/mL ampicillin and grown, shaking at 37°C to an OD_{600} of 0.6-0.8 before addition of IPTG to a final concentration of 0.2mM to induce protein expression. The culture was incubated for a further 3-4 hours at 37°C or overnight at 30°C or, in the case of pRb expression, overnight at room temperature. Cells were pelleted by centrifugation at $5,000 \times g$ for 15 minutes and stored at -20°C for up to one month before using for protein purification.

2.4.2 Preparation of bacterial lysates for protein purification

Cell pellets were resuspended in 30mL TTBS including 1mM EDTA and immediately before

lysis, PMSF was added to a final concentration of 1mM. Cells were lysed by three passages through

an ice-cooled M-110L Microfluidizer® Processor (Microfluidics™) fitted with a H10Z interaction

chamber at 70-80 psi input pressure (approximately 17,000 psi process pressure). Cell debris was

cleared from the lysate by centrifugation at 30,000 × g for 15 minutes at 4°C, followed by filtration

through 0.8μm and 0.45μm filters in series. Lysates were kept on ice where possible an prepared

freshly for each protein purification.

2.4.3 Purification of GST fusion proteins

A glutathione agarose column was equilibrated first in TBS, then TTBS before filtered lysate

(Section 2.4.2) was loaded onto the column. After protein loading, the column was washed with

TTBS (at least 6 column volumes) until no protein could be detected leaving the column using

Bradford reagent. The column was then washed with a further 3 column volumes of TBS. For

purification of intact fusion proteins, protein was eluted in TBS with 10mM glutathione (pH 8.5).

Otherwise, the fusion protein was cleaved by addition of 3-5 units of thrombin per mL of affinity

resin. Thrombin was prepared by addition of $CaCl_2$ to 2.5mM immediately prior. The cleavage

reaction was allowed to proceed overnight at room temperature, before elution of the fusion

partner in TBS, followed by elution of GST with 10mM glutathione (pH 8.5).

2.4.4 Purification of poly-His tagged proteins

An NTA-agarose column was first charged with Ni^{2+} ions by passage of 3 column volumes of

100mM $NiSO_4$ solution and 5 column volumes of MQ water over the column. The charged column

was then equilibrated in TBS (pH 8.0), then TTBS + 5mM imidazole (pH 8.0). Filtered lysates were

supplemented with imidazole to a final concentration of 5mM before loading onto the column.

After protein loading, the column was washed with TTBS + 5mM imidazole (at least 6 column

volumes) until no protein could be detected leaving the column using Bradford reagent. The

column was then washed with a further 3 column volumes of TBS + 5mM imidazole. Protein was

eluted in fractions (approximately one fifth of the column volume) with IMAC elution buffer.

Samples were collected at various stages of the purification and analysed by SDS-PAGE. Fractions

containing the protein of interest were pooled and dialysed.

2.4.5 Concentration of protein samples

Proteins were concentrated by centrifugation at 4,000 × g at 4°C using Amicon Ultra

Centrifugal Filter units (Millipore). 10 or 30kDa nominal molecular weight cut-off membranes were

used depending on the application.

2.4.6 Buffer exchange by size exclusion chromatography

Buffer exchange for NMR samples, to remove unreacted FITC, and other purposes was carried

out using PD-10 size exclusion columns (GE Healthcare) according to the manufacturer's

instructions.

2.4.7 Buffer exchange by dialysis

Buffer exchange to remove imidazole from protein samples after IMAC purification was

carried out by dialysis. Dialysis tubing (Cellu-Sep T2, nominal molecular weight cut-off 6-8 kDa) was

prepared by washing in buffer for 5-10 min. Protein sample was sealed in dialysis tube and

suspended in stirred buffer of at least 200 times the volume of the sample. Equilibration

proceeded for at least 3 hours at room temperature or at least 8 hours at 4°C. Two equilibrations

were carried out for each sample.

2.4.8 Protein quantitation

The concentration of protein samples were determined by UV absorption spectroscopy. The

absorbance of protein samples was measured in the range of 240-320nm wavelength and the

absorbance of the sample buffer alone was measured and subtracted. The appearance of a peak at

280nm was used as a qualitative test that the absorbance being measured was principally due the

presence of protein. The absorbance of the sample at 280nm was then used to determine the

sample concentration using the Beer-Lambert Law (Equation 1):

$$A = \varepsilon c l$$

<div align="right">(Equation 1)</div>

where A is the measured absorbance, l is the path length over which the measurement was

taken (in this case 1cm), c is the sample concentration and ε is the sample extinction coefficient.

Extinction coefficients for the proteins used in this study were estimated on the basis of the

protein's sequence using the ProtParam tool provided by ExPASy

(http://web.expasy.org/protparam/) (Gasteiger, Hoogland et al. 2005)

UV absorbance was employed for its reproducibility, in order to be confident of the relative

concentrations of mutants that were to be directly compared.

2.4.9 Denaturing polyacrylamide gel electrophoresis

Samples were prepared by mixing with concentrated SDS sample buffer and boiling for 5

minutes at 95°C. Samples were loaded onto pre-cast NuPage 4-12% Bis-Tris gels (Invitrogen) and

run in either MOPS or MES buffer (Invitrogen) for 35-45 minutes until the dye front reached the

bottom of the gel. For total protein visualization, gels were stained with Coomassie blue staining

solution for at least 1 hour, followed by washing in Destain Solution for several hours with a

natural fibre destaining aid. Alternatively, gels were analysed by Western blotting (Section 2.4.10).

2.4.10 Western blot analysis

SDS-PAGE was used to separate proteins (Section 2.4.9) which were then transferred onto

Hybond-LFP membrane (GE Healthcare). Transfer was achieved by semi-dry transfer at 40mA for

90 minutes using a Hoefer SemiPhor Western Transfer apparatus (GE Healthcare), with blotting

pads and membrane (after initial wetting in 100% methanol) soaked in Western Transfer Buffer.

Membranes were blocked overnight at 4°C or for 2 hours at room temperature by shaking in

Western Blocking Solution. After blocking, membranes were washed 3 times with PBST before

probing with primary antibody (1 in 1000 dilution in PBST) for 1 hour at room temperature. A

further 3 washes with PBST for 10 minutes each were completed before addition of secondary

antibody (1 in 2000 dilution in PBST) and incubating (with shaking) for a further 1 hour at room

temperature. A final 3 washes with PBST (10 minutes each) were followed by a rinse with PBS then

MQ water, and membranes were air dried (protected from light) at room temperature overnight

or 3 hours at 37°C.

2.4.11 Protein labeling with FITC

2-4mg of protein in 200μL 100mM PBS buffer (pH 7.5) was mixed with a 10-fold molar excess

of FITC (dissolved at 10mg/mL in DMSO) for at least 2 hours, protected from light, with shaking.

Unreacted FITC was removed by twice exchanging the protein into fresh buffer using a PD-10

column (Section 2.4.6). The degree of labeling (DOL) was determined using the absorbance of the

sample at the 280nm and 494nm wavelengths and the following formula:

$$DOL = \frac{A_{494}}{[\text{protein}] \times \epsilon_{FITC}}$$

Where the extinction coefficient of FITC (ϵ_{FITC}) is 68,000 cm^{-1}M^{-1} and the concentration of

protein (in mol/L) has been determined according to the following formula:

$$[\text{protein}] = \frac{A_{280} - A_{494} \times CF}{\epsilon_{protein}}$$

CF, with a value of 0.3, is the ratio of A280 to A494 of free FITC. Protein extinction coefficients

($\epsilon_{protein}$) were estimated on the basis of primary sequence using the ProtParam tool provided by

ExPASy (http://web.expasy.org/protparam/) (Gasteiger, Hoogland et al. 2005)

2.4.12 Solid phase binding assay

Black microtitre plates were incubated overnight at 4°C containing 100μL GST-pRbSP or Trx-

pRbSP (10μg/mL), washed once with TBS, then blocked by incubation for 2 hours at 37C in 5% BSA

blocking solution. The plates were then washed twice with TBS before fluorescein-labeled tracer protein (Trx-G0-6 or S-G0-6) was added, along with competitor if necessary for the experiment, and the plate was incubated at room temperature for 2 hours. Plates were washed five times with TBST and air-dried for 20-30 minutes. Fluorescent signal in each well was measured using a BMG FLUOstar Galaxy plate reader with a 485nm excitation filter and 520nm emission filter.

2.4.13 GST pull-down experiments

50μL binding solution containing 1μM GST-pRbSP, 100μM S-G0-6 and 5% BSA (w/v) in TBST buffer was incubated with gentle mixing for 90 minutes at 4°C. 20μL glutathione-agarose beads were added and incubated for a further 15 minutes with gentle agitation, before slow-speed centrifugation of the beads (2 minutes at 150 × g), aspiration of supernatant, and addition of 1mL TBST to the beads. The beads were washed in this manner 5 times, before elution of protein from the beads with 25μL 10mM reduced glutathione in TBS. Eluted protein was then prepared for SDS-PAGE (Section 2.4.9) and analysed by Western blot (Section 2.4.10).

2.5 NMR Spectroscopy Techniques

2.5.1 NMR sample preparation

Isotope labeled protein samples were produced by conducting protein expression in MinA media containing ^{13}C-glucose and ^{15}NH$_4$Cl as carbon and nitrogen sources, respectively. Protein was purified by the method described in Section 2.4.3. Purified proteins were concentrated to a volume of 1-2 mL and exchanged into 10mM phosphate buffer (pH 7.4) containing 0.01% w/v NaN$_3$ and supplemented with 10% v/v D$_2$O. Sample concentrations were in the range of 0.25 to 0.6 mM. 550-600 μL samples were placed in a 5mm high-resolution thin-walled glass NMR tube for spectroscopy.

2.5.2 NMR experiments

NMR experiments were performed using a Varian Inova 600 MHz spectrometer and Vnmr

software with the BioPack extension. Spectra were recorded at 25°C using a 5mm inverse triple

resonance 1H/13C/15N pfg probe with the carrier frequency centred on the H2O signal. Backbone

resonance assignments were attempted using spectra acquired with 3D ^{1}H-^{15}N NOESY-HSQC, 3D

^{1}H-^{15}N TOCSY-HSQC, HSQC, HNCA, HN(CO)CA, CBCANH and CBCA(CO)NH pulse sequences.

Acquired spectra were processed (including zero-filling and linear prediction) using NMRPipe

(Delaglio, Grzesiek et al. 1995) and analysed using CCPNMR Analysis (Vranken, Boucher et al.

2005).

2.5.3 Ligand titration experiments

^{15}N-labeled NMR samples of G0-3 (L126R) were prepared at a concentration of 0.25mM as

described in Section 2.5.1 and supplemented with DMSO to a concentration of 10% v/v. HSQC

spectra were recorded prior to the addition of ligand, and after each addition of ligand. Ligand was

prepared by dissolving in 100% DMSO and aliquots prepared such that they were between 0.5 and

2µL each in volume and could be added serially to reach the desired ligand concentrations. After

addition of each aliquot to the NMR sample, a corresponding aliquot of 9-fold the volume of

10mM phosphate buffer was also added to maintain a total DMSO concentration in the sample of

10%.

2.6 Computational Techniques

2.6.1 Molecular dynamics simulations

Molecular coordinates were prepared by addition of hydrogen atoms where required and

calculation of residue charge states and force field parameters using XPLOR-NIH v2.25(Schwieters,

Kuszewski et al. 2003) and the CHARMM force field(MacKerell, Bashford et al. 1998). Energy

minimizations were conducted in explicit solvent, using a solvent box, prepared using SOLVATE

v1.3 and the TIP3 water model as employed in VMD v1.8.7b5 (Humphrey, Dalke et al. 1996) to

ensure that non-water atoms were no closer than 10 Å to the edge of the box, and inclusion of

sodium and chloride ions at random to give the system an overall neutral charge and salt

concentration of 150mM. Finally, the system was minimized for 10,000 cycles by conjugate-

gradient minimization using NAMD v2.6 (Phillips, Braun et al. 2005). Periodic boundary conditions

are not enabled during energy minimization in NAMD and thus were not used. When conducted,

molecular dynamics simulations used the output coordinates and identical parameters to energy

minimizations and began by random assignment of particle velocities dependent on the

temperature of the run (specified in text). In order to maintain consistency and comparability

with previous calculations, molecular dynamics calculations considered only a subset of atoms

(described in text), giving the calculation explicit boundaries, and precluding the use of periodic

boundary conditions. Calculations were performed on an SGI Altix XE1300 cluster on Intel

Clovertown quad-core processors using a 64-bit NAMD compilation.

Chapter 3

In silico investigation of the gankyrin-pRb interaction

3.1 Introduction

3.1.1 Gankyrin and the Retinoblastoma protein

The Retinoblastoma protein

The retinoblastoma protein (pRb) is one of the key regulators of cell cycle progression and was one of the first tumour suppressor proteins discovered. pRb – along with the other two members of the RB family, p107 and Rb2/p130 – regulates the passage of the cell through the G1/S-phase transition primarily through its association with the E2F family of transcription factors. Studies have demonstrated that p107 and p130 also have tumour suppressor properties (Howard, Claudio et al. 1998; Robanus-Maandag, Dekker et al. 1998) but pRb is considered the central member of the family. This is supported by the fact that p107 mutations are less common in cancer(Rizzolio, Esposito et al. 2010) while pRb is ubiquitously expressed, in contrast to the more limited expression patterns of p107 and p130(Cobrinik 2005).

Different members of the RB family have been shown to preferentially bind different E2F family transcription factors, which in turn act via different genomic elements, and it is thought, with differing tendencies toward transcriptional activation or suppression(Polager and Ginsberg 2008). pRb interacts with its E2F binding partners via the "large pocket" region of the protein, encompassing the conserved A and B pocket domains as well as the non-conserved C-terminal region(Dick 2007). This region of pRb contacts the transactivation domain of the E2F transcription factor, preventing its activity as an activator, but without disrupting its ability to bind DNA (Flemington, Speck et al. 1993; Helin and Ed 1993). It has subsequently been shown that in addition to inhibition of transactivation by E2Fs, pRb inhibits expression of E2F target genes via the recruitment of chromatin remodeling enzymes that suppress transcriptional activity (Brehm, Miska et al. 1998; Robertson, Ait-Si-Ali et al. 2000; Nielsen, Schneider et al. 2001). Because the interaction between pRb and E2Fs does not inhibit DNA binding of the latter, pRb becomes localized to the E2F-regulated promoters. Along with pRb, chromatin remodeling enzymes are

similarly localized to these promoters because of their association with pRb, via its pocket domains.

As the cell enters the G1/S transition, CDK4/6 and the D-type cyclins, and subsequently CDK2/cyclin E are active, and phosphorylate pRb at multiple sites to "inactivate" it – referred to as "hyperphosphorylation". Inactivation works by disrupting the interaction between pRb and the E2Fs, alleviating the two-fold repression mechanism outlined above. A model is proposed in which pRb is no longer recruited to the promoters of S-phase genes since it possesses no intrinsic DNA-binding properties, and so in turn the recruitment of chromatin remodeling methylases, deacetylases and other negative regulators of transcription ceases. Secondly, the E2Fs, having been freed from their complex with pRb, are now able to act as transcriptional activators of the genes required for entry into S phase. In healthy cells, both pRb and the E2Fs integrate signals from various parts of the cell and can induce apoptosis, cell cycle arrest, or send cells into senescence when DNA damage or signaling from pro-apoptotic pathways is detected (Dick 2007; Polager and Ginsberg 2008). Though with evidence that phosphorylation events at different sites have differing effects on pRb's interactions with its various binding partners (Knudsen and Wang 1996), some consider this an over-simplified model, instead suggesting that phosphorylation of pRb is far more precisely controlled with very specific and varied effects on pRb's behaviour (Burke, Hura et al. 2012; Heilmann and Dyson 2012).

Protein-protein interactions of pRb via the LxCxE motif

One of the first indications of the tumour suppressor function of pRb was the action of the adenovirus E1A, SV40 T antigen and HPV E7 viral oncoproteins, because the oncogenicity of these proteins was linked to their ability to interact with the small pocket region of pRb. Studies indicated that these proteins possessed a conserved Leu-Xxx-Cys-Xxx-Glu (LXCXE) pentapeptide motif, and that it was this motif that was responsible for binding to pRb (DeCaprio, Ludlow et al. 1988; Dyson, Guida et al. 1992).The LXCXE-based interaction of the viral oncoproteins is necessary

but not sufficient to cause dissociation of pRb/E2F complexes (Whyte, Ruley et al. 1988), with

similar effects to the same dissociation caused by hyperphosphorylation of pRb. Since

hyperphosphorylation of pRb is one of the major barriers to entry into S phase, this causes a loss

of cell cycle control.

More than 40 cellular or viral proteins have now been shown to interact with pRb, and 28 of

these are thought to interact via the LXCXE motif. However, the set of proteins that interact via

this motif is broad, and varies significantly in terms of the functions of the pRb partner proteins.

Beside the viral oncoproteins, the transcriptional repressor chromatin remodeling enzymes

mentioned above associate with pRb via an LXCXE motif. Other examples include Hsp75, a variety

of non-E2F transcription factors, a subunit of DNA polymerase and other proteins involved in DNA

replication. In fact, the D-type cyclins responsible (with their partner CDKs) for the inactivation of

pRb also interact via LXCXE interactions. While some of these interaction partners – such as

chromatin remodeling enzymes – appear to be important for cell cycle control, they seem to be at

odds with others – like the D-type cyclins – that are important for the inactivation of pRb. Our lack

of understanding of how these different functions and cellular components interact and compete

via pRb LXCXE binding has obscured a complete picture of the role of this important interaction.

(Dick 2007)

Gankyrin interacts with pRb

The first report of gankyrin as an oncoprotein explained its oncogenicity partly in terms of its

effect on pRb. Higashitsuji *et al* (2000) observed that in cells overexpressing gankyrin, pRb became

hyperphosphorylated, that gankyrin and pRb co-immunoprecipitate and that gankyrin contained

an LXCXE motif between residues L178 and E182, in ankyrin repeat 5 (where the most N-terminal

ankyrin repeat is designated repeat 0 due to its late identification). It was further shown that

mutations within the LXCXE motif reduce the oncogenicity of the protein, along with their ability

to interact with pRb. In particular, the E182A mutation caused complete abrogation of the

interaction with pRb, and reduced colony formation in soft agar by more than 90%, a greater

reduction even than any of the truncation mutants tested. The L178A mutation also reduced, but

did not destroy, the interaction with pRb and reduced soft agar colony formation by 37%.

Subsequent work demonstrated that a peptide derived from gankyrin (residues 176-185) and

containing the LXCXE motif is able to compete with full-length gankyrin for binding to pRb (Li and

Tsai 2002). This was taken as evidence that gankyrin binds pRb in an analogous fashion to other

LXCXE motifs. However, I – and others (Dawson, Higashitsuji et al. 2006) – remain skeptical of the

strength of this evidence. The peptide used in this study is less than a full ankyrin repeat, and in

any case ankyrin repeats are known to require at least three consecutive repeats in order to adopt

their characteristic fold. Therefore the evidence demonstrates that the gankyrin sequence is

(unsurprisingly) *capable* of interacting with pRb in the random peptide conformation characteristic

of other LXCXE peptides and that its binding site overlaps the LXCXE binding cleft, but by no means

demonstrates that this is the true nature of the interaction in the cellular context. Similarly, the

mutagenesis data clearly demonstrates that gankyrin residue E182 is critical for the interaction

with pRb, but not that it interacts at the same site on pRb as other glutamate residues in the

context of the LXCXE motif. The L178A mutagenesis data is even less convincing, since L178 is

found at a position in the ankyrin consensus sequence that has a strong consensus for big,

hydrophobic residues. Therefore, the reduction in pRb binding conferred by the L178A mutation

could equally be explained by a partial loss of structure of gankyrin, if that is important for binding

on the one hand, or a loss of the LXCXE sequence motif on the other.

Adding further information and intrigue to this debate was the solution of the structure of a

complex of gankyrin and the C-terminal domain of the S6-ATPase proteasome subunit, and the

finding that S6-ATPase binding to gankyrin is not competitive with pRb (Nakamura, Nakano et al.

2007). The structure shows that the S6-ATPase C-terminal domain contacts all seven ankyrin

repeats of gankyrin, and forms a very stable complex. Therefore, in combination, these findings

significantly diminish the possibility that gankyrin partially unfolds in order to bind pRb via its LXCXE motif, since doing so would compromise the interaction with S6-ATPase and therefore show evidence of competitive binding, or allosteric inhibition of the interaction.

Supporting this notion is the finding that FIH enzymatically hydroxylates gankyrin in a process that is completely inhibited when gankyrin is supplied in a complex with the S6-ATPase C-terminal domain (see Chapter 6 for further details). Investigations into the hydroxylation of other ankyrin repeat proteins by FIH suggests that unfolding is a necessary pre-requisite for hydroxylation (Coleman, McDonough et al. 2007; Cockman, Webb et al. 2009). Taken together, these data strongly support the notion that gankyrin interacts with pRb in its folded state, not in the extended peptide conformation observed for LXCXE-containing interaction partners.

3.1.2 Protein-protein docking

Protein-protein docking is the use of computational tools in an attempt to predict the structure of a complex of two or more proteins. Increasingly the techniques extend to other biopolymers, but usually not to small molecules. As is the case for all structure prediction algorithms from protein structure prediction (folding) to docking of small molecules or biomolecules, the protein-protein docking problem can be separated into a search method and a scoring method. Thus it can be seen that the problem is two-fold. Firstly, the conformational space of the various relative rotations and translations of the two proteins is very large, and once the conformational freedom of the proteins themselves is also considered is virtually infinite, given that it includes everything from side-chain rotations and backbone perturbations through to entirely new sections of secondary structure or wholesale movement of loops that can be induced upon binding. Secondly, there must be a method for scoring and ranking the various conformations in order to determine which are best, and ideally to assess the likelihood that they bear similarity to the native structure of the complex.

Scoring is often based either on molecular dynamics force-fields, as in the case of HADDOCK(de Vries, van Dijk et al. 2007), or potentials based on the observed frequency and geometry of interactions between particular residue types or atom types, as in the case of RosettaDock(Kortemme, Morozov et al. 2003; Schueler-Furman, Wang et al. 2005) and the RPscore module of 3D-Dock(Gabb, Jackson et al. 1997; Moont, Gabb et al. 1999). The accuracy of such scoring functions depends primarily on access to good data on which to base them, and so improvements in scoring have been incremental and made only marginal differences to the accuracy of protein-protein docking.

On the other hand, a variety of methods are employed to address the search problem. The majority of docking algorithms can be divided into two classes. The first class employ a (Fast) Fourier Transform (FFT) correlation algorithm which allows evaluation of a shape complementarity function for a complete search of rotations and translations about the three Cartesian axes (given defined increment sizes). Thus, given the same input, any two runs of these algorithms will always yield the same output. The Fourier transform vastly increases the efficiency of the algorithm (making it tractable) by allowing the calculation of the function in far fewer steps than would be required for the explicit generation and evaluation of each complex. Algorithms in this class, like FTDock (Katchalski-Katzir, Shariv et al. 1992), its successor 3D-Dock (Gabb, Jackson et al. 1997; Moont, Gabb et al. 1999) and Z-Dock (Chen, Li et al. 2003) typically do not incorporate any methods to explicitly model the flexibility of the proteins involved. However neither do they treat the proteins as rigid bodies. Instead, the edges of the protein are considered soft, or fuzzy – that is, some overlap of the binding partners is not penalized by the algorithm, thereby implicitly allowing for some induced fit and a small amount of conformational change upon binding.

The second class of algorithms are based on Monte Carlo methods, though the precise implementation can vary significantly. Some are based on some form of molecular dynamics, such as the ARIA-based HADDOCK(Dominguez, Boelens et al. 2003; Linge, Habeck et al. 2003; de Vries,

van Dijk et al. 2007) which conducts its search by beginning with randomized relative orientations of the binding partners followed by a simulated annealing protocol first with rigid bodies, then in torsion angle space and finally in Cartesian space, guided by an additional set of energy potentials based on biochemical data about the nature of the interaction. RosettaDock, another popular algorithm from this class, similarly starts with randomized orientations, followed by a Monte Carlo minimization in which random steps of translation and rotation are used to find a local minimum of the energy evaluation function, first in a low-resolution "centroid" mode in which each residue is represented by a single point, and ultimately in a full-atom representation using rotamer libraries and side-chain repacking to conduct the search(Gray, Moughon et al. 2003; Schueler-Furman, Wang et al. 2005). By their nature as Monte Carlo search algorithms, repeated runs of the same job yield different results. Thus, these algorithms are used by repeating the job numerous times (usually thousands to tens of thousands) and looking for convergence among the output – that is, sets of similar structures which all score favourably – rather than necessarily accepting the single best-scoring output as the best solution.

For several years now, a community-wide experiment known as CAPRI (<u>C</u>ritical <u>A</u>ssessment of <u>PR</u>edicted <u>I</u>nteractions) has been underway, allowing the authors and users of protein-protein docking algorithms to assess their techniques (Janin, Henrick et al. 2003). Participants are asked to predict interactions for which the complex structure is solved, but not yet in the public domain. Usually, structures of the components of the target structure are available in their unbound state, though in some cases participants have been required to predict the structure of one of the components as well as the overall complex. It represents the most "real-world" assessment of protein-protein docking techniques possible. While the outcomes of CAPRI can obviously be used in an attempt to assess which single algorithm performs best overall, or for a particular class of problem, arguably a more important outcome is to discern patterns among the types of approaches that work (or not) and the classes of problems which are readily solved, or which remain difficult or impossible to solve with current methods. Some lessons do seem to be fairly

clear from the CAPRI experiment. Firstly, there is evidence that in most cases in which conformational change is absent, the FFT methods perform very well – it is only through their implementation of explicit protein flexibility that the Monte Carlo gain a significant advantage (Janin 2010). Secondly, restraints or filters based on biochemical information – which are often necessary for successful and efficient operation of the Monte Carlo methods – need to be implemented with considerable care. Some of the CAPRI targets for which there was almost universal failure failed because of the inclusion of restraints and filters based on functional and structural conservation which turned out to be false (Mendez, Leplae et al. 2003). Finally – and unsurprisingly – there is room for improvement in both the search methods and scoring methods used for protein-protein docking (Janin 2010).

3.2 Aims and Approaches

Studies investigating the interaction between gankyrin and pRb have suggested that gankyrin interacts with pRb via the LxCxE motif present in the fifth ankyrin repeat (Higashitsuji, Itoh et al. 2000; Li and Tsai 2002). Subsequent structural studies (Krzywda, Brzozowski et al. 2004; Padmanabhan, Adachi et al. 2004; Yuan, Li et al. 2004) have highlighted the inconsistency between the known mechanism by which pRb interacts with an LxCxE peptide (Lee, Chang et al. 2002)(Figure 3.1A) and the helical conformation of the LxCxE of gankyrin and the inaccessibility of the motif other than residue E182 (Figure 3.1B). (Nakamura, Nakano et al. 2007) conclude that the LxCxE motif of gankyrin is unlikely to interact with pRb, or at least not in the same way as the HPV E7 LxCxE peptide on the basis of their finding that the gankyrin/S6-C complex retains the ability to interact with pRb. However, no alternative hypothesis as to the nature of the interaction has been presented.
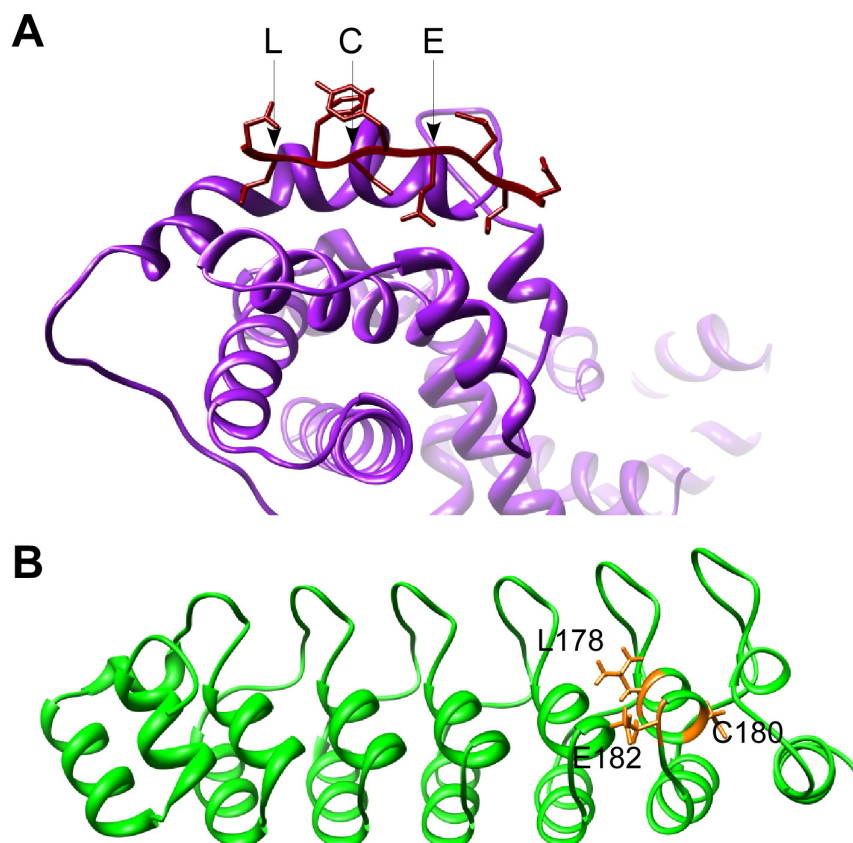
Figure 3.1 (A) An LXCXE peptide (dark red) from the E7 protein of human papilloma virus (HPV) in complex with the small pocket domain of pRb (purple) from the PDB entry 1GUX (Lee, Russo et al. 1998). The critical binding residues are labeled. (B) The "LXCXE motif" (residues 178-182; highlighted orange) as found in the context of the ankyrin repeat protein gankyrin from PDB entry 1UOH.

The aim of this section of work was to establish an alternative to the LxCxE-based model of the gankyrin-pRb interaction via *in silico* methods. This model was to be consistent with existing experimental observations.

Despite the fact that a variety of more advanced protein-docking techniques have been developed since FTDock and subsequent components of 3D-Dock, the approach employed by 3D-Dock offers certain advantages, especially in the context of some of the peculiarities of the particular problem at hand.

55

The advantages offered by other approaches are principally concerned with incorporating representations of protein flexibility (beyond the implicit side-chain flexibility incorporated by FTDock), and more efficient search methods than the uniformly distributed global conformation search employed by FTDock. While these characteristics undoubtedly reflect the general biological situation more accurately, they are not necessarily particularly applicable to this particular problem. By interpreting the fact that the gankyrin/S6-ATPase complex retains the ability to interact with pRb as a suggestion that little or no structural re-arrangement of gankyrin is possible, a rigid-body docking algorithm becomes the appropriate choice for the problem. Similarly, a uniformly distributed global search allows evaluation of precisely the same conformations of the gankyrin/pRb complex in the presence and absence of the S6-ATPase, or the E182A mutation, where other more dynamic search methods would alter their search patterns due to the effects of a third protein chain or a mutation. This repeatability allows simple and explicit comparison of the effects of these additional considerations, which offers a significant advantage in matching the experimental data to the available biological data. Such an advantage is invaluable in this case, given that the objective is to establish a model which better matches the available biological data than that which has already been proposed.

3.3 Results

3.3.1 Parallelisation of 3D-Dock

The FTDock component of the 3D-Dock program operates by maintaining one protein binding partner static, and explicitly rotating the second binding partner in user-defined increments along X, Y and Z axes. The translation of the mobile binding partner is handled by a Fourier Transform based routine, thus dramatically reducing the number of computational steps, and hence the computational time.

The fact that the rotation of the mobile partner is calculated explicitly presents an opportunity for speeding computation (in real time) via parallelization of the algorithm, making finer searches of larger proteins – such as gankyrin and pRb – more accessible. While such parallelization has often been achieved using Message Passing Interface (MPI) programming, to allow parallel threads of the algorithm to execute on different nodes of a cluster computer. However, the increasing use of multi-processor computers, and more particularly CPUs with several computational cores means that significant parallelization (easily up to eight threads) can be achieved in a shared-memory environment, without the need for the complexities of a message-passing parallelization regime. OpenMP (Open Multi-Processing) is a widely used and supported API that has been developed for such tasks, operating via simple compiler directives and environment variables, requiring few, if any, calls to library routines, and thus minimal code alteration.

Conveniently, the rotation of the mobile partner is governed programmatically by a single `for` loop, with independent rotation states calculated on the basis of a single rotation variable that increases in value by 1 with every loop completion. Parallelization of this loop alone is sufficient to achieve significant time savings in calculation, and is relatively easily achieved via OpenMP since each calculation of this loop is independent. Since each loop iteration contains an essentially invariant amount of work, x number of loop iterations can be efficiently split between n processors equally by number, with no further regard to load balancing. OpenMP leaves the precise

implementation of this to the compiler, but for illustrative purposes it can be reasonably conceived as a regime in which processor 1 executes iterations 1 to (x/n), processor 2 executes iterations (x/n + 1) to (2*x/n) in parallel, and so on, with processor n executing the final x/n iterations.

OpenMP (Open Multi-Processing) was used to alter the FTDock code and allow parallel execution. Appendix A details the changes made to ftdock.c for parallel execution, but briefly, parallelization was achieved by:

- Parallelization of the main for loop on a "static" schedule – ie approximately equal numbers of iterations per thread via incorporation of the `#pragma omp for schedule(static)` immediately prior to the start of the main `for` loop.

- Directing output from independent threads into separate scratch files to avoid write conflicts between threads, and then compiling the data from these scratch files after completion of the parallel code loop.

- Similarly, the original FTDock code tracks the maximum (best) score as it passes through each iteration. Taking the safest approach of treating shared/global variables as read-only during parallel execution, this is handled simply by maintaining independent "local maximum" variables in each thread, and resolving these to a global maximum on completion.

- Defining variables that were to be shared between independent threads, such as the electrostatic and surface grids of the static protein before initiation of parallel threads.

- Specifying the simple variables that were to be private to each independent thread – consisting of loop iterators variables associated with the rotation of the mobile partner, and local maximum variables –using the `#pragma omp parallel private` directive. Complex memory allocations such as Fourier Transform matrices for mobile partner surface and electrostatics grids were explicitly declared independently by each thread using the code block initiated by the above directive.

The validity of the parallelized code was established by docking p16 (as the mobile partner) to

CDK6 (static partner) as isolated from PDB entry 1BI7(Russo, Tong et al. 1998) using both the

single-thread and multi-thread versions of FTDock and comparing output. Both versions

successfully and precisely recapitulate the published structure as the most favoured binding

conformation (RMSD 0.0 between predicted and published structures), though this level of

precision was almost certainly due to the fact that docking was carried out using non-randomized,

bound conformation input structures, however this was sufficient to verify a successful

implementation of the method as required here rather than a test of the method itself as

employed in the algorithm. More importantly from this perspective, the output from the original

and modified versions of FTDock were identical when compiled on the same CPU architecture.

Using a 2.0 Å surface grid spacing and 9 degree rotation, execution of the modified code in a two-

processor environment took 64.2% of the time taken for execution of the original program in a

single thread. The additional time over the theoretical optimum of 50% was due to the non-

parallel calculation of static grid parameters and final compilation and sorting of results.

3.3.2 Protein-protein docking of gankyrin and pRb

A representation of the E182A mutant of gankyrin was prepared by simple truncation of

residue E182 at the alpha carbon using PDB entry of 1UOH. This PDB representation in both the

wild-type and E182A forms, as well as the pRb atoms only (residues 380-473, 577-785) of PDB ID

1GUX, representing the small pocket domain of pRb, were prepared for use with FTDock using the

provided pre-processing scripts, and the orientation of pRb was randomized. Both the wild-type

and E182A mutant gankyrin were docked independently as the mobile partners to pRb. The

default values were retained for grid-spacing, surface thickness, and internal deterrence value (0.7

Å, 1.3 Å and -15.0 respectively) and a search angle step of 9 degrees was used. Calculations were

carried out on 8 parallel processors each, taking approximately 90 CPU-hours of computation time each.

Given the primacy being placed on the biological data in this case, the conformations were first filtered on this basis, before any further analysis of the docking run was undertaken. Conformations were filtered on the basis that pRb must contact gankyrin residue E182, and the region of pRb that binds gankryin must substantially overlap the region responsible for LxCxE binding. Using 3D-Dock's `filter` program, these restraints were implemented by specifying that (1) some (any) part of pRb must lie within 4.5 Å (3D-Dock's default cut-off) of gankyrin residue E182, and that (2) some (any) part of gankyrin must lie within 6.0 Å of pRb residue Y709, which lies approximately at the centre of the LxCxE binding site (illustrated in Figure 3.2). Of the 64,320 conformations analysed, 790 (1.2%) met these criteria. Explicit PDB representations of these 790 protein-protein complexes were built. The C-terminal domain of S6 ATPase was included in these models by superimposing the gankyrin/S6 ATPase structure (PDB ID 2DVW) onto the gankyrin only structure used for protein-protein docking. Providing the gankyrin/S6 ATPase complex to the 3D-Dock `build` program in place of that of gankyrin only, results in the creation of a matching hypothetical ternary gankyrin/S6 ATPase/pRb complex.
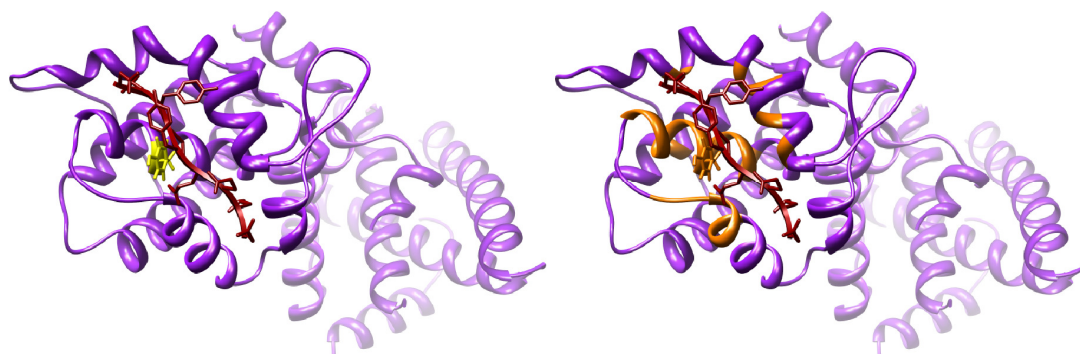
Figure 3.2 Illustration of pose filtering constraint. LxCxE peptides such as that from HPV E7 (dark red) bind a groove on the surface of the pRb pocket domain (purple). *Left:* pRb residue Y709 (yellow) lies approximately at the centre of this binding groove. *Right:* Since gankyrin binding to pRb competes with that of LxCxE peptides, the binding sites on pRb must overlap to some degree. As a simple approximation for this, a 6.0 Å radius zone was defined around pRb Y709 (orange) within which some part of gankyrin must lie for a given pose to pass the filtering criterion.

The ternary complexes were screened for steric clashes between the S6 ATPase and pRb

protein chains using the "Find clashes" tool of UCSF Chimera. Of the 790 complexes built, 43

exhibited no such steric clashes, and of these 9 had a surface complementarity score of over 193,

which represents the 90th percentile cut-off for the surface complementarity score among the

original 64,320 conformations.

Table 3.1 Top ranking poses from FTDock global search after filtering against constraints derived from experimental data

and corresponding surface complementarity, residue potential scores and associated data. Bold font indicates a group of

highly similar poses.

Rotation[†]			Translation[†]			SC Score	SC Score Rank[‡]	RP Score	RP Score (E182A)	ΔRP Score
X	Y	Z	X	Y	Z					
126	**72**	**126**	**-44**	**-60**	**-9**	**285**	**10**	**-1.52**	**-1.95**	**-0.43**
126	**72**	**135**	**-40**	**-64**	**-9**	**263**	**57**	**-1.41**	**-1.51**	**-0.1**
126	**72**	**135**	**-40**	**-64**	**-8**	**254**	**105**	**-0.51**	**-1.19**	**-0.68**
90	117	150	-9	-55	8	253	114	-3.88	-3.6	0.28
171	117	210	-10	-78	11	235	385	-7.3	-8.04	-0.74
144	162	192	-16	-74	4	215	1533	-5.04	-6.11	-1.07
261	36	30	-28	-66	26	213	1775	-6.16	-6.36	-0.2
171	162	216	-18	-73	10	203	3457	-4.2	-4.52	-0.32
126	**72**	**126**	**-43**	**-61**	**-9**	**202**	**3669**	**-0.9**	**n/a**	**n/a**

[†] Rotation and translation of gankyrin
[‡] Rank by surface complementarity within 64,320 poses from initial global search

As can be seen from Table 3.1, the highest ranked conformation of this heavily screened

subset was also very highly ranked (10) among the global search of 64,320 conformations. Further,

there are 3 additional conformations that are slight variations on the first (as evidenced by their

rotational and translational coordinates). This type of convergence is often strongly indicative of

correct prediction of protein structure and the structure of protein complexes (Tovchigrechko and

Vakser 2001; Kozakov, Schueler-Furman et al. 2008). While none of these top 9 conformations

have strongly positive residue potential scores, the top conformation and those similar to it have

substantially more favourable residue potential scores than the other 5, and all within the 75th

percentile of residue potential scores in the global search. Finally, the residue potential scores

deteriorate for all of these complexes when calculated with the E182A gankyrin mutant, as would

be expected based on the experimental data. Thus, the result is the prediction of a novel mode of interaction between pRb and gankyrin (depicted in Figure 3.3) which better fits the available biochemical data (Figure 3.4), and which is strongly supported by theoretical measures.
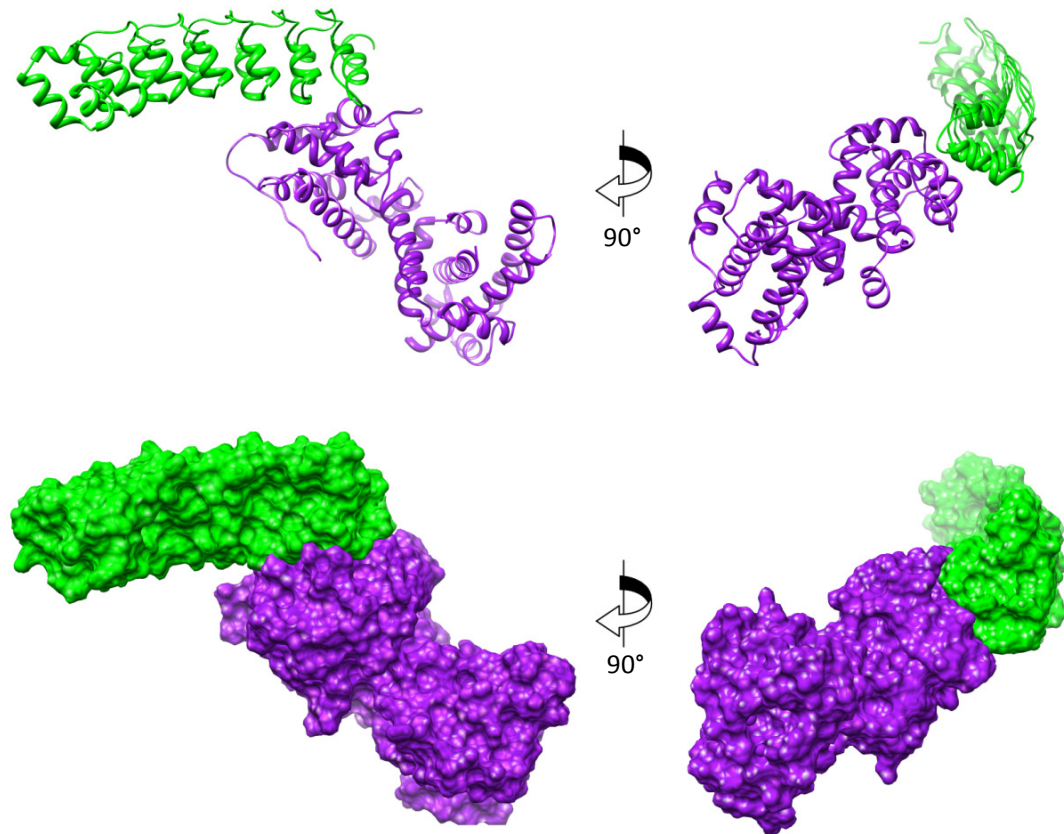
Figure 3.3 Depiction of the top-ranking pose (after filtering) of gankyrin (green) and the pRb pocket domains (purple). Both ribbon representations (top) and molecular surface representations (bottom) are shown.
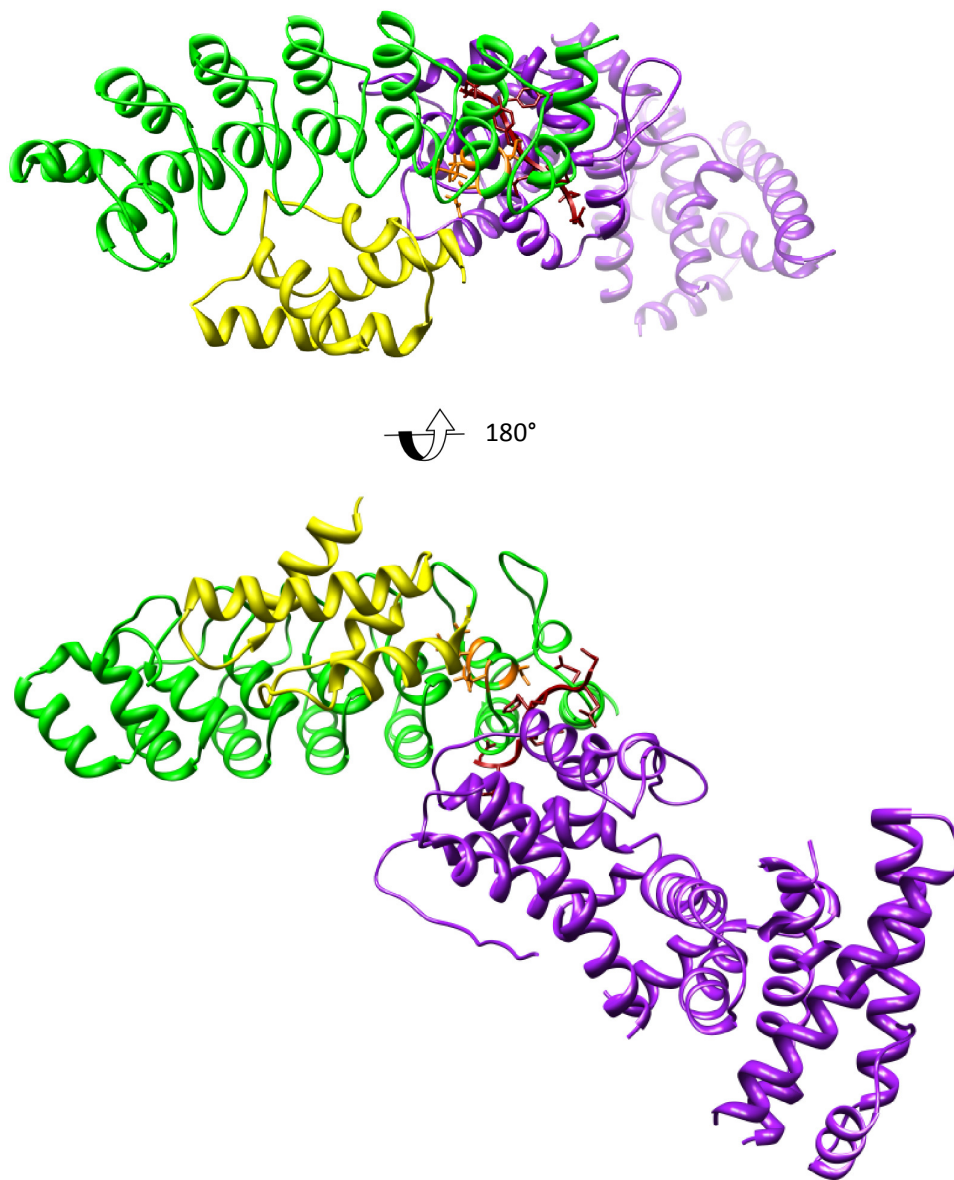
Figure 3.4 The proposed new binding model for gankyrin (green) and pRb (purple) better explain the available experimental data. (1) The respective binding interfaces of gankyrin and the HPV E7 LXCXE peptide (dark red) on pRb overlap extensively, accounting for competition between the two binding modes. (2) The surface areas of gankyrin which contact the C-terminal domain of S6-ATPase (yellow) and pRb are entirely distinct as would be expected from data showing that the S6/gankyrin complex retains the ability to bind pRb. (3) Gankyrin residue E182, of gankyrin's LxCxE motif (orange) is part of the gankyrin/pRb interface, forming a salt bridge with pRb residue K720 (not shown here). This potentially explains the loss of pRb binding in the E182A gankyrin mutant.

3.3.3 Analysis of a novel model for the gankyrin-pRb interaction

The output of FTDock retains the side-chain conformations of the input structures, while it is likely that the complex induces some local rearrangements in the presence of the binding partner. In order to allow such local flexibility and optimize the geometry of partial interactions, conjugate gradient minimization of the complex was performed, to enable easier and more reliable identification of the key interactions at the interface. Hydrogen atom coordinates were calculated and peptide force field parameters prepared for the complex using XPLOR-NIH v2.25(Schwieters, Kuszewski et al. 2003) and the CHARMM force field(MacKerell, Bashford et al. 1998). The complex was subsequently solvated in a box such that non-water atoms were no closer than 10 Å to the edge of the box, using SOLVATE v1.3 as employed in VMD v1.8.7b5 (Humphrey, Dalke et al. 1996). Finally, the system was minimized for 10,000 cycles using NAMD v2.6 (Phillips, Braun et al. 2005) leaving backbone atoms fixed to preserve the overall structure of the complex. Analysis of the minimization trajectory showed that several side-chain interactions at the interface were indeed able to adopt a more favourable geometry as a result of this process.

Following minimization the binding interface was inspected to explore likely binding determinants. The proposed binding interface covers approximately 800 Å^2 and consists of a small area of hydrophobic contact toward the centre of the interface, with numerous charge interactions surrounding it (Figure 3.5). Several hydrogen bonds between both sidechain and backbone atoms also contribute to binding. In addition to those shown in Figure 3.5A, there are several other pairs of charged residues that are in close proximity and may interact, but are not shown to by the specific model generated above. Table 3.2 summarises all possible charged residue interactions. No interactions between residues of like charge are found in the proposed model.

Table 3.2 Summary of putative interactions between charged residues of gankyrin and pRb. Interactions listed as "explicitly modeled" are observed as close (< 3.0 Å), direct interactions in the energy minimized model. Interactions that are not "explicitly modeled" would be made possible by adjustment of torsion angles within the sidechain, yet the charge centres of the respective sidechains are greater than 3.0 Å apart.

Gankyrin residue(s)	pRb residue(s)	Explicitly modeled
K149 (with R184)	D718	Yes
E182	K720	Yes
E186	K765	Yes
E183	K722	No
E187	K713	No
R222	E748	No
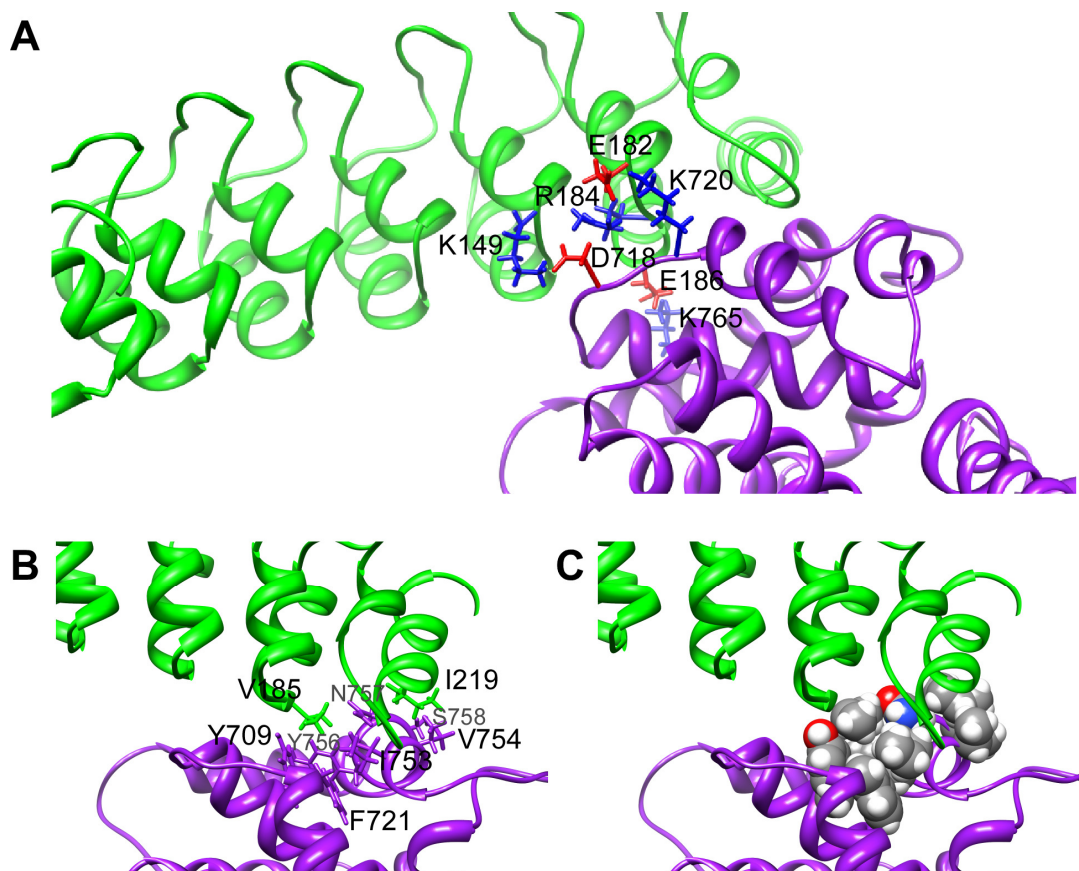
Figure 3.5 Molecular details of the proposed gankyrin/pRb binding model. (A) Several salt-bridges between gankyrin and pRb (residues E182 & K720, E186 & K765, K149/R184 & D718) are likely key determinants of binding. Positively charged residues are shown in blue, negatively charged residues in red. (B and C) Gankyrin residues V185 and I219 bind hydrophobic patches on the pRb surface. V185 fits into a pocket created by pRb residues Y709, Y756, F721 and I753. I219 contacts pRb residue V754, as well as the hydrophobic portions (α and β carbons) of residues S758 and N757. Residues are shown in stick (B) and sphere (C) representations to highlight hydrophobic contacts.

3.3.4 Specificity of binding model

On the basis that the salt bridges noted above in the binding model are the key determinants of binding, bioinformatic approaches were employed to investigate the likely specificity of the interaction. Though a comprehensive theoretical assessment of the specificity of the interaction is impossible with current knowledge, methods and computational resources, in this instance some useful insights can be gained from simple sequence alignments.

Previous studies have demonstrated that the RB family members p107 and p130 also bind peptides containing the LXCXE motif via homologous pocket domains (DeCaprio, Ludlow et al. 1988; Dyson, Guida et al. 1992). Structural studies (Lee, Russo et al. 1998) have further demonstrated which residues of pRb contact the HPV E7 LXCXE peptide, and that most of these residues are conserved – either entirely or in chemical class – with the other pRb family members p107 and p130. Gankyrin, however, binds only pRb and not p107 nor p130 (Higashitsuji, Itoh et al. 2000). Figure 3.6 demonstrates that the residues putatively involved in binding gankyrin are largely distinct from those involved in binding LXCXE peptides, despite their heavily overlapping interfaces. Focusing on the three residues for which explicit charge interactions are observed in the binding model (hereafter referred to as "principal residues" or "principal interactions"), only one of these (D718) exhibits charge conservation with p107 or p130. Similarly, of the three "additional residues" which the model suggests may contribute additional binding energy, only one (K713) exhibits charge conservation with p107 or p130. These observations are consistent with the experimental observation that gankyrin binds pRb but not p107 nor p130.

Turning now to the gankyrin side of the interaction, a similar process can be used to predict whether other ankyrin repeat proteins are able to bind pRb in a similar fashion to that proposed for gankyrin. However, there being 440 human proteins collectively containing 2310 ankyrin repeats (according to the SMART tool (Letunic, Doerks et al. 2009) at the time of analysis) complicates the analysis somewhat, making the problem only tractable by automated methods.

```
                         :  *   .  :      ::* ::  **:**::***:** :**. ** :*   ..
p107    ------ASPKQTNLTKAQEVHSTGINRPKRTGSLALFYRKVYHLASVRLRDLCLKLDVSN    814
p130    QVAIQQISPGGQQQKQGQSVTSS-SNRPRKTSSLSLFFRKVYHLAAVRLRDLCAKLDISD    862
pRb     -------------AETQATSAFQTQKPLKSTSLSLFYKKVYRLAYLRLNTLCERLLSEH    673

         **.: *** *:.:: :   :** ******:::*::* :.**.:  :  .*: *: .*:   *:
p107    -ELRRKIWTCFEFTLVHCPDLMKDRHLDQLLLCAFYIMAKVTKEERTFQEIMKSYRNQPQ    873
p130    -ELRRKIWTCFEFSIIQCPELMMDRHLDQLLMCAIYVMAKVTKEDKSFQNIMRCYRTQPQ    921
pRb     PELEHIIWTLFQHTLQNEYELMRDRHLDQIMMCSMYGICKVKNIDLKFKIIVTAYKDLPH    733
                                          ^      ^ ^ ^
          *  ...:: **:*                        .:: :    .
p107    ANSHVYRSVLLK----------STPREVVAYNKINDDFEMIDCD-------------LE    910
p130    ARSQVYRSVLIKGKRKRRNSGSSDSRSHQNSPTELNKDRTSRDSSPVMRSSSTLPVPQPS    981
pRb     AVQETEKRVLIK----EEEYDSIIVFYNSVFMQRLKTNILQYASTR-----------PP    777
                              ^                ^
```

Figure 3.6 Sequence alignment of pRb, p107 and p130. Purple highlighted residues indicate those that Lee et al found contact the LXCXE peptide in pRb. Full identities of these residues in p107 and p130 are similarly highlighted, while similar residues are highlighted pink. Residues observed to form charged interactions in the proposed binding model are marked with a green highlighted arrow, while those that may form additional salt bridges are marked with a pale green highlighted arrow.

In order to be as open about the problem as possible, and maximize the chance of identifying other possible pRb-binding ankyrin proteins, only the three principal charge interactions and the two ankyrin repeats across which they occur were considered. This is despite the possibility that an additional three residues may contribute substantially to specificity, and the fact that the model suggests that the binding interface crosses three ankyrin repeats of gankyrin. Therefore, the list of 2310 human ankyrin repeats obtained from SMART was used to compile a list of starting and ending residue numbers of 1870 pairs of adjacent ankyrin repeats. These in turn were used to extract the sequences of these ankyrin repeat pairs using the list of ankyrin repeat-containing proteins similarly obtained from SMART. This approach ensures assembly of the correct sequence, since there are often small numbers (1-2) of residues in between SMART's definition of ankyrin repeats that do not appear in its list of ankyrin repeats. One disadvantage of the approach was that where a protein contained more than one cluster of ankyrin repeats, the last repeat of the first cluster and the first repeat of the second cluster were considered "adjacent". With this in mind, ankyrin repeat pairs greater than 68 amino acids in length (where the SMART consensus defines an ankyrin repeat to be 33 amino acids long) were marked as "long" but were retained in

the pool to be analysed, with the intention that "long" pairs would be analysed more closely, should they meet the criteria of the search.

Each member of the ensemble of ankyrin repeat pairs was then aligned against repeats 4 and 5 of gankyrin (residues 138-201) using the T-COFFEE algorithm v1.37 (Notredame, Higgins et al. 2000). On the basis of these alignments, it was then possible to determine the identity of the residues in analogous positions to those of gankyrin residues K149, E182, R184 and E186. It was found that of the 1870 ankyrin repeat pairs, only 3 exhibited charge conservation at all 4 of these sites. Further analysis revealed that despite SMART's attempts to make their database non-redundant, these all corresponded to gankyrin – each entry from different database sources, often with minor sequence variations either due to poor quality data or side-effects of cloning and purification in the case of entries derived from structural databases. However, both K149 and R184 interact with the same negative charge centre on pRb, with K149 providing the most direct interaction and R184 providing an additional, potentially non-essential interaction. If R184 is excluded from the analysis, 30 entries exhibit charge conservation at the 3 positions corresponding to gankyrin residues K149, E182 and E186. After accounting for database redundancy and protein isoforms, this corresponds to 15 entirely distinct proteins, other than gankyrin, which encode residues in such an arrangement that they might be capable of interacting with pRb in a similar fashion to that proposed by the gankyrin/pRb model. For interest's sake, these are listed in Table 3.3. While some of these proteins provide intriguing possibilities for other ankyrin repeat proteins that may exert similar effects on the pRb pathway to that of gankyrin, further investigation of these would be premature without first verifying whether the proposed binding model is accurate.

Table 3.3 Potential pRb-binding ankyrin repeat proteins as determined by sequence alignment analysis. Accession numbers beginning with 'NC_' are TROME accession numbers and those beginning with 'NP_' are RefSeq accession numbers. All others are UniProt accession numbers.

Protein Name	UniProt/RefSeq/TROME ref
Ankyrin 2	Q01484
RNaseL	Q05823
Ankyrin 3	Q59G01
Uncharacterised protein	Q5CZH9
NF-kappa B inhibitor-like protein 1	Q5STV5
Multiple ankyrin repeats single KH domain protein	Q8IWZ2
Ankyrin repeat and KH domain-containing protein 1 Alt. name: HIV-1 Vpr-binding ankyrin repeat protein	Q96G77
Protein phosphatase 1 regulatory inhibitor subunit 16B	Q96T49
Ankyrin repeat domain-containing protein 30B Alt. name: Serologically defined breast cancer antigen NY-BR-1.1	Q9BXX2
Ankyrin repeat domain-containing protein 17 Alt. name: Serologically defined breast cancer antigen NY-BR-16	Q9H288
Ankyrin repeat domain-containing protein 5	Q9NU02
Ankyrin repeat domain-containing protein 53	NP_001108588
Ankyrin repeat domain-containing protein 50	Q9ULJ7
Predicted protein	NC_000005_1040_0

While we cannot know whether any of these particular proteins do indeed directly interact with pRb, the results of the search are nonetheless informative. Because of the ubiquity of ankyrin repeats and because of the fact that certain side-chain chemistries are more often found in particular positions within the consensus sequence, it was possible that the key interactions noted above in the analysis of the binding model would be widely conserved among ankyrin repeat proteins. Given that gankyrin is the only ankyrin repeat protein to have been shown, or even suggested to interact with pRb, conservation of the key determinants of binding would have been a strong suggestion that the proposed binding model was unrealistic in the biological context. However, the finding that of 1870 pairs of ankyrin repeats, only 30 (1.6%) exhibit conservation of the most minimal definition of the key binding determinants suggests that the binding model proposed exhibits an appropriate level of specificity. While the three residues examined here may very well be necessary for the gankyrin/pRb interaction, they are almost certainly not sufficient in their own right. Thus it would be reasonable to expect that many of the 15 proteins identified would in fact not bind pRb at all, even if the premises of the search, including the proposed

binding model, are correct. Nonetheless, this search and its results would be an interesting exercise to revisit at a future time once the nature of the gankyrin/pRb interaction has been verified experimentally.

3.4 Discussion

An interaction between gankyrin and pRb mediated via the classical LXCXE motif provides plausible explanations for the key biological consequences attributed to gankyrin up-regulation. Specifically, recruitment of both pRb and CDK4 into a common complex, coupled with gankyrin's ability to de-repress CDK4 by competition with the INK4A proteins can explain the increased phosphorylation and degradation of pRb observed when gankyrin is up-regulated. It is also thought that pRb binds histone modifying enzymes via their LXCXE motifs and that these enzymes contribute to silencing of E2F target genes when pRb is in its hypophosphorylated state. Part of the oncogenic mechanism of viral proteins such as HPV E7 is hypothesized to be to compete with these enzymes for binding to pRb and therefore removing this layer of control of the genes important to the G1/S phase transition (Dick 2007). A similar mechanism could very easily be ascribed to gankyrin. Yet either the binding model proposed here or an interaction via the canonical LXCXE binding motif could fulfill these roles.

One possibility, as alluded to earlier, is that a binding mode that does not rely on the LXCXE motif allows the observed pattern of specificity within the RB family of proteins to be more easily achieved. However, examples of LXCXE peptides that bind pRb but not p107 or p130 have been previously reported (Dick 2007) meaning that this is not a compelling argument in its own right. A further possibility is that binding of pRb and CDK4 to gankyrin orients them in a way that promotes or facilitates phosphorylation of pRb – or a specific residue thereof – by CDK4. Indeed, Higashitsuji et al (2000) report that overexpression of gankyrin in U2-OS cells results in increased phosphorylation of pRb at residues Ser249/252, Thr393 and Ser795 but not Ser780 nor Ser807/811. Such specificity might be more readily achieved with a more rigid binding mode than that afforded by the random peptide conformation of the LXCXE.

A distinct binding mode for gankyrin that still competes with LXCXE peptide binding also has implications for the nature of this competition. A variety of cellular proteins contain LXCXE motifs through which they interact with pRb. The exact purpose and role of these interactions is still the subject of considerable discussion. However, there is almost certainly some level of competition between these binding partners, and while some of this will be resolved or affected by sub-cellular localization, the relative strength of binding among different partners is likely to have significant affects on cellular outcomes. The non-consensus residues within and immediately adjacent to the LXCXE motif will provide some scope for modulation of binding affinities, but within a relatively limited scope. An entirely separate mode of binding potentially broadens this scope, and allows greater possibilities for biological decision-making via competitive binding.

Developing this line of thinking further, a separate mode of binding also allows for distinct methods of *modifying* interaction affinities. That is, post-translational modifications of pRb could potentially be used to disrupt or enhance the interaction described above in such a way that it does not affect the affinity of partners interacting via the LXCXE motif. The opposite is also possible. Indeed there is existing evidence that phosphorylation of Thr 821/826 is able to disrupt the interaction of LXCXE peptides (Knudsen and Wang 1996), but the effects of these phosphorylations on the gankyrin-pRb interaction are unknown.

Of course, all of this discussion is rather speculative without further experimental evidence to support the proposed binding model presented here. The existing evidence is inconclusive in terms of whether the gankyrin/pRb interaction occurs via a canonical LXCXE mechanism, but short of the resolution of the complex structure via xray crystallography (potentially difficult due to the low affinity of the interaction) or NMR (made even more challenging by the size of the complex concerned) further insight into the problem is difficult. The development of a specific proposal for how the interaction may occur such as that presented here allows the design of simpler, more targeted experiments to be designed to probe the hypothesis.

Chapter 4

In vitro investigation of the gankyrin-pRb interaction

4.1 Introduction

4.1.1 Gankyrin-pRb binding model

In the previous chapter, a structural model of the interaction between pRb and gankyrin was developed via *in silico* methods. One of the predictions of this model was that a series of salt bridges (interactions between residues of opposite charge) rather than the residues of a supposed LXCXE motif in gankryin, are critical to the formation of the complex with pRb. Specifically, gankyrin residues E182, E186 and K149 are predicted to interact with pRb residues K720, K765 and D718, respectively (see Section 3.3.3 for details). Previous studies have demonstrated that the E182A gankyrin mutant loses that ability to bind pRb, both alone, and when in complex with the C-terminal domain of the S6 ATPase (Higashitsuji, Itoh et al. 2000; Nakamura, Nakano et al. 2007). The effect of the E182A mutant is consistent with both the canonical LXCXE-based model of the interaction coupled with partial unfolding of gankyrin to facilitate the interaction, as well as the "rigid-body" interaction model proposed in Chapter 3. However, the inability of the S6-ATPase C-terminal domain to compete with pRb for binding to gankyrin is largely inconsistent with a model that requires partial or complete unfolding of gankyrin to allow the interaction to take place. The S6 ATPase-gankyrin complex is extremely stable, allowing purification of gankryin via a His-tagged S6 ATPase, and enabling solubilisation of the otherwise insoluble S6 ATPase (Nakamura, Nakano et al. 2007). This, together with the fact that the S6 ATPase contacts all seven ankyrin repeats of gankyrin, implies that at the very least any equilibrium between a folded and (partially) unfolded form of gankyrin would be shifted toward the folded form in the presence of S6 ATPase, and that this would result in a reduction in the interaction between gankyrin and pRb if that interaction is dependent upon unfolding. In essence, there would be some evidence of at least this allosteric form of competition exerted by the S6 ATPase, which has not been detected thus far, though admittedly the pull-down and immunoprecipitation techniques that have thus far been employed to probe the question are only semi-quantitative.

4.2 Aims and Approaches

On the basis of the previously described gankyrin-pRb binding model and the observation that mutation of gankyrin residue E182 is sufficient to disrupt the interaction, the aim of this section of work is to interrogate the binding model by mutagenesis of gankyrin and pRb. Table 4.1 summarises the planned mutants and their effect on the interaction as predicted by the binding model.

Table 4.1 A series of mutants of gankyrin and pRbSP will be created and assayed in order to interrogate the model of binding presented in Chapter 3.

Gankyrin mutant	pRbSP mutant	Predicted effect
Wild-type	Wild-type	Normal binding
E182K	Wild-type	Complete disruption of binding (based on existing data)
E186K	Wild-type	Partial or complete disruption of binding
K149D	Wild-type	Partial or complete disruption of binding
Wild-type	K720E	Complete disruption of binding
Wild-type	K765E	Partial or complete disruption of binding
Wild-type	D718K	Partial or complete disruption of binding
E182K	K720E	Partial or complete restoration of binding
E186K	K765E	Partial or complete restoration of binding
K149D	D718K	Partial or complete restoration of binding

The use of charge-swap mutations maximizes that likelihood that a single mutation will be capable of causing a detectable reduction in binding affinity. Furthermore, mutating a given residue to the amino acid with which it interacts in the binding model means that the interaction should be able to be restored by using the appropriate pair of mutants (as in the last three rows of Table 4.1). The effectiveness of these compensatory mutants would demonstrate not only that, for instance, residues E182 (gankyrin) and K720 (pRb) are each important for the interaction, but that those two residues specifically interact. If all three of the proposed residue-level interactions can be proven, as three separate points in space, they will constitute strong evidence for the binding model.

Of course, determining the effects of these mutations on the protein interaction requires an interaction assay with which to test binding. Therefore, the first aim of this section of work was to establish such an assay. In addition to using this assay for interrogating the binding model, it would be desirable to have an assay that could also be used for testing competition between different binding partners of gankyrin, and ultimately, for testing the effectiveness of potential inhibitors of the gankyrin/pRb interaction. Thus, a desirable assay would at a minimum be able to be used in competition format, and allow calculation of the dissociation constant (K_d) of the interaction in question, as well as a quantitative measure of the effectiveness of competitors, such as the IC_{50}.

4.3 Results

4.3.1 Sub-cloning, expression and purification of gankyrin and mutants

A DNA fragment encoding full-length gankyrin (G0-6) was generated by PCR from a HepG2 cDNA preparation (generously donated by Tiffany Liu (University of Adelaide)) using oligonucleotides Gank-M1-5′ and Gank-G226-3′ as per the method described in Section 2.3.3. This DNA fragment was sub-cloned into the pET32a(+) vector using the *Eco*RI and *Xho*I restriction sites. The sequence and location of the DNA insert was verified by BigDye sequencing according to the method described in Section 2.3.11 using the T7-Term primer (data not shown). Mutants of gankyrin were created via the QuikChange method, using the primer pairs E182K-QC-fwd/E182K-QC-rev, E186K-QC-fwd/E186K-QC-rev, or K149D-QC-fwd/K149D-QC-rev, and the sequence of the gankyrin-coding region was verified by BigDye Terminator sequencing.

The resulting plasmid (pET32a-G0-6), encoding Trx-6His-G0-6 (or mutants thereof), was transformed into *E. coli* BL21 DE3 cells for IPTG-inducible expression (see Section 2.4.1). Recombinant protein was purified by affinity purification on a Ni-NTA-agarose column. After elution with buffer containing 100mM imidazole, protein-containing fractions were pooled. The resulting protein was either used in this form (Trx-6H-G0-6) or subjected to in-solution proteolysis with thrombin and dialysis against imidazole-free buffer. The resulting protein solution was passed

over a fresh Ni-NTA-agarose column to capture the Trx-6His fragment resulting in a purified

solution of G0-6 with an N-terminal S-tag (S-G0-6). Samples were collected at various stages of

purification for analysis by SDS-PAGE. A representative example is shown in Figure 4.1. The

purified protein migrated at a position consistent with its expected molecular weight of 28.5 kDa

and was reactive to anti-S-tag antibody (see Figure 4.9 or Figure 4.10 for examples), indicating that

the expected protein had been purified.
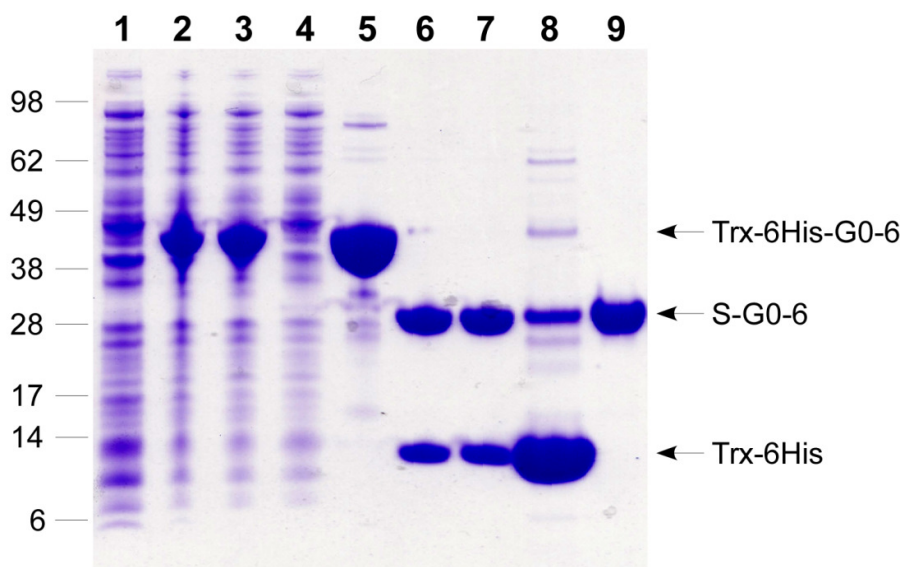
Figure 4.1 Expression and purification of S-G0-6. Lanes: (1) Expression culture whole-cell extract (WCE) prior to induction with IPTG. (2) Expression culture WCE after induction with IPTG. (3) Soluble fraction of expression culture lysate. (4) Protein fraction that did not bind Ni-NTA affinity column. (5) Protein eluted from Ni-NTA affinity column. (6) Protein solution after thrombin digestion. (7) Protein solution after dialysis. (8) Protein fraction that bound to Ni-NTA column (second pass). (9) Purified S-G0-6 protein. Arrows indicate the species of interest, with expected molecular weights of 42.4 kDa (Trx-6His-G0-6), 28.5 kDa (S-G0-6) and 13.9 kDa (Trx-6His). The migration of the molecular weight markers (Invitrogen See-Blue Plus 2) are shown to the left.

4.3.2 Sub-cloning, expression and purification of pRbSP and mutants

A DNA fragment encoding the "small pocket domain" (residues 379-787) of retinoblastoma

protein (pRbSP) was generated by PCR from a pCMV plasmid vector encoding full-length pRb with

an N-terminal HA tag (generously donated by Dr Jo White (University of Adelaide)). The PCR was

carried out as per the method described in Section 2.3.3 using oligonucleotides Rb-M379-5' and

pRb-R787-3′. The resulting DNA fragment was subcloned into the pGEX-4T1 vector using the *Eco*RI and *Xho*I restriction sites. The sequence and location of the DNA insert was verified by sequencing according to the method described in 2.3.11 using both the pGEX-SP-5′ and pGEX-SP3′ primers in separate sequencing reactions. Mutants of pRbSP were created via the Quikchange method, using the primer pairs Rb-K720E -fwd/Rb-K720E -rev, Rb-K765E -fwd/Rb-K765E -rev, or Rb-D718K-fwd/Rb-D718K -rev, and the sequence of the pRb-coding region was verified by BigDye Terminator sequencing.
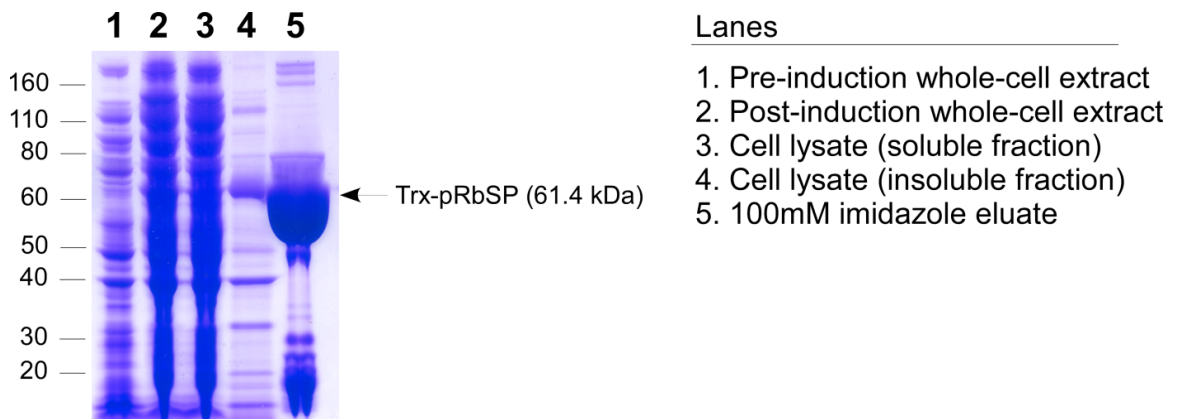
The resulting plasmid, encoding GST-pRbSP, was transformed into *E. coli* BL21 cells for IPTG-inducible expression. *E. coli* BL21 cells harbouring the pGEX-4T2-pRbSP plasmid were grown at 30°C to an OD_{600} of 0.6-0.8, after 1 in 100 dilution from an overnight culture. Expression was induced by addition of IPTG to a final concentration of 0.2mM and carried out by incubation in shake flasks at room temperature for 16 hours. Recombinant protein was purified by affinity purification on a glutathione-agarose column (Section 2.4.3). After elution with buffer containing 10mM reduced glutathione, protein-containing fractions were pooled, concentrated and exchanged into fresh TBS buffer using a PD-10 column (see Section 2.4.6). Typical yields were 4-6 mg purified fusion protein per litre of expression culture. Samples were collected at various stages of purification for analysis by SDS-PAGE. A representative example is shown in Figure 4.2. A fraction of expressed protein appears to be present in the insoluble fraction. However given that protein was successfully purified, and that the method utilizes low temperature expression and is based on that used for purification of pRbSP for crystallization, optimization of the method was not pursued on the assumption that improvements were unlikely. Eluted protein typically showed low levels (<5% estimated by SDS-PAGE) of contaminants. The purified protein product, which migrated at a position consistent with its expected molecular weight of 73.5 kDa and was reactive to anti-GST antibody as detected by Western blot analysis (see Figure 4.9 or Figure 4.10 for examples).
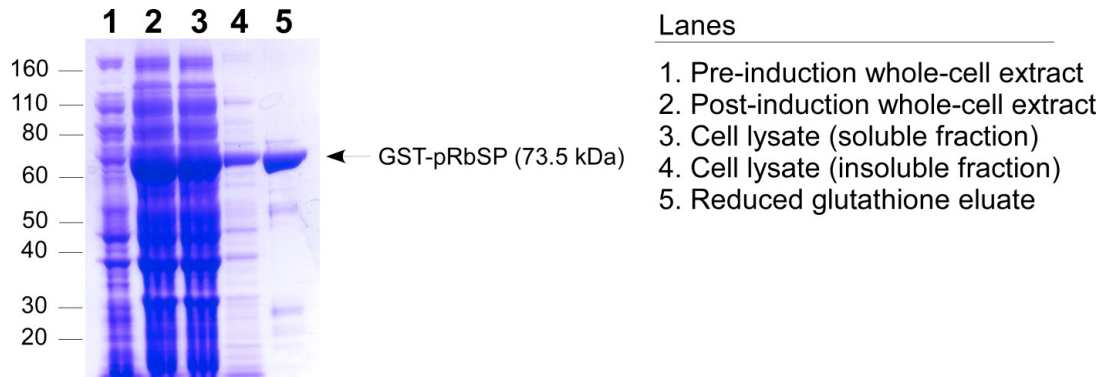
Figure 4.2 Expression and purification of GST-pRbSP protein. Lanes are as indicated in the right panel. The arrow indicates the purified protein product, which migrated at a position consistent with its expected molecular weight of 73.5 kDa.

Protease digestion of the GST-pRbSP fusion protein with thrombin was attempted in TBS buffer containing 2.5mM $CaCl_2$ at various pH and varying ratios of protease to fusion protein, however all conditions resulted in precipitation of protein product. Analysis of the soluble fraction by SDS-PAGE revealed that thrombin digestion yielded additional products over those expected at lower molecular weights, indicating that non-specific cleavage of the fusion protein was occurring (see Section 4.3.4 for further details).

pRbSP was also sub-cloned from pGEX-4T1 into the *EcoRI* and *XhoI* sites of the pET32a plasmid, and transformed into BL21 DE3 *E. coli* cells for expression, using the same method as that for expression of GST-pRbSP (see above). Recombinantly expressed Trx-pRbSP was purified by Ni-NTA affinity chromatography (Section 2.4.4). Purified protein was eluted in buffer containing 100mM imidazole. Typical yields were 5-6 mg purified protein per litre of expression culture. The purified product migrated at a position consistent with its expected molecular weight of 61.4 kDa (Figure 4.3).

Lanes

1. Pre-induction whole-cell extract
2. Post-induction whole-cell extract
3. Cell lysate (soluble fraction)
4. Cell lysate (insoluble fraction)
5. 100mM imidazole eluate

Figure 4.3 Expression and purification of Trx-pRbSP protein. Lanes are as indicated in the right panel. The arrow indicates the purified protein product, which migrated at a position consistent with its expected molecular weight of 61.4 kDa.

4.3.3 Fluorescent labeling of gankyrin

Trx-6His-G0-6 or S-G0-6 were labelled with the amine-reactive fluorescein isothiocyanate (FITC) by incubation of protein and FITC at a ratio of 1:10 (w/w) with protein concentrations in the range of 10-20 mg/mL. Reactions were carried out over 2 hours at room temperature in PBS buffer at pH 7.5 in order to favour reaction at the N-terminal amine over side-chain amine groups. Unreacted FITC was separated from the reaction mixture by two passes over a PD-10 desalting column. The degree of labeling was determined according to the recommendations of the FITC supplier, using the absorbance of the sample at the 280nm and 494nm wavelengths and the following formula:

$$DOL = \frac{A_{494}}{[\text{gankyrin}] \times \in_{FITC}}$$

Where the extinction coefficient of FITC ($\in_{FITC}$) is 68,000 $cm^{-1}M^{-1}$ and the concentration of gankyrin (in mol/L) has been determined according to the following formula:

$$[\text{gankyrin}] = \frac{A_{280} - A_{494} \times CF}{\in_{gankyrin}}$$

Where the correction factor (CF) – the ratio of A280 to A494 of free FITC – is 0.3, and the extinction coefficient ($\in_{gankyrin}$) of gankyrin is 33,920 $cm^{-1}M^{-1}$ (Trx-6His-G0-6) or 20,190 $cm^{-1}M^{-1}$ (S-G0-6) as estimated by the ExPasy ProtParam tool on the basis of amino acid sequence.

The degree of labeling was typically 70-90%, which is indicative of a relatively efficient reaction, and being less than 100%, is consistent with the major product being a single label conjugation at the N-terminus as desired.

## 4.3.4 Solid-phase gankyrin-pRb binding assay

A solid-phase binding assay, based on the method of (Chen 2008) was developed to investigate the binding of pRbSP and gankyrin. Initially, GST-Rb was immobilized on the walls of a 96-well microtitre plate before blocking the wells with BSA, and varying concentrations of FITC-labelled Trx-6His-G0-6 (F-Trx-6H-G0-6) were added and allowed to bind for 1 hour at room temperature before washing the plates four times with TTBS (see Section 2.4.12 for further details). As a negative control, the same procedure was carried out for wells coated with GST only, and the fluorescence of the wells was compared (Figure 4.4A). While this assay did successfully demonstrate binding of gankyrin to GST-RbSP above that to GST alone, there were aspects of the result that were less than ideal. Specifically, the GST-only data series did not fit a linear pattern as would be expected from non-specific binding. Rather, the data exhibited the characteristics of a specific, saturable binding similar to that of GST-Rb. Associated with this was that background subtraction resulted in a degradation, not improvement of the quality of fit to a single-site saturable binding model. Finally, the K_d of the interaction as determined using the raw data and a statistical model that incorporates non-specific binding (K_d = 9.82 μM) was significantly different to that determined by explicitly subtracting the non-specific component and subsequently fitting using a specific binding only statistical model (K_d = 1.89 μM to 5.12 μM, 95% confidence interval).

In an effort to reduce the possibility of binding other than the specific interaction between pRbSP and gankyrin complicating the analysis, the assay was repeated using the thrombin-cleaved species of gankyrin (yielding an N-terminally S-tagged gankyrin, here named S-G0-6) that was subsequently labeled with fluorescein (F-G0-6). In addition, a thioredoxin-tagged RbSP (Trx-RbSP)

was used as an alternative to the GST-fusion protein. Attempts to isolate the pRb small pocket domain alone by thrombin digestion of the fusion protein were unsuccessful, with losses of protein to precipitation, and a number of non-specific products of digestion evident in the supernatant (Figure 4.5). Use of Trx-free G0-6 improved the assay – particularly in terms of the fit of the background-subtracted data series ($R^2 = 0.98$) – but still exhibited undesirable characteristics of saturable binding to GST alone (Figure 4.4B). Similarly, the K_d of the interaction determined using the raw data and a fit incorporating non-specific binding ($K_d = 3.51$ µM) was still significantly different to that determined using a specific binding fit and the background-subtracted data series ($K_d = 1.05$ µM to 3.31 µM; 95% confidence interval) though an improvement was also evident here.

Figure 4.4C demonstrates that use of the Trx-free gankyrin tracer together with a Trx-RbSP binding partner reduced background binding to the expected linear trend and yielded a background-subtracted data series with a compelling fit to a single-site saturable binding model ($R^2 = 0.99$). Finally, the fit of the raw data yielded a K_d of 6.86 µM, which fell within the 95% confidence interval of that determined using background-subtracted data (6.78 µM to 9.70 µM). This suggests that the two methods are in agreement that the explicitly measured non-specific binding is the only source of non-specific binding evident in the raw Trx-RbSP data set.

As mentioned above, attempts to isolate thrombin-cleaved RbSP from the fusion protein were unsuccessful. Cleavage under ordinary conditions (TBS pH 8.5, 1U thrombin/mg substrate, 4 hour cleavage at room temperature; Figure 4.5 Lane 4) resulted in significant loss of protein to precipitation and non-specific protein cleavage (marked as products X,Y and Z in the figure). In an attempt to reduce non-specific cleavage, less thrombin was used. This reduced, but did not eliminate the production of the products of non-specific digestion (particularly "Product X", Lanes 5 and 6). Also of note was the fact that despite greater consumption of GST-RbSP substrate with
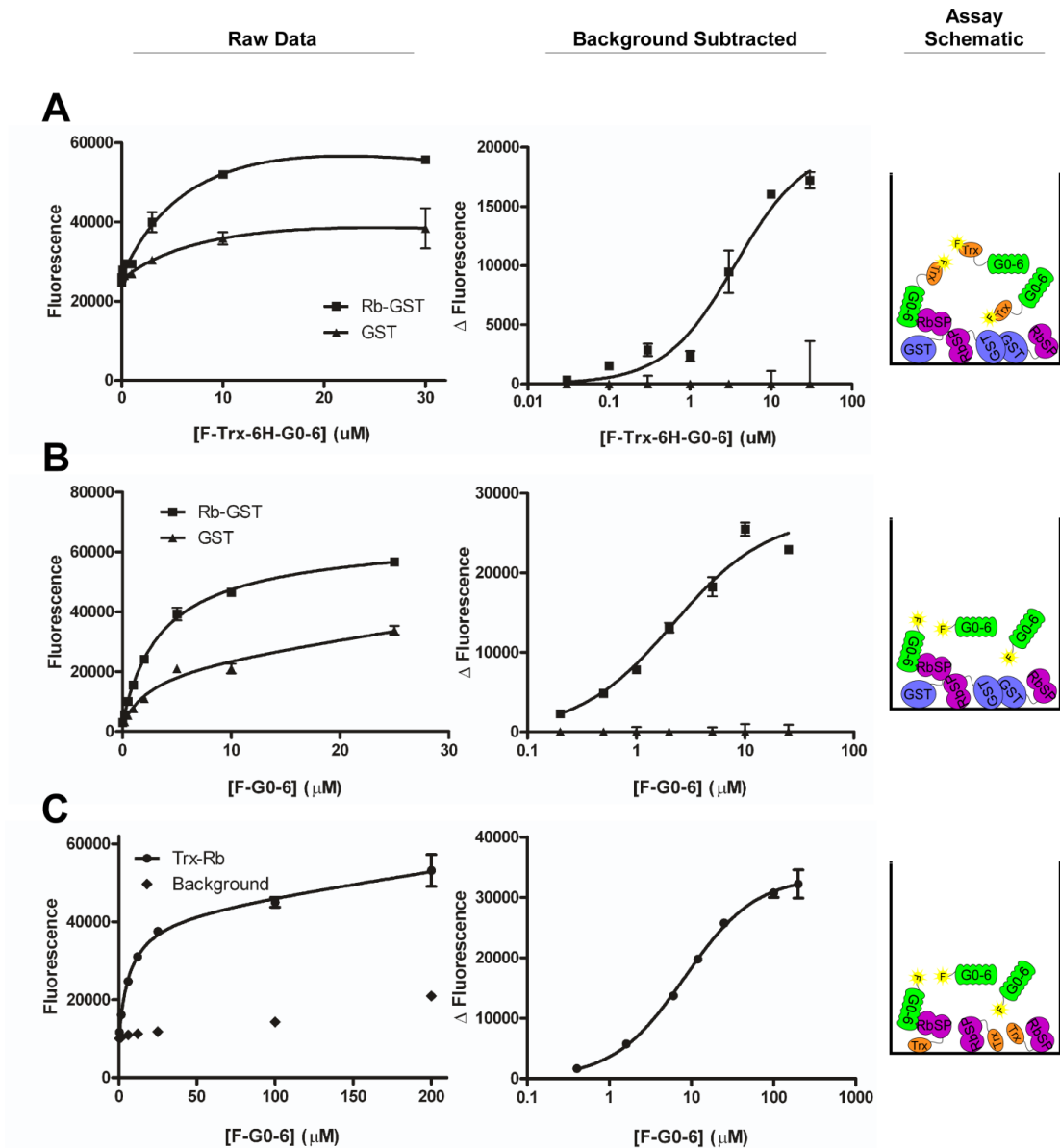
Figure 4.4 Comparison of various versions of a solid-phase interaction assay for pRbSP-gankyrin binding. The left-most column shows the raw fluorescence detection data on a linear scale. Trend lines are shown where a one-site specific, saturable binding plus linear non-specific binding model produces an accurate fit to the data ($R^2 \geq 0.90$). The middle column shows a log-scale plot of the background-corrected data, obtained by subtracting corresponding data points of the GST-only/background series from the GST-Rb/Trx-Rb series. Trend lines represent the fit to a one-site specific, saturable binding model. The right-most column shows a schematic of the version of the assay used. (A) Data obtained using immobilized GST-Rb, and F-Trx-6H-G0-6. (B) Data obtained using immobilized GST-Rb and F-G0-6. (C) Data obtained using Trx-Rb and F-G0-6. Error bars on all plots represent standard deviation of 3 or more replicates.

higher levels of thrombin, the yield of RbSP peaks with the use of 0.5 or 0.2 U thrombin per mg substrate. On this basis it seems that a yield higher than approximately 50% would be difficult or impossible to achieve. Non-specific digestion is likely to be occurring in the unstructured linker region of the RbSP (residues 577-645). Purification of the RbSP for crystallization utilized two additional thrombin sites at each end of the linker region to specifically remove this region (Lee, Russo et al. 1998). Though this was justified on the basis of facilitating protein crystallization, it may also have had advantages in protein preparation as the results presented here allude to.

In order to address the issue of protein precipitation, the addition of 1mM DTT, or adjustment of the reaction pH to 9.0 were also investigated, protein precipitation may have been due to oxidation or a pH close to the pI of the digestion product (RbSP). If this were the case, a pH change in either direction should reduce precipitation; a pH of 9.0 was chosen as thrombin activity is higher at higher pH. Neither DTT nor a higher pH reduced precipitation, nor non-specific digestion, though they both appeared to marginally reduced thrombin activity (Figure 4.5 Lanes 2 and 3).

On the basis of the above studies, large scale preparation of RbSP was attempted using 0.2U thrombin per mg substrate and subsequent size exclusion chromatography using a Superdex 75 resin. Losses during sample concentration and incomplete separation by this method together with sample dilution during size exclusion chromatography meant that detectable quantities of purified RbSP were not isolated. Larger quantities of input protein and further optimization of the purification strategy may well have yielded useful quantities of RbSP. However, this avenue was not pursued since by this stage the available data (above) suggested that the alternative of a solid-phase assay utilizing Trx-RbSP fusion protein was a viable option, and significantly more accessible.
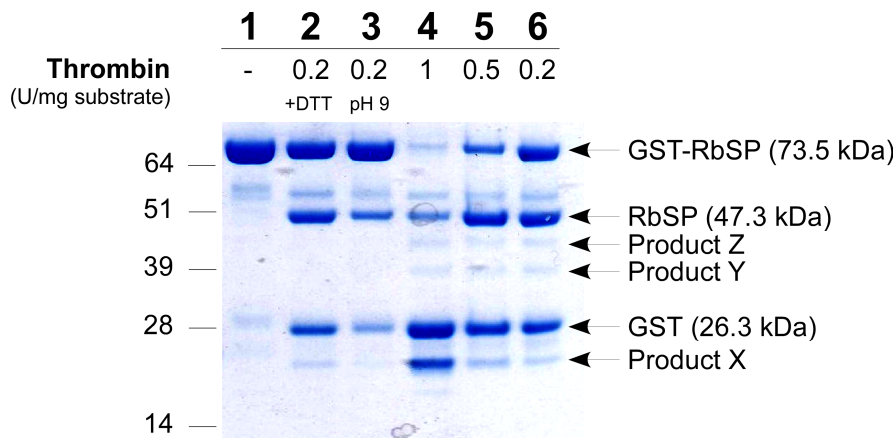
Figure 4.5 Thrombin digestion trials of GST-RbSP. Results of thrombin digestion under various conditions were cleared of precipitate by centrifugation and the supernatant analysed by SDS-PAGE on a 4-12% Bis-Tris gel in MOPS running buffer. The migration of the molecular weight markers (Invitrogen See-Blue Plus 2) are shown to the left. Bands of interest are marked with their identity (where known) to the right, together with their expected molecular weight in parentheses.

While solid-phase interaction data using Trx-RbSP and F-G0-6 exhibited a more conventional pattern of non-specific binding than those using GST-RbSP, optimization to reduce the level of non-specific binding was still desirable. Incorporation of Tween-20 detergent, additional salt or 1mM DTT during the binding step were investigated for their effect on the binding of F-G0-6 on both Trx and Trx-RbSP (Figure 4.6). Of these, 0.1% Tween-20 was the only additive to reduce F-G0-6 binding to Trx, and as would be expected in the case of reduced non-specific binding, reduced binding to Trx-RbSP by a similar quantity. DTT, on the other hand, reduced binding to Trx-RbSP but not Trx. Neither the interaction model proposed here, nor a canonical LxCxE-pRbSP interaction would be expected to be inhibited in any way by a reducing agent such as DTT. In fact, both of them rely on cysteine residues in both pRb and gankyrin that are close to or at the proposed interfaces appearing in their reduced form. However, it is possible that intermolecular disulfide bonds between cysteine residues in gankyrin and pRbSP are able to form at interfaces other than the biologically relevant interface, thus increasing the detected level of binding to Trx-pRbSP but not Trx alone in a manner unrelated to the "true" biological interaction. On this basis, it was decided to include 1mM DTT in the binding step of the optimized solid phase assay.

84
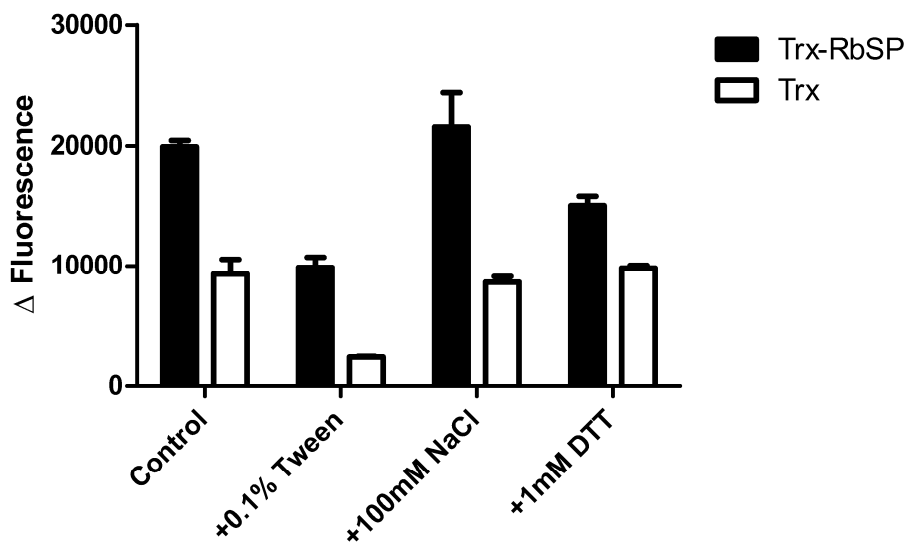
Figure 4.6 The effect of various additives on binding of 10µM F-G0-6 to immobilized Trx or Trx-pRbSP was investigated using the solid phase binding assay. Control assay represents binding in TBS buffer, followed by washes with TTBS. All additives were added to binding buffer only. Plot represents mean and standard deviation of the experiment in triplicate. The fluorescence of untreated, BSA-blocked wells was also recorded and subtracted as background from the data presented.

Having optimized the solid phase binding assay to eliminate as much non-specific binding as possible, the final test of the assay was to demonstrate that unlabeled gankyrin (S-G0-6) is able to compete fluorescently labeled gankyrin (F-G0-6) off of Trx-pRbSP in the context off the assay (homologous competition). This is important for two reasons. First, it further demonstrates that the binding interaction being detected by the assay is directly due to the interaction between pRbSP and gankyrin. Secondly, the competition observed in this experiment will provide an important baseline for comparison to other experiments, since competition assays were to be used for comparing the binding affinities of mutants of pRbSP and gankyrin, in order to avoid the confounding effects of potential differences in labeling of different gankyrin mutants. Statistical analysis of the homologous competition experiments (representative example in Figure 4.7, filled circles) demonstrated a close fit to a single site competitive binding model (R^2 of three repeat experiments between 0.985 and 0.995) with best fit IC_{50} values between 29µM and 67µM.

Given that gankyrin E182 mutants have previously been demonstrated to abrogate binding to pRb (Higashitsuji, Itoh et al. 2000), the ability of the competition assay to recapitulate this result was also tested using the E182K gankyrin mutant. Surprisingly, however, these experiments demonstrated that S-G0-6 E182K competes just as effectively with F-G0-6 as the wild-type S-G0-6 (Figure 4.7). Statistical analysis of the competition curves revealed that the IC_{50} values for wt and E182K S-G0-6 were not significantly different (P value 0.57). The experiment was repeated with three independent protein preparations, and expression clones were re-sequenced directly to ensure the identity of the mutant protein. All repeats exhibited the same effect - Figure 4.7 shows a representative example. Given the difficulties already experienced with non-specific binding, however, it was unclear whether this effect was an artifact of the assay, or a genuine finding that unexpectedly, the E182A mutation, but not the E182K mutation disrupts the G0-6/pRb interaction. Beyond the theoretical expectations of the effect of the E182K charge reversal, the competition curve itself provided some suggestions that the former was the more likely scenario. While the overall fit to a single-site competitive binding model is shared, the individual points at 750nM and 3μM competitor concentration exhibit a statistically significant difference between the competitive binding observed, with wt S-G0-6 proving the more effective competitor. It is possible that most of the binding detected in this assay is non-specific, and that the non-specific signal obscures the true signal of specific binding.

To address this possibility, a data fit using a two-site competition model was conducted on the data shown in Figure 4.7 (fit not shown) in which the lower affinity interaction IC_{50} and was constrained to be equal between the two data sets, as was the proportion of sites available for high affinity binding. This model improved the fit of both wild-type and E182K data points, though this could be expected for any model that introduces additional variable parameters as this one did. However it yielded overlapping 95% confidence intervals for the IC_{50} values of wild-type gankyrin than the E182K mutant, while formal statistical tests suggested the difference between them was not statistically significant (p = 0.35). It was therefore concluded that even if the phenomenon observed was a mixture of specific and non-specific competition, the assay would not yield data that could reliably identify variations in specific binding.
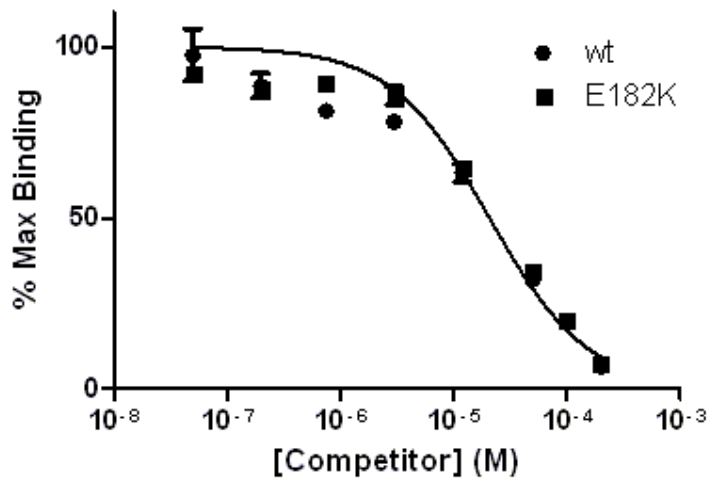
Figure 4.7 The gankyrin/pRbSP competition binding assay exhibits indistinguishable levels of competition by wt and

E182K mutant S-G0-6. 10μM F-G0-6 and varying concentrations of competitor S-G0-6 (wt or E182K) were incubated with

immobilized Trx-pRbSP, washed, and fluorescence was read. Data points represent the mean of triplicate samples; error

bars represent sample standard deviation.

Recalling the observed effect of 1mM DTT on the detected level of binding (Figure 4.6) in the

context of the competitive binding results, the effect of additional DTT in the binding assay

warranted additional investigation. Therefore, the effect of various concentrations of DTT on

binding of F-G0-6 to Trx-pRbSP was investigated (Figure 4.8). Binding was compared to that of Trx

alone, in the absence of DTT, since it had already been shown that DTT had no effect on binding to

Trx. This analysis revealed that concentrations of DTT greater than or equal to 5mM in the binding

step of the assay reduced binding of F-G0-6 to Trx-RbSP to the level of that to Trx alone, while also

increasing the variability between repeat samples. While the data suggested that 5mM DTT had

reduced binding to undetectable levels, the use of 5mM DTT in competition assays was attempted,

in case a small magnitude binding event might be detectable via this method (data not shown).

There was also no evidence of any trend within the data, partly due to the increased variability

observed in the prior experiment.
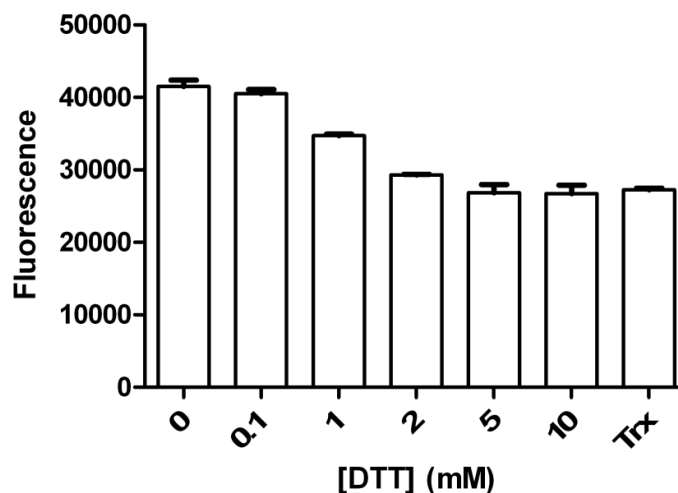
Figure 4.8 DTT reduces binding to Trx-pRbSP as detected in the solid phase binding assay in a dose-dependent fashion.

10μM F-G0-6 together with varying concentrations of DTT was incubated with immobilized Trx-pRbSP, washed, and

fluorescence was detected. A control sample containing no DTT and immobilized Trx in place of Trx-RbSP was also

included. Data represent mean and standard deviation of triplicate samples.

4.3.5 Interrogation of gankyrin-pRb interaction by GST-pulldown

In order to clarify some of the results of the previous section of work, a GST-pulldown

approach was employed to investigate the gankyrin/pRbSP interaction, since this approach has

previously been used by others to demonstrate the intermolecular interaction (Li and Tsai 2002;

Nakamura, Nakano et al. 2007). The results presented so far do not convincingly demonstrate a

specific interaction between pRb and gankyrin. Reproduction of published GST-pulldown results

would eliminate the possibility that the proteins used in this study are for some reason incapable

of interacting (whether due to undetected mutation or the presence/identity of any N-terminal

additions to the protein originating from the expression vector), as well as a method to probe the

binding of the various mutants of pRb and gankyrin. Earlier efforts had focused on the

development of an assay system that would be useful for testing competition between gankyrin

interaction partners or novel peptide or small molecule inhibitors of the interaction. However, at

this point, a positive control for the gankyrin/pRb interaction was a higher priority, and GST pull-

down experiments provided the means to test the most immediate question – that of the proposed binding model.

The same purified proteins that had been used for the solid-phase assays were employed for GST pull-down experiments based on a previously published method (Nakamura, Nakano et al. 2007). Briefly, 50pmol GST-RbSP fusion proteins were incubated with 5nmol S-G0-6 (or mutants) in a total reaction volume of 50μL at 4°C for 90 minutes before addition of 20μL glutathione-agarose beads, washing (five times) and elution with 25μL 10mM reduced glutathione in TBS. In my hands, inclusion of 5% BSA in the binding incubation was necessary to reduce non-specific binding to a level that allowed detection of specific binding. This was consistent with data from the solid-phase binding assay, but was not included in the method described by Nakamura *et al* (2007). Eluate was analysed by SDS-PAGE and Western blot, utilizing Cy3-labelled and Cy5-labelled secondary antibodies to simultaneously detect GST/GST-RbSP and S-G0-6, respectively. The benefit of this approach is that fluorescent detection has a more linear response than other detection methods, which together with the ability to simultaneously probe for and detect different signals with different fluorophores, means that differences in the level of signal can be interpreted more reliably.
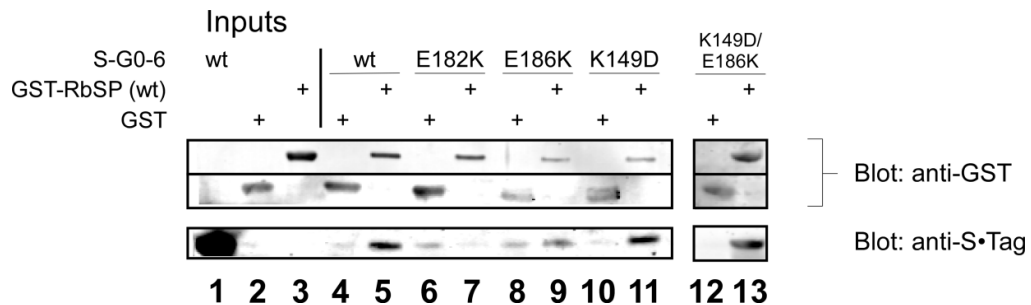
Figure 4.9 GST pull-down experiments testing the binding of various mutants of S-G0-6 to GST-RbSP. Note that lanes 12 & 13 are from a separate gel and transfer than lanes 1-11, so levels of signal should not be directly compared between these two sets.

The results of GST pull-down experiments using wild-type GST-RbSP and S-G0-6 confirmed the interaction (Figure 4.9 lanes 4 & 5), and crucially, demonstrated that G0-6 (E182K) is unable to bind RbSP (lanes 6 & 7). Thus the results of this assay can be interpreted with much greater confidence than the previously attempted solid-phase assay, as it correlates closely with the results presented in the literature. Extending the assay to test the predicted interaction via other mutants of gankyrin demonstrated that mutants E186K and K149D both retain the ability to bind RbSP, though there were indications that the E186K mutant exhibited reduced binding (Figure 4.9 lanes 8-11). This effect was not evident in experimental repeats, however, and is likely an artifact of the particular experiment shown. The fact that the K149D/E186K double mutant does not exhibit any defect in binding (lanes 12 & 13) shows further that the E186K mutation does not reduce binding to RbSP, and that there is no additive effect of possible small reductions in binding of the individual E186K and K149D mutants which may not be evident by this method of detection.

Similar experiments investigating the effect of mutations on the RbSP side of the interaction also suggested that neither the K720E, K765E, D718K nor D718K/K765E mutants exhibited reduced pull-down of wild-type S-G0-6 (Figure 4.10 lanes 5-10), contrary to the predictions of the binding model being investigated.
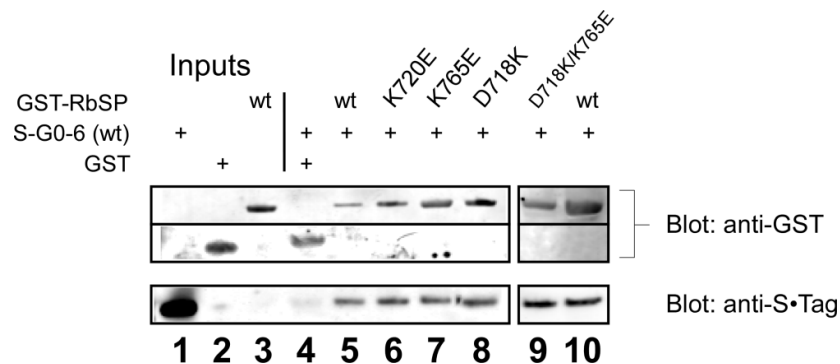
Figure 4.10 GST pull-down experiments testing the binding of various mutants of GST-RbSP to S-G0-6. Note that lanes 9 & 10 are from a separate gel and transfer than lanes 1-8, so levels of signal should not be directly compared between these two sets

While these results were not supportive of the proposed binding model, as they stand, they are also insufficient to disprove it. In particular, the fact that the assays are carried out in the presence of vast excesses of S-G0-6 means that a reduced K_d may be difficult to detect. By way of illustration, if a relatively weak wild-type binding affinity is assumed ($K_d = 10\mu M$) and the protein concentrations described above are used, then a 5-fold reduction in binding affinity would produce less than a 30% reduction in S-G0-6 pull-down. With a slightly higher wild-type affinity ($K_d = 1\mu M$) that figure falls to less than 5%.

Use of lower concentrations of S-G0-6 would help to alleviate this problem. However, the signal of the above assays was already extremely low, making quantitative analysis in particular very difficult and unreliable. Attempts to improve the signal resulted in the detection of non-specific binding to GST (already visible to some extent in Figure 4.9). The use of different concentrations of BSA and Tween-20 during binding or washing, salt and DTT concentrations and pH of the wash buffers, the quantity of glutathione agarose used, and the use of spin columns for washing were pursued in an effort to reduce non-specific binding without success, so ultimately other methods of assessing the mutants were pursued.

4.3.6 Semi-quantitative analysis of binding using a bacterial two-hybrid system

Having established that large changes in binding are not observed for any of the mutants of RbSP, nor for any of the mutants of S-G0-6 except E182K, but without access to an assay capable of detecting more subtle changes in binding affinity, one final assay method was pursued. A two-hybrid system – in this case based on a bacterial host – allows for rapid and simple detection of relative binding affinities between mutants of the same interaction partners. Like the GST pull-down assay, it was not the first choice for the investigation of the gankyrin-RbSP interaction because it cannot determine absolute K_d values, nor is it suitable for direct competition assays with other interaction partners of gankyrin, nor potential synthetic small molecule or peptide

inhibitors. It does, however, give access to an assay system that has previously allowed detection

of relatively minor reductions in binding due to mutation (Hao, Whitelaw et al. 2011). Given that

based on the data obtained thus far it appears unlikely that the proposed binding model is an

accurate model of the interaction, the other advantage of this bacterial two-hybrid system is that

it can be readily harnessed as a reverse two-hybrid system and used to conduct a screen for

mutants of RbSP or gankyrin that exhibit reduced binding (Hao, Whitelaw et al. 2011).

The basis of this system is the creation of fusion proteins of the λ phage repressor protein CI

and the N-terminal region of the *E. coli* RNA polymerase α subunit (rpoA) with each of the proteins

under investigation (usually denoted bait and prey proteins, respectively). A reporter construct

that places the LacZ gene under the control of an intrinsically weak promoter with a copy of the

OR2 operator (λ-CI binding site) a small distance upstream of the -35 RNA polymerase binding site.

Thus the λ-CI-bait protein is recruited to the promoter, from where it can aid recruitment of RNA

polymerase complexes containing the rpoA-prey fusion in instances where there is binding

between the bait and prey proteins. Thus in response to an interaction between bait and prey, the

expression of the LacZ gene is increased, and is detectable by β-galactosidase assay.

As with all two-hybrid systems, the arrangement of the interaction partners as bait or prey can

influence the detection of an interaction, and of the level of background activity of the reporter

assay. Therefore, the wild-type gankyrin and RbSP ORFs were each sub-cloned from pET32a into

both pBT and pTRG via the BamHI and XhoI sites (for details see Section 2.2.3) and verified by

BigDye Terminator sequencing using primers pBT-5'-SP, pBT-3'-SP, pTRG-5'-SP and pTRG-3'-SP. The

resulting plasmid constructs were then transformed into the KS1 *E. coli* strain which contains a

chromosomal copy of the above-described reporter gene and assayed for β-galactosidase activity

according to the method described in Section 2.3.13, along with positive and negative control

strains (Figure 4.11).
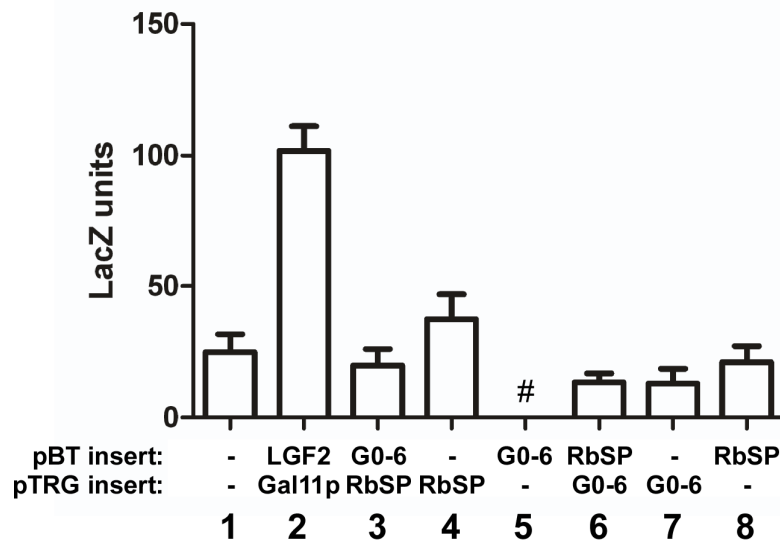
Figure 4.11 Use of a bacterial two-hybrid system to investigate the gankyrin-pRb interaction. Shown are the mean and standard deviation results (n=6) for β-galactosidase assay in the presence of 20μM IPTG of the KS1 reporter strain harbouring pBT and pTRG or their derivatives, as indicated. LGF2 and Gal11p (Bar 2) represent the positive control supplied as part of the BacterioMatch II system. The # symbol (Bar 5) indicates that this strain did not achieve adequate growth to be assayed.

Analysis of these results revealed that the gankyrin-pRbSP interaction was not detectable by this method. The strains harbouring G0-6- and RbSP-enconding plasmids, in either orientation (Figure 4.11 Bars 3 and 6) exhibited reporter activity approximately equal to or less than that of the negative control strain harbouring pBT and pTRG plasmids encoding no bait or prey (Bar 1). Even more tellingly, these strains also exhibited lower activity than the negative control strains in which the G0-6 coding region was absent (Bars 4 and 8). To ensure the veracity of the above results, the identity of the plasmids after transformation into reporter strains was verified by colony PCR.

4.4 Discussion

The work presented above strongly suggests that the structural model presented in Chapter 3 for the binding of gankyrin and pRb is deficient. However this conclusion is tempered by the fact that no suitable method for the quantitative analysis of the binding of these two proteins was

found. In fact, even the GST pull-down experiments presented here did not produce the level of signal (versus non-specific binding) that seems to be evident in similar experiments in the literature. This may have been due to the detection method used – the experiments here are the only ones to use fluorescent detection, compared to the chemiluminescence (ECL) methods seen in the literature. Fluorescent detection methods are known to have lower sensitivity than ECL methods, but were chosen here because of their more linear response, making it a better choice for the detection of changes in the signal level.

An alternative, or additional, explanation lies in the protein constructs used for the experiments. Previous experiments (Higashitsuji, Itoh et al. 2000; Dawson, Apcher et al. 2002) used the reverse "orientation" of the experiment compared to the experiments above, using a GST-gankyrin fusion to pull down full-length pRb from cell lysates. This would have avoided the problems with non-specific binding of gankyrin to the glutathione resin observed here, and the use of lysates potentially means that a higher affinity binding was observed due the formation of multi-factor complexes in the pull-down. The most directly comparable experimental setup to that described above was the one employed by Nakamura *et al* (2007), who similarly observed higher levels of non-specific binding, though still less than that seen here, possibly due to the detection method (see above).

Ideally, a quantitative, label-free method of determining the binding affinity – such as Surface Plasmon Resonance or thermophoresis – of the purified proteins would be pursued to verify the findings presented here. Even disregarding the question of whether the predicted binding model presented here is correct, measurement of the affinity of the gankyrin/pRb interaction – not to mention the other interactions of gankyrin – would be useful in its own right, and would be a critical component of any attempt to fully understand the role that gankyrin plays in oncogenesis. Gankyrin seems to lie at the centre of a complex interaction network; it is already known that competition between gankyrin and p16INK4A for binding to CDK4 is central to its activity, and it is

likely that in the process of elucidating the interaction network around gankyrin, other competitive or co-operative interactions will be discovered. If we are to understand the various input signals and read-outs from this network, a quantitative understanding of the interactions involved will be essential.

Of course without verification of a specific binding model, the question of whether gankyrin interacts with pRb in a folded or unfolded form remains live. Probing this question directly remains difficult. One option is to use a screening approach (as alluded to in the final parts of this chapter) that does not assume one model or the other, but simply looks for mutants that disrupt the intermolecular interaction. The nature and location of a set of such mutations may be more consistent with one model more than the other.

Alternatively, rational design of mutants that change the stability of gankyrin, rather than its binding epitopes, may also be informative. Even this approach, however, brings its own difficulties. A presumably unstructured peptide from gankyrin has been shown to interact with pRb, but this does not necessarily indicate that this is the biologically relevant interaction, or that gankyrin cannot also interact in its structured form. If these two modes of interaction do exist (but remain unverified), any analysis of the effect of stabilizing or destabilizing mutations on the interaction will be complex at best, and may even obscure any identifiable results.

Chapter 5

Small molecule inhibitors of gankyrin

5.1 Introduction

5.1.1 Protein-protein interactions as drug targets

The inhibition of the active sites of enzymes, or the binding clefts of signaling receptors has long been the mainstay of rational drug discovery attempts, as well as forming the molecular basis for many drugs that have been discovered by less targeted means. More recently, however, attention has turned to the application of drug discovery techniques to the design of inhibitors of protein-protein interactions, with some success (Pagliaro, Felding et al. 2004; Zhao and Chmielewski 2005) including recent examples of progress into clinical trials. The expansion of viable drug targets to protein-protein interactions greatly increases the number of possible targets, and brings with it other potential advantages. Catalytic sites are structurally very highly conserved, given that they often need to bind the same small molecule substrate. This in turn can lead to significant problems in the development of selective inhibitors of enzymes, particularly kinases that are both very common and very biologically important. Protein interaction, however, is often designed to be highly specific to facilitate the selective interaction of, for instance, two proteins within a signalling cascade, thus preventing signal from 'leaking' into another cascade via a similar signalling component. Thus, there is potentially a greater structural diversity among protein-protein interactions that can be capitalised upon in the design of specific inhibitors.

Of course, the transition from targeting active sites to sites of protein-protein interaction has not been without its challenges. By their very nature, enzymes and signaling receptors have evolved to tightly bind substrates that are often small molecules or small peptides, leading to deep clefts that are strongly amenable to the design or discovery of novel, exogenous small-molecule ligands. In contrast, protein-protein interactions typically occur across large, relatively flat protein surfaces, typically thousands of square angstroms in size. Some of the presumed difficulty posed by this has been overcome as it has become clear that targeting the entire interaction surface with a small molecule is unnecessary (Fletcher and Hamilton 2007). In fact, it has been shown that

there is no relationship between the size of an interaction interface and the binding energy of the interaction – the key property determining the effectiveness of a competitive inhibitor. Further, interaction "hotspots" can typically be identified that impart a large portion of the binding energy (Bogan and Thorn 1998), making it possible to target these critical sites with molecules of drug-like size. While identifying such hotspots is considered difficult without mutagenic or structural information, it has been shown that there are certain characteristics that are common to many hotspots, such as a prevalence of tryptophan, tyrosine and arginine residues, and the exclusion of bulk solvent by hydrophobic residues (Bogan and Thorn 1998).

Some more recent work has focused on determining whether small molecule inhibitors occupy a different region of chemical space than active site inhibitors (Sperandio, Reynes et al. 2010) finding that (within the confines of typical ADME/Tox filters) protein-protein interaction inhibitors tend to be larger and more hydrophobic, which may aid discovery of novel inhibitors through the construction of more appropriate compound libraries, whether physical or virtual. Other efforts have focused on identifying those target sites that are more similar to active sites, thus making them more "druggable." It is commonly observed that many successful inhibitors of protein-protein interactions bind in relatively deep clefts or pockets which are more similar to active sites, and typically have random-coil peptides as their native ligands (Arkin and Wells 2004; Fry and Vassilev 2005; Whitty and Kumaravel 2006). Given that the field is relatively young, however, it should be kept in mind that both these analyses of target suitability and the chemical nature of successful inhibitors may be biased, favouring targets and ligands that bear greater similarity to active site targets simply because current efforts employ very similar approaches to those that have proven successful in the past. With greater knowledge and a wider variety of examples of protein-protein interaction inhibitors, targets with less similarity to active sites may become more tractable, and the regions of chemical space employed may differ more.

5.1.2 In silico drug discovery

Another area of recent progress in drug discovery is the use of *in silico* techniques. The aim of this endeavour is to deploy our knowledge of the various elements of drug discovery via the considerable computational resources that are now often accessible in an attempt to reduce the time and cost of drug discovery. The use of *in silico* techniques in drug discovery can encompass a variety of tasks from structural modeling of the target protein, to cheminformatics, analysis of potential target site characteristics, or prediction of ADME/toxicity properties of lead compounds or, more importantly, derivatives thereof prior to synthesis. Principally, however, *in silico* techniques are applied to the docking of virtual representations of small molecule compounds to a structure of the biomolecular target, and derivation of a score of this pose in order to either rank a library of compounds in sequence of affinity for binding to the target (in the case of virtual screening) or as a tool for modeling the interaction between a validated ligand and its target in order to develop QSAR models, plan derivatisations, linkages of multiple small ligands or other chemical changes in an effort to further progress a lead molecule.

The problem of virtual screening can be broken down into two parts: (1) search of the conformational space accessible by the ligand and target; and (2) estimation/evaluation of binding efficacy (usually known as scoring). In the process of the conformational search, a docking algorithm will typically generate hundreds of conformations, or poses, of a given pair of ligand and target. Each of these is then scored with a function that aims to identify which of these poses is most favourable. Assuming that the molecule is a genuine ligand and that the true binding pose has been sampled during the conformational search, a perfectly performing scoring function would identify this pose as having a more favourable score than the other poses sampled. In addition, scoring functions aim to rank the scores of the native poses of multiple ligands in order of binding affinity, either by training them against a set of known ligands with known binding affinities, or by a direct estimation of the binding energy (though a training set is often used in these cases in order to refine or discover various parameters of the energy equation). In reality, of

course, neither of these processes is perfect. The true binding pose is often not sampled, due to the sheer size of the search space and the small amount of time available in the context of a virtual screen of tens or hundreds of thousands of molecules. Simplifications, such as no or limited treatment of receptor flexibility are commonly used to reduce the complexity of the problem, but in themselves can prevent sampling of the true conformation. Even where the true pose or a similar pose is sampled, other poses can out-rank the near-native poses due to approximations, simplifications or other imperfections in the scoring function, though studies disagree about whether this (Warren, Andrews et al. 2006) or the conformational search (Kontoyianni, Sokol et al. 2005) represents the greater problem.

It has been observed that "energy funnels" typically exist in the conformational space surrounding the true native conformation (Rejto and Verkhivker 1996; Verkhivker, Rejto et al. 1999). Thus one of the best methods of discriminating between near-native poses and near-native false positives is via the presence of clusters of similar conformations with similarly favourable scores. In the context of virtual screening, however, this method is undesirably computationally intensive, requiring generation of (at least) hundreds of optimized poses, scoring, clustering and ranking of clusters. An alternative approach that is more conducive to deployment in the problem of virtual screening is consensus scoring. Consensus scoring is a process by which several *in silico* measures are combined by "voting" to identify those compounds that are scored favourably by all of the input measures (scoring algorithms). The success of this approach is predicated on the assumption that false positives identified by different scoring functions are going to be relatively independent, whereas there should be considerably more overlap in identifying true positives since all functions are optimised for this purpose. It is likely that this assumption is a good one, since we would expect that false positives are predominantly the product of imperfect assumptions, approximations or training of the scoring functions which will all vary between the various functions. Indeed studies probing the effectiveness of consensus scoring have found it to improve enrichment, and to a lesser extent, the prediction of binding poses and energetics (Feher 2006).

Scoring functions

Many different scoring functions are available for use in structure-based ligand design. This work makes use of the SCORER algorithm (Branson 2005) which implements several existing scoring functions in a simple re-scoring tool, including DOCK, PMF, PK, SMoG, ChemScore and AutoDock, the details of which are discussed below. It is important to note that some of these scoring functions have several different implementations, and that only the one used in SCORER is described below.

DOCK

The DOCK scoring function (shown below) is based on the non-bonded interaction terms of the AMBER force field. The score aims to evaluate interaction energy by calculation of the van der Waals and electrostatic components of interaction, using Lennard-Jones parameters to describe the van der Waals component.

$$E = \sum_{i=1}^{lig}\sum_{j=1}^{rec}\left(\frac{A_{ij}}{r_{ij}^{a}} - \frac{B_{ij}}{r_{ij}^{b}} + 332\frac{q_1 q_2}{Dr_{ij}}\right)$$

(Meng, Shoichet et al. 1992)

AutoDock

The AutoDock function (below) is another example of an energy evaluation function. It also includes van der Waals and electrostatic terms, as well as an angle-dependent (*E(t)*) hydrogen bonding term, penalties for restriction of rotatable bonds ($\Delta G_{tor}N_{tor}$), and a desolvation term based on the proportion of the volume of the area surrounding the ligand that is occupied by protein atoms.

$$\Delta G = \quad \Delta G_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \Delta G_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right)$$

$$+\Delta G_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}} + \Delta G_{tor}N_{tor}$$

$$+\Delta G_{sol}\sum_{i,j}\left(S_i V_j + S_j V_i\right)e^{\left(-r_{ij}^2/2\sigma^2\right)}$$

(Morris, Goodsell et al. 1998)

ChemScore

ChemScore also utilises a binding energy evaluation function determined by the summation of several energy terms representing different types of interactions. Of note is that while it includes similar terms to other functions, it does not account for electrostatic interactions in any way. The function *f(r)* essentially measures whether the interaction between two atoms occurs within the accepted range of distances for that type of interaction. The ΔG coefficients were then calculated by multiple linear regression to best fit scores to experimental binding affinities in the training set.

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} \sum_{iI} g_1(\Delta r) g_2(\Delta \alpha)$$
$$+ \Delta G_{metal} \sum_{aM} f(r_{aM}) + \Delta G_{lipo} \sum_{IL} f(r_{IL})$$
$$+ \Delta G_{rot} H_{rot}$$

(Eldridge, Murray et al. 1997)

PMF

PMF takes a different approach to the same aim of estimating interaction free energies. Rather than using a theoretical approach, PMF is based on an empirical approach in which a simplified function is derived to directly predict and fit existing structural data. This approach in principle results in a more generally applicable function since no theoretical basis for an interaction is required. The function is principally composed of a volume correction function and a quotient comparing the frequency that an interaction between receptor atom *i* and ligand atom *j* occur in the training set with the frequency of no interaction. The volume correction function accounts for ligand volume and thus implicitly accounts for solvent effects. The energies of interaction for each atom pair are then summed to determine the PMF score.

$$A_{ij}(r) = -k_B T \ln\left[f_{vol\_corr}^{j}(r) \frac{p_{seg}^{ij}(r)}{p_{bulk}^{ij}} \right]$$

(Muegge and Martin 1999)

SCORE (PK)

The SCORE algorithm (subsequently referred to as PK, since it provides an estimate of the pK_d of an interaction) is another empirical scoring function. It takes a similar form to many of the other functions described above, simply calculating the summation of terms covering the energies of van der Waals interactions, metal-ligand interactions, hydrogen bonding, ligand desolvation and the deformation of the ligand from its ideal 3-dimensional shape. The term K_0 is a constant derived purely by regression analysis of the training set of 170 protein-ligand complexes, though the authors speculate that it predominantly describes the entropic effects of binding.

$$pK_d = K_0 + K_{vdw} + K_{metal} + K_{hbond} + K_{desolvation} + K_{deformation}$$

(Wang, Liu et al. 1998)

SMoG

Another statistical analysis of published structural data forms the basis of the SMoG scoring function, except that in this case, the frequency of occurrence of a particular interaction is compared with a reference state. This reference state is essentially the average frequency of formation of interaction, so that interactions that occur more frequently than average are favoured, while less frequent interactions are penalised.

$$g_{ij} = -kT \log\left[\frac{p_{ij}}{\overline{p}}\right]$$

(DeWitte and Shakhnovich 1996)

5.1.3 Gankyrin as a drug target

The oncogenic effect of gankyrin and the ability to reverse this oncogenicity by mutation

(Higashitsuji, Itoh et al. 2000), together with several studies showing that gankyrin is up-regulated

early in the development of nearly all hepatocellular carcinomas (Fu, Tan et al. 2004; Llovet, Chen

et al. 2006; Umemura, Itoh et al. 2008) make gankyrin an attractive target from a biological

perspective for therapeutic intervention. The fact that it exerts its effect by a series of protein-

protein interactions, rather than any enzymatic function makes it a less orthodox target for

therapeutic development as discussed above, but by no means intractable. As well as its

therapeutic appeal, it represents an academically appealing problem as an arena in which to

further explore and expand our knowledge of this new class of targets and inhibitors.

While this thesis was in preparation, a similar attempt at developing an inhibitor of gankyrin

was reported (Thakur and Hassan 2011), targeting the pRb interaction site rather than that of

CDK4. However, this report lacks any experimental verification of the efficacy of the designed

ligands or of their biological effects, and so contributes little to our understanding of either

gankyrin or the use of *in silico* drug discovery techniques in the search for inhibitors of protein-

protein interactions.

As highlighted above, some of the key characteristics for the successful development of

protein-protein interaction inhibitors are the targeting of so-called hot-spots and a target pocket

or cavity of sufficient depth. This information can be used to aid the selection of a target site for

development of a gankyrin inhibitor to maximize the chances of success. All three of the

interactions between gankyrin and pRb, CDK4 or HDM2 represent plausible targets from a

biological perspective. The interaction with pRb is backed by the best biological data, since

elimination of this interaction via the E182A mutation abolishes the oncogenic effect of gankyrin

(Higashitsuji, Itoh et al. 2000). Equally, however, this interaction is poorly understood from a

structural perspective (the subject of Chapters 3 and 4) making targeting of hotspots and cavities
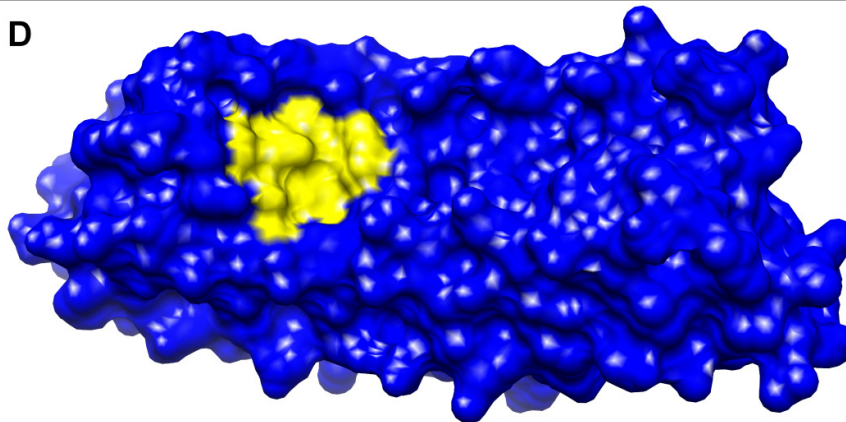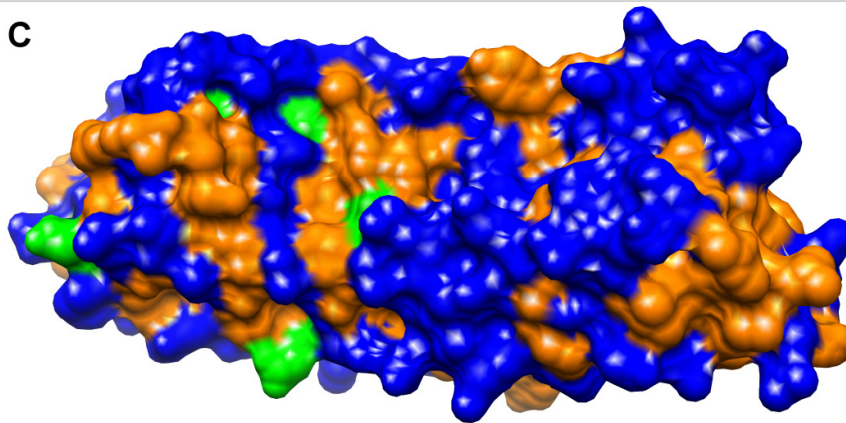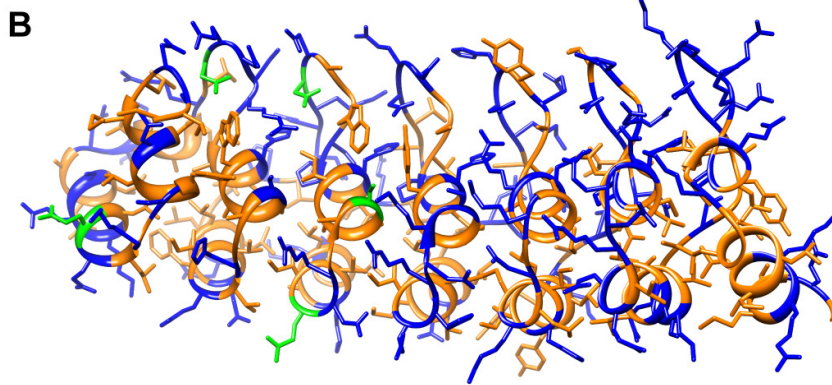
problematic. The same is true of the interaction with HDM2, but to an even greater extent. The interaction site with CDK4, on the other hand, is sufficiently well understood from both structural and functional perspectives to make it a plausible target. Inhibition of this interaction can at the very least be expected to reduce the activity of CDK4 by removing competition between gankyrin and the INK4 family of CDK4 inhibitors. Though a less certain effect of inhibiting the gankyrin/CDK4 interaction, it is also likely to reduce phosphorylation of pRb by any remaining active CDK4 since they will no longer bind to gankyrin as a common scaffold.

5.1.4 Characteristics of target site

Structurally, CDK4 is known to bind the concave surface of ankyrin repeats 0 to 3 of gankyrin, and the important residues for this interaction have been inferred from the sequence, structure and functional homology of other known CDK4-binding ankyrin repeat proteins p16^{INK4A} and p18^{INK4C} (Yuan, Li et al. 2004). Evidence for this argument by homology has been bolstered further by its use to engineer a mutant of gankyrin (I79D/L62H) which gains the ability to inhibit CDK4 like the other INK4 family members (Mahajan, Guo et al. 2007). Thus despite the lack of any direct evidence regarding hot-spots in the gankyrin/CDK4 interaction site, they can similarly be inferred with a high degree of confidence from the relative wealth of data concerning the CDK4/6 interaction with the INK4 inhibitors.

On the other hand, the existence of other proteins with significant homology both in sequence and structure raises the spectre of off-target interactions – most likely with INK4 proteins, or potentially with other ankyrin repeat proteins. Thus selection of a site to target for development of a small molecule inhibitor involves ensuring it encompasses likely interaction hotspots (which are conserved among at least a few related proteins) while also ensuring that sufficient heterogeneity exists in the other residues in the targeted region, and that the surface of the area in question is as concave as possible. Based on homology with p16 and structures of the p16/CDK6 complex(Russo, Tong et al. 1998), it is expected that the side-chains of gankyrin residues K18, E21,

D37, D70, E87 and possibly S82 will make important contacts with CDK4 (Figure 5.1A). Mapping these onto the tertiary structure of gankyrin reveals that of these, residue D70 lies in a concave pocket on the surface of gankyrin with D37 and S82 at its periphery (Figure 5.1B). Analysis of the residues that contribute to this portion of surface area shows that of the 15 residues in that region (Figure 5.1A and C – yellow highlights) there are only 5 identities with the corresponding residues of p16. As would be expected, there are a number of residues with similar properties (a further 7) since they need to interact favourably with the same regions of the CDK4 surface. These similar residues provide sufficient variation in the surface shape and properties that they can likely be exploited to achieve binding specificity, especially given the relative rigidity of small molecule ligands. Achieving selectivity for binding to gankyrin as opposed to other targets is of course not a given, but the level of divergence described here between gankyrin and p16 suggests that the constraints are lower in this case than that of some highly conserved enzyme active sites that have been successfully targeted for drug discovery and development, including kinases such as specific isoforms of PI3K (Fruman and Rommel 2011), or Btk, a single member of the Tec family of kinases (Harrison 2012; Tai, Chang et al. 2012). Regardless, the site targeted for drug discovery in this section of work is clearly the best choice in terms of the structural data confirming its biological relevance and the properties of the region related to the likelihood of success.

A

```
Gankyrin    MEGCVSNLMVCNLAYSGKLEELKESILADKSLATRTDQDSRTALHWACSAGHTEIVEFLLQLGVPVN
p16         --MEPSADWLATAAARGRVEEVRALLEAGANPNAP-NSYGRRPIQVMM-MGSARVAELLLLHGAEPN

Gankyrin    DKDDAGWS-PLHIAASAGRDEIVKALLGKGAQVNAVNQNGCTPLHYAASKNRHEIAVMLLEGGANPD
p16         CADPATLTRPVHDAAREGFLDTLVVLHRAGARLDVRDAWGRLPVDLAEELGHRDVARYLRAAA----
```

B

C

D

Figure 5.1

(A) Sequence alignment of the CDK4 interacting proteins gankyrin and p16INK4A (p16). Residues of p16 that have been shown by crystal structure to make significant contacts with CDK6 (Russo, Tong et al. 1998) are highlighted in green, and are largely conserved between the proteins. Yellow highlighted residues constitute the region of gankyrin targeted for virtual screening and are divergent in sequence from p16. Surface representation of these residues is shown in **(D)**. **(B)** & **(C)** Ribbon and surface representations, respectively, of gankyrin showing the key residues for interaction with CDK4 (green, matching highlights in panel A). Hydrophobic residues are shown in orange to demonstrate the presence and clustering of hydrophobic surface patches around the charged key interaction residues, consistent with the solvent exclusion properties expected of protein-protein interaction "hotspots."

5.2 Aims and Approaches

Inhibition of the protein interactions of gankyrin has been demonstrated to abrogate its oncogenicity via studies of gankyrin mutants (Higashitsuji, Itoh et al. 2000). This evidence, combined with various studies demonstrating that up-regulation of gankyrin is a key and early step in the onset of hepatocellular carcinoma makes the development of small molecules that inhibit the protein-protein interactions of gankyrin an attractive prospect. This section of work aims to discover and develop small molecules that bind to gankyrin and inhibit its interaction with CDK4, given its obvious therapeutic potential. In addition to this, it provides a useful case in which to explore the discovery and development of small molecule protein-protein interaction inhibitors, particularly with regard to targeting interaction sites with non-peptide (surface) native interaction partners and the use of *in silico* techniques to aid discovery of protein-protein interaction inhibitors as these topics are under-represented in the literature.

A virtual screening approach was taken to screen the "lead-like" subset of the ZINC compound library (version 5) using UCSF Dock, followed by consensus scoring. The ZINC library provides *in silico* descriptions of commercially available chemical compounds, thus providing a convenient method of screening a wide variety of compounds without the need for chemical synthesis expertise.

The compounds identified by this method will be assessed for binding to gankyrin via an NMR-based *in vitro* assay. By titrating the candidate compound into a sample of 15N-labelled gankyrin, both the binding affinity of the compound and the residues involved in binding can be determined based on changes to the HSQC spectrum. In practice, only the first four ankyrin repeats of gankyrin was used (G0-3) since this has previously been shown to be structured and bind to CDK4 (Li and Tsai 2002), and it reduces potential problems with spectral crowding. This is especially important given the repeat nature of gankyrin leading to similar backbone shifts for residues at analagous

positions in the various repeats. Further exacerbating the problem is that I have access only to a 600MHz field NMR spectrometer, where the NMR structure of gankyrin was solved using a 900MHz spectrometer (Yuan, Li et al. 2004), meaning that while TROSY-HSQC methods could be used to reduce peak widths, such methods are not useful in a 600MHz field.

Once compounds that bind to gankyrin at the appropriate site have been identified, it will be necessary to verify that they do indeed inhibit the interaction with CDK4, and to investigate the biological effects of this, however this chapter deals only with the discovery of small molecule ligands.

5.3 Results

5.3.1 In silico *small molecule screening*

A library of 1,378,666 commercially available "lead-like" chemical entities were docked against both the NMR (PDB entry 1TR4) and X-ray crystallography (PDB entry 1UOH) derived structures of gankyrin, using UCSF Dock v5.1. These lead-like entities represented the distinct, physiologically relevant protonation states of 1,014,276 chemicals with calculated logP values between –2 and 4, molecular weight between 150 and 350, and containing less than 4 hydrogen bond donors and less than 7 hydrogen bond acceptors. The rationale in using lead-like compounds was to restrict the chemical space that needed to be searched, thus making the search practical, while ensuring that those hits that were identified could be readily optimized to improve binding affinity while remaining in the regions of chemical space from which therapeutic small molecules are typically derived.

UCSF Dock uses a set of spheres for the initial placement of ligands at the site of interest, before incorporating ligand flexibility, energy minimization and scoring of the final pose or orientation of the small molecule against the "receptor" (in this case gankyrin). Spheres generated using the SPHGEN utility provided with UCSF Chimera were clustered and selected for proximity to

residues in ankryin repeats 1 and 2. A small number of spheres were also manually excluded as they did not reside on the concave surface of gankyrin being targeted. The sphere clusters used for each docking run are shown in Figure 5.2.
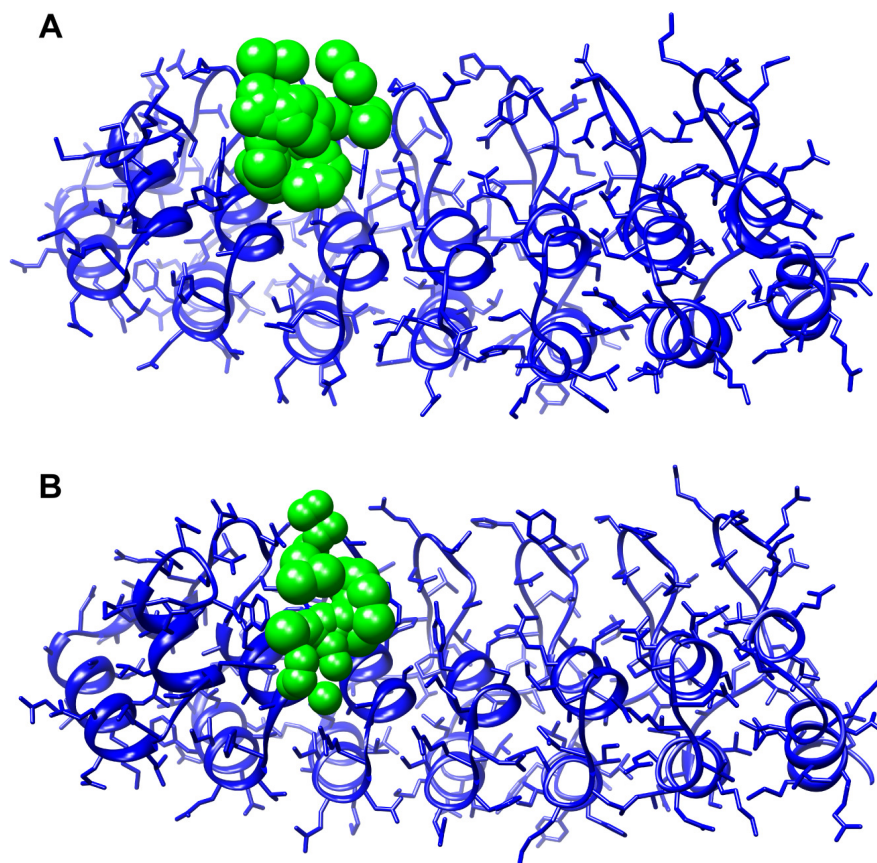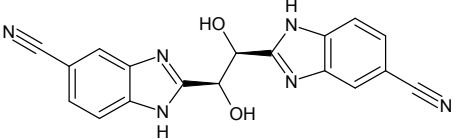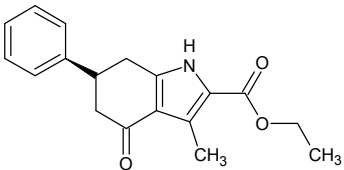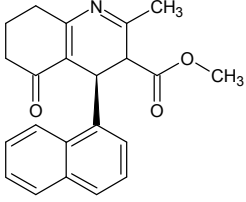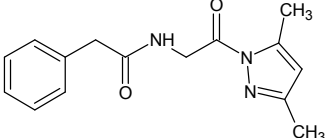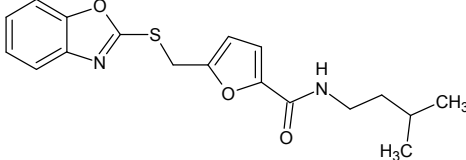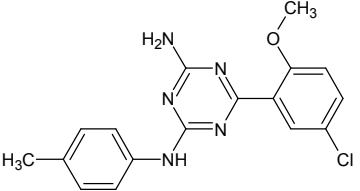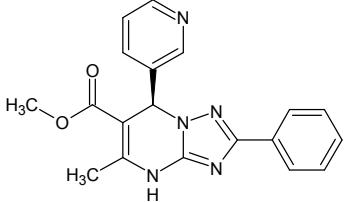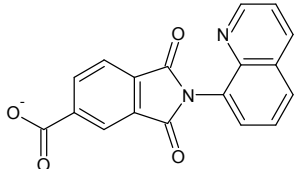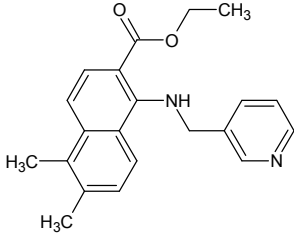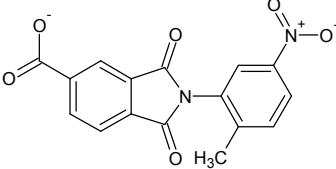

Figure 5.2 NMR structure (PDB entry 1TR4) *(A)* and crystal structure (PDB entry 1UOH) *(B)* of gankyrin used for virtual screening. Protein backbone is shown as a ribbon representation, with sidechains shown as sticks. Green spheres represent the spheres used by the DOCK algorithm for the initial placement of ligands in the target site.

Docking generated over 34,466,650 small molecule poses against each target structure, which were subsequently re-scored using Scorer, a utility program that implements the scoring algorithms of 6 different docking programs. In this way, it was possible to select those poses that appeared in the 90th percentile or higher for each of the 6 scoring algorithms. These top compounds were further ranked by their DOCK scores, and the top 50 poses against each target were inspected visually for surface complementarity and the presence of several favourable receptor interactions (salt bridges, hydrogen bonds and hydrophobic interactions). Ultimately, 5

compounds from each round of screening were selected for *in vitro* testing (Table 5.1). By

reference to their ZINC identifiers, compounds suppliers and catalogue numbers for these

compounds were obtained and the compounds purchased.

Table 5.1 Candidate compounds selected from virtual screen for further *in vitro* testing

Abbreviation	Supplier and ID	Chemical Formula
GC1	Ryan Scientific: SPB07794	
GC2	ChemDiv: C276-0190	
GC3	ChemDiv: 1630-1442	
GC4	Asinex: BAS02973992	
GC5	Asinex: BAS00917497	
GC6	Sigma-Aldrich: R745790	
GC7	ChemBridge: 6920482	

Abbreviation	Supplier and ID	Chemical Formula
GC8	ChemBridge: 5685647	
GC9	ChemBridge: 7790662	
GC10	ChemBridge: 5160664	

5.3.2 NMR investigation of gankyrin and potential ligands

Cloning, expression and purification of G0-3

A DNA fragment encoding the first four ankyrin repeats of gankyrin (G0-6) was generated by PCR from a HepG2 cDNA preparation using oligonucleotides Gank-M1-5′ and Gank-Y138-3′ as per the method described in Section 2.3.3. This DNA fragment was sub-cloned into the pGEX4T1 vector using the *Eco*RI and *Xho*I restriction sites. The sequence and location of the DNA insert was verified by BigDye sequencing using the pGEX-5'-SP and pGEX-3'-SP primers. The resulting plasmid, pGEX-4T1-G0-3, was transformed into BL21 cells for protein expression. Overnight cultures of BL21 pGEX-4T1-G0-3 were diluted 1/100 and grown to an OD_{600} of 0.6-0.8 in MinA minimal media with ^{15}N or ^{13}C supplied as required via labeled NH_4Cl or glucose. At this point, protein expression was induced with IPTG at a final concentration of 0.2mM for 4 hours at 37°C or overnight at 30°C.

G0-3 protein was purified by glutathione agarose affinity chromatography (Section 2.4.3) and on-column thrombin cleavage, before being exchanged by PD-10 column (Section 2.4.6) into 10mM phosphate buffer pH 7.4 for NMR. Typical yields were 6-8mg purified G0-3 protein per litre

112

of expression culture, and yielded excellent sample purity. Figure 5.3 shows a sample SDS-PAGE

following the various stages of purification.
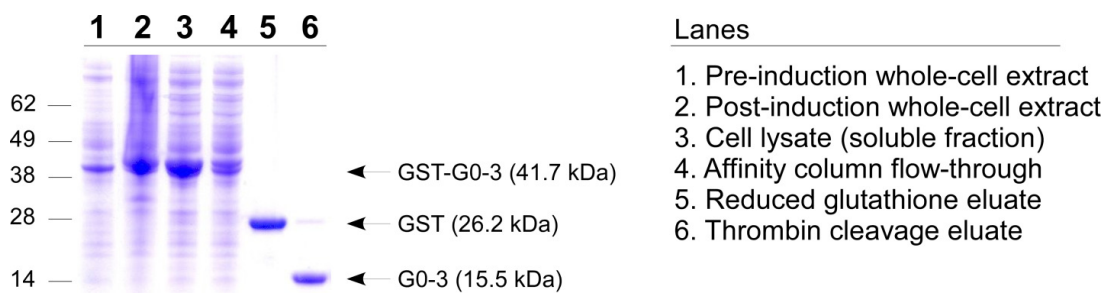
Figure 5.3 Expression and purification of the first four ankyrin repeats of gankyrin (G0-3) for NMR studies.

NMR assignment of G0-3

Initially, ^{15}N-HSQC, HSQC-NOESY and HSQC-TOCSY spectra of the G0-3 protein were collected,

with the aim of assigning HSQC cross-peaks only (primarily backbone resonances), to enable ligand

titrations to be monitored by HSQC and the data fully interpreted. 150 peaks were picked from the

HSQC spectrum, where 162 were expected (Figure 5.4), including the expected 12 pairs of

glutamine/asparagine side-chain peaks. The remaining 12 peaks were most likely unable to be

picked due to spectral overlap, though the weakness of a few backbone resonances (for example

42T and 108T) suggest that a small number of peaks may be weak enough to be difficult to

distinguish from noise. The dispersion of the spectrum in the proton dimension is evidence that

the protein is folded, consistent with published data (Yuan, Li et al. 2004). Further, the pattern of

peaks is very similar to that expected based on the published chemical shifts for full-length

gankyrin that were deposited in the BioMagResBank (accession number 5898).

Figure 5.4 HSQC spectrum for G0-3. Picked peaks are marked with crosses. Peak assignments are marked where known, with the residue number and single letter amino acid code.

The shifts were used as a reference for the assignment of HSQC peaks, and where possible, spin system linkages were confirmed via through-space correlations with adjacent backbone protons in the collected HSQC-NOESY spectra. This yielded assignments for 66 of the expected 138 backbone peaks (48%). As would be expected, the positions of the backbone peaks in ankyrin repeats 2 and 3 of the G0-3 construct differed significantly from the corresponding peaks in the full-length gankyrin due to the change in chemical environment caused by the removal of the subsequent ankyrin repeats. Changes in side-chain packing, and backbone hydrogen bonding network as a result of the truncation allow the changes in the HSQC spectrum to propagate to some extent into the penultimate ankyrin repeat. The combination of this effect with marginal 3D spectrum quality made full backbone assignment from these spectra impossible. Signal strength could not be improved, due to the solubility limit of the G0-3 protein. It was decided that other

experiments should be prioritized until the installation of a cold-probe at the University of

Adelaide enabled collection of new spectra with greater signal, including triple-resonance spectra

to help alleviate the issues with spectral overlap. Screening of potential ligands by HSQC titration

could continue in the interim, since HSQC resonance assignment is predominantly required to

determine the nature (as opposed to the affinity) of ligand binding. Since no evidence of ligand

binding was found (see below), further resonance assignment was ultimately not pursued to

completion.

HSQC titrations of potential ligands

Potential ligands were assayed by step-wise titration of two compounds together into samples

of ^{15}N-labelled G0-3 at protein concentrations of 0.25-0.3mM in 10mM phosphate buffer with 10%

DMSO. Stock solutions of each compound were made to 400mM in DMSO and added step-wise to

concentrations of 25μM, 50μM, 100μM, 200μM, 500μM, 1mM and 2mM of each compound.

Phosphate buffer was added independently at each step to maintain a total DMSO concentration

of 10% and HSQC spectra were obtained at each step. The resulting spectra were overlaid in order

to identify changes in chemical shift or peak intensity as indicators of fast- or slow-exchange ligand

binding, respectively. An example is shown in Figure 5.5. In several cases (as below), obvious

precipitation resulted from the addition of compounds to 2mM each, and these spectra were

excluded from analysis.
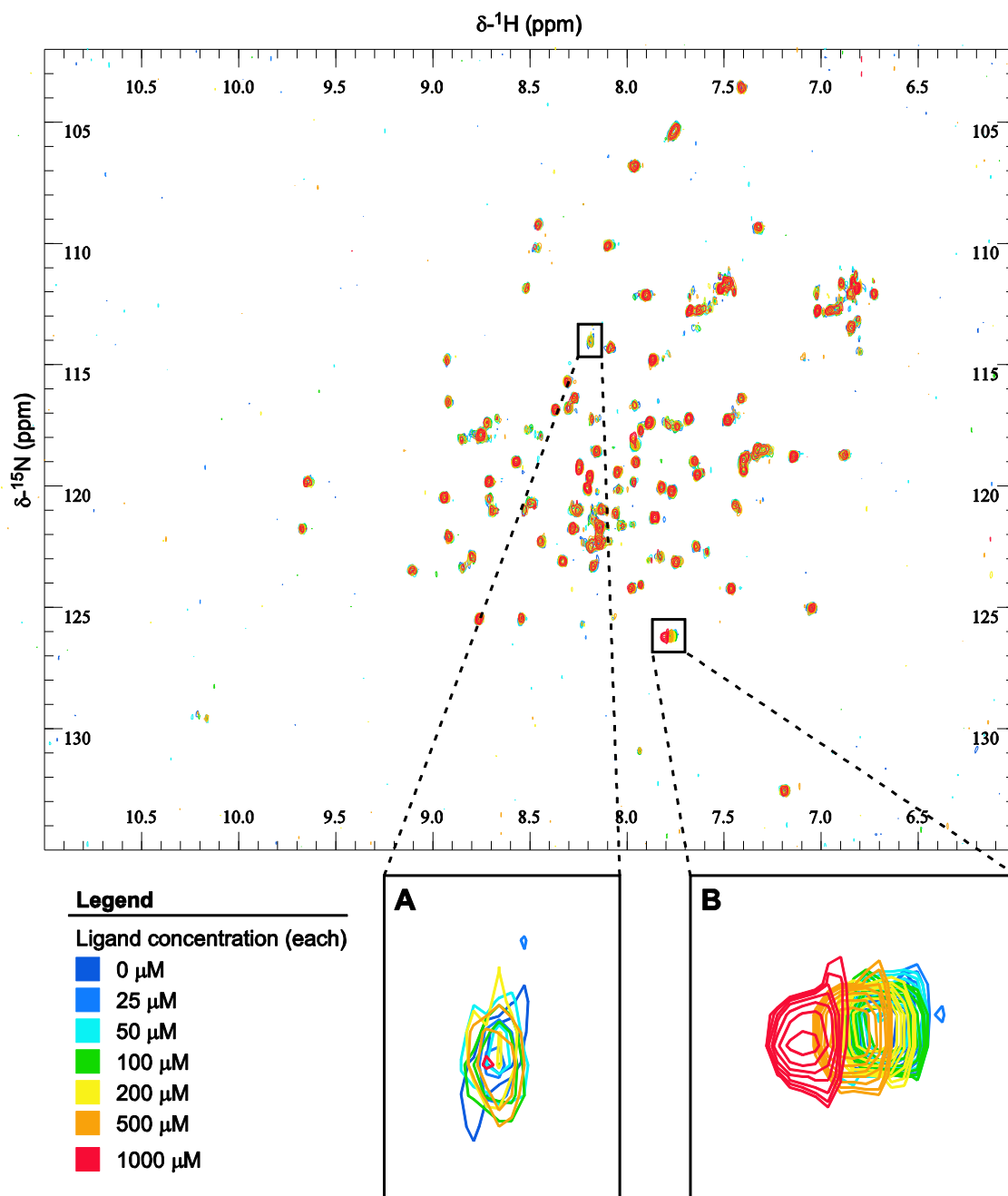
Figure 5.5 HSQC titration of compounds GC7 and GC8 (refer Table 5.1) into G0-3. Insets A and B represent enlargements of the indicated regions of the spectrum.

As can be seen above, the chemical shifts of a single peak were observed to change (inset B), while other peaks appeared to decrease in intensity (an example is shown in inset A). However, the peak decrease occurs across all peaks - it is simply the weaker peaks for which this is most obvious, since they decrease to near or below the level of noise. This phenomenon is most likely due to the early effects of sample aggregation and precipitation, before it became directly visible in the

sample itself. Over the course of the titration, the sample was diluted by 10-12% (see Section 2.5.3 for further details of the method) so would not have been sufficient to explain all the loss of signal. Consistent with this theory, the loss of signal was less pronounced in samples that did not end with precipitation.

The single peak showing chemical shift changes is highly unlikely to be due to ligand binding. Genuine ligand binding would be expected to induce chemical shift changes in several peaks, as binding inevitably involves and changes the chemical environment of several residues. Smaller changes to the same peak were observed in titrations of GC9 and GC10. Given that GC8 and GC10 are both acidic, while GC9 is weakly basic, this is consistent with this change being due to changes in pH, rather than ligand binding, suggesting that in these cases, the buffering capacity of the 10mM phosphate buffer used was insufficient. If this assay is to be pursued in future, it may be wise to increase the buffering capacity slightly and collection of a pH titration series of HSQC spectra as a control would be crucial.

5.3.3 A positive control for G0-3 ligand binding

In order to verify that interactions between G0-3 and ligands can be detected via this method, the 58-amino acid N-terminal fragment of CDK4 (termed C5), which had previously been shown to interact with p16 (Byeon, Li et al. 1998), was analysed for binding to gankyrin. This work was accomplished in collaboration with Phillippa Smith, who cloned, expressed and purified C5, and prepared ^{15}N-labelled gankyrin samples for HSQC titration experiments. I provided support in the form of assistance with experimental design, day-to-day laboratory supervision and assistance with NMR data acquisition and analysis. This work has been previously presented in greater detail in Phillippa Smith's Honours thesis (Smith 2007).

C5 was expressed as a GST-fusion protein and purified by glutathione-affinity chromatography and on-column thrombin digestion. Protein samples were buffer exchanged out of TBS into MQ water using a PD-10 column (GE Healthcare) and freeze-dried in aliquots in preparation for NMR titration. Freeze-dried aliquots were added sequentially to a sample of uniformly ^{15}N-labeled G0-3

(0.3mM) and ^{1}H-^{15}N HSQC spectra recorded between each addition, shown in Figure 5.6 (selected peaks also highlighted in Figure 5.7).
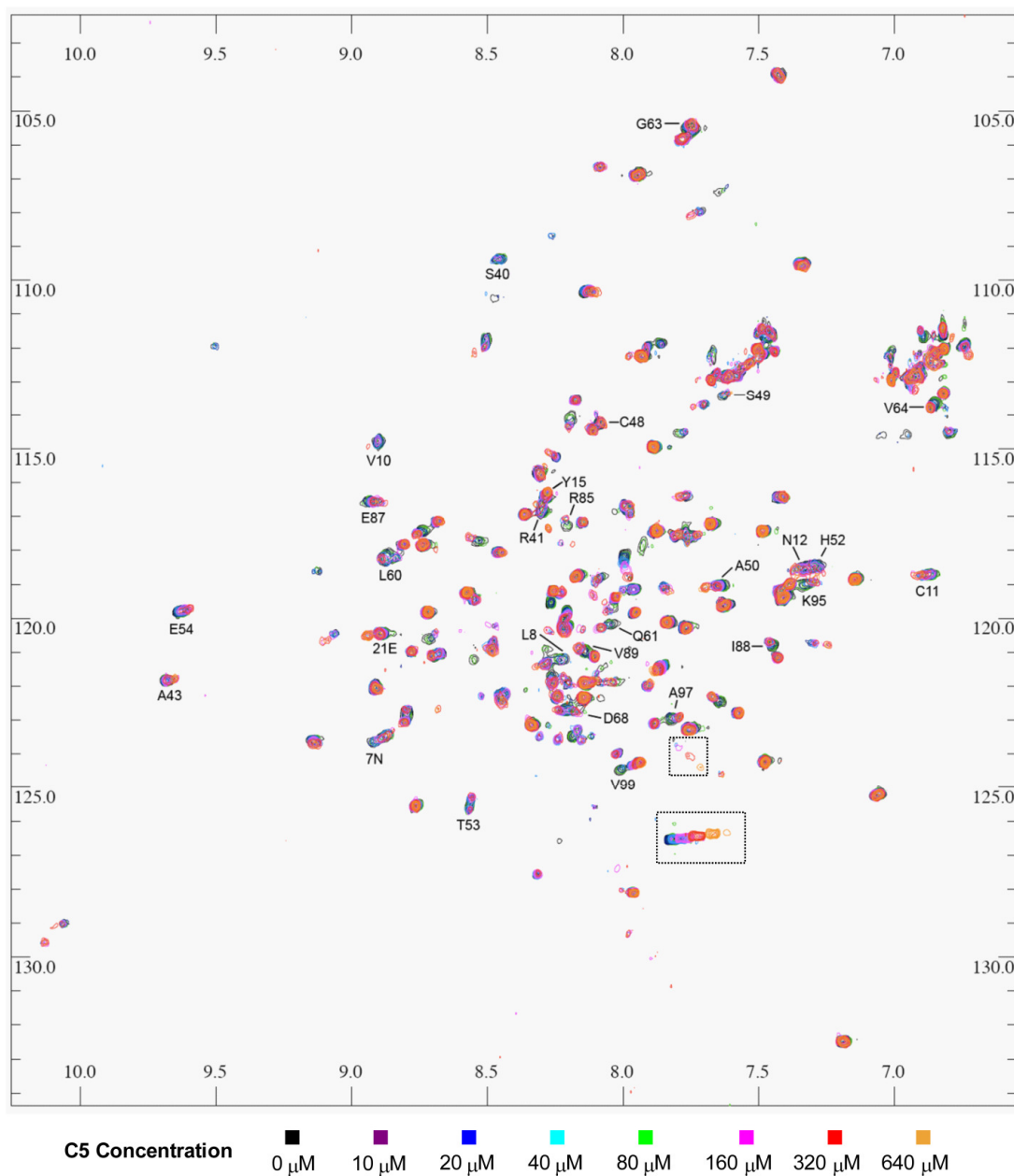
Figure 5.6 HSQC titration of purified C5 protein into 0.3mM ^{15}N-labeled G0-3. Two effects are visible with increasing concentrations of C5. A fast-exchange process is visible for a small number of resonances (boxed). Peak assignments are shown where a slow- or intermediate-exchange, C5 concentration-dependent change is evident and peak assignments are known (28 residues, selected residues shown in more detail in Figure 5.6). Similar effects are observable for some unassigned peaks.
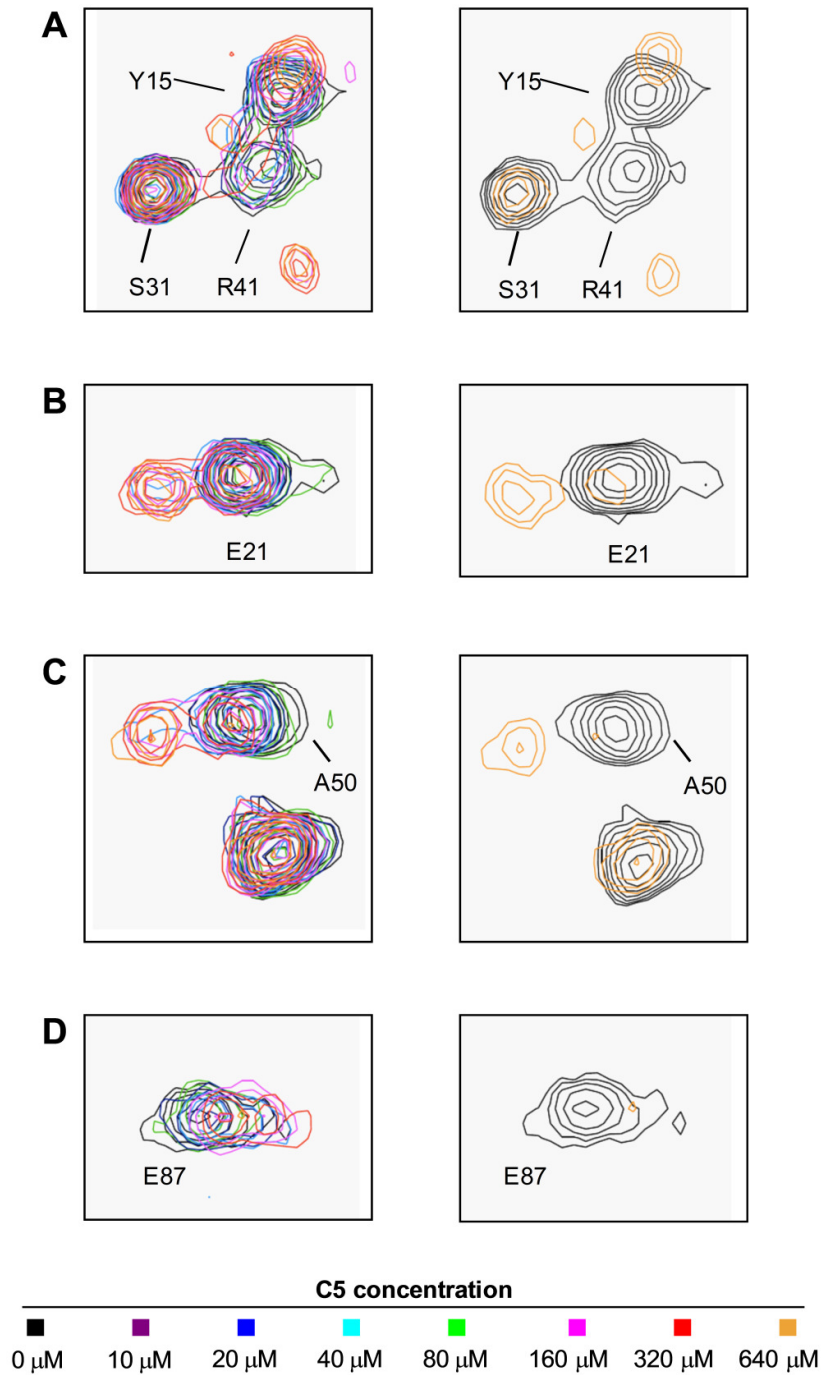
Figure 5.7 Detailed view of selected peaks from titration experiment (Figure 5.6). All titration points are shown in the panels on the left, while only the start- and end-points of the titration experiment are shown in the right-hand panels. Figure prepared by Phillippa Smith.

There is evidence in these spectra for interactions occurring on two different time-scales. It is worth noting that the peak most clearly exhibiting the effects of a fast-exchange process is the same peak that was observed to undergo chemical shift changes during titrations with some candidate small molecule ligands (Figure 5.5). It is likely that this, and the other similar change is a pH-dependent effect due to incomplete removal of Tris buffer before freeze-drying of C5 samples, leading to alteration of the sample pH concomitantly with addition of C5. This in turn may lead to the observed effect on a particularly pH-sensitive region of gankyrin, such as the backbone resonance of a histidine residue, in which the protonation state of the side chain changes with pH, leading to a change in the chemical environment (and hence chemical shift value) of the backbone resonance.

More interesting, however, is the evidence of a slow- or intermediate-exchange process of numerous other residues, as demonstrated by peak broadening and disappearance in a C5 concentration dependent manner. In a slow-exchange regime, receptor residues whose chemical environments are altered by ligand binding are in essence present in two populations – bound and unbound – for the time over which the NMR spectrum is acquired, and are thus expected to give rise to two distinct resonances. As the proportion of sample moves from the unbound state to the bound state with higher ligand concentrations, the relative strengths of the two resonances change accordingly, until only the "bound" resonance is visible at saturation. This effect, including the observation of new resonance peaks arising from the bound state, is evident in the titration of C5 into G0-3 (Figure 5.7 A, B & C) though in some cases (Figure 5.7D) there is no clearly identifiable appearance of a new bound-state peak. This, combined with the fact that most of the bound-state peaks that can be observed are weak, suggests that binding of C5 to G0-3 has not reached saturation.

These results should be treated with a degree of caution, given that there were some indications from mass spectrometry data that the C5 peptide was partially degraded over the

course of the experiment. However, the fact that the results were repeatable across three independent titration experiments using independent preparations of C5 provides additional assurance that the observed effects are due to specific binding of C5.

Mapping those residues which exhibit the effects of slow-exchange binding onto the structure of gankyrin reveals that the residues are predominantly located in the concave surface of G0-3 as would be expected based on the expected similarity between the binding of CDK4 to $p16^{INK4A}$ and gankyrin.
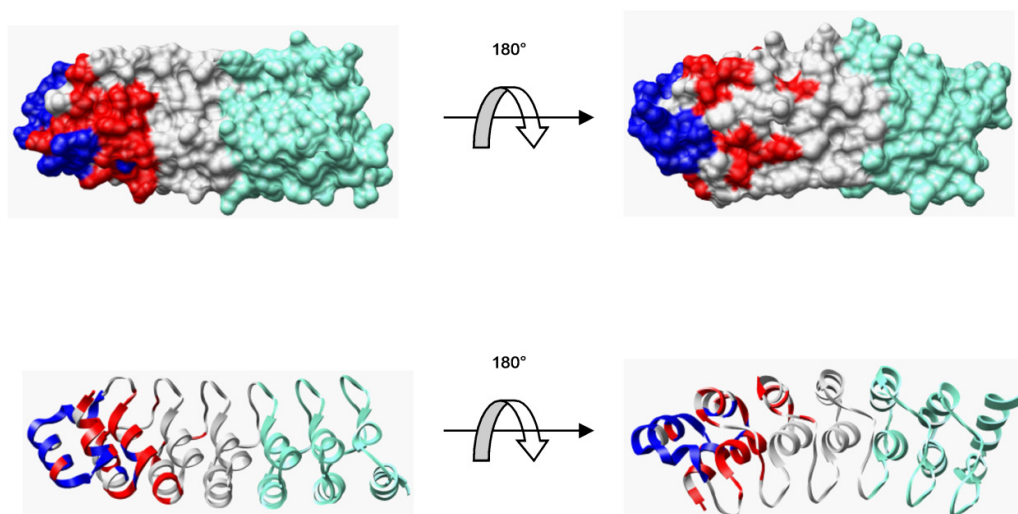
Figure 5.8 Surface (top) and ribbon (bottom) representations of gankyrin, highlighting assigned residues with perturbed HSQC resonances (red) and assigned residues with unperturbed HSQC resonances (blue). The portion of gankyrin shown in cyan are residues 139-226, which are included for illustrative purposes only, since the experiment was conducted with G0-3. Figure prepared by Phillippa Smith.

The appearance of a few residues on the "back" side of gankyrin, opposite the main cluster in the expected region is likely due to relatively small but significant changes in the global conformation of G0-3, such as the relative positioning of adjacent repeats, which occur upon ligand binding. Indeed this sort of global flexibility is precisely the phenomenon that is observed when a de-stabilizing mutation to a synthetically designed ankyrin repeat increases its affinity for a protein binding partner (Zahnd, Wyler et al. 2007). Additionally, a single point mutation in gankyrin

has previously been observed to induce widespread changes to its HSQC spectrum while retaining

its ability to bind CDK4 (Mahajan, Guo et al. 2007).

These experiments successfully demonstrated that ligand binding by G0-3 is detectable by

HSQC titration experiments. Of 16 residues that lie at the site targeted for virtual screening, 8

were assigned and showed C5 concentration-dependent changes in the HSQC spectrum, while the

remaining 8 were unassigned. Thus the experiment also served to confirm that the region being

targeted for drug design did, in fact, have significant overlap with the binding site of CDK4.

5.3.4 A test case for in silico discovery of ankyrin-repeat protein ligands

At this point, it was decided that continuing to screen down the ranks of compounds was not

the best use of the available time. With several scoring functions made available by the Scorer

program, a question remains as to whether some of them perform better than others in the

context of docking to a protein-protein interaction interface. Analysis of the docking and scoring of

a similar "receptor" target would help answer this question and clarify whether continued *in vitro*

screening of potential inhibitors of gankyrin is likely to be fruitful. That is, it would be desirable to

know whether an *in silico* docking approach is *capable* of identifying small molecule ligands of an

ankyrin-repeat protein, as we have asked it to do here. To identify a test case that might help to

answer this question, the PDB was searched for proteins with sequence similarity to gankyrin that

also contained at least one small molecule ligand. Of the search results returned, the only result

that was a suitable example for exploring this question was the entry 1WDY: a crystal structure of

the ankyrin-repeat domain of human Ribonuclease L (RNase L) in complex with its natural ligand

5'-phosphorylated 2',5'-linked oligoadenylate (2-5A). More importantly, however, Thakur *et al*

(2007) report the high-throughput screening of the ChemBridge DiverSet small molecule library

against RNaseL and the discovery of 12 ligands within this library with K_d values in the range of 20-

100μM. This system thus provides a very appropriate test case for the exploration of whether *in*

silico techniques are able to recapitulate the *in vitro* results of the discovery of small-molecule ligands of an ankyrin-repeat protein.
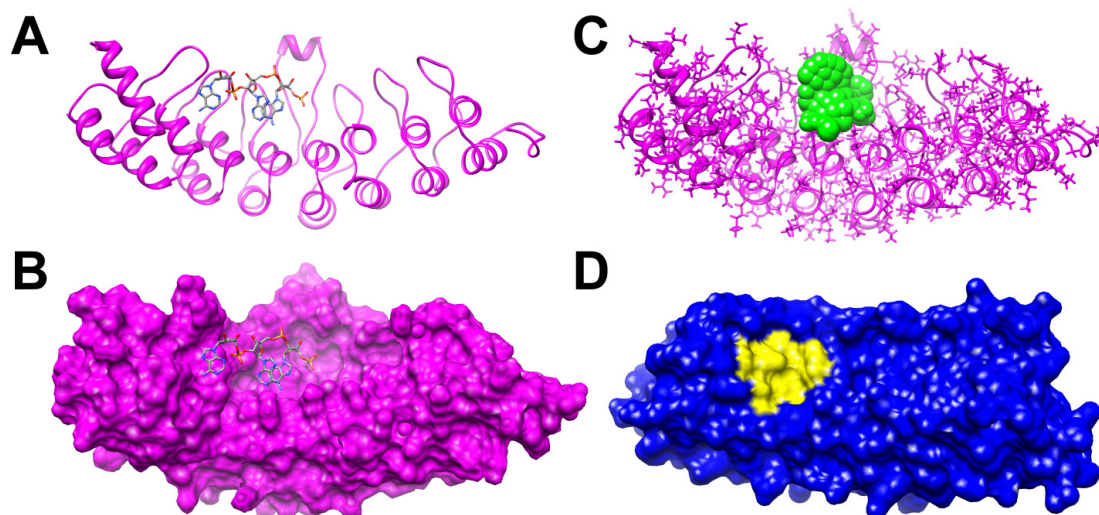
Figure 5.9 Ribbon **(A)** and surface **(B)** representations of the crystal structure of ankyrin-repeat ligand-binding domain of RNaseL (PDB ID 1WDY). The natural ligand, 2',5'-linked adenosine (2-5A) is shown in stick representation, coloured according to element. **(C)** Full-atom representation of the same RNaseL crystal structure without the ligand shown. The spheres used for the initial placement of ligands during docking are shown in green, chosen to overlap the main binding pocket of 2-5A and encompassing the region previously predicted to be the binding site of a number of small-molecule ligands found by high-throughput screening (Thakur, Jha et al. 2007). **(D)** Surface depiction of gankyrin (targeted site highlighted in yellow) is shown for comparison of the targeted regions of the two proteins.

To this end, representations of the DiverSet small molecule library were obtained from Chembridge and prepared for docking to the crystal structure of the ankyrin-repeat domain of RNaseL (Figure 5.9). It was found that the bond-lengths implied by the coordinates provided by Chembridge were physically implausible as they appeared to be scaled to roughly (but not consistently) 60-70% of their canonical lengths, meaning that the represented bond lengths were not accurate, and nor could accurate bond lengths be reliably derived from them. Therefore, the library was converted to SLN (Sybyl Line Notation) linear format using Sybyl 7.3, before regeneration of 3-dimensional coordinates and calculation of partial charges via the Gasteiger-Marsilli method and conversion to mol2 format in readiness for docking. PDB entry 1WDY was

prepared for docking by removal of the 2-5A molecule, and passage through the UCSF Chimera

DockPrep utility as was described previously for gankyrin. Spheres targeting the 2-5A binding site

were generated (pictured in Figure 5.9B) and energy potential grids prepared for docking with

UCSF Dock v5.1. The chosen spheres do not cover the extremities of the 2-5A binding pocket, since

the evidence presented by Thakur *et al* (2007) suggests that the small molecule inhibitors do not

bind to this region. The use of these spheres also doesn't preclude the discovery of poses that

commence within but extend beyond the sphere region, given the nature of the Dock algorithm

and the fact that the scoring grids covered the entire 2-5A binding pocket Docking was carried out

as described above for gankyrin.

5.3.5 Comparison of scoring methods using the RNaseL test case

While all of the scoring algorithms employed by Scorer have, of course, been subjected to

various tests of their ability to predict the binding of small molecules to protein receptors, there is

a case to be made that the design of inhibitors of protein-protein interactions represents a distinct

class of problem and that the applicability of the different scoring algorithms to this problem has

not been directly assessed. The dearth of potent, small-molecule inhibitors of protein-protein

interactions with matching structural and affinity data makes the assessment of docking and/or

scoring algorithms specifically with regard to this class of problem very difficult to achieve on the

same scale and with the same rigour as previous, broader tests of their efficacy and accuracy

(Warren, Andrews et al. 2006). However a comprehensive and thoroughly objective comparison of

the performance of the scoring algorithms employed by Scorer is not the primary objective in this

case.

Comparison of the success of different strategies at correctly identifying hits is often assessed

by enrichment factors. That is, the fold change in the number of true positives (hits) contained

within a given fraction (x%) of top-ranking results that can be achieved by docking, relative to the

number that could be expected by random chance. This method of comparison is useful insofar as

it yields a quantitative measure of success for direct comparison. However, the choice of a hard

boundary condition at x% can quite dramatically alter the conclusions drawn, particularly in cases

– such as RNaseL in consideration here – in which there are relatively few true positives. Few true

positives means that a single result falling either side of the boundary results in a large change in

the enrichment factor, and hence the conclusions drawn. A better alternative, used here, is to plot

the number of true positives found as a function of the proportion of the library screened. When

compared to the theoretical number of hits identified at random, represented by a diagonal line

through the plot, in essence a continuous plot of the enrichment factor for all boundary conditions

(variations of x%) can be drawn, where greater enrichment is indicated by the plot tracking a

greater distance above the diagonal.

Conducting this type of analysis on the RNaseL docking test set (Figure 5.10) demonstrates

significant variability in the performance of the different scoring algorithms. Overall, PK and PMF

clearly perform the best of all the algorithms. Smog, ChemScore and ChemClash all perform

similarly (and quite poorly) while AutoDock, which also performs quite poorly, yields a distinctly

different pattern of hit discovery. Strikingly, while most of the algorithms produce an *overall*

enrichment (100% of the hits are found after screening approximately 60% of the compounds for

all algorithms except AutoDock), they equally perform very poorly with respect to discovery of the

first hit – again, curiously, with the exception of AutoDock, no algorithm out-performs random

chance in this respect. AutoDock overall only convincingly out-performs chance over the discovery

of the first four ligands. However none of them, AutoDock included, demonstrate a success rate

that would allow one to confidently adopt a virtual screening approach. Even the best-performing

algorithms would require the screening of more than 10% of the library in order to identify more

than a single hit, which represents an *in vitro* screen of around 5,000 compounds. While limiting

the cost of the undertaking in terms of the purchase of chemical compounds, it still represents an

undertaking that requires a well-developed and validated high-throughput screen in order to be

feasible, which in itself represents a significant investment. Importantly, none of the hits are

represented in the set of compounds in the 90th percentile or higher for each of the 6 scoring

algorithms (excluding ChemClash). Thus it is reasonable to assume that the approach adopted for

the discovery of gankyrin ligands is highly unlikely to succeed, and that search was terminated.
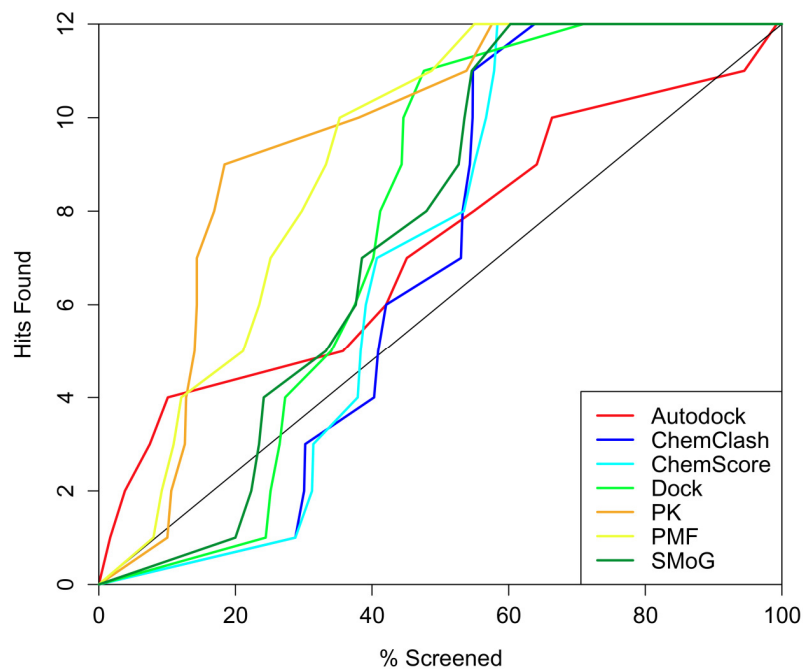
Figure 5.10 Comparison of the performance of different scoring functions in identifying the 12 known ligands (hits)

of RNaseL in a virtual screen of the ChemBridge DiverSet library. The solid diagonal represents the theoretical rate at

which hits would be found by a non-predictive, random score. Lines tracking furthest into the top-left region of the plot

are the best performing.

5.3.6 *Comparing ligand burial of RNaseL ligands with a docking validation set*

Beyond its use in this manner as a proof-of-principle, however, the RNaseL test system

potentially allows us to explore what differences there may be between the targets to which *in

silico* drug discovery has been commonly applied in the past and protein-protein interaction

targets. Fry and Vassilev (2005) have previously explored this question, noting that in order to be

amenable to drug targeting, a receptor must have a cavity of sufficient depth and surface area.

This being the most obvious – and proven – qualitative difference between the more common

enzymatic targets of *in silico* discovery and protein interaction interfaces, exploration of this

concept in the more specific context of gankyrin and the RNaseL test case may aid assessment of whether they are "druggable" targets, and help determine strategies for improving the efficiency of *in silico* identification of small molecule ligands.

Unfortunately, the computational tools used by Fry and Vassilev (2005) are unavailable to other scientists. Rather than attempting to develop a new implementation of the algorithm described by Fry and Vassilev, other available tools that enable examination of similar concepts were sought. As a corollary to the depth and surface area of a ligand binding pocket, Fry and Vassilev (2005) note that "amenable sites were all able to sequester a larger portion of the surface area of a small molecule away from accessibility by solvent than were the non-amenable sites" (referring to amenability to drug targeting). Thus, rather than analyzing the receptor proteins directly, similar information can be accessed via analysis of ligand poses and the amount of surface area sequestered away from solvent accessibility – defined by Pattabirama and Fleming (1995) as the "occluded surface." This approach was particularly suitable in this instance given that a large set of small molecule poses docked to gankyrin had already been generated, and a similar set had also been generated for RNaseL in order to analyze the performance of different scoring algorithms. The tool *OS* (v7.2)(Pattabiraman, Ward et al. 1995; Fleming and Richards 2000) allowed calculation of the occluded surface of a given ligand pose.

Determining whether RNaseL and gankyrin represent a distinct class of problem in terms of ligand burial requires a reference set of examples against which to compare the burial of the RNaseL ligands (or putative ligands in the case of gankyrin). Such reference sets have already been compiled as the training sets upon which the various docking and scoring algorithms have been trained and/or assessed. One such set – the CCDC/Astex test set – comprises 305 crystallographic solutions of protein/small-molecule complexes, and was used to investigate ligand burial with the OS tool. Analysis of the burial of the 305 ligands suggested that there was no identifiable lower limit for the absolute amount of ligand surface area that was solvent excluded. However, a clear correlation between the occluded surface area and the total surface area of the ligand was

apparent (Figure 5.11). This clearly suggests that there is a requirement for a minimum *proportion* of the ligand surface area to be occluded.
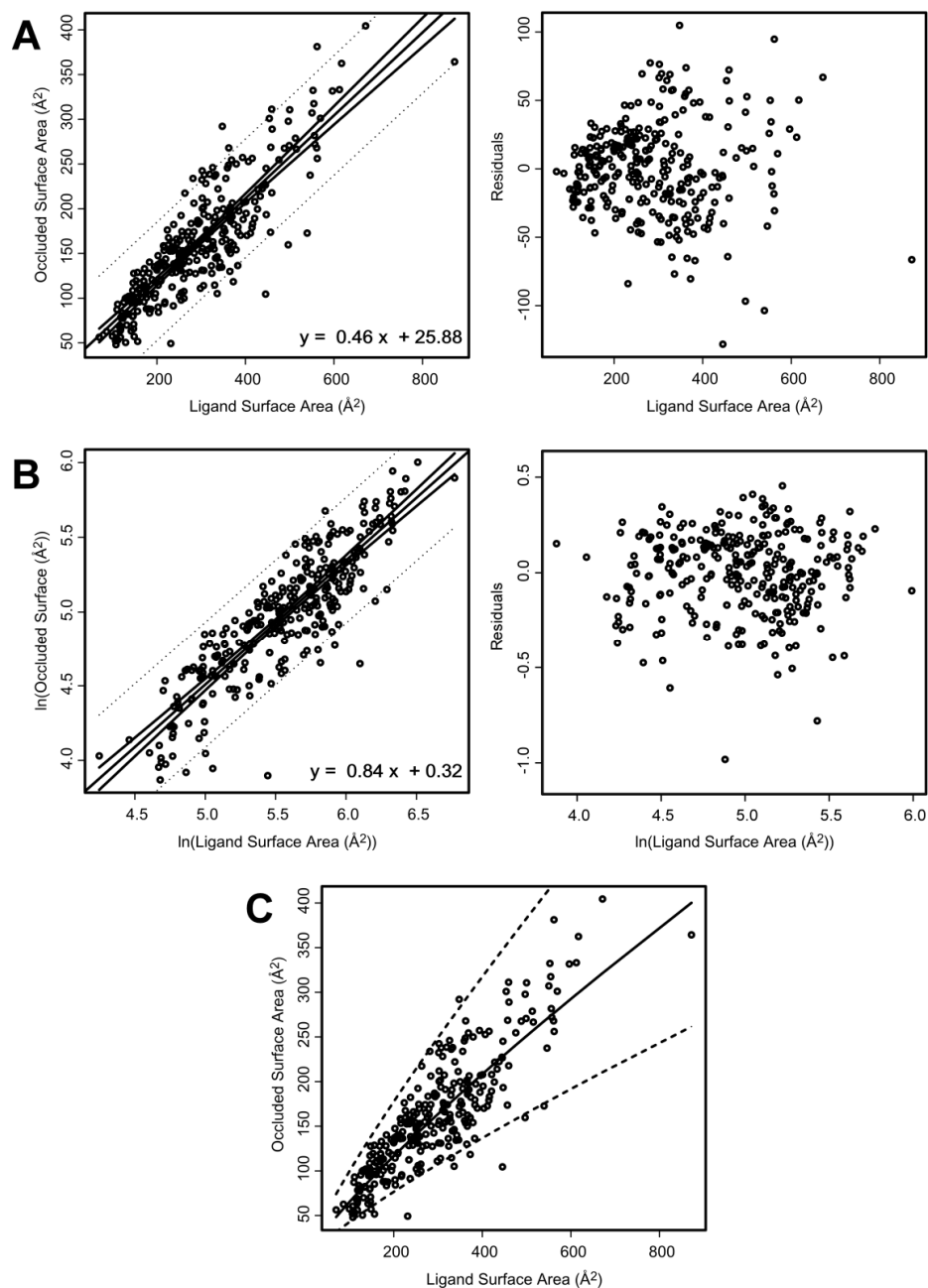
Figure 5.11 Correlation and regression analysis of ligand burial among members of the CCDC/Astex test set. *(A)* Linear regression analysis of the relationship between total ligand surface area and occluded surface area. $R^2 = 0.7506$ *(B)* Log-log regression analysis. $R^2 = 0.7586$; 95% confidence interval (prediction) = mean $\pm$ 0.84 *(C)* Transform of log-log regression analysis onto linear plot. *(All plots)* Solid lines: Best-fit regression curve with 95% confidence interval of the mean. Dashed lines: 95% confidence prediction interval.

Regression analysis of this data allows the definition of a prediction interval, within which 95%

of all ligands can be expected to be found. A simple linear regression (Figure 5.11A) was deemed

unsatisfactory, as the residuals of the fit increase with rising ligand surface areas. The existence of

a trend in the residuals (right panel) suggest that a different type of regression would be more

suitable, particularly when it is to be used for the calculation of a prediction interval. In order to

address this, a log-log regression was performed (Figure 5.11B on log-log scale with residual plot,

while Figure 5.11C shows the results of log-log regression transformed back onto linear scale),

which successfully eliminated any observable trend in the residuals. From a purely theoretical

perspective, this is quite unsurprising, since it reflects the idea that there is a constant variation in

the *proportion* of surface area that is occluded, rather than a constant variation in the *absolute*

occluded surface area, consistent with the initial observation of the trend. In fact, plotting the

distribution of ligand burial as a proportion of total surface as a histogram shows that ligand burial

within the reference set is approximately normally distributed ($p = 0.10$ by the Pearson chi-square

test of normality) about a mean of 57.2% with a standard deviation of 12.1% (Figure 5.12).
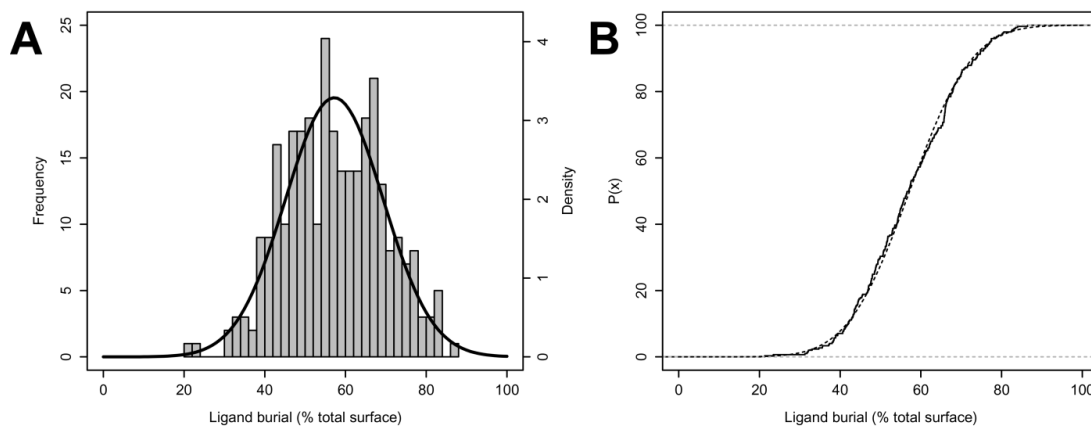
Figure 5.12 (A) Histogram representing the ligand burial (expressed as the percentage of the total surface area that

is occluded) of the members of the CCDC/Astex test set. Bold black line represents a normal distribution model of the

data. The Density axis (right) is defined such that the area under both the normal distribution and frequency plot are

equal to 1. **(B)** Empirical cumulative distribution function for ligand burial of the CCDC/Astex test set (solid line) and

theoretical cumulative distribution function for the fitted normal distribution (dashed line).

Including the ligand burial data for the 12 RNaseL small molecule ligands identified by Thakur

et al (2007) together with the above reference set and regression analysis would demonstrate

whether these ligands are less buried than those in the CCDC/Astex test set. Unfortunately,

crystallographic data is not available for the RNaseL ligands, so *in silico* docked poses from the

docking of the Chembridge Diverset needed to be used as an estimate of the native pose.

Analysing the burial of the best-scoring pose for each ligand shows that while the ligands don't

occupy an entirely distinct area of the graph, they certainly are among the least buried ligands in

the set (Figure 5.13, blue circles). Further, more than half of them lie outside the 95% confidence

prediction interval, suggesting that on average, they represent a "less buried" class of ligand as

compared to the aggregate CCDC/Astex test set.
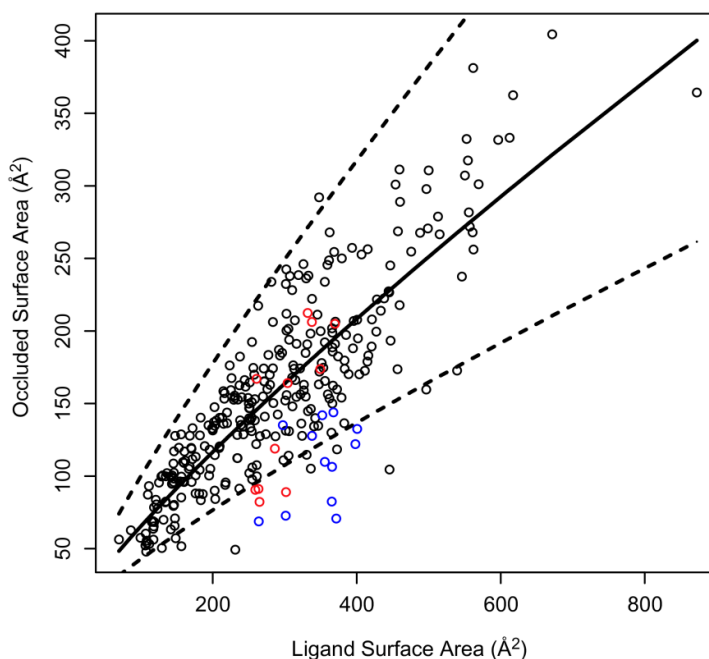
Figure 5.13 Ligand burial of the RNaseL small molecule ligands selected by best score (blue circles) or greatest

surface occlusion (red circles) compared to the CCDC/Astex test set (black circles) and regression analysis (black lines;

solid line: best-fit mean value; dashed line: 95% confidence prediction interval)

On the one hand, this may indicate that RNaseL and its small molecule ligands represent a distinct class of protein-ligand interactions. However, because this data is derived from *docked* poses of the ligands rather than experimental structure data, it may equally indicate that the docking and scoring algorithms underestimate the importance of ligand burial, precisely because the majority of the training sets are dominated by receptors in which the target site is a relatively deep pocket. In such circumstances, ligand burial is unlikely to prove to be a significant discriminator between the true pose and decoy (false) poses, since all are likely to exhibit significant burial. This can lead to an underestimation of the importance of ligand burial in the scoring algorithm, which when applied to binding sites that present as more of an open face, lead to more favourable scores for less buried ligands than is desirable. Indeed, if the most buried poses of the RNaseL ligands are selected rather than the best scoring poses, and similarly compared to the burial of the ligands in the CCDC/Astex test set, the extent of ligand burial is broadly comparable (Figure 5.13, red circles). This demonstrates at the very least that a similar level of ligand burial is *accessible* given the surface contours of RNaseL, and calls into question the notion that RNaseL and its ligands represent a distinct class of receptor-ligand interactions.

5.3.7 Can ligand burial analysis improve enrichment?

The above data renders the hypothesis that the DOCK scoring algorithm takes inadequate account of ligand burial entirely plausible. A corollary of that hypothesis is that consideration of ligand burial along with scoring data should be able to more successfully identify the correct pose out of many when considering an "open faced" receptor such as RNaseL. While structural data for the true poses of the RNaseL inhibitors are unavailable, a small-scale analysis of ligand burial in the context of pose selection can be carried out using 2-5A, the natural ligand of RNaseL, which was co-crystallised in PDB entry 1WDY. If the above hypothesis is correct, one would expect consideration of ligand burial information along with docking scores to favour accurate poses more so than the docking scores alone. However the use of DOCK v5.1 to dock 2-5A into the RNaseL pocket yields only poses with an RMSD of 5Å or greater compared with the native pose. This is probably due to the large number of rotatable bonds in this particular ligand, which makes

the necessary conformational sampling – and hence accurate docking – a huge computational problem. Consequently there are essentially no "near-native" poses even generated. Analysing the ability of a scoring algorithm to correctly rank or identify native or near-native poses is relatively meaningless unless a variety of poses – some similar to and some distant from the native pose – can be evaluated. Further, any regime which improves the selection of near-native poses via the use of ligand burial information for 2-5A would need to be tested on other ligands, for which no structural data exists.

An alternative is to examine whether ligand burial information can improve the performance of virtual screening of the ChemBridge DiverSet for ligands of RNaseL. Unsurprisingly, the proportion of ligand surface area that is occluded is, alone, a relatively poor measure for the identification of ligands of RNaseL (Figure 5.14; grey line). Despite this, it does appear to perform better than a number of the scoring functions surveyed, and similarly to the AutoDock score.
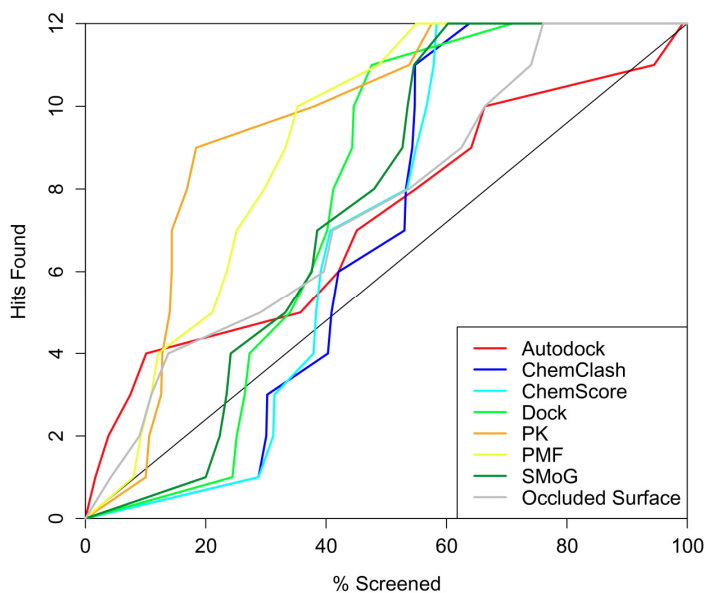
Figure 5.14 Comparison of the performance of occluded surface (ligand burial) as a docking score with that of the scores implemented by Scorer. Performance is measured by the identification of the 12 known ligands (hits) of RNaseL in a virtual screen of the ChemBridge DiverSet library. The solid diagonal represents the theoretical rate at which hits would be found by a non-predictive, random score. Lines tracking furthest into the top-left region of the plot are the best performing.
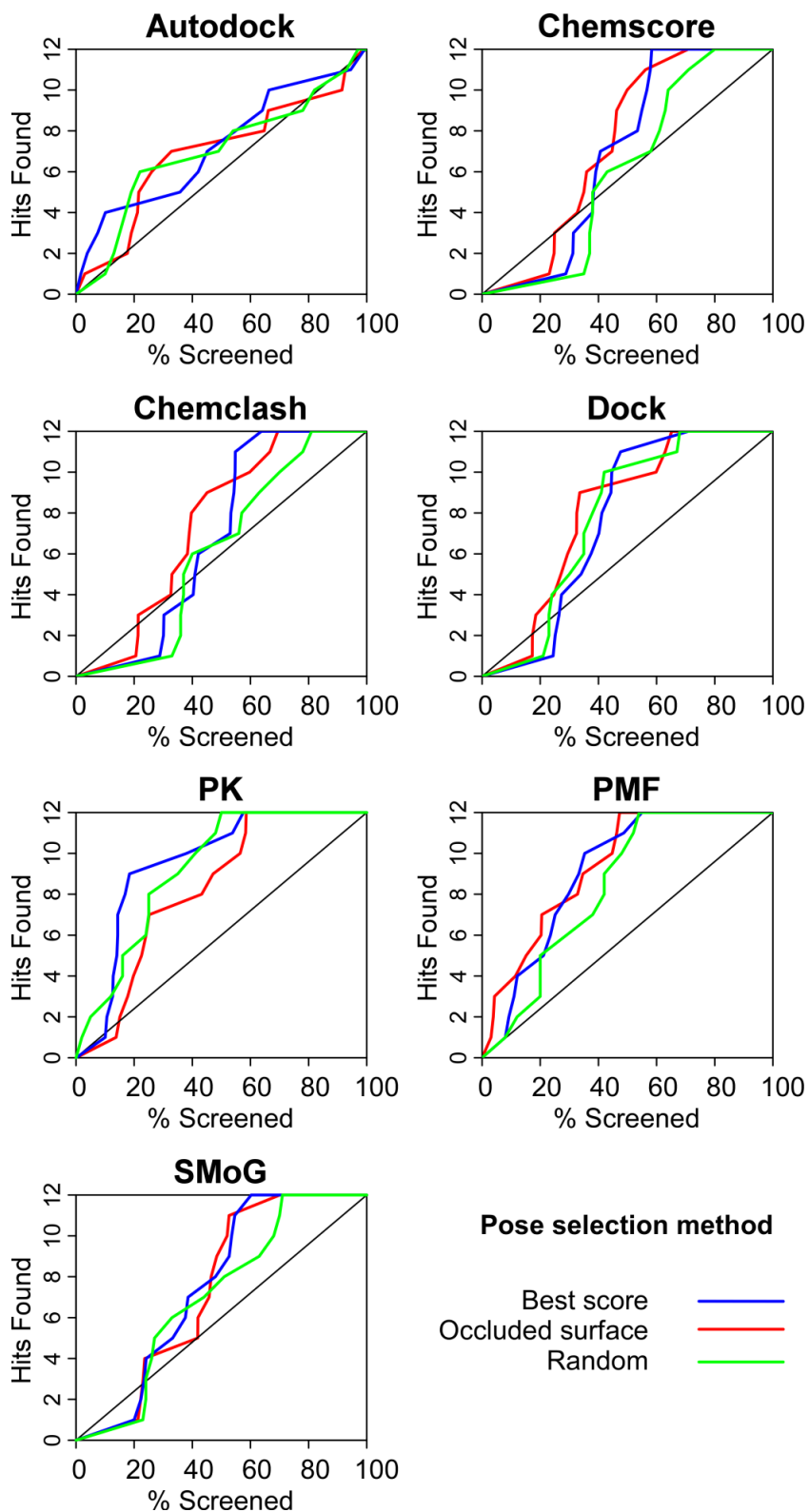
Figure 5.15

Comparison of the effects of pose selection method on the performance of Scorer algorithms. A single pose of each compound in the DiverSet library was selected by one of three methods (best score, greatest occluded surface, or random choice) before being ranked by score and enrichment plots calculated.

A comparison of different methods of pose selection is shown in Figure 5.15, including random pose selection for comparison to help distinguish *systematic* changes in enrichment from those that happen randomly due to any change in the pose selection regime. Most of the scores investigated demonstrate little effect on enrichment by selecting poses based on their occluded surface. Revealingly, random pose selection also performs comparably to selection by score in most cases (green vs blue lines). This could simply demonstrate that there is little variation in the scores of the different poses of a given ligand, and therefore that changing which of those poses is used in ranking can only have marginal effects. This is undoubtedly true to some extent. However it is also interesting to note that those scores that produced better enrichment in general (PK, PMF) were also those that exhibited the clearest impairment in enrichment due to random pose selection. Correspondingly, selection by occluded surface for these scores either degraded their performance or made no definitive difference. By contrast, selection by occluded surface appears to marginally improve the performance of those scores which performed most poorly in the first instance (ChemScore, ChemClash and to a lesser extent Dock), while it made little difference to the middling scores (Autodock, Smog). Taken together, these results suggest that a key difference between the poorly performing scores and those that perform better are that the better scores incorporate ligand burial more appropriately. Thus, selection by occluded surface contributes no additional information or adds too much weight to ligand burial and leads to equal or poorer performance for ChemScore and ChemClash, while the opposite is true for those at the other end of the spectrum.

Looking to the theoretical bases of the various scoring functions bears this out to some extent. In an empirical scoring function, ligand burial is accounted for by a desolvation term. ChemScore/ChemClash, which perform poorly, contain no explicit desolvation term. DOCK includes desolvation only implicitly and incompletely by the use of a dielectric constant (D) to moderate the electrostatic term. SMoG, which performs only marginally better, similarly takes no account of desolvation or ligand burial, looking only at the pairwise favourability of atom-atom

interactions. On the other hand, the other statistically based scoring function, PMF, performs

much better and accounts for desolvation effects by the inclusion of a volume correction function.

PK, the other best-performing function, but based on free energy estimation, includes explicit

energy terms for both desolvation and ligand deformation (shape), which together represent the

most detailed analysis of the energy not just of the interaction, but of the change of state of the

ligand itself in the bound versus unbound states. The only scoring function that fits this analysis

less well is AutoDock, which despite including an explicit desolvation energy term, is not among

the best performing algorithms.

5.3.8 Burial analysis of gankyrin in silico *docked compounds*

Having established that the burial of RNaseL ligands is at or near the lower limits of those

typical of other "drug-able" targets, the question remains whether this information can be applied

to the discovery of ligands for gankyrin. While RNaseL and gankyrin are very similar targets insofar

as they are both ankyrin repeat protein-protein interaction domains/molecules, the atypical

addition of residues including a short helix between ankyrin repeats four and five leads to the

creation of a ligand-binding pocket that simply does not exist to the same degree in the surface of

gankyrin. By comparing the levels of ligand burial that were accessible by docking ligands to

RNaseL with those accessed by compounds docked to gankyrin, it is possible to ascertain whether

the lack of such a pocket on the surface of gankyrin precludes the discovery of ligands with a

reasonable affinity for gankyrin. More specifically, if poses exist in the database of compounds

docked to gankyrin that exhibit similar levels of ligand burial to that of the known ligands of

RNaseL, then we can conclude that a pocket of sufficient depth and surface area exists for genuine

ligands of gankyrin to be found. This is true regardless of whether the poses in question represent

true ligands, since adequate ligand burial is *necessary* but not *sufficient* for ligand binding.

Therefore, the mere existence of poses with adequate burial even of false positive compounds

demonstrates the existence of a pocket in which other ligands could achieve similar levels of

burial, but more favourable contacts, yielding a true ligand. Calculating the occluded surface for

the entire set of gankyrin poses would be computationally very expensive. However, analysis of

the burial of RNaseL docking showed that there the distribution of ligand burial does not

systematically change depending on the DOCK scores of the poses analysed, consistent with the

earlier finding that ligand burial alone is a poor scoring method. Indeed, among the top 500 or top

5% (approximately 2500) poses, the distribution is only slightly different to that for all compounds,

with a trend to a very slight increase in ligand burial as the set is restricted to those most

favourable according to the DOCK score (Figure 5.16).
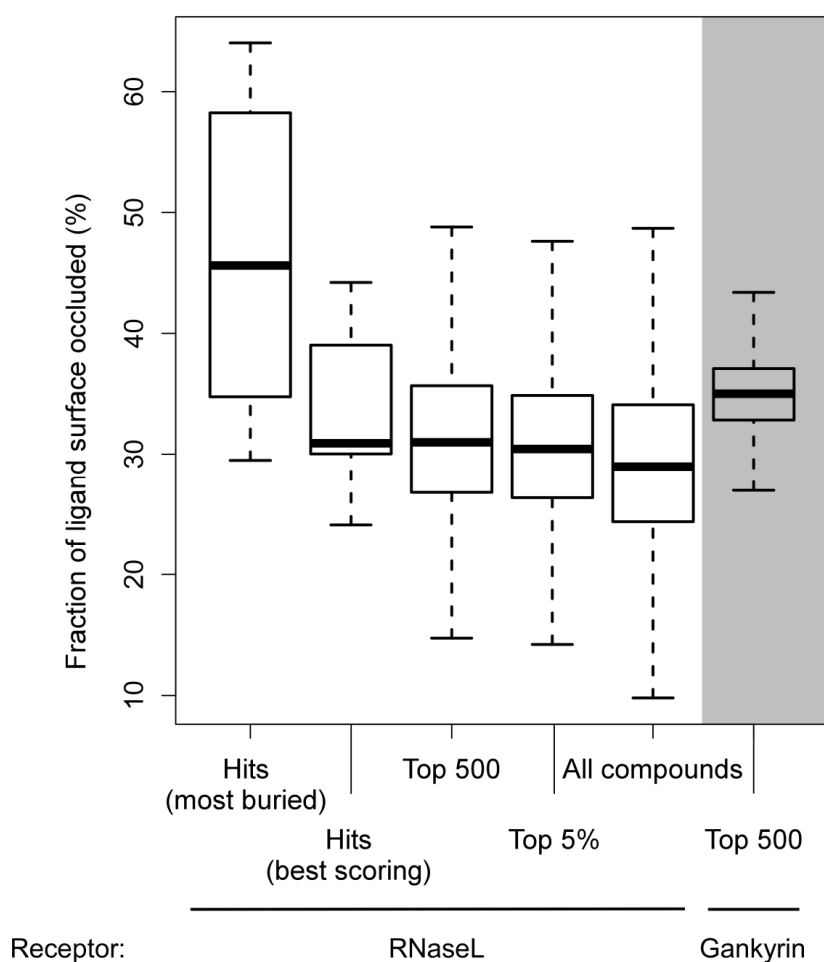
Figure 5.16 Comparison of the ligand burial (measure by the proportion of ligand surface area that is occluded) of

various subsets of the total set of docked poses against RNaseL or Gankyrin. Box plots represent mean and the inter-

quartile ranges after automatic exclusion of outliers which lie more than 1.5 multiples of the length of the box from its

edges.

On this basis, then, the distribution of ligand burial among the top 500 poses of gankyrin docked compounds can be taken to be broadly indicative of the distribution of ligand burial among the entire set of compounds, and therefore provides an insight into the level of ligand burial that is accessible with regard to this target site of gankyrin. Calculating the ligand burial of these poses and plotting the distribution beside those calculated for RNaseL (Figure 5.16) shows that on average, ligands docked to gankyrin achieved greater burial than those docked to RNaseL, and that the gankyrin top 500 set of ligands exhibit a similar level of burial to the *in vitro* verified RNaseL ligands (hits) – provided one assumes that the best-scoring pose is representative of the true nature of the interaction. As noted previously, the "most buried" poses of the RNaseL hits exhibit greater burial. Based on the discussion above, it is likely that neither of these two sets of poses represent the native interactions of the ligands, and it is possible or even likely that the "best scoring" set underestimates the extent of ligand burial in the native states. However, it is also likely that the distribution of ligand burial observed for the top 500 gankyrin poses equally underestimates the accessibility of greater levels of ligand burial, given that they were obtained by the same method. Together, these data strongly suggest that the targeted region of the concave surface of gankyrin has the necessary shape to support binding of a small molecule ligand.

5.4 Discussion

Despite being unable to demonstrate *in vitro* the binding of any small molecules to gankyrin, the use of RNaseL and its small molecule ligands as a test case has nonetheless made it possible to better evaluate the druggability of gankyrin and gain some insights into the reasons for the failure of the *in silico* screen. The obvious and certain conclusion of the analysis of virtual screening is that this method does not perform well on this class of target, and for most of the scoring functions provides little enrichment over random screening. While this work was in progress, other investigators have come to similar conclusions (Betzi, Guerlesquin et al. 2009). Equally, however,

analysis of the extent of ligand burial achievable against both gankyrin and RNaseL shows no significant differences, and suggests that the relatively open interface of gankyrin that was targeted is sufficiently concave to expect that small molecule ligands could be found, given that such ligands have been found for RNaseL.

This too, accords with the findings and opinions of other researchers (Betzi, Guerlesquin et al. 2009; Mullard 2012), who suggest that the traditionally defined limits of druggable targets and productive chemical spaces are consistently being breached. They suggest that such limits can be self-reinforcing, and that in some cases such limits may represent the limits of certain techniques more than the targets themselves. Along similar lines, some recent research has suggested that screens for protein-protein interaction inhibitors should search a different region of space than that typically represented in today's chemical libraries (Sperandio, Reynes et al. 2010), though the discovery of 12 hits in the search for ligands of the RNaseL ankyrin repeat domain certainly suggests that such libraries are certainly not completely devoid of hits.

The reasons for the failure of the above docking and consensus scoring approach are less clear. The evidence presented does suggest that the *in silico* tools used would probably have performed better with improved treatment of ligand burial or desolvation energy. Desolvation energy is thought to be the least accurate term of the empirical scoring functions and others have remarked that improvements therein are critical to the further improvement of *in silico* drug discovery (Zoete, Grosdidier et al. 2010). It is likely that the performance of the consensus scoring approach could be improved by using only the best-performing scores identified in this work (at least in the context of this type of target), given that it has been shown that the performance of the input scores is essential to the success of consensus scoring (Yang, Chen et al. 2005). However, it is unlikely on its own to result in sufficient improvement to yield the desired result.

Another likely reason for the overall failure is the docking algorithm (search) itself. As noted previously, the DOCK algorithm used, and indeed the majority of available algorithms incorporate

little or no treatment of receptor flexibility. Some have argued that this is a more severe failure in the context of virtual screening for inhibitors of protein-protein interactions, on the basis that protein flexibility is more common and more important at protein-protein interaction interfaces than the active sites or ligand binding sites of traditional targets (Corradi, Mancini et al. 2010; Mullard 2012). Additionally, the flatter, more open target sites of protein-protein interaction sites present a more difficult search problem insofar as the conformational space of ligands is less restricted, meaning that a larger search must be conducted to ensure a reasonable likelihood of sampling the correct pose. Additional sampling also increases the likelihood of sampling false positive poses, which due to imperfections in scoring are harder to distinguish from the correct pose than is usually the case. Previous work from our group supports this notion. Studying 2-amino-quinoline, a small-molecule ligand of the Tec SH3 domain(Inglis, Stojkoski et al. 2004), it was firstly found that *in silico* docking techniques are unable to correctly rank-order derivatives of the lead according to affinity. Imposing an effective constraint on the docking search by selecting only the pose of each derivative that positioned the 2-amino-quinoline "core" in the correct orientation (as determined by NMR) dramatically improved the rank-ordering of derivatives (Stojkoski, C. and Booker, G.W. unpublished), demonstrating that the identification of the correct pose (ie the search) was the major impediment to the performance of *in silico* tools in that case.

It may in part be the importance of interface flexibility which is driving the successful application of pharmacophore and fragment-based approaches to the discovery of protein-protein interaction inhibitors (Corradi, Mancini et al. 2010; Voet and Zhang 2012). Pharmacophore screening tends to result in a "fuzzier" view of the target site than attempting to dock small molecules directly into a full-atom representation of the target site, thus capturing a degree of ligand flexibility. Similarly, fragment-based screening, by separating the steps of identifying chemical classes or epitopes that can bind a target and the subsequent optimization of their geometry, provides greater allowance for the discovery of induced fit rather than assuming that the geometry that docks best to a crystal structure will ultimately bind with the greatest affinity.

When combined with the fact that researchers have now identified that the hit rate of a chemical fragment screen (typically defined as compounds with a molecular weight below 150-200 Da) is a good indicator of druggability (Hajduk, Huth et al. 2005; Edfeldt, Folmer et al. 2011) and the importance of the question of druggability to efforts to inhibit protein-protein interactions, fragment-based approaches a particularly attractive.

Combining *in silico* methods with the rationale of a fragment-based approach adds further to the attractiveness of this approach. One of the major limitations of the fragment-based approach is chemical solubility (Konteatis 2010), which limits a fragment approach on two fronts. First, because researchers are looking for fragments with relatively low affinities for the target, they must be assayed at high concentration and so must be highly soluble. Secondly, by limiting the search to only very small chemical fragments, there is less opportunity to offset large hydrophobic regions of the molecule with other hydrophilic functionality that in the context of a larger molecule might make it sufficiently soluble to be assayed. *In silico* screening is not hampered by such physical properties. The reduction in the size of the library to be screened by virtual screening also frees computational resources to be directed at other aspects of the problem, such as better conformational searching (of ligand and/or receptor) or cluster analysis (see below), while also reducing the size of the search problem itself by the fact that fragments contain fewer rotatable bonds.

Of course in the context of the findings presented here, it may not seem wise to rely on virtual screening to effectively identify fragments that are likely to bind. However several factors suggest that a fragment virtual screen, employing the right techniques, could be more successful than what is presented here. As mentioned above, additional computational resources allow the deployment of better searching and additional techniques to identify the correct pose. Several studies have suggested clustering and filtering techniques as a step after docking to better implement a search for convergence of solutions (Bottegoni, Cavalli et al. 2006; Bottegoni, Rocchia

et al. 2012; Zeifman, Stroylov et al. 2012), since this is usually the strongest indicator of the correct pose. Employing similar techniques to the docking results of multiple search algorithms would represent a more rigorous version of that draws on the theory of consensus scoring to try to limit false positives.

Taking all this together, the present work and available literature strongly suggests that gankyrin remains a valid drug target, but that it is not an attractive problem to approach via traditional *in silico* methods. A combination of *in vitro* and *in silico* fragment screening via a relatively high-throughput primary screening technique such as thermal denaturation seem likely to provide the most attractive path forward. The use of a relatively low-resolution assay such as thermal denaturation is often not considered attractive because of its inability to detect whether ligands bind at the target site, a greater tendency that alternatives to produce false-positive results, and the superiority of the affinity data obtained from some other techniques. However, the fact that the G0-3 construct could once again be used for this assay means that virtually the only potential binding pocket on the protein is the target site, making the identification of "off-target" ligands very unlikely. The simplicity of the thermal denaturation assay and the small quantities of protein sample required make it ideal for use as a primary screen of a relatively large chemical collection when supported by a secondary assay technique providing greater information such as the NMR-based technique employed above. The use of an *in silico* screen alongside the *in vitro* screen would enable training of the *in silico* techniques based on those chemicals that are shared between the two screens, to then allow use of these techniques to probe those regions of space that are inaccessible to the *in vitro* screen. It may also prove informative to compare the results of these screens and the regions of chemical space that they identify as favourable with the methods that aim to tailor a chemical library to be more suited to its target (Orry, Abagyan et al. 2006; Sperandio, Reynes et al. 2010).

Chapter 6

Ankyrin repeat hydroxylation by FIH-1

6.1 Introduction

6.1.1 Factor Inhibiting HIF-1 (FIH-1)

HIF-1α and HIF-2α are transcriptional regulators that control the expression of many genes in response to low cellular oxygen levels (hypoxia). It follows, then, that the HIFs must have a mechanism for "sensing" the levels of oxygen being experienced by a given cell. One such mechanism is hydroxylation by FIH-1, a 2-oxoglutarate (2OG)-dependent asparaginyl hydroxylase that modifies residues in the C-terminal activation domains (CADs) of HIF-1α and HIF-2α. This oxygen-dependent modification prevents their interaction with p300/cAMP-response element-binding protein coactivators. Thus, only under hypoxic conditions where oxygen availability is limited and FIH-1 is therefore inactive can HIF-1 α and HIF-2 α successfully transcriptionally activate target genes as part of the hypoxic response.

6.1.2 Hydroxylation of ankyrin-repeat proteins

For some years after the discovery of FIH-1, HIF-1α and HIF-2α were its only two known substrates. More recently, however, it has been discovered that FIH-1 also hydroxylates a variety of ankyrin repeat proteins (Cockman, Webb et al. 2009; Cockman, Webb et al. 2009). The biological significance of ankryin repeat hydroxylation is still unclear. Studies into the effect of hydroxylation of Notch ankyrin repeat domains reveal no effect on Notch signaling, and only marginal structural effects(Coleman, McDonough et al. 2007). While ankyrin repeat proteins compete with the HIF-CAD for binding to FIH-1, the enzyme kinetics of ankyrin hydroxylation is different to that of the HIF-CAD modification(Cockman, Webb et al. 2009). This has led to the hypothesis that ankyrin repeat proteins act as decoys or sinks, reducing the availability of FIH-1 – dependent on the hydroxylation status of the ankyrin repeats – and thus providing an additional mechanism for modulating HIF-CAD hydroxylation by FIH-1 (Schmierer, Novak et al. 2010).

### *6.1.3 Specificity studies of the FIH/ankyrin interaction*

The prevalence of ankyrin-repeat substrates of FIH-1 in part reflects some similarity between

sequence elements known to be important for the HIF-CAD/FIH-1 interaction and commonly

occurring sequence elements of ankyrin repeats. In particular, the presence of a leucine residue at

the -8 position, alanine at -3, valine at -1 and a second leucine in the +10 position are highly

conserved among FIH-1 substrates. However, it has also been demonstrated that the presence of

these residues are insufficient to predict whether a given asparagine within an ankyrin repeat will

actually be hydroxylated. It is thought, then, that more subtle local sequence effects involving

other contacts with FIH-1 are likely to be further determinants of ankyrin repeat hydroxylation.

Additionally, the structural stability of a given ankyrin repeat is also thought to affect its

hydroxylation, given that ankyrin repeats appear to be hydroxylated in at least a partially unfolded

state (Coleman, McDonough et al. 2007; Cockman, Webb et al. 2009).

Collaborators in the laboratory of Dr Dan Peet set out to further investigate the specificity of

ankyrin repeat hydroxylation by FIH-1, and in particular to separate the effects of local sequence

and structural stability. The approach taken was to analyse the hydroxylation of the ankyrin repeat

domain of mouse Notch-1 (mNotch-1) and Notch-4 (mNotch-4). mNotch-1 is hydroxylated at two

sites, while mNotch-4 is not hydroxylated by FIH-1, despite having an appropriately placed

asparagine residue (N1656) with leucine residues at the -8 and +10 positions, in the equivalent

position to site 1 of mNotch1 (Wilkins, Hyvarinen et al. 2009). By mutating the sequence of

mNotch-4 towards that of mNotch-1 until the former is able to be hydroxylated, insights can be

gained into the mechanism/s of specificity. Assuming relatively few mutations are required, it is

likely that the effects observed will be predominantly due to substrate recognition, and

independent of tertiary structure or structural stability. This can also be further verified

experimentally via circular dichroism and thermal denaturation studies. The complementary study

of abrogating FIH-1 hydroxylation of mNotch-1 by mutating its sequence toward that of mNotch-4

was also carried out. To focus the analysis on site 1 in mNotch 1, the asparagine residue in site 2 was mutated to a glutamine (N2012Q) for these experiments.

The hydroxylation of the ankryin repeat domains (ARDs) of mNotch-1, mNotch-4 and their mutants was assessed using the previously published CO_2 capture assay (Linke, Stojkoski et al. 2004). Mutating only the residues adjacent to the hydroxylated asaparagine residue (ie -1 and +1 residues) proved sufficient to either enable or abrogate hydroxylation of mNotch-4 and mNotch-1, respectively, by FIH-1 (Figure 6.1 and summarized in Table 6.1).

The magnitude of the change in hydroxylation seems to suggest that changes to the sequence surrounding the hydroxylation site are strong determinants of whether the reaction proceeds. However, the results do not provide a clear picture of what the sequence determinants of recognition by FIH-1 are, specifically the contributions of individual positions and sequence changes. For instance, most of the data suggest that a proline residue in the -1 position prohibits hydroxylation of the adjacent asparagines. However, the mNotch-4 Q1657I mutant encodes a proline at the -1 position, yet is efficiently hydroxylated. Similarly, most of the data suggest that the identity of the residue at the +1 position is not a strong determinant of hydroxylation, yet the mNotch-4 Q1657I mutation modifies only the +1 residue, with the result of enabling hydroxylation whereas the wild-type mNotch-4 cannot be hydroxylated. Thus it is not only the nature of the amino acid at each position, but their context relative to other surrounding amino acids that determine the efficiency of substrate hydroxylation.
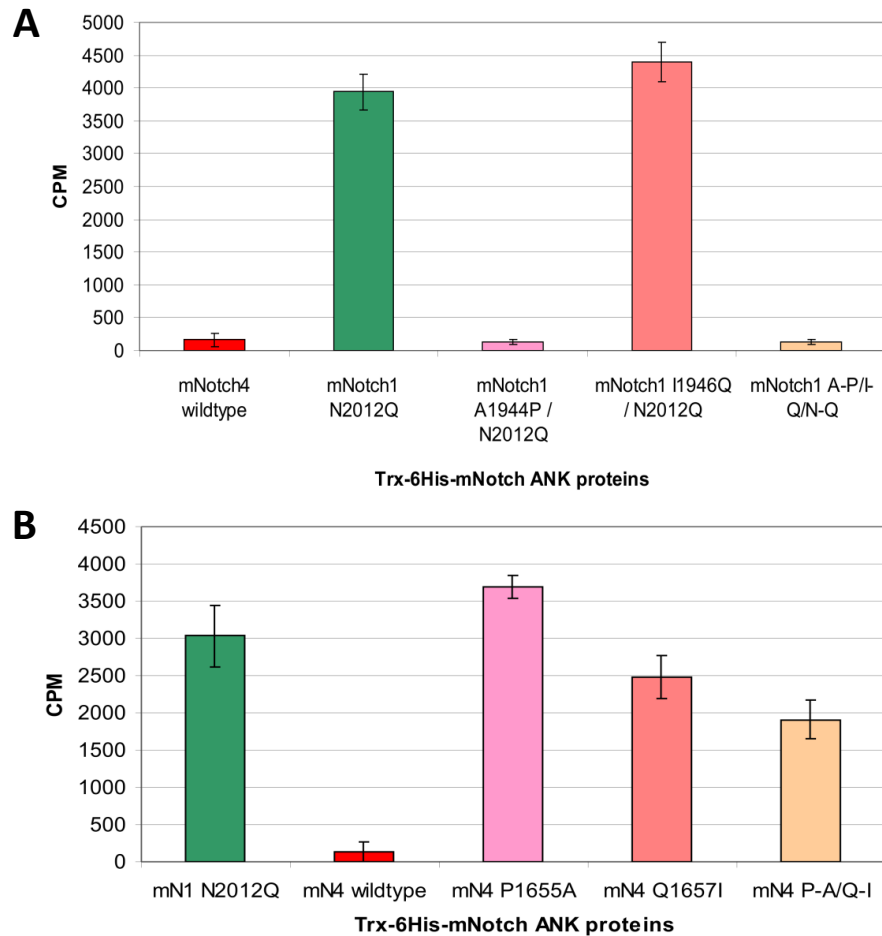
Figure 6.1 ARD hydroxylation of mNotch-1 derivatives **(A)** and mNotch-4 derivatives **(B)** determined by CO_2 capture assay (Linke, Hampton-Smith et al. 2007). Data are the mean of triplicate reactions +/- SD and are representative of >3 independent experiments. Figure and data from Sarah Wilkins.

Table 6.1 Summary of hydroxylation assays for mNotch-1 and mNotch-4 and derivative mutants, based on data provided by Sarah Wilkins (Figure 6.1). "Local sequence" gives the sequence that is the subject of mutation; the central asparagine is the substrate residue of FIH-1. Mutated residues are underlined. The table is arranged so that the central tri-peptide is identical for the two mutants on a given row.

Origin	Mutation	Local Sequence	Hydrox-ylated	Origin	Mutation	Local Sequence	Hydrox-ylated
mNotch-1	N2012Q	DANIQ	Yes	mNotch-4	P1655A/ Q1657I	NANIP	Yes
mNotch-1	A1944P/ N2012Q	DPNIQ	No	mNotch-4	Q1657I	NPNIP	Yes
mNotch-1	I1946Q/ N2012Q	DANQQ	Yes	mNotch-4	P1655A	NANQP	Yes
mNotch-1	A1944P/ I1946Q/ N2012Q	DPNQQ	No	mNotch-4	wt	NPNQP	No

6.2 Aims and Approaches

While it has been demonstrated that local sequence effects are the likely explanation for the observed pattern of hydroxylation, the primary sequence alone seems unable to provide a coherent explanation. An understanding of the molecular details of the interaction is still dependent on the primary sequence, but may provide greater insight into the precise phenomena at play. Therefore, in an effort to explain the pattern of selectivity of FIH-1 for the mNotch-1 and mNotch-4 derivatives, the first aim of this section of work was to develop structural models of these complexes and investigate molecular dynamics simulations of the complexes.

A second aim was to investigate the role of the structural stability of ankyrin repeats in determining whether they can be bound and hydroxylated by FIH-1. In addition to direct studies of the stability of ankyrin repeat FIH-1 substrates by my collaborators, the obligate interaction between gankyrin and the C-terminal domain of the S6 ATPase proteasomal regulatory subunit (S6C) presented a unique opportunity to alter the stability of an ankyrin fold without altering its primary sequence. Therefore, in an attempt to separate the effects of primary sequence and tertiary structure stability, we set out to compare separately both the binding and the hydroxylation of gankyrin in the presence and absence of its S6C binding partner.

6.3 Results

6.3.1 Modelling of mNotch-1 and mNotch-4 peptides

Experimentally derived crystal structures exist for the complexes of FIH-1 with either a peptide derived from the HIF C-terminal Activation Domain (HIF-CAD)(Elkins, Hewitson et al. 2003) or peptides derived from hydroxylation sites 1 (residues 1930-1949) or 2 (residues 1997-2016) from mNotch-1(Coleman, McDonough et al. 2007). The structure of the mNotch-1 site 1 peptide in complex with FIH-1 (kindly provided by the authors) provides co-ordinates only for residues 1937-1945, and so is insufficient for studying the impact of residues more C-terminal to Asn1945 on the recognition and hydroxylation of the peptide by FIH-1. However, the conformation of the mNotch-

1 peptide bears strong similarity to that in the HIF-1α/FIH-1 structure, exhibiting a backbone RMSD

of 0.547 Å across the resolved residues of mNotch-1 and the homologous residues of HIF-1α

(Figure 6.2). This similarity strongly suggests that hydroxylated ankyrin repeats adopt a very similar

conformation to that of the HIF-CAD for at least the hydroxylation step of the reaction. On this

basis, additional residues 1946-1951 of mNotch-1 site 1 were modeled to allow examination of the

role of all residues in the proximity of the hydroxylated Asn1945.
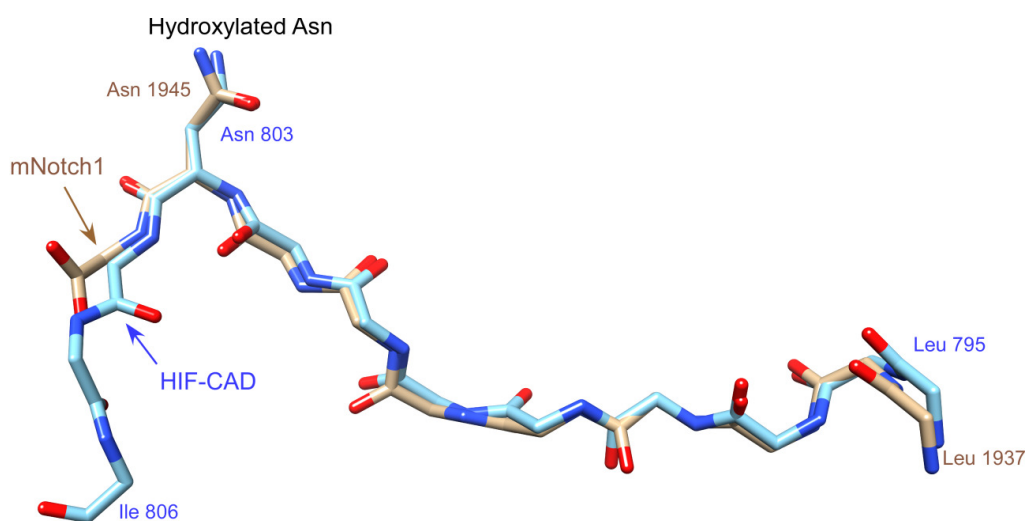
Figure 6.2 Comparison of the backbone conformations of HIF-CAD residues 796-806 (PDB ID 1H2K)(light blue) and
mNotch-1 residues 1937-1945 (tan) as determined in complex with FIH-1. Only backbone heavy atoms are shown. The
calculated RMSD for HIF-CAD residues 795-803 to mNotch-1 residues 1937-1945 is 0.547Å.

The backbone atoms of residues 1946-1948 were assumed to adopt the same conformation as

those of HIF-CAD residues 804-806 in PDB 1H2K, while the backbone conformation of Notch

residues 1949-1951 were based on those modeled for HIF residues 807-809 in earlier work (Linke,

Stojkoski et al. 2004).

Where crystallographic coordinates were unavailable, side-chain conformations were modeled

on the basis of the backbone conformation using Scwrl4 (Krivov, Shapovalov et al. 2009). Hydrogen

atom coordinates were calculated and peptide force field parameters prepared using XPLOR-NIH

v2.25(Schwieters, Kuszewski et al. 2003) and the CHARMM force field(MacKerell, Bashford et al.

1998). The complex was then subjected to 30,000 cycles of conjugate gradient minimization leaving all atoms fixed except modeled side-chain atoms.

In order to assess the accuracy of this side-chain modeling technique, crystallographic coordinates of mNotch-1 residues 1937-1945 (62 atoms) were compared to those derived via the above protocol, and various alternatives. That is, side-chain atoms were removed from the extracted mNotch-1 peptide structure, then rebuilt by prediction with Scwrl4 and subjected to energy minimization. For evaluation of the technique, all heavy-atom RMSDs were calculated between the crystallographic structure and the modeled peptide at various points throughout the protocol (Table 6.2).

Table 6.2 All heavy-atom RMSD comparison of mNotch-1 residues 1937-1945 at various points throughout side-chain modeling protocol, as well as with alternative methods, demonstrating the accuracy of the chosen technique.

RMSD pairing		RMSD (Å) (all heavy atoms)
Published crystal structure	Scwrl4 output	1.216
Published crystal structure	Scwrl4 output after energy minimisation	0.964
Published crystal structure after energy minimisation	Scwrl4 output after energy minimisation	0.888
Published crystal structure	Published crystal structure after energy minimisation	0.655
Alternative rotamer reconstruction methods:		
Published crystal structure	DockPrep output (Dunbrack rotamer library)	1.711
Published crystal structure	DockPrep output (Richardson common-atom library)	1.600
Published crystal structure	DockPrep output (Richardson mode library)	1.596

It can be clearly seen that Scwrl4 out-performs other similar and accessible methods, and it is self-evident that the resulting heavy-atom RMSD of 1.2 Å over 62 atoms is considerably better than could be expected for random rotamer assignment. Further, energy minimization of the modeled complex improves similarity of the modeled peptide to the experimental data, while minimization of the crystal structure yields greater similarity still. While this improvement comes at the cost of moving the structure further from that indicated by experimental data, it is an

important pre-cursor to any molecular dynamics simulation, as well as reducing the likelihood that differences seen during molecular dynamics simulations will be merely a consequence of differences between the starting structures of respective simulations. In any case, energy minimization of the crystal structure yields a RMSD of only 0.655 Å to its original structure. It is worth remembering, also, that the coordinates in the crystal structure to which we are comparing have uncertainty in their own right – its 2.4 Å resolution implies an uncertainty of approximately $\pm$ 0.4 Å in atomic coordinates (Fersht 1985). Taken together, the models produced via the above protocol can be used for molecular dynamics with a high degree of confidence that they are representative of the natural state(s) of the complex.

Using this protocol, models were constructed for four Notch peptides for which a puzzling pattern of hydroxylation has been determined by collaborators (Table 6.3).

Table 6.3 Peptides of interest were modeled based on the crystal structures of complexes of FIH-1 with HIF-CAD and mNotch-1 peptides in order to investigate structural effects that may explain observed hydroxylation patterns via molecular dynamics.

Origin	Mutation	Amino Acid Sequence	Hydroxylated	Short Name
mNotch-1	wt	RLLEASADANIQDNMG	Yes	mN1 DANIQ
mNotch-1	A1944P	RLLEASADPNIQDNMG	No	mN1 DPNIQ
mNotch-4	wt	RLLEAGANPNQPDRAG	No	mN4 NPNQP
mNotch-4	Q1657I	RLLEAGANPNIPDRAG	Yes	mN4 NPNIP

For the sake of convenience and clarity, the different peptides will be referred to by the above listed short names, which specify the Notch isoform from which the peptide originated, and the sequence of the penta-peptide centering on the hydroxylated peptide, since these comprise the region explored experimentally via mutagenesis by our collaborators.

6.3.2 Molecular dynamics of FIH/Notch peptide complexes

For molecular dynamics simulation of the complex, a similar approach was taken to that previously described (Linke, Stojkoski et al. 2004). Briefly, hydrogen atom coordinates were calculated and peptide force field parameters prepared using XPLOR-NIH v2.25 (Schwieters, Kuszewski et al. 2003) and the CHARMM force field (MacKerell, Bashford et al. 1998). Force field

parameters for the Fe^{2+} ion and 2-oxoglutarate were copied from previous work. Molecular

dynamics were carried out using NAMD v2.6 (Phillips, Braun et al. 2005). The structure was

solvated in a box such that non-water atoms were no closer than 10 Å to the edge of the box,

using SOLVATE v1.3 as employed in VMD v1.8.7b5 (Humphrey, Dalke et al. 1996). An 18-Å radius
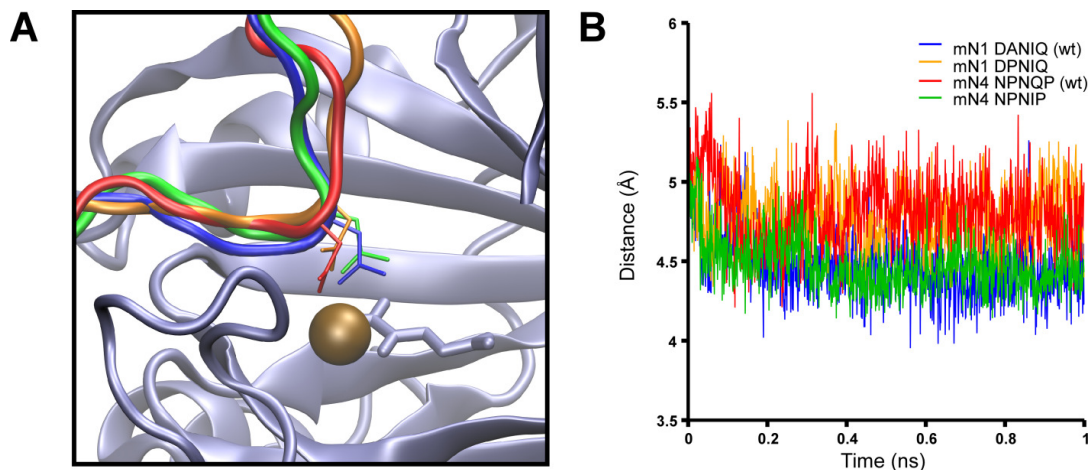
around the hydroxylated Asn residue was used in 1-ns MD simulations with a time step of 1

femtosecond (fs). Simulation was commenced at a temperature of 0K, followed by heating to 300K

in 25K increments of 1000fs duration. Figure 6.3 shows a representative plot of the total system

energy over the course of simulation, demonstrating that the system successfully equilibrated to

an essentially stable total energy after 0.05ns (50,000 steps). The time-scales over which

hydroxylation by FIH-1 occurs are unknown, however the aim of the simulation was not to attempt

simulation of the reaction itself, since MD is incapable of this in any case. Rather, the aim was to

analyse the stability of the peptide conformation that is required for hydroxylation while in

complex with FIH-1, as an indicator of the likelihood that it would ever adopt the necessary

conformation itself, allowing formation of the complex.
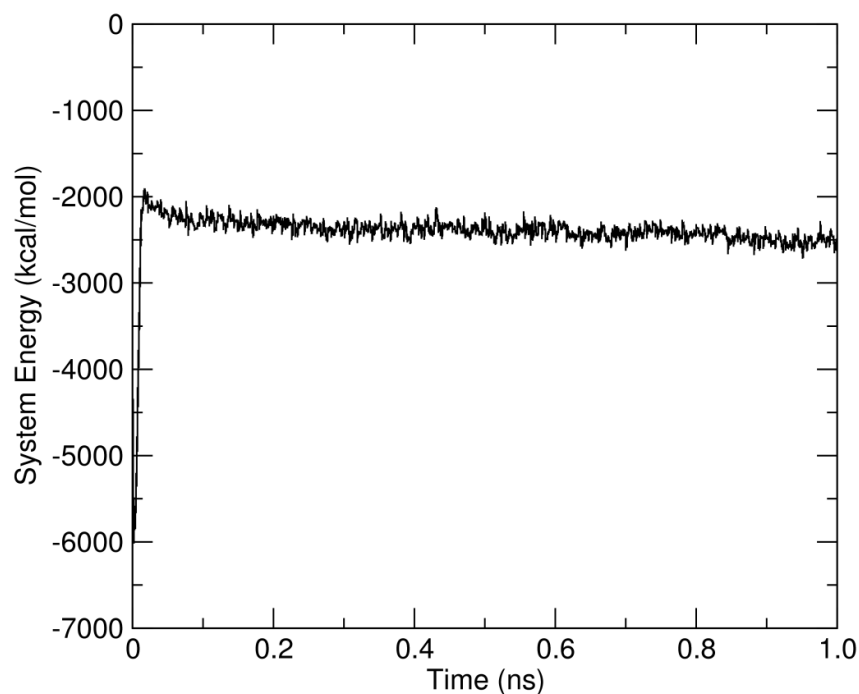
Figure 6.3 System energy plot for 1ns molecular dynamics simulation of FIH-1/mNotch-1 (wt) complex, showing

energy equilibration after 0.05ns (50,000 steps).

In visual comparison of the various simulations, the most striking difference was the conformation adopted by the peptide backbone. While the mN1 DANIQ and mN4 NPNIP peptides maintained a tight turn across residues -1 to +1, this was less evident for the mN1 DPNIQ mutant and mN4 NPNQP peptides (Figure 6.4A). The extent of this difference and its maintenance throughout the duration of the simulation is demonstrated by comparing the distance across the turn. Specifically, the distance between the C_α of the residue at the -1 position, and the backbone N of the +1 residue was measured every 1000 steps (fs) of the simulation and plotted against time (Figure 6.4B). It is clear from this analysis that despite starting from very similar conformations, the hydroxylated peptides (mN1 DANIQ and mN4 NPNIP) maintain a tighter and more stable turn with the distance remaining around 4.5Å for the duration of the simulation, in comparison to 4.8-5.0Å for the non-hydroxylated peptides.

Figure 6.4 (A) Schematic representation shows that the hydroxylated peptides mN1 DANIQ (shown in *blue*) and mN4 NPNIP (*green*) exhibit a tighter turn around the hydroxylated Asn residue, as compared to non-hydroxylated peptides mN1 DPNIQ (*orange*) and mN4 NPNQP (*red*). **(B)** This characteristic can be shown quantitatively throughout the 1ns duration of molecular dynamics simulation via the measured distance between the C_α of the residue at the -1 position, and the backbone N of the +1 residue. This research was originally published in JBC. Wilkins,S.E., S. Karttunen, et al. Factor inhibiting HIF (FIH) recognises distinct molecular features within hypoxia inducible factor (HIF)-alpha versus ankyrin repeat substrates. Journal of Biological Chemistry. 2012; 287:8769-81. © the American Society for Biochemistry and Molecular Biology. Used with permission.

The turn surrounding the hydroxylated Asn has been previously noted for its importance in allowing positioning of the Asn sidechain within the active site (Elkins, Hewitson et al. 2003). In simulations of the non-hydroxylated peptides, the Asn sidechain does escape its key position in the active site, while the wider path of the peptide also deforms the active site to a small extent. It must be noted, however, that our simulations begin with a pre-formed complex and that the true biochemical consequence of a more extended backbone conformation may simply be that the peptide cannot dock appropriately into the active site at all. The simulations also reveal that for the wild-type mNotch-1 peptide, the turn is maintained by hydrogen bonding between the amide hydrogen atoms of the -1 and +1 residues with a water molecule between them (Figure 6.5A). Introduction of a proline at the -1 position removes the requisite amide hydrogen, explaining the lack of hydroxylation of wild-type mNotch-4 (Figure 6.5C). One would expect the same argument to apply to the mN4 NPNIP peptide. However, the proline and isoleucine residues at -1 and +1, respectively, form a close hydrophobic interaction which performs a similar function to the aforementioned hydrogen bond (Figure 6.5D). Finally, while the mN1 DPNIQ mutant similarly has hydrophobic proline and isoleucine residues at -1 and +1, the deviation of the peptide caused by the proline residue at the -1 position promotes an interaction between mNotch-1 Asp1943 and FIH-1 Lys107. Combined with an interaction between mNotch-1 Gln1947 and FIH-1 Glu105 and Arg120, this restrains the whole peptide in its more extended form (Figure 6.5B).
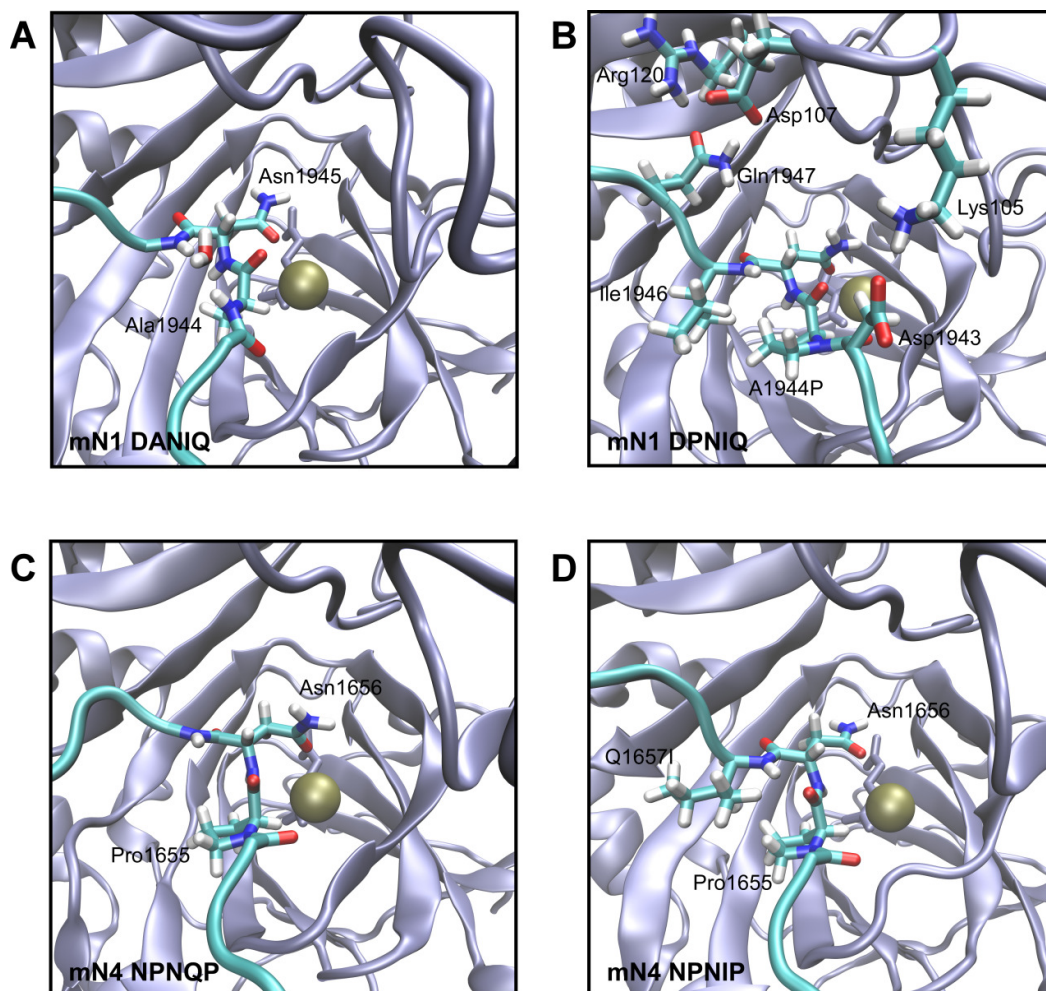
Figure 6.5 The hydroxylated Asn residue and Fe^{2+} ion (tan) are centred in each field of view. Notch peptides are shown in ribbon (cyan) with atoms/side-chains of interest shown in full, while FIH-1 is shown in cartoon format (grey). **(A)** *mN1 DANIQ* A water-bridged hydrogen bond between the backbone N-H in the -1 and +1 positions maintains the turn conformation around Asn 1945. **(B)** *mN1 DPNIQ* and **(C)** *mN4 NPNQP* A proline residue in the -1 position prevents this hydrogen bonding pattern, and the peptide adopts a more extended conformation. **(D)** *mN4 NPNIP* The unrestrained peptide forms a hydrophobic interaction between Pro and Ile residues at -1 and +1, compensating for the lack of hydrogen bonding. This research was originally published in JBC. Wilkins,S.E., S. Karttunen, et al. Factor inhibiting HIF (FIH) recognises distinct molecular features within hypoxia inducible factor (HIF)-alpha versus ankyrin repeat substrates. Journal of Biological Chemistry. 2012; 287:8769-81. © the American Society for Biochemistry and Molecular Biology. Used with permission.

6.3.3 Binding and hydroxylation of gankyrin/S6C complex by FIH-1

A DNA fragment encoding the C-terminal domain of the S6 ATPase proteasomal regulatory subunit (S6C) was generated by PCR from a HepG2 cDNA preparation using oligonucleotides S6-D337-5' and S6-K418-3' as per the method described in Section 2.3.3. This DNA fragment was sub-cloned into the pET32a-G0-6 vector using the *Bam*HI and *EcoRI* restriction sites. The sequence and location of the DNA insert was verified by BigDye sequencing according to the method described in Section 2.3.11 using the T7_Term primer. The resulting plasmid, pET32a-S6C/G0-6, which encodes a Trx-6His-S6C fusion protein and the G0-6 open reading frame dicistronically, was transformed into BL21 cells for protein expression.

Collaborators Sarah Wilkins, Sarah Karttunen and Rachel Hampton-Smith expressed and purified the the S6C-gankyrin complex by Ni^{2+}-affinity column chromatography and analysed its ability to bind FIH-1 relative to Trx-6His-G0-6 alone by pull-down experiments. It should be noted that if gankyrin separates from its complex with S6C in order to bind FIH-1 it will not be detected in this experiment, since FIH-1-gankyrin complexes are only pulled down indirectly via the 6His tag of S6C. Figure 6.6 shows that FIH-1 can bind to gankyrin equally well in the presence or absence of S6C. Unfortunately, the negative control showing that FIH-1 does not bind directly to S6C itself could not be carried out, because S6C (as well as full-length S6) is insoluble (Nakamura, Umehara et al. 2007)(S6C insolubility was confirmed by expression, cell lysis, fractionation and SDS-PAGE analysis, data not shown).
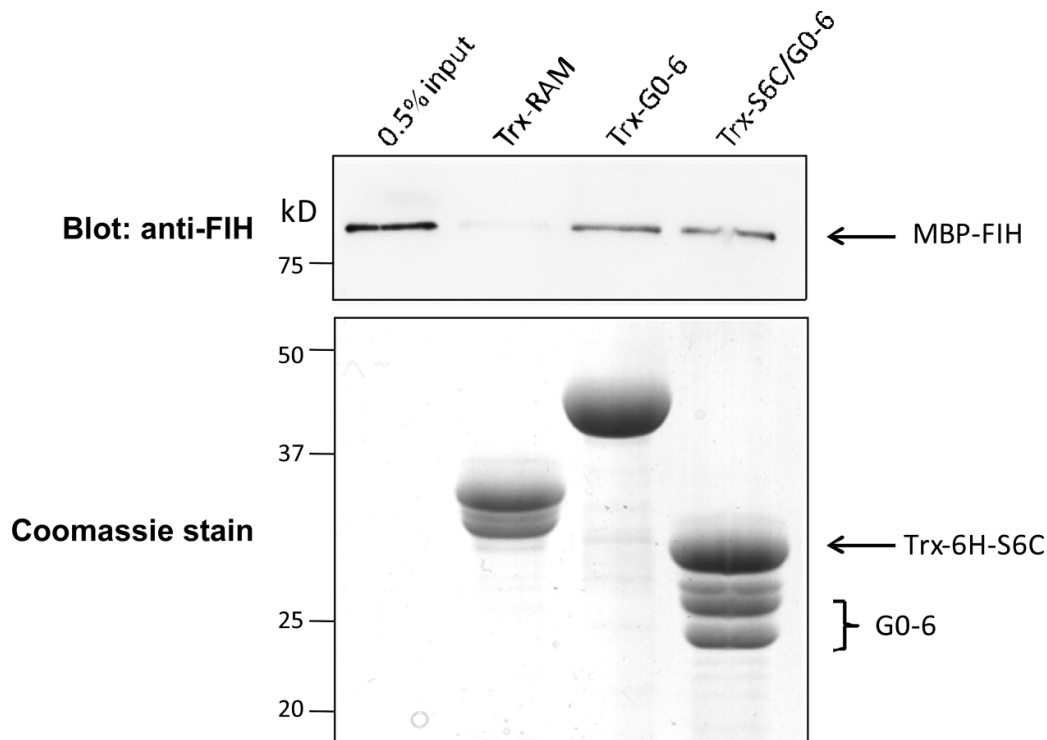
Figure 6.6 Pull-down experiments testing the binding of MBP-FIH to gankyrin alone (Trx-G0-6) or the gankyrin-S6 C-terminal domain complex (Trx-S6C/G0-6). Trx-RAM is a thioredoxin-6His fusion protein of the RAM domain of Notch, which has previously been shown not to interact with or be hydroxylated by FIH-1 (Zheng, Linke et al. 2008). Figure and data from Sarah Wilkins, Sarah Karttunen and Rachel Hampton-Smith.

The banding pattern of co-purified Trx-6His-S6C and G0-6 when analysed by SDS-PAGE (Figure 6.6, right-most lane) is not as expected, exhibiting an extra band migrating at a molecular weight between that of gankyrin alone and Trx-6His-S6C. Analysis of the two lower bands by mass spectrometry after tryptic digest revealed only peptides derived from gankyrin were present in both bands. No differences in the presence or absence of peptides from near the N- and C-termini of gankyrin were observed, but nor was full sequence coverage achieved. Thus there is no evidence that the lower band represents a degradation product, but nor is there conclusive evidence against this hypothesis. An alternative hypothesis was that the band migrating at the lower molecular weight was the result of an alternative translation initiation site. A GST-fusion cassette was inserted at the N-terminus of G0-6 in order to alter the availability of any

alternative initiation sites, however the resulting Trx-S6C/GST-G0-6 complex still yielded two gankyrin-derived bands when analysed by SDS-PAGE (data not shown), providing evidence against this hypothesis and any N-terminal degradation. It appears likely that the banding pattern observed is simply an artifact of this particular complex.

Also examined by these collaborators was the *in vitro* hydroxylation of Trx-6His-G0-6 and the Trx-6His-S6C/G0-6 complex by FIH-1. Figure 6.7 shows that Trx-6His-G0-6, but not the Trx-6His-S6C/G0-6 complex can undergo hydroxylation by FIH-1. Because the S6C-gankyrin complex has been demonstrated to retain the ability to bind to FIH-1, it can be concluded that inhibition of the hydroxylation reaction by S6C cannot be simply a result of obscuring the interaction interface and inhibiting the interaction between gankyrin and FIH-1. Thus we can conclude that hydroxylation of ankyrin repeats is not simply a function of primary sequence, nor of the ability of FIH-1 to bind its substrate. Combined with existing structural data showing that ankyrin repeats unfold to undergo hydroxylation, these data point strongly toward a model in which S6C prevents unfolding of gankyrin, which in turn prevents the hydroxylation of gankyrin by FIH-1. Because of the insolubility of S6C alone and the resulting inability to conduct a control experiment to test whether S6C affects the hydroxylation of other ankyrin repeat proteins, we cannot absolutely rule out the possibility that S6C inhibits hydroxylation via a different mechanism, such as allosteric inhibition of FIH-1 or sequestration of one of the co-factors required for the reaction. Nonetheless, these must be considered less likely explanations, especially in the context of preliminary data (not shown) that hydroxylation of mNotch1 is no less efficient in the presence of the Trx-6His-S6C/G0-6 complex than its absence.
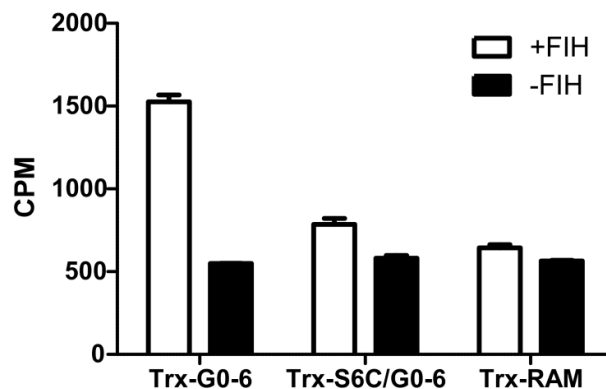
Figure 6.7 Hydroxylation of gankyrin alone (Trx-G0-6) or the gankyrin-S6 C-terminal domain complex (Trx-S6C/G0-6) as detected indirectly by CO_2 capture assay (Linke, Hampton-Smith et al. 2007). Trx-RAM is a thioredoxin-6His fusion protein of the RAM domain of Notch, which has previously been shown not to interact with or be hydroxylated by FIH-1 (Zheng, Linke et al. 2008). Data are the mean of triplicate reactions +/- SD and are representative of >3 independent experiments. Data from Sarah Wilkins, Sarah Karttunen and Rachel Hampton-Smith.

6.4 Discussion

Alongside the work presented here, our collaborators also analyzed the structural stability of the Notch proteins by thermal denaturation (Figure 6.8). These data show that an alternative or complementary explanation for the observed pattern of hydroxylation of mNotch1, mNotch4 and mutants thereof was that mutations which increase the structural stability of these proteins resulted in a reduction in their hydroxylation by FIH-1, while mutations which destabilize the proteins promote hydroxylation (Wilkins, Karttunen et al. 2012). However, it was only the *change* in protein stability of a mutant relative to its wild-type that could explain the pattern of hydroxylation; there was no threshold T_m value at which hydroxylation became allowed/disallowed. Specifically, mNotch1 DPNIQ is not hydroxylated, yet has a lower apparent T_m (ie is less stable) than mNotch4 NPNIP, which is hydroxylated. However, mNotch1 DPNIQ exhibited a 5°C increase in T_m relative to wild-type mNotch1 (DANIQ), consistent with the abrogation of hydroxylation by this mutation. Similarly, the mNotch4 NPNIP mutation promotes hydroxylation while decreasing the apparent T_m relative to the mNotch4 wild-type (NPNQP).
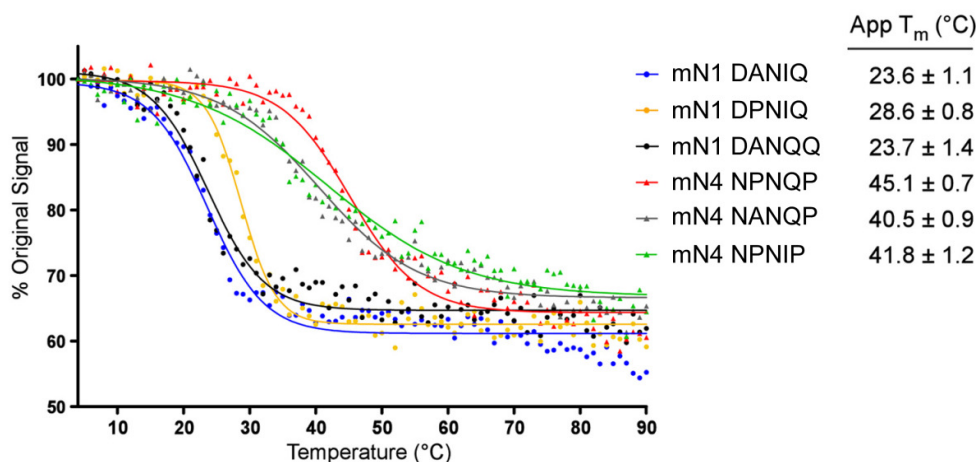
Figure 6.8 CD spectroscopy was employed to analyze the thermal denaturation of Trx-6H-tagged Notch ankyrin repeat proteins. The ellipticity at 220 nm (θ_{220}) was monitored continuously as the temperature increased from 4 to 90 °C. Data are expressed as a percentage of the θ_{220} value at 4 °C, and apparent T_m values were determined using Graphpad PRISM software. A representative denaturation curve is shown for each protein, and apparent T_m values are the average of three independent experiments ± S.D. This research was originally published in JBC. Wilkins,S.E., S. Karttunen, et al. Factor inhibiting HIF (FIH) recognises distinct molecular features within hypoxia inducible factor (HIF)-alpha versus ankyrin repeat substrates. Journal of Biological Chemistry. 2012; 287:8769-81. © the American Society for Biochemistry and Molecular Biology. Used with permission.

It is possible that the stability of individual ankyrin repeats, obscured by the measurement of global protein stability here, is the principal determinant of ankryin repeat hydroxylation, or it could be that the sequence effects described in Section 6.3.2 have the greater effect. However at present we have insufficient data to discern the relative contributions of these two explanations to observed changes in hydroxylation. An analysis of the *in vitro* hydroxylation of peptides derived from the Notch proteins and mutants described here would make a valuable contribution to this effort. Supporting information could be gathered by the inverse experiment, of analyzing the hydroxylation of Notch mutants that have been mutated at sites more distant from the hydroxylation site, that have been mutated towards the ankyrin consensus sequence (see Section 1.1.1) in an attempt to increase their stability without altering their interaction with FIH-1.

The inability of FIH-1 to hydroxylate gankryin that is in complex with S6C, however, strongly suggests that while these two phenomena may contribute to varying extents in some cases at the margin, there are equally situations in which stabilization of an ankyrin fold alone can cause dramatic changes to its hydroxylation. In the broader context, this is the first experimental

evidence that has been obtained in support of the hypothesis that gankyrin cannot partially unfold while in complex with S6C.

As mentioned earlier, two hypotheses about the function of ankyrin hydroxylation by FIH-1 are favoured. First is the hypothesis that hydroxylation of ankyrin repeats gives them additional structural stability, while the second proposes that binding of FIH-1 to ankyrin repeat proteins contributes to the regulation of FIH-1 activity. The work presented here did not aim to provide any direct contribution to resolving this question. Nonetheless, it must be said that the data presented are more easily reconciled with the second hypothesis than the first. The phenomenon of FIH-1 binding to a folded, naturally-occurring ankyrin repeat protein without hydroxylating it could conceivably play a role in its regulation. Whereas if the importance of hydroxylation of ankyrin repeats is its effect on the substrate, one would expect that the evolution of FIH-1 substrate recognition would favour binding only to the unfolded form to avoid large amounts of FIH-1 being sequestered into non-functional complexes with stable ankyrin repeats. Of course binding to folded ankyrin repeats in this scenario would make sense if FIH-1 or a closely associated factor had the ability to actively unfold its substrates (as opposed to relying on spontaneous unfolding), but at present there is no evidence of such a phenomenon. It must also be said that the notion of using hydroxylation to stabilize ankyrin repeats seems a rather complex way of solving a relatively simple problem, given that the evidence to date suggests that the majority of naturally-occurring ankyrin repeat proteins could be made more stable with just a few changes to its sequence.

Chapter 7

Final Discussion

7.1 Gankyrin as a drug target

Unsurprisingly, significant developments have occurred in the field of structure-based drug discovery since the commencement of this work in 2005. As such, there are a variety of advances which would provide valuable inputs into any attempt to discover inhibitors for gankyrin, but were unable to be deployed in this study due to the timing of these advances.

In particular, two pieces of work (Cheng, Coleman et al. 2007; Halgren 2009) have developed methods for directly predicting the druggability of a target based on structural properties of the protein, before investment in any discovery effort. Such tools, now that at least the work of Halgren (2009) is available commercially via the SiteMap product from Schrödinger, should be used not only to try to get a better sense of the feasibility of designing small-molecule inhibitors of gankyrin, but would also more rigorously define the most attractive sites on the surface of gankyrin.

Another key development in the field has been an increased appreciation of the importance of monitoring and tailoring the chemical diversity of screening libraries, and along with that has come greater availability of tools to do so. As mentioned previously, some recent work claims that chemical diversity can be tailored towards inhibition of protein-protein interactions (Sperandio, Reynes et al. 2010), while combining the computed properties of a site (for example using SiteMap) with those of a lead compound or fragment library presents the possibility of tailoring a compound library to the specific target of choice.

Despite the failure of this work to meet the aim of discovering small molecule ligands of gankyrin, it nonetheless has made some important contributions to our understanding of the challenges inherent in this problem and other *in silico* attempts to develop protein-protein interaction inhibitors, which was a secondary aim of the exercise. Via a different mechanism than

those discussed above, it was identified that the selected site on the surface of gankyrin can support the binding of ligands with levels of solvent inaccessibility equivalent or greater than those exhibited by ligands with affinities in the range of 20-100µM, suggesting that the shape of the molecular surface alone should not be a barrier to discovery of ligands with affinities in at least the micromolar range.

More importantly, however, this study identified that a conventional *in silico* docking and scoring approach is not sufficiently effective to be attractive against the targets examined, and in the case of some scoring algorithms is almost totally ineffective. Two different indicators – the characteristics of effective versus ineffective scoring algorithms, and the effect of different pose selection methods – suggest that the treatment of ligand solvation/desolvation requires improvement to address this. This finding is in keeping with the broader observation that the importance of solvent effects is the most clearly identifiable difference between targeting protein-protein interaction surfaces and the types of sites that have been used to develop and validate these scoring algorithms.

Recent work has also highlighted the role that convergence of solutions can play in identifying the correct binding mode rather than relying on scores alone, and the improvements that this can bring to the overall performance of virtual screening (Bottegoni, Cavalli et al. 2006; Bottegoni, Rocchia et al. 2012; Zeifman, Stroylov et al. 2012). The inevitable increase in the accessibility of computational power in recent years, particularly in concert with a better filtered chemical library, makes the application of such techniques relatively easily achievable. Given the greater conformational space to be searched in the shallower target sites of protein-protein interaction sites, there is reason to suspect that such analysis may be of particular use in cases like gankyrin.

Combining these advances with the advantages of parallel *in vitro* and *in silico* fragment-based screening as discussed in Section 5.4 would, based on both the advances in the field and the

findings of this work, present the most productive way forward for the discovery of small molecule inhibitors of gankyrin.

7.2 Molecular functions of gankyrin

Several distinct mechanisms for the action of gankyrin have been identified and described in the literature, each relying on its interaction with a particular cellular factor. However the relationship between these intermolecular interactions and thus the various oncogenic mechanisms remain largely unknown. Given that a low level of gankyrin appears not to stimulate cellular proliferation, and that its oncogenic effects are only observed at elevated levels, it is likely that cooperativity or competition between some of these interactions lead to ultrasensitivity in the cell's response to gankyrin concentrations.

Unfortunately, attempts to meet the aim of developing a quantitative *in vitro* gankyrin interaction assay that may have been able to shed light on at least a subset of these interactions were not successful. The data presented here highlight a number of challenges that not only impede our ability to undertake quantitative analysis of gankyrin's various interactions, but are also indicative of the broader challenges faced by biochemists trying to map molecular interaction networks that are dominated by protein-protein interactions. Specifically, the detection of low affinity interactions between two folded domains presents challenges in terms of limits of detection, the life-time of the bound complex, and mechanisms and efficiencies of labeling or detection that are typically less acute for higher interaction assays, or assays involving peptides, DNA or small molecules that can be more readily and specifically customized via synthetic techniques. An additional challenge in this case (and potentially other low affinity protein-protein interactions) is an apparent tendency for gankyrin to interact non-specifically, adding reduced signal-to-noise ratio to the challenges of developing an assay. Use of surface plasmon resonance (SPR) technology may well allow this aim to be achieved, by leveraging several advantages. First, it has a demonstrated ability to detect interactions with K_d values in the range of

0.1-1mM (Zhang and Oglesbee 2003), and while it is most often used for direct-binding experiments, competition assays based on the technique have been reported (Jarsch, Brandt et al. 2008; Zhang, Beaudet et al. 2012). Crucially, while non-specific binding may still be observed with this technique, the fact that it measures binding kinetics allows the detection and (usually) de-convolution of distinct binding modes much more directly and reliably than assay systems based on the detection of binding equilibrium.

The aim of defining a structural model for the interaction (Chapter 3) between gankyrin and pRb was a complementary way to gain insights into the relationship between gankyrin's various intermolecular interactions, while also contributing to the pool of knowledge available to be exploited to treat hepatocellular carcinoma and to our understanding of the pRb system in general. Again, obtaining structural information for low affinity protein-protein interactions remains a challenging area of the field, given their relative unsuitability for protein crystallization due to the instability of the complex and the difficulty of using NMR to probe large molecular weight complexes. Small angle X-ray scattering (SAXS) has significant potential to contribute to this area, but still poses challenges in either isolating a homogeneous preparation of the complex or de-convolution of data arising from multiple species. In this context, computational techniques provide a valuable addition to the toolkit in trying to derive structural models from low-resolution, ambiguous data such as mutagenesis data. The work presented here (Chapter 4) demonstrates that neither residues E186 nor K149 of gankyrin are essential to the interaction with pRb, nor residues D718, K720 or K765 of pRb. While showing that the particular model derived in this work is not accurate, the evidence obtained here will also provide additional constraints to future efforts in this area.

Additionally, the inability of the K765E mutant of pRb to perturb the interaction with gankyrin provides further evidence that the interaction between gankyrin and pRb occurs by a different mechanism than other LXCXE-containing proteins. K765 is observed to bind the backbone of the

E7 LXCXE peptide in the crystal structure (Lee, Russo et al. 1998) as well as exhibiting chemical shifts when E7 peptide is titrated into a ^{15}N lysine labeled NMR sample (Singh 2006). Thus the K765E mutation could be expected to reduce binding of LXCXE peptides, though this conclusion is admittedly weak given that the control experiment of analyzing E7 binding to a K765 mutant has not been reported.

The stronger evidence obtained in this work for a qualitatively different mode of interaction between gankyrin and pRb is that binding of the S6C to gankyrin is sufficient to totally prevent hydroxylation of gankyrin (Chapter 6). When combined with the structural evidence that ankyrin repeats undergo hydroxylation in an unfolded form (Coleman, McDonough et al. 2007), this constitutes the first experimental evidence to support the widely-held belief that the S6C-gankyrin interaction would inhibit gankyrin unfolding. This, along with the fact that the S6C-gankyrin interaction does not interfere with the gankyrin-pRb interaction (Nakamura, Nakano et al. 2007), provides the clearest evidence yet that gankyrin interacts with pRb in its folded form.

It is notable that much of the evidence in the literature for gankyrin's interactions comes from co-immunoprecipitation experiments, or GST pull-downs from cellular lysates. When combined with the difficulty implementing a quantitative binding assay reported here and the inability to detect the weak gankyrin-pRb interaction using purified proteins *in vitro* by others leads one to speculate that there is a material difference in the interaction between the different assay types. There are several possible reasons this could be true. Firstly, it may be due to cooperative interactions with other factors increasing the apparent affinity of the interaction in a whole-cell context relative to the *in vitro* context. Second, it could be due to post-translational modifications of gankyrin when expressed in a mammalian cell line that are absent when expressed and purified recombinantly for bacterial cells. Or thirdly, it could be due to differences in gankyrin's folding equilibrium in the cellular versus aqueous solution contexts.

A limited number of post-translational modifications of gankyrin have been observed and reported in the literature. Namely, N-terminal acetylation (Wang, Chen et al. 2007), phosphorylation of residue T207 (Olsen, Vermeulen et al. 2010), and ubiquitylation at residues K23, K30 and K90 (Kim, Bennett et al. 2011; Wagner, Beli et al. 2011), however neither the biological effects of these modifications nor their abundance is clear.

Resolving the problems with the detection of some of gankyrin's interactions, and the possibilities which that provides to explore competition and cooperativity between these interactions and to probe their structural details is clearly necessary and would have been the principal focus of further work, had time constraints permitted. Aside from the use of surface plasmon resonance for the establishment of a viable competition assay as discussed above, I propose that two other techniques should be pursued in parallel.

The use of a mammalian two-hybrid system to conduct a random mutagenesis screen of gankyrin and, initially, pRb (but potentially other binding partners) offers a way of probing the interaction in its natural environment. If a strong, detectable interaction is dependent on the cellular environment, cooperativity with other interactions, or post-translational modification, these will all be present in this experimental system. Such a system will probably bring with it a number of challenges beyond the technical. Interpretation of the resulting mutation data would be complicated by a wide variety of factors, from alterations to the expression and stability of the proteins, to disruption of cooperative interactions, to the effects of competition between bait/prey proteins and their endogenous counterparts. Even true positive identification of residues at the interface may exert their effect either through changing the interface directly, or through altering the state of post-translational modification. However, in combination with an *in vitro* system for analyzing the interaction such as SPR, it provides a valuable source of lead mutations for more thorough analysis. A broader set of biological data would also be invaluable in guiding and interpreting further efforts at *in silico* modeling of the interactions.

A complementary, more technically challenging, but less fraught approach is a mass spectrometry-based analysis of gankyrin and the factors that are isolated with it when immunoprecipitated from mammalian cell culture could shed light on both the prevalence/role of post-translationally modified gankyrin and potentially also the cooperativity and competition between some of its binding partners if sequential immunoprecipitations of gankyrin, then its partners was conducted. In this way one could detect not only the total pool of cellular proteins associated with gankyrin, but also whether it is present in sub-populations in which particular pairings of associated factors are favoured or disfavoured.

The question of whether the folding of gankyrin, and ankyrin proteins more broadly, differs significantly *in vivo* versus *in vitro* remains a fairly fundamental question to the study of this class of protein and its intermolecular interactions. The work presented in this thesis in combination with the observation that hydroxylation of ankyrin repeat proteins is a widespread phenomenon signals the potential for the development of new techniques for indirectly probing the foldedness of ankyrin repeat proteins *in vivo* – an ordinarily very challenging problem. If, upon further examination, the stability of ankyrin repeat proteins is shown to be the dominant effect in the *in vitro* hydroxylation of ankyrin repeat proteins, then this fact could potentially be harnessed to probe whether mutations that affect an ankyrin repeat protein's *in vitro* folding equilibrium result in corresponding changes to the extent of its hydroxylation *in vivo* (technically in cell culture) when co-expressed with FIH-1. For instance, if a particular mutation of a given ankyrin repeat protein has a lower T_m *in vitro* but exhibits no increase in the extent of its hydroxylation *in vivo* (or vice-versa), that would signify that for that protein, *in vitro* T_m measurements are probably not indicative of its foldedness *in vivo*. That could be due to a variety of factors, the most likely of which is the ankyrin repeat protein's association with a binding partner *in vivo*. Nonetheless, such a method can provide additional insight that has thus far been unavailable, into the mechanisms by which a mutation exerts an effect, and provide tools to help researchers determine the source of any observed disconnect between ankyrin repeat stability *in vitro* and *in vivo*.

7.3 Conclusion

Quantitative and structural studies of gankyrin's intermolecular interactions remain an important way for us to understand the functions of the protein itself, and to target it for the development of novel therapies for hepatocellular carcinoma. The work presented here has contributed to these broad aims through the proposal and testing of a model of the gankyrin-pRb interaction and an investigation into the effects of the hydroxylation of gankyrin. Together these investigations provide strong evidence that gankyrin does not unfold in order to interact with pRb. Therefore gankyrin interacts with pRb in a different way than with LXCXE-containing peptides, in a manner yet to be elucidated. An *in silico* screen for inhibitors of the gankyrin-CDK4 interaction did not yield any lead compounds, but provided evidence that the use of a straight-forward docking and scoring (or consensus scoring) approach is unlikely to be successful in finding small molecule ligands for ankyrin repeat proteins, and that virtual screening for inhibitors of protein-protein interactions is more sensitive to errors in the way docking and scoring algorithms treat solvent effects. Finally, the evidence presented suggests that gankyrin remains a valid, if somewhat challenging, drug target.

References

Arkin, M. R. and J. A. Wells (2004). "Small-molecule inhibitors of protein-protein interactions: progressing towards the dream." Nat Rev Drug Discov **3**(4): 301-317.

Baldin, V., J. Lukas, et al. (1993). "Cyclin D1 is a nuclear protein required for cell cycle progression in G1." Genes Dev **7**(5): 812-821.

Bammidi, L. S., G. N. Neerukonda, et al. (2012). "p16 gene alterations in human ovarian cancers: comparison between tissue and blood samples." Int J Gynecol Cancer **22**(4): 553-560.

Barrick, D., D. U. Ferreiro, et al. (2008). "Folding landscapes of ankyrin repeat proteins: experiments meet theory." Curr Opin Struct Biol **18**(1): 27-34.

Basseres, D. S. and A. S. Baldwin (2006). "Nuclear factor-kappaB and inhibitor of kappaB kinase pathways in oncogenic initiation and progression." Oncogene **25**(51): 6817-6830.

Betzi, S., F. Guerlesquin, et al. (2009). "Protein-protein interaction inhibition (2P2I): fewer and fewer undruggable targets." Comb Chem High Throughput Screen **12**(10): 968-983.

Binz, H. K., M. T. Stumpp, et al. (2003). "Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins." J Mol Biol **332**(2): 489-503.

Bipatnath, M., P. P. Dennis, et al. (1998). "Initiation and velocity of chromosome replication in Escherichia coli B/r and K-12." J Bacteriol **180**(2): 265-273.

Bjorklund, A. K., D. Ekman, et al. (2006). "Expansion of protein domain repeats." PLoS Comput Biol **2**(8): e114.

Bogan, A. A. and K. S. Thorn (1998). "Anatomy of hot spots in protein interfaces." J Mol Biol **280**(1): 1-9.

Bottegoni, G., A. Cavalli, et al. (2006). "A comparative study on the application of hierarchical-agglomerative clustering approaches to organize outputs of reiterated docking runs." J Chem Inf Model **46**(2): 852-862.

Bottegoni, G., W. Rocchia, et al. (2012). "Application of conformational clustering in protein-ligand docking." Methods Mol Biol **819**: 169-186.

Branson, K. (2005). Development and application of computer aided drug design methods Thesis (Ph.D.), University of Melbourne.

Brehm, A., E. A. Miska, et al. (1998). "Retinoblastoma protein recruits histone deacetylase to repress transcription." Nature **391**(6667): 597-601.

Buchler, N. E. and M. Louis (2008). "Molecular titration and ultrasensitivity in regulatory networks." J Mol Biol **384**(5): 1106-1119.

Burke, J. R., G. L. Hura, et al. (2012). "Structures of inactive retinoblastoma protein reveal multiple mechanisms for cell cycle control." Genes Dev **26**(11): 1156-1166.

Burstein, E. and C. S. Duckett (2003). "Dying for NF-kappaB? Control of cell death by transcriptional regulation of the apoptotic machinery." Curr Opin Cell Biol **15**(6): 732-737.

Byeon, I. J., J. Li, et al. (1998). "Tumor suppressor p16INK4A: determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4." Mol Cell **1**(3): 421-431.

Cammett, T. J., L. Luo, et al. (2003). "Design and characterization of a hyperstable p16INK4a that restores Cdk4 binding activity when combined with oncogenic mutations." J Mol Biol **327**(1): 285-297.

Chen, H.-H. (2008). Ca^{2+} and Phosphoinositides Regulations in α-actinin-4 F-actin Binding PhD Thesis, University of Adelaide.

Chen, J., R. Peters, et al. (1996). "A 39 amino acid fragment of the cell cycle regulator p21 is sufficient to bind PCNA and partially inhibit DNA replication in vivo." Nucleic Acids Res **24**(9): 1727-1733.

Chen, R., L. Li, et al. (2003). "ZDOCK: an initial-stage protein-docking algorithm." Proteins **52**(1): 80-87.

Chen, Y., H. H. Li, et al. (2007). "Oncoprotein p28 GANK binds to RelA and retains NF-kappaB in the cytoplasm through nuclear export." Cell Res **17**(12): 1020-1029.

Cheng, A. C., R. G. Coleman, et al. (2007). "Structure-based maximal affinity model predicts small-molecule druggability." Nat Biotechnol **25**(1): 71-75.

Cheng, M., P. Olivier, et al. (1999). "The p21(Cip1) and p27(Kip1) CDK 'inhibitors' are essential activators of cyclin D-dependent kinases in murine fibroblasts." EMBO J **18**(6): 1571-1583.

Cheng, X., W. Xia, et al. (2010). "Activation of p21(CIP1/WAF1) in mammary epithelium accelerates mammary tumorigenesis and promotes lung metastasis." Biochem Biophys Res Commun **403**(1): 103-107.

Cobrinik, D. (2005). "Pocket proteins and cell cycle control." Oncogene **24**(17): 2796-2809.

Cockman, M. E., J. D. Webb, et al. (2009). "Proteomics-based identification of novel factor inhibiting hypoxia-inducible factor (FIH) substrates indicates widespread asparaginyl hydroxylation of ankyrin repeat domain-containing proteins." Mol Cell Proteomics **8**(3): 535-546.

Cockman, M. E., J. D. Webb, et al. (2009). "FIH-dependent asparaginyl hydroxylation of ankyrin repeat domain-containing proteins." Ann N Y Acad Sci **1177**: 9-18.

Coleman, M. L., M. A. McDonough, et al. (2007). "Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor." J Biol Chem **282**(33): 24027-24038.

Corradi, V., M. Mancini, et al. (2010). "Identification of the first non-peptidic small molecule inhibitor of the c-Abl/14-3-3 protein-protein interactions able to drive sensitive and Imatinib-resistant leukemia cells to apoptosis." Bioorg Med Chem Lett **20**(20): 6133-6137.

Dawson, S., S. Apcher, et al. (2002). "Gankyrin is an ankyrin-repeat oncoprotein that interacts with CDK4 kinase and the S6 ATPase of the 26 S proteasome." J Biol Chem **277**(13): 10893-10902.

Dawson, S., R. Hastings, et al. (1997). "The 26S-proteasome: regulation and substrate recognition." Mol Biol Rep **24**(1-2): 39-44.

Dawson, S., H. Higashitsuji, et al. (2006). "Gankyrin: a new oncoprotein and regulator of pRb and p53." Trends Cell Biol **16**(5): 229-233.

de Vries, S. J., A. D. van Dijk, et al. (2007). "HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets." Proteins **69**(4): 726-733.

DeCaprio, J. A., J. W. Ludlow, et al. (1988). "SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene." Cell **54**(2): 275-283.

Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.

DeWitte, R. S. and E. I. Shakhnovich (1996). "SMoG: de Novo design method based on simple, fast, and accurate free energy estimates .1. Methodology and supporting evidence." J Am Chem Soc **118**(47): 11733-11744.

Di Conza, G., F. Mancini, et al. (2012). "MDM4 enhances p53 stability by promoting an active conformation of the protein upon DNA damage." Cell Cycle **11**(4): 749-760.

Dick, F. A. (2007). "Structure-function analysis of the retinoblastoma tumor suppressor protein - is the whole a sum of its parts?" Cell Div **2**: 26.

Dodd, I. B., A. J. Perkins, et al. (2001). "Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny." Genes Dev **15**(22): 3013-3022.

Dominguez, C., R. Boelens, et al. (2003). "HADDOCK: a protein-protein docking approach based on biochemical or biophysical information." J Am Chem Soc **125**(7): 1731-1737.

Dove, S. L., J. K. Joung, et al. (1997). "Activation of prokaryotic transcription through arbitrary protein-protein contacts." Nature **386**(6625): 627-630.

Dyson, N., P. Guida, et al. (1992). "Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins." J Virol **66**(12): 6893-6902.

Edfeldt, F. N., R. H. Folmer, et al. (2011). "Fragment screening to predict druggability (ligandability) and lead discovery success." Drug Discov Today **16**(7-8): 284-287.

Eldridge, M. D., C. W. Murray, 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**(5): 425-445.

Elkins, J. M., K. S. Hewitson, et al. (2003). "Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha." J Biol Chem **278**(3): 1802-1806.

Feher, M. (2006). "Consensus scoring for protein-ligand interactions." Drug Discov Today **11**(9-10): 421-428.

Fersht, A. R. (1985). Enzyme Structure and Mechanism. New York, Freeman & Co.

Fleming, P. J. and F. M. Richards (2000). "Protein packing: dependence on protein size, secondary structure and amino acid composition." J Mol Biol **299**(2): 487-498.

Flemington, E. K., S. H. Speck, et al. (1993). "E2F-1-mediated transactivation is inhibited by complex formation with the retinoblastoma susceptibility gene product." Proc Natl Acad Sci U S A **90**(15): 6914-6918.

Fletcher, S. and A. D. Hamilton (2007). "Protein-protein interaction inhibitors: small molecules from screening techniques." Curr Top Med Chem **7**(10): 922-927.

Forrer, P., M. T. Stumpp, et al. (2003). "A novel strategy to design binding molecules harnessing the modular nature of repeat proteins." FEBS Lett **539**(1-3): 2-6.

Fruman, D. A. and C. Rommel (2011). "PI3Kdelta inhibitors in cancer: rationale and serendipity merge in the clinic." Cancer Discov **1**(7): 562-572.

Fry, D. C. and L. T. Vassilev (2005). "Targeting protein-protein interactions for cancer therapy." J Mol Med (Berl) **83**(12): 955-963.

Fu, X., L. Tan, et al. (2004). "A novel diagnostic marker, p28GANK distinguishes hepatocellular carcinoma from potential mimics." J Cancer Res Clin Oncol **130**(9): 514-520.

Gabb, H. A., R. M. Jackson, et al. (1997). "Modelling protein docking using shape complementarity, electrostatics and biochemical information." J Mol Biol **272**(1): 106-120.

Gartel, A. L. and S. K. Radhakrishnan (2005). "Lost in transcription: p21 repression, mechanisms, and consequences." Cancer Res **65**(10): 3980-3985.

Gasteiger, E., C. Hoogland, et al. (2005). Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook. J. M. Walker, Humana Press**:** 571-607.

Gordon, G. M. and W. Du (2011). "Conserved RB functions in development and tumor suppression." Protein Cell **2**(11): 864-878.

Gray, J. J., S. Moughon, et al. (2003). "Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations." J Mol Biol **331**(1): 281-299.

Guo, Y., A. Mahajan, et al. (2009). "Comparisons of the conformational stability of cyclin-dependent kinase (CDK) 4-interacting ankyrin repeat (AR) proteins." Biochemistry **48**(19): 4050-4062.

Hajduk, P. J., J. R. Huth, et al. (2005). "Predicting protein druggability." Drug Discov Today **10**(23-24): 1675-1682.

Halgren, T. A. (2009). "Identifying and characterizing binding sites and assessing druggability." J Chem Inf Model **49**(2): 377-389.

Hao, N., M. L. Whitelaw, et al. (2011). "Identification of residues in the N-terminal PAS domains important for dimerization of Arnt and AhR." Nucleic Acids Res **39**(9): 3695-3709.

Harrison, C. (2012). "Trial watch: BTK inhibitor shows positive results in B cell malignancies." Nat Rev Drug Discov **11**(2): 96.

He, G., Z. H. Siddik, et al. (2005). "Induction of p21 by p53 following DNA damage inhibits both Cdk4 and Cdk2 activities." **24**(18): 2929-2943.

Heilmann, A. M. and N. J. Dyson (2012). "Phosphorylation puts the pRb tumor suppressor into shape." Genes Dev **26**(11): 1128-1130.

Helin, K. and H. Ed (1993). "The retinoblastoma protein as a transcriptional repressor." Trends Cell Biol **3**(2): 43-46.

Higashitsuji, H., K. Itoh, et al. (2000). "Reduced stability of retinoblastoma protein by gankyrin, an oncogenic ankyrin-repeat protein overexpressed in hepatomas." Nat Med **6**(1): 96-99.

Higashitsuji, H., K. Itoh, et al. (2005). "The oncoprotein gankyrin binds to MDM2/HDM2, enhancing ubiquitylation and degradation of p53." Cancer Cell **8**(1): 75-87.

Higashitsuji, H., Y. Liu, et al. (2007). "The oncoprotein gankyrin interacts with RelA and suppresses NF-kappaB activity." Biochem Biophys Res Commun **363**(3): 879-884.

Higashitsuji, H., Y. Liu, et al. (2005). "The Oncoprotein Gankyrin Negatively Regulates Both p53 and RB by Enhancing Proteasomal Degradation." Cell Cycle **4**(10).

Hirai, H., M. F. Roussel, et al. (1995). "Novel INK4 proteins, p19 and p18, are specific inhibitors of the cyclin D-dependent kinases CDK4 and CDK6." Mol Cell Biol **15**(5): 2672-2681.

Hori, T., S. Kato, et al. (1998). "cDNA cloning and functional analysis of p28 (Nas6p) and p40.5 (Nas7p), two novel regulatory subunits of the 26S proteasome." Gene **216**(1): 113-122.

Howard, C. M., P. P. Claudio, et al. (1998). "Retinoblastoma-related protein pRb2/p130 and suppression of tumor growth in vivo." J Natl Cancer Inst **90**(19): 1451-1460.

Humphrey, W., A. Dalke, et al. (1996). "VMD: visual molecular dynamics." J Mol Graph **14**(1): 33-38, 27-38.

Hussein, Y. M., A. F. Ghareib, et al. (2011). "MAGE-3 and MAGE-4 genes as possible markers for early detection of metastases in hepatitis C virus Egyptian patients complicated by hepatocellular carcinoma." Med Oncol **29**(2): 994-999.

Hussein, Y. M., F. E. Morad, et al. (2012). "MAGE-4 gene m-RNA and TGF in blood as potential biochemical markers for HCC in HCV-infected patients." Med Oncol.

Inglis, S. R., C. Stojkoski, et al. (2004). "Identification and specificity studies of small-molecule ligands for SH3 protein domains." J Med Chem **47**(22): 5405-5417.

Iwai, A., H. Marusawa, et al. (2003). "Role of a novel oncogenic protein, gankyrin, in hepatocyte proliferation." J Gastroenterol **38**(8): 751-758.

Janin, J. (2010). "Protein-protein docking tested in blind predictions: the CAPRI experiment." Mol Biosyst **6**(12): 2351-2362.

Janin, J., K. Henrick, et al. (2003). "CAPRI: a Critical Assessment of PRedicted Interactions." Proteins **52**(1): 2-9.

Jarsch, M., M. Brandt, et al. (2008). "Comparative erythropoietin receptor binding kinetics of C.E.R.A. and epoetin-beta determined by surface plasmon resonance and competition binding assay." Pharmacology **81**(1): 63-69.

Kamb, A., N. A. Gruis, et al. (1994). "A cell cycle regulator potentially involved in genesis of many tumor types." Science **264**(5157): 436-440.

Kang, Y. H., N. Y. Ji, et al. (2012). "ESM-1 regulates cell growth and metastatic process through activation of NF-kappaB in colorectal cancer." Cell Signal **24**(10): 1940-1949.

Katchalski-Katzir, E., I. Shariv, et al. (1992). "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques." Proc Natl Acad Sci U S A **89**(6): 2195-2199.

Kim, W., E. J. Bennett, et al. (2011). "Systematic and quantitative assessment of the ubiquitin-modified proteome." Mol Cell **44**(2): 325-340.

Knudsen, E. S. and J. Y. Wang (1996). "Differential regulation of retinoblastoma protein function by specific Cdk phosphorylation sites." J Biol Chem **271**(14): 8313-8320.

Komives, E. A. (2012). "Consequences of fuzziness in the NFkappaB/IkappaBalpha interaction." Adv Exp Med Biol **725**: 74-85.

Konteatis, Z. D. (2010). "In silico fragment-based drug design." Expert Opin Drug Discov **5**(11): 1047-1065.

Kontoyianni, M., G. S. Sokol, et al. (2005). "Evaluation of library ranking efficacy in virtual screening." J Comput Chem **26**(1): 11-22.

Kortemme, T., A. V. Morozov, et al. (2003). "An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes." J Mol Biol **326**(4): 1239-1259.

Kozakov, D., O. Schueler-Furman, et al. (2008). "Discrimination of near-native structures in protein-protein docking by testing the stability of local minima." Proteins **72**(3): 993-1004.

Krivov, G. G., M. V. Shapovalov, et al. (2009). "Improved prediction of protein side-chain conformations with SCWRL4." Proteins **77**(4): 778-795.

Krzywda, S., A. M. Brzozowski, et al. (2004). "The crystal structure of gankyrin, an oncoprotein found in complexes with cyclin-dependent kinase 4, a 19 S proteasomal ATPase regulator, and the tumor suppressors Rb and p53." J Biol Chem **279**(2): 1541-1545.

Kucharczak, J., M. J. Simmons, et al. (2003). "To be, or not to be: NF-kappaB is the answer--role of Rel/NF-kappaB in the regulation of apoptosis." Oncogene **22**(56): 8961-8982.

LaBaer, J., M. D. Garrett, et al. (1997). "New functional activities for the p21 family of CDK inhibitors." Genes Dev **11**(7): 847-862.

Lee, C., J. H. Chang, et al. (2002). "Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor." Genes Dev **16**(24): 3199-3212.

Lee, J. O., A. A. Russo, et al. (1998). "Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7." Nature **391**(6670): 859-865.

Lee, W., X. Zeng, et al. (2012). "Mechanical anisotropy of ankyrin repeats." Biophys J **102**(5): 1118-1126.

Letunic, I., T. Doerks, et al. (2009). "SMART 6: recent updates and new developments." Nucleic Acids Res **37**(Database issue): D229-232.

Letunic, I., T. Doerks, et al. (2012). "SMART 7: recent updates to the protein domain annotation resource." Nucleic Acids Res **40**(Database issue): D302-305.

Li, H., X. Fu, et al. (2005). "Use of adenovirus-delivered siRNA to target oncoprotein p28GANK in hepatocellular carcinoma." Gastroenterology **128**(7): 2029-2041.

Li, J., T. J. Knobloch, et al. (2011). "Gankyrin, a biomarker for epithelial carcinogenesis, is overexpressed in human oral cancer." Anticancer Res **31**(9): 2683-2692.

Li, J. and M. D. Tsai (2002). "Novel insights into the INK4-CDK4/6-Rb pathway: counter action of gankyrin against INK4 proteins regulates the CDK4-mediated phosphorylation of Rb." Biochemistry **41**(12): 3977-3983.

Li, X., Y. Zhang, et al. (2009). "Overexpression of a new gene P28GANK confers multidrug resistance of gastric cancer cells." Cancer Invest **27**(2): 129-139.

Linge, J. P., M. Habeck, et al. (2003). "ARIA: automated NOE assignment and NMR structure calculation." Bioinformatics **19**(2): 315-316.

Linke, S., R. J. Hampton-Smith, et al. (2007). "Characterization of ankyrin repeat-containing proteins as substrates of the asparaginyl hydroxylase factor inhibiting hypoxia-inducible transcription factor." Methods Enzymol **435**: 61-85.

Linke, S., C. Stojkoski, et al. (2004). "Substrate requirements of the oxygen-sensing asparaginyl hydroxylase factor-inhibiting hypoxia-inducible factor." J Biol Chem **279**(14): 14391-14397.

Llovet, J. M., Y. Chen, et al. (2006). "A molecular signature to discriminate dysplastic nodules from early hepatocellular carcinoma in HCV cirrhosis." Gastroenterology **131**(6): 1758-1767.

Low, C., U. Weininger, et al. (2007). "Folding mechanism of an ankyrin repeat protein: scaffold and active site formation of human CDK inhibitor p19(INK4d)." J Mol Biol **373**(1): 219-231.

Lowe, A. R. and L. S. Itzhaki (2007). "Biophysical characterisation of the small ankyrin repeat protein myotrophin." J Mol Biol **365**(4): 1245-1255.

Luedde, T., N. Beraza, et al. (2007). "Deletion of NEMO/IKKgamma in liver parenchymal cells causes steatohepatitis and hepatocellular carcinoma." Cancer Cell **11**(2): 119-132.

MacKerell, A. D., D. Bashford, et al. (1998). "All-atom empirical potential for molecular modeling and dynamics studies of proteins." Journal of Physical Chemistry B **102**(18): 3586-3616.

Maeda, S., H. Kamata, et al. (2005). "IKKbeta couples hepatocyte death to cytokine-driven compensatory proliferation that promotes chemical hepatocarcinogenesis." Cell **121**(7): 977-990.

Mahajan, A., Y. Guo, et al. (2007). "Dissection of protein-protein interaction and CDK4 inhibition in the oncogenic versus tumor suppressing functions of gankyrin and P16." J Mol Biol **373**(4): 990-1005.

Marti-Renom, M. A., A. C. Stuart, et al. (2000). "Comparative protein structure modeling of genes and genomes." Annu Rev Biophys Biomol Struct **29**: 291-325.

Martin-Caballero, J., J. M. Flores, et al. (2001). "Tumor susceptibility of p21(Waf1/Cip1)-deficient mice." Cancer Res **61**(16): 6234-6238.

Mendez, R., R. Leplae, et al. (2003). "Assessment of blind predictions of protein-protein interactions: current status of docking methods." Proteins **52**(1): 51-67.

Meng, E. C., B. K. Shoichet, et al. (1992). "AUTOMATED DOCKING WITH GRID-BASED ENERGY EVALUATION." Journal of Computational Chemistry **13**(4): 505-524.

Meng, Y., L. He, et al. (2010). "Gankyrin promotes the proliferation of human pancreatic cancer." Cancer Lett **297**(1): 9-17.

Miller, J. H. (1972). Experiments in molecular genetics. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Moont, G., H. A. Gabb, et al. (1999). "Use of pair potentials across protein interfaces in screening predicted docked complexes." Proteins **35**(3): 364-373.

Morris, G. M., D. S. Goodsell, et al. (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function." Journal of Computational Chemistry **19**(14): 1639-1662.

Mosavi, L. K., T. J. Cammett, et al. (2004). "The ankyrin repeat as molecular architecture for protein recognition." Protein Sci **13**(6): 1435-1448.

Mosavi, L. K., D. L. Minor, Jr., et al. (2002). "Consensus-derived structural determinants of the ankyrin repeat motif." Proc Natl Acad Sci U S A **99**(25): 16029-16034.

Mosavi, L. K., S. Williams, et al. (2002). "Equilibrium folding and stability of myotrophin: a model ankyrin repeat protein." J Mol Biol **320**(2): 165-170.

Moustakas, D. T., P. T. Lang, et al. (2006). "Development and validation of a modular, extensible docking program: DOCK 5." J Comput Aided Mol Des **20**(10-11): 601-619.

Muegge, I. and Y. C. Martin (1999). "A general and fast scoring function for protein-ligand interactions: a simplified potential approach." J Med Chem **42**(5): 791-804.

Mullard, A. (2012). "Protein-protein interaction inhibitors get into the groove." Nat Rev Drug Discov **11**(3): 173-175.

Nagao, T., H. Higashitsuji, et al. (2003). "MAGE-A4 interacts with the liver oncoprotein gankyrin and suppresses its tumorigenic activity." J Biol Chem **278**(12): 10668-10674.

Nakamura, Y., K. Nakano, et al. (2007). "Structure of the oncoprotein gankyrin in complex with S6 ATPase of the 26S proteasome." Structure **15**(2): 179-189.

Nakamura, Y., T. Umehara, et al. (2007). "Structural basis for the recognition between the regulatory particles Nas6 and Rpt3 of the yeast 26S proteasome." Biochem Biophys Res Commun **359**(3): 503-509.

Nielsen, S. J., R. Schneider, et al. (2001). "Rb targets histone H3 methylation and HP1 to promoters." Nature **412**(6846): 561-565.

Nobori, T., K. Miura, et al. (1994). "Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers." Nature **368**(6473): 753-756.

Notredame, C., D. G. Higgins, et al. (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." J Mol Biol **302**(1): 205-217.

Oku, T., S. Ikeda, et al. (1998). "Functional sites of human PCNA which interact with p21 (Cip1/Waf1), DNA polymerase delta and replication factor C." Genes Cells **3**(6): 357-369.

Olsen, J. V., M. Vermeulen, et al. (2010). "Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis." Sci Signal **3**(104): ra3.

Orry, A. J., R. A. Abagyan, et al. (2006). "Structure-based development of target-specific compound libraries." Drug Discov Today **11**(5-6): 261-266.

Ortiz, C. M., T. Ito, et al. (2008). "Gankyrin oncoprotein overexpression as a critical factor for tumor growth in human esophageal squamous cell carcinoma and its clinical significance." Int J Cancer **122**(2): 325-332.

Padmanabhan, B., N. Adachi, et al. (2004). "Crystal structure of the homolog of the oncoprotein gankyrin, an interactor of Rb and CDK4/6." J Biol Chem **279**(2): 1546-1552.

Pagliaro, L., J. Felding, et al. (2004). "Emerging classes of protein-protein interaction inhibitors and new tools for their development." Curr Opin Chem Biol **8**(4): 442-449.

Parry, D., S. Bates, et al. (1995). "Lack of cyclin D-Cdk complexes in Rb-negative cells correlates with high levels of p16INK4/MTS1 tumour suppressor gene product." EMBO J **14**(3): 503-511.

Pattabiraman, N., K. B. Ward, et al. (1995). "Occluded molecular surface: analysis of protein packing." J Mol Recognit **8**(6): 334-344.

Pettersen, E. F., T. D. Goddard, et al. (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.

Phillips, J. C., R. Braun, et al. (2005). "Scalable molecular dynamics with NAMD." Journal of Computational Chemistry **26**(16): 1781-1802.

Polager, S. and D. Ginsberg (2008). "E2F - at the crossroads of life and death." Trends Cell Biol **18**(11): 528-535.

Punta, M., P. C. Coggill, et al. (2012). "The Pfam protein families database." Nucleic Acids Res **40**(Database issue): D290-301.

Qian, X., J. Hulit, et al. (2012). "p21CIP1 mediates reciprocal switching between proliferation and invasion during metastasis." Oncogene.

Qin, J. M., X. Y. Fu, et al. (2003). "Gene and protein expressions of p28GANK in rat with liver regeneration." World J Gastroenterol **9**(11): 2523-2527.

Rejto, P. A. and G. M. Verkhivker (1996). "Unraveling principles of lead discovery: from unfrustrated energy landscapes to novel molecular anchors." Proc Natl Acad Sci U S A **93**(17): 8945-8950.

Resnitzky, D. and S. I. Reed (1995). "Different roles for cyclins D1 and E in regulation of the G1-to-S transition." Mol Cell Biol **15**(7): 3463-3469.

Rickman, D. S., R. Millon, et al. (2008). "Prediction of future metastasis and molecular characterization of head and neck squamous-cell carcinoma based on transcriptome and genome analysis by microarrays." Oncogene **27**(51): 6607-6622.

Rizzolio, F., L. Esposito, et al. (2010). "RB gene family: genome-wide ChIP approaches could open undiscovered roads." J Cell Biochem **109**(5): 839-843.

Robanus-Maandag, E., M. Dekker, et al. (1998). "p107 is a suppressor of retinoblastoma development in pRb-deficient mice." Genes Dev **12**(11): 1599-1609.

Robertson, K. D., S. Ait-Si-Ali, et al. (2000). "DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters." Nat Genet **25**(3): 338-342.

Rocco, A., L. Schandl, et al. (2002). "Loss of expression of tumor suppressor p16(INK4) protein in human primary gastric cancer is related to the grade of differentiation." Dig Dis **20**(1): 102-105.

Roch, N., A. Kutup, et al. (2010). "Coexpression of MAGE-A peptides and HLA class I molecules in hepatocellular carcinoma." Anticancer Res **30**(5): 1617-1623.

Russo, A. A., L. Tong, et al. (1998). "Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a." Nature **395**(6699): 237-243.

Sakurai, T., K. Itoh, et al. (2004). "A cleaved form of MAGE-A4 binds to Miz-1 and induces apoptosis in human cells." J Biol Chem **279**(15): 15505-15514.

Schmierer, B., B. Novak, et al. (2010). "Hypoxia-dependent sequestration of an oxygen sensor by a widespread structural motif can shape the hypoxic response--a predictive kinetic model." BMC Syst Biol **4**: 139.

Schueler-Furman, O., C. Wang, et al. (2005). "Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility." Proteins **60**(2): 187-194.

Schultz, J., F. Milpetz, et al. (1998). "SMART, a simple modular architecture research tool: identification of signaling domains." Proc Natl Acad Sci U S A **95**(11): 5857-5864.

Schwieters, C. D., J. J. Kuszewski, et al. (2003). "The Xplor-NIH NMR molecular structure determination package." J Magn Reson **160**(1): 65-73.

Sherr, C. J. and J. M. Roberts (1995). "Inhibitors of mammalian G1 cyclin-dependent kinases." Genes Dev **9**(10): 1149-1163.

Singh, M. (2006). Biochemical and structural investigations of the retinoblastoma protein, its binding partners, and the BRG1 protein - a subunit of human SWI/SNF remodeling complexes. Doktors der Naturwissenschaften, Technischen Universität München.

Singleton, R. S., D. C. Trudgian, et al. (2011). "Quantitative mass spectrometry reveals dynamics of factor-inhibiting hypoxia-inducible factor-catalyzed hydroxylation." J Biol Chem **286**(39): 33784-33794.

Smith, P. J. (2007). An Investigation into the Interaction Between Gankyrin and CDK4. B. Sc (Hons), University of Adelaide.

Sneppen, K., S. Krishna, et al. (2010). "Simplified models of biological networks." Annu Rev Biophys **39**: 43-59.

Sonnhammer, E. L., S. R. Eddy, et al. (1997). "Pfam: a comprehensive database of protein domain families based on seed alignments." Proteins **28**(3): 405-420.

Sperandio, O., C. H. Reynes, et al. (2010). "Rationalizing the chemical space of protein-protein interaction inhibitors." Drug Discov Today **15**(5-6): 220-229.

Tai, Y. T., B. Y. Chang, et al. (2012). "Bruton's tyrosine kinase inhibition is a novel therapeutic strategy targeting tumor in the bone marrow microenvironment in multiple myeloma." Blood.

Tang, K. S., B. J. Guralnick, et al. (1999). "Stability and folding of the tumour suppressor protein p16." J Mol Biol **285**(4): 1869-1886.

Tang, S., G. Yang, et al. (2010). "Overexpression of a novel gene gankyrin correlates with the malignant phenotype of colorectal cancer." Cancer Biol Ther **9**(2): 88-95.

Thakur, C. S., B. K. Jha, et al. (2007). "Small-molecule activators of RNase L with broad-spectrum antiviral activity." Proc Natl Acad Sci U S A **104**(23): 9585-9590.

Thakur, P. K. and M. I. Hassan (2011). "Discovering a potent small molecule inhibitor for gankyrin using de novo drug design approach." Int J Comput Biol Drug Des **4**(4): 373-386.

Tong, J., X. Sun, et al. (2011). "Expression of p16 in non-small cell lung cancer and its prognostic significance: a meta-analysis of published literatures." Lung Cancer **74**(2): 155-163.

Tovchigrechko, A. and I. A. Vakser (2001). "How common is the funnel-like energy landscape in protein-protein interactions?" Protein Sci **10**(8): 1572-1583.

Tripp, K. W. and D. Barrick (2003). "Folding by consensus." Structure **11**(5): 486-487.

Tripp, K. W. and D. Barrick (2007). "Enhancing the stability and folding rate of a repeat protein through the addition of consensus repeats." J Mol Biol **365**(4): 1187-1200.

Umemura, A., Y. Itoh, et al. (2008). "Association of gankyrin protein expression with early clinical stages and insulin-like growth factor-binding protein 5 expression in human hepatocellular carcinoma." Hepatology **47**(2): 493-502.

Uniprot-Consortium (2012). "Reorganizing the protein space at the Universal Protein Resource (UniProt)." Nucleic Acids Res **40**(Database issue): D71-75.

Verkhivker, G. M., P. A. Rejto, et al. (1999). "Towards understanding the mechanisms of molecular recognition by computer simulations of ligand-protein interactions." J Mol Recognit **12**(6): 371-389.

Voet, A. and K. Y. Zhang (2012). "Pharmacophore modelling as a virtual screening tool for the discovery of small molecule protein-protein interaction inhibitors." Curr Pharm Des.

Vranken, W. F., W. Boucher, et al. (2005). "The CCPN data model for NMR spectroscopy: development of a software pipeline." Proteins **59**(4): 687-696.

Wagner, S. A., P. Beli, et al. (2011). "A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles." Mol Cell Proteomics **10**(10): M111 013284.

Wang, R., L. Liu, et al. (1998). "SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex." Journal of Molecular Modeling **4**(12): 379-394.

Wang, X., C. F. Chen, et al. (2007). "Mass spectrometric characterization of the affinity-purified human 26S proteasome complex." Biochemistry **46**(11): 3553-3565.

Warren, G. L., C. W. Andrews, et al. (2006). "A critical assessment of docking programs and scoring functions." J Med Chem **49**(20): 5912-5931.

Werbeck, N. D. and L. S. Itzhaki (2007). "Probing a moving target with a plastic unfolding intermediate of an ankyrin-repeat protein." Proc Natl Acad Sci U S A **104**(19): 7863-7868.

Whitty, A. and G. Kumaravel (2006). "Between a rock and a hard place?" Nat Chem Biol **2**(3): 112-118.

Whyte, P., H. E. Ruley, et al. (1988). "Two regions of the adenovirus early region 1A proteins are required for transformation." J Virol **62**(1): 257-265.

Wilkins, S. E., J. Hyvarinen, et al. (2009). "Differences in hydroxylation and binding of Notch and HIF-1alpha demonstrate substrate selectivity for factor inhibiting HIF-1 (FIH-1)." Int J Biochem Cell Biol **41**(7): 1563-1571.

Wilkins, S. E., S. Karttunen, et al. (2012). "Factor inhibiting HIF (FIH) recognises distinct molecular features within hypoxia inducible factor (HIF)-alpha versus ankyrin repeat substrates." J Biol Chem **287**(12): 8769-8781.

Yang, J., S. Kantrow, et al. (2012). "INK4a/ARF inactivation with activation of the NF-kappaB/IL-6 pathway is sufficient to drive the development and growth of angiosarcoma." Cancer Res.

Yang, J. M., Y. F. Chen, et al. (2005). "Consensus scoring criteria for improving enrichment in virtual screening." J Chem Inf Model **45**(4): 1134-1146.

Yang, W., K. S. Klos, et al. (2003). "ErbB2 overexpression in human breast carcinoma is correlated with p21Cip1 up-regulation and tyrosine-15 hyperphosphorylation of p34Cdc2: poor responsiveness to chemotherapy with cyclophoshamide methotrexate, and 5-fluorouracil is associated with Erb2 overexpression and with p21Cip1 overexpression." Cancer **98**(6): 1123-1130.

Yuan, C., J. Li, et al. (2004). "Solution structure of the human oncogenic protein gankyrin containing seven ankyrin repeats and analysis of its structure--function relationship." Biochemistry **43**(38): 12152-12161.

Zahnd, C., F. Pecorari, et al. (2006). "Selection and characterization of Her2 binding-designed ankyrin repeat proteins." J Biol Chem **281**(46): 35167-35175.

Zahnd, C., E. Wyler, et al. (2007). "A designed ankyrin repeat protein evolved to picomolar affinity to Her2." J Mol Biol **369**(4): 1015-1028.

Zeeb, M., H. Rosner, et al. (2002). "Protein folding and stability of human CDK inhibitor p19(INK4d)." J Mol Biol **315**(3): 447-457.

Zeifman, A. A., V. S. Stroylov, et al. (2012). "Hit clustering can improve virtual fragment screening: CDK2 and PARP1 case studies." J Mol Model **18**(6): 2553-2566.

Zhang, B. and Z. Peng (2000). "A minimum folding unit in the ankyrin repeat protein p16(INK4)." J Mol Biol **299**(4): 1121-1132.

Zhang, F., J. M. Beaudet, et al. (2012). "Analysis of the Interaction between Heparin and Follistatin/Follistatin-ligand complexes Using Surface Plasmon Resonance." Biochemistry.

Zhang, X. and M. Oglesbee (2003). "Use of surface plasmon resonance for the measurement of low affinity binding interactions between HSP72 and measles virus nucleocapsid protein." Biol Proced Online **5**: 170-181.

Zhao, L. and J. Chmielewski (2005). "Inhibiting protein-protein interactions using designed molecules." Curr Opin Struct Biol **15**(1): 31-34.

Zhao, W., C. C. Huang, et al. (2012). "Altered p16(INK4) and RB1 Expressions Are Associated with Poor Prognosis in Patients with Nonsmall Cell Lung Cancer." J Oncol **2012**: 957437.

Zheng, X., S. Linke, et al. (2008). "Interaction with factor inhibiting HIF-1 defines an additional mode of cross-coupling between the Notch and hypoxia signaling pathways." Proc Natl Acad Sci U S A **105**(9): 3368-3373.

Zoete, V., A. Grosdidier, et al. (2010). "Use of the FACTS solvation model for protein-ligand docking calculations. Application to EADock." Journal of Molecular Recognition **23**(5): 457-461.

Zweifel, M. E., D. J. Leahy, et al. (2003). "Structure and stability of the ankyrin domain of the Drosophila Notch receptor." Protein Sci **12**(11): 2622-2632.

Appendix A

FTDock Modifications

The text below describes the technical differences between the distributed version of ftdock.c and the OpenMP parallelized version described in this work. To generate a complete, parallelized version of ftdock.c for compiling, copy the text below into a file called `ftdock.patch`, place it in the same directory as the distributed version of ftdock.c and run `patch -i ftdock.patch`. You may like to make a backup of the distributed ftdock.c first.

```
*** ftdock.c        2001-03-22 00:40:25.000000000 +1030
--- ../../3D_Dock_multicore/progs/ftdock.c    2006-08-11
15:14:54.000000000 +0930
***************
*** 27,32 ****
--- 27,33 ----
   */

   #include "structures.h"
+ #include <omp.h>

   int main( int argc , char *argv[] ) {

***************
*** 57,65 ****
--- 58,72 ----
     char            *default_electrostatics ;
     char            *default_keep_per_rotation ;

+    /* OMP stuff */
+    int              num_threads ;
+    int              thread_num ;
+
     /* File stuff */

     FILE            *ftdock_file ;
+    FILE            *ftdock_scores ;
+    char             file_name[25] ;
     char            line_buffer[100] ;
     int             id , id2 , SCscore ;
     float           RPscore ;
***************
*** 74,113 ****

     struct Structure   Static_Structure , Mobile_Structure ;
     struct Structure   Origin_Static_Structure ,
Origin_Mobile_Structure ;
-    struct Structure   Rotated_at_Origin_Mobile_Structure ;

     /* Co-ordinates */

-    int             xyz , fx , fy , fz , fxyz ;
```

```
    /* Grid stuff */

    float          grid_span , one_span ;

    fftw_real    *static_grid ;
-   fftw_real    *mobile_grid ;
-   fftw_real    *convoluted_grid ;

    fftw_real    *static_elec_grid = ( void * ) 0 ;
-   fftw_real    *mobile_elec_grid = ( void * ) 0 ;
-   fftw_real    *convoluted_elec_grid = ( void * ) 0 ;

    /* FFTW stuff */

    rfftwnd_plan p , pinv ;

    fftw_complex  *static_fsg ;
-   fftw_complex  *mobile_fsg ;
-   fftw_complex  *multiple_fsg ;

    fftw_complex  *static_elec_fsg = ( void * ) 0 ;
-   fftw_complex  *mobile_elec_fsg = ( void * ) 0 ;
-   fftw_complex  *multiple_elec_fsg = ( void * ) 0 ;

    /* Scores */

-   struct Score *Scores ;
    float          max_es_value ;

  /************/

--- 81,110 ----

    struct Structure   Static_Structure , Mobile_Structure ;
    struct Structure   Origin_Static_Structure ,
Origin_Mobile_Structure ;

    /* Co-ordinates */


    /* Grid stuff */

    float          grid_span , one_span ;

    fftw_real    *static_grid ;

    fftw_real    *static_elec_grid = ( void * ) 0 ;

    /* FFTW stuff */

    rfftwnd_plan p , pinv ;

    fftw_complex  *static_fsg ;

    fftw_complex  *static_elec_fsg = ( void * ) 0 ;

    /* Scores */
```

```
      float        max_es_value ;
+     float        partial_max ;

  /************/

***************
*** 311,317 ****
        if( strncmp( line_buffer , "Electrostatics
::    on" , 44 ) == 0 ) electrostatics = 1 ;
        if( strncmp( line_buffer , "Electrostatics
::    off" , 44 ) == 0 ) electrostatics = 0 ;
        if( strncmp( line_buffer , "Global keep per rotation" , 25 )
== 0 ) sscanf( line_buffer , "Global keep per rotation :: %d" ,
&keep_per_rotation ) ;
!
      }

      fclose( ftdock_file ) ;
--- 308,314 ----
        if( strncmp( line_buffer , "Electrostatics
::    on" , 44 ) == 0 ) electrostatics = 1 ;
        if( strncmp( line_buffer , "Electrostatics
::    off" , 44 ) == 0 ) electrostatics = 0 ;
        if( strncmp( line_buffer , "Global keep per rotation" , 25 )
== 0 ) sscanf( line_buffer , "Global keep per rotation :: %d" ,
&keep_per_rotation ) ;
!      if( strncmp( line_buffer , "OMP thread number" , 17 ) == 0 )
sscanf( line_buffer , "OMP thread number :: %d" , &num_threads ) ;
      }

      fclose( ftdock_file ) ;
***************
*** 416,454 ****

    /* Memory Allocation */

-   if( ( Scores = ( struct Score * ) malloc ( ( keep_per_rotation + 2
) * sizeof( struct Score ) ) ) == NULL ) {
-     GENERAL_MEMORY_PROBLEM
-   }
-
    if(
      ( ( static_grid = ( fftw_real * ) malloc
       ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
!     ||
!     ( ( mobile_grid = ( fftw_real * ) malloc
!      ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
!     ||
!     ( ( convoluted_grid = ( fftw_real * ) malloc
!      ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
!     ) {
      printf( "Not enough memory for surface grids\nUse (sensible)
smaller grid size\nDying\n\n" ) ;
      exit( EXIT_FAILURE ) ;
    }
```

184

```
      static_fsg = ( fftw_complex * ) static_grid ;
-     mobile_fsg = ( fftw_complex * ) mobile_grid ;
-     multiple_fsg = ( fftw_complex * ) convoluted_grid ;

      if( electrostatics == 1 ) {

        if(
          ( ( static_elec_grid = ( fftw_real * ) malloc
            ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
-         ||
-         ( ( mobile_elec_grid = ( fftw_real * ) malloc
-           ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
-         ||
-         ( ( convoluted_elec_grid = ( fftw_real * ) malloc
-           ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
          ) {
          printf( "Not enough memory for electrostatic grids\nSwitch off
electrostatics or use (sensible) smaller grid size\nDying\n\n" ) ;
          exit( EXIT_FAILURE ) ;
--- 413,433 ----

      /* Memory Allocation */

      if(
        ( ( static_grid = ( fftw_real * ) malloc
          ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
!       ) {
        printf( "Not enough memory for surface grids\nUse (sensible)
smaller grid size\nDying\n\n" ) ;
        exit( EXIT_FAILURE ) ;
      }

      static_fsg = ( fftw_complex * ) static_grid ;

      if( electrostatics == 1 ) {

        if(
          ( ( static_elec_grid = ( fftw_real * ) malloc
            ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
          ) {
          printf( "Not enough memory for electrostatic grids\nSwitch off
electrostatics or use (sensible) smaller grid size\nDying\n\n" ) ;
          exit( EXIT_FAILURE ) ;
**************
*** 458,465 ****
        }

        static_elec_fsg = ( fftw_complex * ) static_elec_grid ;
-       mobile_elec_fsg = ( fftw_complex * ) mobile_elec_grid ;
-       multiple_elec_fsg = ( fftw_complex * ) convoluted_elec_grid ;

      }
```

```
--- 437,442 ----
***************
*** 469,477 ****

    printf( "Creating plans\n" ) ;
    p    = rfftw3d_create_plan( global_grid_size , global_grid_size ,
global_grid_size ,
!                               FFTW_REAL_TO_COMPLEX , FFTW_MEASURE |
FFTW_IN_PLACE ) ;
    pinv = rfftw3d_create_plan( global_grid_size , global_grid_size ,
global_grid_size ,
!                               FFTW_COMPLEX_TO_REAL , FFTW_MEASURE |
FFTW_IN_PLACE ) ;

  /***********/

--- 446,454 ----

    printf( "Creating plans\n" ) ;
    p    = rfftw3d_create_plan( global_grid_size , global_grid_size ,
global_grid_size ,
!                               FFTW_REAL_TO_COMPLEX , FFTW_MEASURE |
FFTW_IN_PLACE | FFTW_THREADSAFE ) ;
    pinv = rfftw3d_create_plan( global_grid_size , global_grid_size ,
global_grid_size ,
!                               FFTW_COMPLEX_TO_REAL , FFTW_MEASURE |
FFTW_IN_PLACE | FFTW_THREADSAFE ) ;

  /***********/

***************
*** 498,503 ****
--- 475,485 ----
    printf( "  done\n" ) ;

  /***********/
+ #pragma omp parallel
+ {
+   #pragma omp master
+   num_threads = omp_get_num_threads() ;
+ }

    /* Store paramaters in case of rescue */

***************
*** 523,528 ****
--- 505,511 ----
      fprintf( ftdock_file, "Electrostatics                      ::
off      %s\n" , default_electrostatics ) ;
    }
    fprintf( ftdock_file, "Global keep per rotation          :: %6d
%s\n" , keep_per_rotation , default_keep_per_rotation ) ;
+   fprintf( ftdock_file, "OMP thread number                 ::
%6d\n" , num_threads ) ;

    fprintf( ftdock_file, "\nCalculated values\n" ) ;
```

```
    fprintf( ftdock_file, "Global rotations                    ::
%6d\n" , Angles.n ) ;
***************
*** 538,546 ****
    max_es_value = 0 ;

    printf( "Starting main loop through the rotations\n" ) ;

!   for( rotation = first_rotation ; rotation <= Angles.n ; rotation
++ ) {

      printf( "." ) ;

      if( ( rotation % 50 ) == 0 ) printf( "\nRotation number %5d\n" ,
rotation ) ;
--- 521,596 ----
    max_es_value = 0 ;

    printf( "Starting main loop through the rotations\n" ) ;
+ #pragma omp parallel
private(partial_max,i,x,y,z,ftdock_file,thread_num)
+ {
+   struct Structure   Rotated_at_Origin_Mobile_Structure ;
+   fftw_real    *mobile_grid ;
+   fftw_real    *convoluted_grid ;
+   fftw_real    *mobile_elec_grid = ( void * ) 0 ;
+   fftw_real    *convoluted_elec_grid = ( void * ) 0 ;
+   fftw_complex  *mobile_fsg ;
+   fftw_complex  *multiple_fsg ;
+   fftw_complex  *mobile_elec_fsg = ( void * ) 0 ;
+   fftw_complex  *multiple_elec_fsg = ( void * ) 0 ;
+   struct Score *Scores ;
+   int          xyz , fx , fy , fz , fxyz ;
+   char          sfile_name[25] ;

!   thread_num = omp_get_thread_num();
!
!   printf( "This is thread %d of %d\n" , thread_num , num_threads ) ;
!
!   sprintf( sfile_name , "scratch_scores%02d.dat" , thread_num ) ;
!   printf( "%s\n" , sfile_name ) ;
!   if( ( ftdock_file = fopen( sfile_name , "w" ) ) == NULL ) {
!     printf( "Could not open scratch_scores%02d.dat for
writing.\nDying\n\n" , thread_num ) ;
!     exit( EXIT_FAILURE ) ;
!   }
!
!   printf( "Successfully opened %s\n", sfile_name ) ;
!
!   fclose( ftdock_file );
!   if( ( Scores = ( struct Score * ) malloc ( ( keep_per_rotation + 2
) * sizeof( struct Score ) ) ) == NULL ) {
!     GENERAL_MEMORY_PROBLEM
!   }
!
!   if (
!     ( ( mobile_grid = ( fftw_real * ) malloc
```

```
!       ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
!      ||
!      ( ( convoluted_grid = ( fftw_real * ) malloc
!       ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
!      ) {
!     printf( "Not enough memory for surface grids\nUse (sensible)
smaller grid size\nDying\n\n" ) ;
!     exit( EXIT_FAILURE ) ;
!    }
!
!   mobile_fsg = ( fftw_complex * ) mobile_grid ;
!   multiple_fsg = ( fftw_complex * ) convoluted_grid ;
!
!   if( electrostatics == 1 ) {
!
!      if(
!         ( ( mobile_elec_grid = ( fftw_real * ) malloc
!          ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
!         ||
!         ( ( convoluted_elec_grid = ( fftw_real * ) malloc
!          ( global_grid_size * global_grid_size * ( 2 * (
global_grid_size / 2 + 1 ) ) * sizeof( fftw_real ) ) ) == NULL )
!         ) {
!        printf( "Not enough memory for electrostatic grids\nSwitch off
electrostatics or use (sensible) smaller grid size\nDying\n\n" ) ;
!        exit( EXIT_FAILURE ) ;
!      } else {
!        /* all ok */
!        printf( "Electrostatics are on\n" ) ;
!      }
!      mobile_elec_fsg = ( fftw_complex * ) mobile_elec_grid ;
!      multiple_elec_fsg = ( fftw_complex * ) convoluted_elec_grid ;
!
!   }

+   partial_max = 0 ;
+ #pragma omp for schedule(static)
+   for( rotation = first_rotation ; rotation <= Angles.n ; rotation
++ ) {
      printf( "." ) ;

      if( ( rotation % 50 ) == 0 ) printf( "\nRotation number %5d\n" ,
rotation ) ;
***************
*** 570,576 ****
        fourier grid with other (raw) one
        hence the sign changes from a normal complex number
multiplication
      */
-
      for( fx = 0 ; fx < global_grid_size ; fx ++ ) {
        for( fy = 0 ; fy < global_grid_size ; fy ++ ) {
          for( fz = 0 ; fz < global_grid_size/2 + 1 ; fz ++ ) {
--- 620,625 ----
***************
```

```
*** 665,685 ****
        }
      }

!     if( rotation == 1 ) {
!       if( ( ftdock_file = fopen( "scratch_scores.dat" , "w" ) ) ==
NULL ) {
!         printf( "Could not open scratch_scores.dat for
writing.\nDying\n\n" ) ;
!         exit( EXIT_FAILURE ) ;
!       }
!     } else {
!       if( ( ftdock_file = fopen( "scratch_scores.dat" , "a" ) ) ==
NULL ) {
!         printf( "Could not open scratch_scores.dat for
writing.\nDying\n\n" ) ;
!         exit( EXIT_FAILURE ) ;
!       }
      }

      for( i = 0 ; i < keep_per_rotation ; i ++ ) {

!       max_es_value = min( max_es_value , Scores[i].rpscore ) ;
        fprintf( ftdock_file, "G_DATA %6d   %6d    %7d      %.0f
%4d %4d %4d      %4d%4d%4d\n" ,
                  rotation , 0 , Scores[i].score ,
(double)Scores[i].rpscore , Scores[i].coord[1] , Scores[i].coord[2] ,
Scores[i].coord[3 ] ,
                  Angles.z_twist[rotation] , Angles.theta[rotation]
, Angles.phi[rotation] ) ;
--- 714,730 ----
        }
      }

!     sprintf( sfile_name , "scratch_scores%02d.dat" , thread_num ) ;
!     if( ( ftdock_file = fopen( sfile_name , "a" ) ) == NULL ) {
!       printf( "Could not open scratch_scores%02.dat for
writing.\nDying\n\n" , thread_num ) ;
!       exit( EXIT_FAILURE ) ;
      }

      for( i = 0 ; i < keep_per_rotation ; i ++ ) {

! /*     max_es_value = min( max_es_value , Scores[i].rpscore ) ;  */
!     if  (Scores[i].rpscore < partial_max)
!           partial_max = Scores[i].rpscore;
        fprintf( ftdock_file, "G_DATA %6d   %6d    %7d      %.0f
%4d %4d %4d      %4d%4d%4d\n" ,
                  rotation , 0 , Scores[i].score ,
(double)Scores[i].rpscore , Scores[i].coord[1] , Scores[i].coord[2] ,
Scores[i].coord[3 ] ,
                  Angles.z_twist[rotation] , Angles.theta[rotation]
, Angles.phi[rotation] ) ;
***************
*** 695,704 ****
      free( Rotated_at_Origin_Mobile_Structure.Residue ) ;

    }
```

189

```
-
    /* Finished main loop */

! /************/

    /* Free the memory */

--- 740,776 ----
       free( Rotated_at_Origin_Mobile_Structure.Residue ) ;

    }
    /* Finished main loop */
+ #pragma omp critical (max_es)
+   if ( partial_max < max_es_value )
+     max_es_value = partial_max ;
+ }
+ /* Finished parallel section */
+ /************/
+ /* Compile scores */
+   if( ( ftdock_scores = fopen( "scratch_scores.dat" , "w" ) ) ==
NULL ) {
+     printf( "Could not open scratch_scores.dat for
writing.\nDying\n\n" ) ;
+     exit( EXIT_FAILURE ) ;
+   }

!   for ( i = 0 ; i < num_threads ; i ++ ) {
!     sprintf( file_name , "scratch_scores%02d.dat" , i ) ;
!     if( ( ftdock_file = fopen( file_name , "r" ) ) == NULL ) {
!       printf( "Could not open scratch_scores.dat for
reading.\nDying\n\n" ) ;
!       exit( EXIT_FAILURE ) ;
!     }
!     while( fgets( line_buffer , 99 , ftdock_file ) ) {
!       fprintf ( ftdock_scores , line_buffer ) ;
!     }
!     fclose( ftdock_file ) ;
!   }
!
!   fclose( ftdock_scores );
! /*
!   for ( i = 0 ; i <= sizeof(max_es_values)/sizeof(max_es_values[0])
; i ++ ) {
!     max_es_value = min( max_es_value , max_es_values[i] );
!   }
! */

    /* Free the memory */

**************
*** 706,718 ****
    rfftwnd_destroy_plan( pinv ) ;

    free( static_grid ) ;
!   free( mobile_grid ) ;
!   free( convoluted_grid ) ;

    if( electrostatics == 1 ) {
```

```
      free( static_elec_grid ) ;
!     free( mobile_elec_grid ) ;
!     free( convoluted_elec_grid ) ;
    }

    for( i = 1 ; i <= Origin_Static_Structure.length ; i ++ ) {
--- 778,790 ----
    rfftwnd_destroy_plan( pinv ) ;

    free( static_grid ) ;
! /*  free( mobile_grid ) ;
!   free( convoluted_grid ) ; */

    if( electrostatics == 1 ) {
      free( static_elec_grid ) ;
! /*    free( mobile_elec_grid ) ;
!     free( convoluted_elec_grid ) ; */
    }

    for( i = 1 ; i <= Origin_Static_Structure.length ; i ++ ) {
***************
*** 736,742 ****
      exit( EXIT_FAILURE ) ;
    }

!   if( ( Scores = ( struct Score * ) realloc ( Scores , ( 1 +
keep_per_rotation ) * Angles.n * sizeof( struct Score ) ) ) == NULL )
{
      printf( "Not enough memory left for storing scores\nProbably
keeping too many per rotation\nDying\n\n" ) ;
      exit( EXIT_FAILURE ) ;
    }
--- 808,816 ----
      exit( EXIT_FAILURE ) ;
    }

!   struct Score *Scores ;
!
!   if( ( Scores = ( struct Score * ) malloc ( ( 1 + keep_per_rotation
) * Angles.n * sizeof( struct Score ) ) ) == NULL ) {
      printf( "Not enough memory left for storing scores\nProbably
keeping too many per rotation\nDying\n\n" ) ;
      exit( EXIT_FAILURE ) ;
    }
```